REVIEW

Editorial Process: Submission:12/12/2024 Acceptance:07/13/2025

Bioinformatics and Omics-based Perspectives on Breast Cancer: Advancing Target Gene Identification for the Development of Anticancer Agents

Irmasari Irmasari[1,2], Alim Khodimul Rahmat[3], I Made Bayu Kresna Yoga[1,2], Dini Maharani[4], Wasita Rachma Widayanti[4], Monica Hana Widyardhita[4], Bakti Wahyu Saputra[1], Inas Haidar[1], Amalia Amalia[1], Sri Rahmayani Ningsih[1], Dyaningyas Dewi Pamungkas Putri[2,5], Adam Hermawan[2,6]*

Abstract

Objective: Breast cancer remains the leading cause of mortality among women worldwide. Innovative strategies, particularly bioinformatics and omics-based approaches, play a crucial role in identifying potential target genes for breast cancer treatment. This review aims to highlight the future acceleration of drug discovery from single natural compounds and plant-derived natural products, including Traditional Chinese Medicines (TCMs), and to explore their potential in enhancing the efficacy of conventional drugs when combined, through the application of bioinformatics tools and various omics-based databases, web servers, or software platforms. **Methods:** This review focuses on research conducted over the past five years, utilizing three major scientific databases: PubMed, ScienceDirect, and Scopus. Using the Rayyan platform, we systematically narrowed 3,800 original studies down to 70 relevant articles. **Result:** The findings present a comprehensive overview of key bioinformatics approaches and omics-based resources, including databases, web servers, and software tools, covering data mining, Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, Protein-Protein Interaction (PPI) network construction and hub gene selection, genetic alteration, and survival analysis. These tools have been employed to identify potential target genes that contribute to the acceleration of drug discovery from single natural compounds and plant-derived natural products, including Traditional Chinese Medicines (TCMs), as well as their role in improving the therapeutic efficacy of conventional drugs when used in combination. **Conclusion:** A comprehensive understanding of omics-based databases, web servers and software can facilitate the acceleration of new drug discovery and enhance the effectiveness of existing conventional drugs in breast cancer therapy. This approach supports validation in preclinical models, both *in vitro* and *in vivo*, ensuring clinical applicability.

Keywords: breast cancer- bioinformatics- omics-based- databases- drug discovery

Asian Pac J Cancer Prev, **26 (7)**, 2291-2308

Introduction

Breast cancer is a complex disease and remains one of the leading causes of death among women worldwide. It is classified into four main subtypes based on hormone receptor expression: estrogen receptor-positive (ER+), progesterone receptor-positive (PR+), human epidermal growth factor receptor-2-positive (HER2+), and triple-negative breast cancer (TNBC) [1, 2]. According to GLOBOCAN 2020 data, breast cancer ranks as the second most commonly diagnosed cancer after lung cancer, accounting for approximately 2.3 million new cases (11.7%) [3]. Current treatment strategies include neoadjuvant chemotherapy (administered prior to surgery), surgical tumor removal, and radiotherapy. Despite these available treatments, significant challenges remain in

*[1]Master's Student in Pharmaceutical Sciences, Faculty of Pharmacy, Universitas Gadjah Mada, Sekip Utara II, 55281 Yogyakarta, Indonesia. [2]Laboratory of Advanced Pharmaceutical Sciences, Faculty of Pharmacy, Gadjah Mada University, Yogyakarta, Indonesia. [3]Doctoral Student in Pharmaceutical Sciences, Faculty of Pharmacy, Universitas Gadjah Mada, Sekip Utara II, 55281 Yogyakarta, Indonesia. [4]Undergraduate Student in Pharmaceutical Sciences, Faculty of Pharmacy, Universitas Gadjah Mada, Sekip Utara II, 55281 Yogyakarta, Indonesia. [5]Laboratory of Pharmacology and Toxicology, Department of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, Universitas Gadjah Mada Sekip Utara II, 55281 Yogyakarta, Indonesia. [6]Laboratory of Macromolecular Engineering, Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada Sekip Utara II, 55281 Yogyakarta, Indonesia. *For Correspondence: adam_apt@ugm.ac.id*

breast cancer therapy, particularly in overcoming drug resistance, accelerating the discovery of new drugs from novel compounds, and predicting patient responses [4, 5].

Bioinformatics approaches, together with omics-based methods, provide powerful tools for the in-depth analysis of genes, pathways, and molecular networks involved in disease progression and therapeutic response. These strategies also play a pivotal role in accelerating the drug discovery process and enhancing the efficacy of breast cancer treatments by combining existing conventional drugs with single natural compounds or plant-derived natural products through the identification of potential molecular targets. Unlike conventional laboratory methods, which are often time-consuming and resource-intensive, bioinformatics utilizes advanced computational tools to efficiently analyze large-scale genomic and proteomic datasets [4, 6, 7]. This computational approach enables researchers to predict drug efficacy, uncover mechanisms of resistance, and prioritize candidate genes or pathways for further experimental validation, thus significantly reducing the time and cost associated with preclinical research. For example, the integration of bioinformatics with omics-based techniques allows the identification of key signaling pathways implicated in breast cancer progression, thereby guiding the development of more targeted and effective therapeutic strategies [8].

A bioinformatics workflow that begins with data mining, followed by GO enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis, protein-protein interaction (PPI) network construction and hub gene selection, genetic alteration assessment, and survival analysis can be applied to explore the molecular mechanisms underlying breast cancer therapeutic responses. Additionally, this approach facilitates the identification of potential target genes to support the rapid discovery of new drugs from single natural compounds and plant-derived natural products, including TCMs, as well as their potential to enhance the efficacy of conventional drugs when used in combination [9–11].

Therefore, this review aims to highlight the application of bioinformatics and omics-based approaches in identifying potential target genes to accelerate the drug discovery process and improve the effectiveness of breast cancer therapy through the combination of existing conventional drugs with single natural compounds or plant-derived natural products. The primary focus is on recent studies from the past five years that have successfully identified relevant potential target genes using bioinformatics approaches, including data mining, GO and KEGG pathway analysis, PPI network construction and hub gene selection, genetic alteration, and survival analysis. Furthermore, this review discusses how this information can serve as a valuable foundation for future research in developing effective breast cancer therapies using bioinformatics tools and various omics-based databases, web servers, or software.

Materials and Methods

Search strategy

The literature search was performed using three major scientific databases: PubMed, ScienceDirect, and Scopus. Boolean operators "OR" and "AND" were applied to combine the search items, which included: "Bioinformatics approaches, target gene, breast cancer, identification, data mining, gene ontology enrichment, and KEGG pathway, protein-protein interaction (PPI), hub gene, genetic alteration, and survival analysis." These keywords were used to identify original research articles published within the last five years, up to August 2024.

Study screening and selection

A total of 3,882 articles were retrieved from PubMed, ScienceDirect, and Scopus (Figure 1). After removing duplicates (n = 82), a total of 3,800 original articles remained for further screening. The compilation of articles and removal of duplicates were conducted using the Rayyan web-based reference management tool (https://www.rayyan.ai) [12].

Data extraction

The data extraction process focused on highlighting the use of omics-based databases, web servers, and software in the identification of potential target genes for breast cancer treatment. This involved exploring single natural compounds and natural products derived from plants or TCMs, as well as their impact on enhancing the effectiveness of conventional drugs when used in combination. The bioinformatics analysis included data mining, GO, and KEGG pathway analysis, PPI network and hub gene selection, genetic alteration and survival analysis across various breast cancer subtypes. Following the screening process, a total of 70 primary articles were selected for review (Figure 1). The extracted data were then organized into a narrative according to the specified writing guidelines and format requirements.

Results

Bioinformatics analysis involves the application of informatics techniques, derived from fields such as applied mathematics, computer science, and statistics, to process and interpret biological data. It also integrates concepts from macromolecular physics, biology, and chemistry, applying bioinformatics approaches to generate large-scale, detailed information about these molecules [13]. In general, several studies have focused on bioinformatics approaches (Figure 2) for identifying potential target genes in breast cancer treatment, utilizing single natural compounds, natural products derived from plants, or TCMs. These studies also examine the effects of conventional drugs when combined with natural products, including data mining, GO and KEGG pathway analysis, PPI network and hub gene analysis, genetic alteration and survival analysis, all performed using various omics-based databases, web servers, and software tools [9–11].

Data Mining

Data mining is the process of extracting valuable information from large and complex data sets [14]. It can serve as a supporting method in the search for target genes using various databases to improve the success of breast cancer treatment, either through single or combination therapy [15]. In the context of bioinformatics and breast cancer research, data mining is used to explore various chemical compounds, drugs, and natural products with potential anticancer activity. It also facilitates the identification of differentially expressed genes (DEGs) relevant to specific breast cancer subtypes, thereby increasing therapeutic effectiveness [16]. This stage typically utilizes omics-based databases or web servers (Table 1). Examples of such resources include the Gene Expression Omnibus from the National Center for Biotechnology Information (NCB GEOI), Swiss Target Prediction, SuperPred, GeneCards, Search Tool for Interactions of Chemicals (STITCH), Traditional Chinese Medicine System Pharmacology (TCMSP), The Cancer Genome Atlas (TCGA), and others.

NCBI GEO is an omics-based database that collects data related to gene expression and gene function analysis. However, a limitation of NCBI GEO is that its metadata is sometimes insufficiently detailed for specific studies. Information such as experimental conditions, techniques applied, or the design of experiments is not always described in sufficient detail, making the interpretation of results challenging without additional information [17]. This database is useful for identifying potential target genes and is frequently employed in breast cancer research involving various natural compounds or traditional medicines, both as single agents and in combination therapies (Supplementary Table 1). These include trastuzumab, Compound Kushen Injection (CKI), paclitaxel, honokiol, and tamoxifen [9,18–21]. The use of NCBI GEO dominates data searches and is applied across multiple breast cancer subtypes, such as luminal [10,22–27] and TNBC [28–36].

Swiss Target Prediction is a database designed to predict the molecular targets of bioactive compounds based on their chemical structure. However, this database does not include all potential chemical compounds, which may result in missing new targets or compounds with undetected therapeutic effects [37]. This database has been used in several studies, including research on trastuzumab, schisandrin, honokiol, and propolis [9, 38–40].

GeneCards provides detailed information on human genes, including their functions, expression patterns, and disease associations. However, unlike databases such as GTEx or TCGA, GeneCards does not offer real-time gene-expression data, which allows for a more detailed analysis across various tissues and conditions [41]. Despite this limitation, GeneCards serves as an important resource for identifying target genes of specific compounds in breast cancer research, including compounds such as anhydroicaritin and schisandrin [39, 42]. A total of 22 studies (Supplementary Table 1) that include various natural compounds or traditional medicine involved in breast cancer research both single and in combination using the GeneCards database. These include compounds such as anhydroicaritin, schisandrin, CKI, Xihuang Pill, borneol, and trastuzumab combinations (Supplementary Table 1).

Discussion

TCGA is an omics-based database that provides extensive genomic information for cancer research, including gene-expression profiles, mutations, and epigenetic data from various tumor types [43]. This database supports the linkage of cancer genome data with investigational therapies, including relevant compounds with potential anticancer properties such as [10]-Gingerol, γ-Mangostin, Sulfasalazine, and CKI [19, 29, 44–46]. However, the data is specific to cancer and does not encompass the entire human genome.

PubChem is a chemical database that contains a wide range of molecules, from small compounds to large macromolecules such as nucleotides, carbohydrates, lipids, peptides, and chemically modified macromolecules. It supports target identification by providing data on bioactive compounds, their molecular structures, biological activities, and interactions with various proteins and genes. However, the database is limited to compounds that have already been published, with no inclusion of new data [47]. There are 13 studies (Supplementary Table 1) on various single compounds such as anhydroicaritin and Paclitaxel [20, 42] as well as combinations of antibreast cancer agents such as honokiol and trastuzumab on the HER2+ breast cancer subtype that used this database to identify differentially expressed target genes [48].

The STITCH is a web server designed to integrate and analyze interactions between chemical compounds and proteins, as well as their interaction networks, based on data collected from several other major databases. Similar to PubChem, STITCH is also limited to compounds that have already been published [49]. Through the analysis of these interaction networks, researchers can explore the relationships between compounds and their gene or protein targets. This approach is highly useful in the discovery of potential therapeutic targets, particularly in cancer research and drug discovery. A total of 14 studies have utilized the STITCH database in the process of identifying potential target genes using various compounds across specific cancer subtypes (Supplementary Table 1).

In addition, several omics-based databases and web servers are utilized during the data mining stage. These include DisGeNET, which focuses on the relationship between genes and diseases, making it particularly useful for target gene identification by providing comprehensive data on genes and genetic variants associated with various diseases, including cancer. However, this makes it is less suitable for data analysis that requires detailed information on biological pathways, protein interactions, or specific molecular mechanisms [50]. OMIM focuses on the relationship between genes and hereditary diseases and provides important information about genes associated with certain pathological conditions; however, its data is limited to known genetic diseases [51]. DrugBank provides detailed information on drugs, including pharmacokinetics, pharmacodynamics, drug interactions,

Table 1. A Summary of the Bioinformatics Approaches and Omics-Based Databases, Web Servers, and Software Used in the Search for Potential Target Genes for Breast Cancer Treatment from Single Natural Compounds, Natural Products from Plants, or TCMs. It also examines their impact on enhancing the effectiveness of conventional drugs when used in combination.

Bioinformatics approaches	Omics-based resources	Category	Link
Data Mining	ArrayExpress	Database	https://www.ebi.ac.uk/ArrayExpress/
	Bioinformatics Analysis Tool for Molecular mechanism of Traditional Chinese Medicine (BATMAN-TCM)	Web server	http://bionet.ncpsb.org.cn/batman-tcm
	BindingDB	Database	https://bindingdb.org/bind/
	BioCrick	Web server	http://www.biocrick.com/
	Comparative Toxicogenomics Database (CTD)	Database	https://ctdbase.org/
	Cancer Cell Line Encyclopedia (CCLE)	Database	https://portals.broadinstitute.org/ccle
	ChEMBL	Database	https://www.ebi.ac.uk/chembl/
	ChemMapper	Web server	http://www.chemmapper.com/
	Developmental Therapeutics Program (DTP)	Database	https://dtp.cancer.gov/
	DisGeNET	Database	http://www.disgenet.org/
	Drug Bank	Database	https://go.drugbank.com/
	Existing Traditional Chinese Medicine (ETCM)	Database	http://www.etcm.info/
	European Genome-Phenome Archive (EGA)	Database	https://ega-archive.org/
	National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO)	Database	https://www.ncbi.nlm.nih.gov/geo/
	Gene Set Enrichment Analysis (GSEA)	Web server	http://software.broadinstitute.org/gsea/index.jsp
	GeneCards	Database	https://www.genecards.org/
	GreenMolBD		http://www.greenmolbd.com/
	HERB	Database	https://herb.ac.cn/
	Herbal Ingredients Targets 2.0 (HIT 2.0)	Web server	http://lifecenter.sgst.cn:8000/
	Indian Medicinal Plants, Phytochemistry and Therapeutics (IMPPAT)	Web server	http://www.imppat.com/
	KNApSACK	Database	http://www.knapsackfamily.com/
	MalaCards	Database	https://www.malacards.org/
	Online Mendelian Inheritance in Man (OMIM)	Database	https://omim.org/
	PharmMapper	Web server	http://www.pharmmapper.org/
	PharmGkb	Database	https://www.pharmgkb.org/
	PubChem	Database	https://pubchem.ncbi.nlm.nih.gov/
	Similarity Ensemble Approach (SEA)	Web server	http://sea.bkslab.org/
	Search Tool for Interactions of Chemicals (STITCH)	Database	http://stitch.embl.de/
	STRING	Web server	https://string-db.org/
	SuperPred	Web server	http://prediction.charite.de/
	Swiss Target Prediction	Web server	http://www.swisstargetprediction.ch/
	SwissADME	Web server	http://www.swissadme.ch/
	SymMap	Web server	http://symmap.ulti-map.com/
	The Cancer Genome Atlas (TCGA)	Database	https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga
	Traditional Chinese Medicine Integrated Database (TCMID)	Database	http://www.megabionet.org/tcmid/
	Traditional Chinese Medicine System Pharmacology (TCMSP)	Database	http://www.cuilab.cn/tcmsp.php
	Terapeutic Target Database	Database	http://bidd.nus.edu.sg/group/cjttd/
	TargetMol	Database	http://www.targetmol.com/
	TargetNet	Database	http://targetnet.scbdd.com/
	UALCAN	Web server	http://ualcan.path.uab.edu/
	UCSC Xene database	Database	https://xenabrowser.net/datapages/
	UniProt	Database	https://www.uniprot.org/
	canSAR Black	Database	https://cansar.icr.ac.uk/
GO and KEGG pathway enrichment analysis	BATMAN-TCM	Web server	https://cran.rproject.org/
	WebGestalt	Web server	https://www.webgestalt.org/
	Database for Annotation, Visualization, and Integrated Discovery (DAVID)	Web server	https://david.ncifcrf.gov/
	GEN ONTOLOGY Database	Database	http://geneontology.org/

Table 1. Continued

Bioinformatics approaches	Omics-based resources	Category	Link
GO and KEGG pathway enrichment analysis	KEGG Database	Database	https://www.kegg.jp/
	ShinyGO	Web server	http://bio.tools/shinygo
	Metascape	Web server	https://metascape.org/
	The Harmonizome software	Software	https://www.harmonizome.org/
	Enrichr	Web server	https://maayanlab.cloud/Enrichr/
	CluePedia	Software	https://apps.cytoscape.org/apps/cluepedia
	ClueGO plug in Cytoscape	Software	https://apps.cytoscape.org/apps/cluego
	SRplot	Web server	http://bio.tools/srplot
	PANTHER	Web server	http://pantherdb.org/
	Bioconductor (clusterProfiler)	Software	https://www.bioconductor.org/packages/release/bioc/html/clusterProfiler.html
	Bioconductor (DOSE)	Software	https://bioconductor.org/packages/release/bioc/html/DOSE.html
	Bioconductor (GSEABase)	Software	https://bioconductor.org/packages/release/bioc/html/GSEABase.html
	Bioconductor (GSVA)	Software	https://bioconductor.org/packages/release/bioc/html/GSVA.html
	SIGNOR	Web server	https://www.signor.dbcls.jp/
PPI network and hub gene selection	STRING	Web server	https://string-db.org/
	Cytoscape	Software	https://cytoscape.org/
	GeneMANIA	Software	https://genemania.org/
Genetic Alteration	cBioPortal	Web server	https://www.cbioportal.org/
	TIMER	Web server	http://timer.cistrome.org/
Survival analysis	Kaplan–Meier Plotter	Web server	http://gepia.cancer-pku.cn/
	Breast Cancer Gene-Expression Miner version 4.7r	Web server	https://bcgenex.ico.unicancer.fr/
	GEPIA	Web server	https://kmplot.com/analysis/
	OncoLnc	Web server	http://www.oncolnc.org/

and chemical structures. Nevertheless, its primary focus is confined to drugs used in medical practice [52]. PharmMapper is designed to identify potential drug targets from chemical compounds based on pharmacological data, enabling rapid mapping of the relationships between chemical compounds and drug targets. Despite this, its accuracy is limited when identifying potential targets in newly discovered compounds [53]. SEA applies a chemical structure similarity approach to predict molecular targets, meaning its analysis results are confined to compounds that share structural similarities with known compounds, making it less effective for compounds with unique or unknown structures [54].

There are also databases and web servers, such as the Therapeutic Target Database which provide information on therapeutic targets and drug molecules related to cancer. Similar to other databases, there may be a delay in data updates, meaning that the most recent information on a specific drug target or compound may not yet be available [55]. BindingDB offers binding affinity data of small compounds to relevant target proteins in cancer; however, it only includes interactions that have been experimentally verified [56]. TargetNet and TargetMol utilize computational engines to predict drug targets, primarily relying on the chemical similarity of compounds to known targets. As a result, their effectiveness is reduced when predicting targets for compounds with unique or novel chemical structures [57]. canSAR Black integrates information from multiple fields including biology, chemistry, pharmacology, structural biology, cellular networks, and clinical data, to support the identification of therapeutic targets in cancer. Nevertheless, its scope is limited to specific cancer types and lacks comprehensive in-depth genetic studies [58]. PharmGkb provides data on genetic associations with drug response and is valuable for understanding treatment responses in cancer. However, access to certain detailed data or advanced analytical tools may require special permissions, limiting general users in conducting deeper data exploration [59]. The Comparative Toxicogenomics Database links chemical compounds with genes and diseases to aid in the identification of relevant targets. However, it does not include all chemical compounds or molecular targets within its database [60]. The UCSC Xena Database offers genomic data for cancer gene expression and mutation analysis, but it lacks detailed information on specific genetic mechanisms [61].

UALCAN enables the analysis of breast cancer gene expression, but it does not cover all cancer types and is limited to gene-expression data [62]. STRING analyzes protein-protein interactions, helping to understand molecular networks in cancer therapy. Therefore, to conduct an in-depth data mining analysis, STRING should be used in conjugation with other tools that can provide additional biological context or quantitative

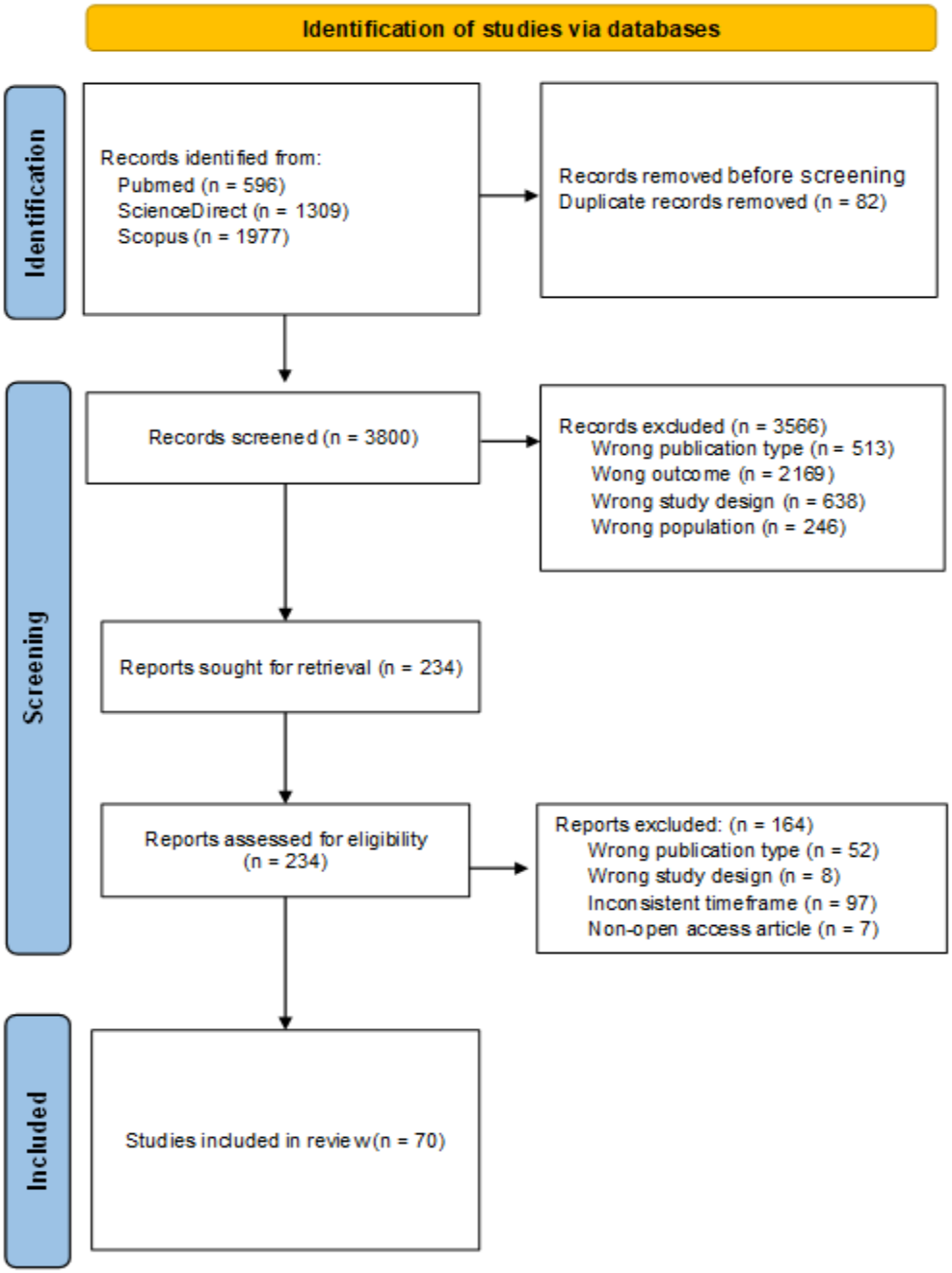

Figure 1. PRISMA Flow Diagram of the Literature Search Strategy and Selection Process in This Narrative Review.

data, such as TCGA, NCBI GEO, or DisGeNET [63]. IMPPAT supports the discovery of potential targets from herbal compounds, but it is limited to anticancer drugs and lacks detailed information on other pathways [64]. HERB integrates data on herbal-based therapies and active molecules, but its scope is restricted to specific herbal medicines, and the knowledge in this field is still limited [65]. SwissADME aids in assessing the pharmacokinetic properties of cancer compounds, but the predictions are based solely on computational data and require laboratory validation [66]. ChEMBL provides information on bioactive compounds and their activities, including chemical, bioactivity, and genomic data, which helps translate genomic information into effective drugs. However, the data in ChEMBL predominately relate to the preclinical activity of chemical compounds, and data on clinical effectiveness or advanced-phase studies are often unavailable [67]. BioCrick is a database that collects information related to bioactive compounds used in pharmacology and toxicology, offering detailed information on chemical structures, biological activity, and molecular targets. However, interpretating chemical structure data requires additional expertise. The Cancer Cell Line Encyclopedia provides data on gene expression and genetic mutations in cancer cells, but the data are limited to specific cancer cell lines, not covering all genes or pathways [68]. The Developmental Therapeutics Program is a database from the National Cancer Institute that provides information on chemical compounds and their anticancer activity. The data are available for research on anticancer compounds and their effects on various cancer cells, but the focus is primarily on anticancer

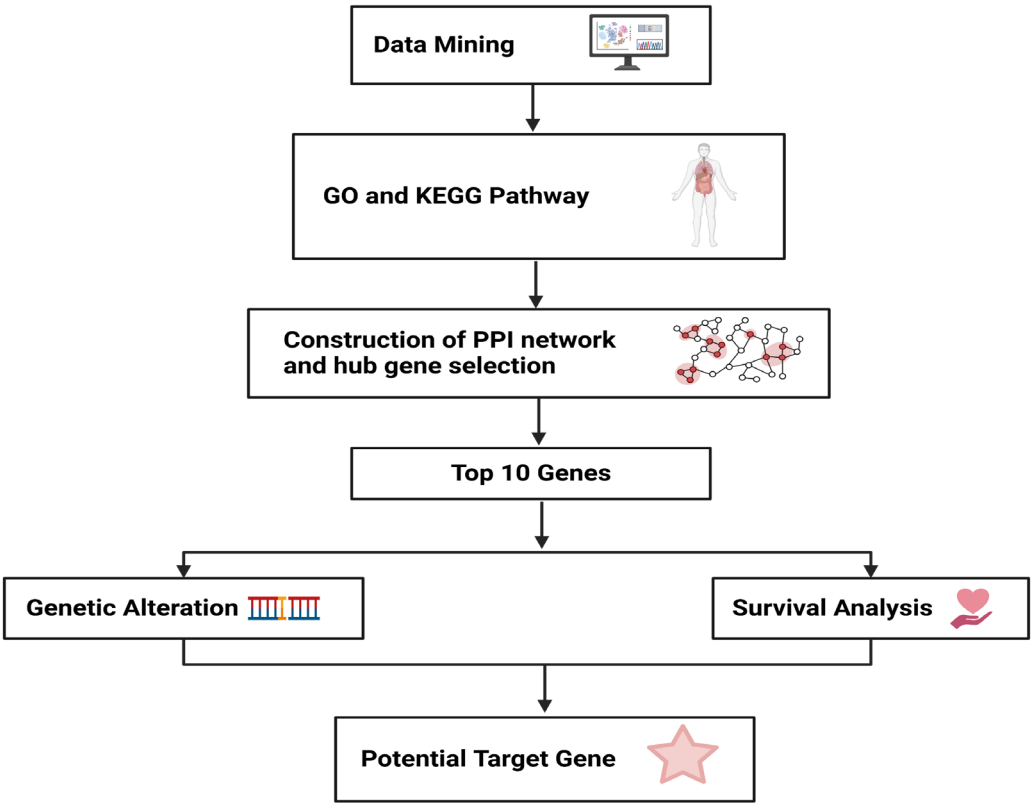

Figure 2. The Workflow of Bioinformatics Approaches Applied in the Search for Potential Target Genes for Breast Cancer Treatment Using Various Omics-based Databases, Web Servers and Software. This process involves the analysis of single natural compounds, natural products derived from plants, or TCMs, as well as the evaluation of their combined effects with conventional drugs to enhance treatment efficacy across different types of breast cancer.

assays and does not cover the full range of biological activity [69]. DrugBank provides detailed information on drugs, including pharmacokinetics, pharmacodynamics, drug interactions, and chemical structure, but is limited to drugs used in medical practice [70].

UniProt was also utilized in the data mining stage (Table 1), providing protein-related data, including those associated with cancer, such as information on protein structure, function, and interactions. However, it lacks detailed data on gene expression and mutations [71]. KNApSACK offers data on plant metabolites that may be relevant to cancer targets, but it does not include other types of chemical compounds [72]. SuperPred predicts compound targets based on structural similarity, which limits its ability to predict the activity of compounds with unique or previously unstudied structures [73]. MalaCards compiles data on the association of genes with diseases, including cancer, from various sources. However, due to the large volume of data, it can sometimes be challenging to locate specific information [74]. HIT 2.0 provides data on medicinal plants and their targets; however, its data is limited to published interactions and does not consider potential interactions that have yet to be reported [75]. ChemMapper predicts molecular targets of compounds through pharmacophore similarity. If a compound does not share pharmacophore similarity with previously tested compounds, its predictive accuracy may be reduced or unavailable [76]. Gene Set Enrichment Analysis (GSEA) enables the analysis of gene expression and cancer-related pathways but requires advanced bioinformatics knowledge for accurate interpretation of the results [77]. ArrayExpress stores gene-expression data for target analysis; however, it relies on existing gene-expression data, meaning that if data related for a particular compound or gene is not yet available in the database, the analysis may be incomplete or inaccurate [78]. GreenMolBD provides data on plant bioactive compounds, but it does not include all chemical compounds with pharmacological activity [79]. The European Genome-Phenome Archive offers genotype and phenotype data for target analysis in breast cancer, but its data is limited to studies that have been deposited in this database [80].

Some databases also offer information related to TCM, such as TCMSP, which facilitate drug discovery from herbal medicines by integrating information on pharmacochemistry, Absorption, Distribution, Metabolism, and Excretion properties, drug feasibility, drug targets, related diseases, and interaction networks. Similarly, BATMAN-TCM, SymMap, Existing Traditional Chinese Medicine (ETCM), and the Traditional Chinese Medicine Integrated Database (TCMID) also intregated TCM data to explore the molecular mechanisms of herbal compounds, including potential cancer targets. However, the data in some of these web servers are limited to specific medicinal plants and their interactions, with less emphasis on compounds used in modern medicine [81–84].

Data mining plays a vital role in advancing the understanding of target genes within compounds and

molecular mechanisms that were previously unknown, thereby supporting progress in the field of molecular oncology. By utilizing various omics-based databases or web servers, researchers can analyze the molecular mechanisms of single natural compounds and natural products derived from plants or TCMs, enabling the identification of more effective strategies for breast cancer treatment. As more data become available, this approach will continue to evolve, facilitating the discovery of more potential targets and improving treatment outcomes in the future.

GO and KEGG Pathway Analysis

Complex diseases, such as breast cancer, are often driven by mutations in multiple genes. Identifying disease-associated genes is essential for understanding the biological mechanisms underlying the disease. GO and KEGG pathway enrichment analysis provide information on the biological processes, metabolic pathways, molecular functions, and disease classifications of DEGs [85]. GO classifies genes into three main categories: biological process, molecular function, and cellular component. This analysis helps to determine the roles of cancer-related genes in specific biochemical activities, cellular structures, or molecular interactions [86]. KEGG focuses on metabolic pathways and molecular signaling networks that are disrupted in certain diseases, including cancer. By mapping DEGs to KEGG pathways, researchers can gain insights into the cellular and molecular mechanisms involved in the development and progression of breast cancer [87].

Several databases, web servers, or software tools are commonly used to perform GO and KEGG pathway enrichment analyses, including DAVID, WebGestalt, ShinyGO, ClusterProfiler, and others (Table 1). These tools are selected based on the specific research needs related to various compounds and breast cancer subtypes. For instance, the DAVID database is frequently used for GO and KEGG pathway enrichment analysis of various single natural compounds, natural products derived from plants, or TCMs. It is also applied to study their potential to enhance the efficacy of conventional drugs when used in combination for breast cancer treatment (Supplementary Table 1). Examples include compounds such as [10]-Gingerol, Naringenin, CKI, Oleanolic acid with tamoxifen, Paeoniflorin with tamoxifen, and Berberine combined with tamoxifen [19, 23, 29, 88, 89]. This is because DAVID offers a complete set of functional annotation tools that enables researchers to interpret the biological significance of large gene lists, integrating information from various functional annotation sources. However, DAVID also has limitations, such as offering less interactive and less visually engaging representations of analysis results compared to tools like WebGestalt or ShinyGO [90].

WebGestalt is a bioinformatics platform used for gene functional analysis, including GO, KEGG pathway analysis, and GSEA of omics-based datasets [91,92]. However, it offers fewer options for advanced analysis or parameter customization compared to tools such as Bioconductor. This platform has been used to analyze various single natural compounds or natural products derived from plants, either individually or in combination, to enhance the sensitivity of conventional drugs in breast cancer treatment (Supplementary Table 1). Examples include the use of Tangeretin alone or in combination with tamoxifen, as well as borneol with trastuzumab, and other compounds such as Hesperetin, Hesperidin, Naringenin, and Nobiletin [93–99].

ShinyGO is a web-based tool designed for gene annotation analysis, including GO and pathway analysis such as KEGG, while also providing interactive visualization of analysis results through graphs and functional networks. This allows users to explore results in real time with interactive and visually appealing graph representations. However, ShinyGO is limited in flexibility compared to R-based tools like clusterProfiler, particularly for analyzing more complex analysis [92]. There are two studies that have employed ShinyGo in the GO and KEGG pathway analysis stages, including research on Aporphine Alkaloid and Epigallocatechin gallate (EGCG), as well as Epicatechin (EC), Epigallocatechin (EGC), and Epicatechin gallate (ECG) [100,101].

In addition, several databases, web servers, or software tools are commonly used in GO and KEGG pathway analysis (Table 1). The GO Database provides structured information on gene function, biological processes, and cellular components. However, it has limitations when applied to complex analyses, as it requires a good knowledge of GO terminology and bioinformatics [102]. Disease Ontology Semantic Enrichment (DOSE) is used to visualize the association of genes with diseases based on the disease ontology database. It is suitable for rapid GO and pathway enrichment analysis, as it provides live graphical representation of the results. However, it requires knowledge of genomic analysis and proficiency in R programming [103]. The KEGG Database integrates information on biological pathways, drugs, and genomes, supporting pathway analysis relevant to breast cancer. However, accurate gene annotation is essential to avoid misinterpretation of pathways, and access to the full data requires a subscription [104]. Metascape is an online tool used for genomic data analysis, including GO and KEGG, offering visualization and clustering features. However, its analysis results depend on the frequency and accuracy of database updates [105].

Signaling Network Open Resources provides signaling pathway interaction data relevant to cancer, including GO and KEGG enrichment results, along with access to various genetic interaction datasets. However, regular data updates are necessary to maintain accuracy [106]. Enrichr is a web-based tool designed for genetic analysis, allowing for quick and easy enrichment of GO, KEGG, and other pathways. It is particularly suitable for rapid initial analysis, as it requires no programming skills and offers clear, easy-to-interpret outputs. However, it lacks the flexibility of R-based tools for more in-depth analysis, and its results depend on database updates, which may not always include the most recent information [107].

CluePedia and ClueGO, which are Cytoscape plugins, can also be used for the visualization and analysis of biological pathway networks. These tools support PPI

analysis and visualization of GO networks. However, they may be challenging to use without a good understanding of genetic pathways [108, 109]. Srplot is a web-based tool designed for visualizing gene expression, survival data, and pathway analysis. It integrates omics-based data to provide insights into gene-expression patterns, survival outcomes, and their association with functional annotations such as GO. Although Srplot can display the association of genes and biological processes, it may not be as comprehensive as GO-specific databases like DAVID or Bioconductor [110].

Protein Analysis Through Evolutionary Relationships (PANTHER) offers genetic analysis, including GO and KEGG enrichment, as well as PPI mapping across multiple species. However, database updates are required to ensure result accuracy [111]. Biological Annotation Tool for Molecular mechanism of Traditional Chinese Medicine (BATMAN-TCM) is a web-based tool that integrates omics-based data, including gene expression, protein-protein interactions, and pathway databases such as GO and KEGG. It facilitates the identification of molecular mechanisms, potential targets, and related pathways in TCM applications. BATMAN-TCM is particularly useful for exploring the biological functions of herbal medicines and identifying potential biomarkers for drug development. However, its data limitations may restrict its application to certain species or disease conditions, which means it may not always provide relevant results for all studies [84]. Harmonizome provides easy access to thousands of genetic datasets, including GO and KEGG annotations, and protein-protein interactions, making it a user-friendly platform for data exploration. However, its data may not always reflect the latest findings and may lack the depth of detailed analysis available in specialized tools [112].

ClusterProfiler, DOSE, GSEABase, and GSVA are tools within Bioconductor that provide R-based analysis for GO and KEGG pathway enrichment, each offering specific strengths and limitations. ClusterProfiler delivers high flexibility in analysis and advanced visualization; however, it requires a steep learning curve and a basic understanding of R programming. DOSE is more user-friendly for performing quick GO and pathway enrichment analysis, although it also presents a steep learning curve and offers limited flexibility. GSEABase provides high flexibility in gene-expression analysis and allows for customization of settings, but it demands strong knowledge of R programming and genomic analysis, and its are highly dependent on the quality of the input data. GSVA enables unsupervised pathway analysis, making it effective for evaluating time-series data; however, its performance relies on accurate annotation and it may lack sensitivity in detecting small changes.

GO and KEGG pathway analysis play a crucial role in understanding the molecular complexity of breast cancer. The use of databases such as DAVID, along with web servers and other bioinformatics software, assist researchers in identifying biological pathways and potential target genes that could serve as the basis for developing new therapies across various subtypes of breast cancer. Notably, the pathways identified through these analyses, including apoptosis, NF-B, and hormone signaling pathways, may become key focal points in future drug development and therapeutic strategies.

PPI Network and Hub Gene Selection

Proteins play a vital role in determining biological functions, as most cellular activities are regulated through interactions between proteins. Understanding these protein-protein interactions is essential for understanding their roles within the cell. PPI analysis also enables more accurate predictions of potential interactions between proteins, facilitating interaction mapping [113]. This approach not only helps to validate experimental results but also assists in selecting potential targets for further investigation. Additionally, PPI analysis allows for the exploration of protein interactions across multiple levels, ranging from metabolic pathways and cellular processes to the organism as a whole [114]. Molecular interaction networks provide valuable visualization of complex cellular processes, which are often difficult to interpret due to their complexity. Cancer-specific PPIs are also useful in linking and mapping both common and rare mutations, demonstrating that key mutations in breast cancer can significantly alter interaction patterns [115]. Through PPI network analysis, utilizing omics-based databases and software tools such as STRING, Cytoscape, and GeneMANIA (Table 1), researchers have identified hub genes. These include proteins characterized by a high degree of connectivity within the network and their involvement in many interactions, suggesting their central role in the biological processes underlying cancer. In the context of breast cancer, PPI networks have been employed to investigate the effects of various single natural compounds, plant-derived natural products, and TCMs. Furthermore, these networks have been used to assess the potential of such compounds to enhance the effectiveness of conventional drugs when used in combination for breast cancer treatment (Supplementary Table 1).

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) is a database that provides a comprehensive resource for both known and predicted protein-protein interactions [63]. STRING is widely used for mapping protein-protein interactions and is often complemented by Cytoscape software to further analyze hub genes within a study [116]. Cytoscape is a powerful tool for visualizing molecular interaction networks, integrating gene-expression profiles and other datasets, and identifying hub genes based on specific parameters. It enables users to construct interaction networks between chemical compounds and biological targets. Additionally, this software facilitates the analysis of the significance of nodes within the network using parameters such as degree, betweenness, and closeness centrality. It also supports the integration of data from multiple sources, enabling more in-depth analysis, while its interactive visualizations help simplify the understanding of complex biological networks [116].

Various single natural compounds, plant-derived natural products, and TCMs, along with their effects on enhancing the efficacy of conventional drugs when

used in combination for breast cancer treatment, have been studied using STRING for mapping protein-protein interactions, followed by Cytoscape for hub gene analysis. These compounds include anhydroicaritin, propolis, cardiac glycosides from *Vernonia amygdalina*, schisandrin, Hesperetin, Curcumin, *Zingiber officinale* (Ginger) and *Allium sativum* (Garlic) extract, Xiaoyaosan, CKI, Si-Wu-Tang, and combinations such as Tangeretin with tamoxifen, borneol with tamoxifen, and Honokiol with trastuzumab [11, 19, 22, 28, 39, 40, 42, 46, 48, 93, 117–120]. STRING provides extensive and reliable interaction data, while Cytoscape offers in-depth network analysis and visualization tools. The combination of both platforms enable detailed analysis using network metrics such as, degree, closeness, and betweenness to identify key genes within biological systems. However, optimal analysis in Cytoscape often requires additional plugins, such as cytoHubba, for optimal functionality, requiring users to develop proficiency in both tools for effective analysis.

GeneMANIA is a Cytoscape plugin that enables rapid gene function prediction using a guilt-by-association approach. It identifies genes related to a query set by leveraging extensive functional interaction networks from various organisms, with each prediction traceable to its source network [121]. Some of the natural compounds analyzed using GeneMANIA include the combination of honokiol with tamoxifen, and γ-Mangostin with doxorubicin for the TNBC subtype, to map PPIs through extensive data integration capabilities and additional gene predictions. Its main advantages lie in its ease of use and predictive functionality, making it a popular choice in bioinformatics. However, the results require experimental validation, and in-depth analysis often requires complementary tools such as Cytoscape in the context of cancer.

PPI network analysis provides a powerful method for understanding the complex molecular mechanisms involved in breast cancer. By focusing on hub genes and using a combination of tools such as STRING and Cytoscape, researchers can identify potential therapeutic targets. This highlights the utility of this approach in identifying promising single natural compounds and natural products derived from plants or TCMs, as well as their ability to enhance the effectiveness of conventional drugs when combined in the treatment of breast cancer and its associated gene targets across various subtypes. This approach is crucial in the development of more precision medicine, where treatment is tailored based to the specific molecular profile of a patient's breast cancer.

Genetic Alteration

Genetic alteration allows researchers to understand how specific genetic mutations affect gene expression and molecular pathways, and how they contribute to the development of or resistance to breast cancer [23]. Omics-based databases or web servers used for genetic alteration analysis, such as cBioPortal and Tumor Immune Estimation Resource (TIMER) (Table 1), enable in-depth exploration of mutations, copy number variations (CNVs), and other changes in the profiles of breast cancer patients. These omics-based databases or web servers can analyze large-scale clinical data from various sources; such as, cBioPortal is frequently used to investigate the effects of various single natural compounds and natural products derived from plants or TCM, as well as their ability to enhance the efficacy of conventional drugs when used in combination for breast cancer treatment and their effects on breast cancer subtypes (Supplementary Table 1).

cBioPortal offers interactive visualization of genetic data, including mutations and gene expression, to uncover associations with drug resistance [122]. For example, the combination of tangeretin and tamoxifen utilized cBioPortal to identify specific genetic alterations linked to critical molecular pathways in breast cancer [93]. This analysis enabled researchers to map genes involved in the therapeutic response, particularly in distinct subtypes. Similar studies involved the combination of Honokiol and trastuzumab focused on the HER2+ subtype, examining the HER2 signaling pathway in which genes such as *ERBB2* are frequently amplified or mutated, presenting potential targets for combination therapy [48]. Other research, including the combination of borneol and tamoxifen, employed cBioPortal to investigate genetic changes associated with estrogen signaling pathways in the luminal A subtype [22]. Additionally, brazilin, studied in the *HER2*+ subtype, demonstrated potential in modulating key genes such as *CCND1* and *ERBB2* [123]. In TNBC, analyses of compounds like cannabidiol analogs and cardiac glycosides from *Vernonia amygdalina* using cBioPortal aimed to identify mutations in genes that could be targeted therapeutically [11].

While tools like cBioPortal provide extensive data on genetic alterations, some compounds, such as curcumin and epigallocatechin gallate (EGCG) [28, 101], are not directly linked to genetic analyses. However, their investigation is often complemented by PPI data from tools like STRING, which help connect genetic alterations to relevant biological pathways. A major limitation of these studies is that not all compounds or their combinations have complete data on specific gene mutations. However, the focus on PPI networks and molecular pathways still continues to offer valuable insights into the therapeutic potential of these compounds. By utilizing tools like cBioPortal, this research significantly contributes to identifying potential target genes, such as ERBB2 in HER2+ subtypes and PIK3CA in various breast cancer subtypes. This approach supports the development of more targeted and effective cancer therapies.

TIMER, used to analyze the correlation between gene mutations and immune cell infiltration in tumor tissue, plays a vital role in understanding the immune response to cancer, especially in subtypes like TNBC. It is used to study the interaction between tumor cells and the immune system within the tumor microenvironment. The web server also maps gene expression and its associations with genetic mutations [124]. Quercetin, a well-known natural compound, has been extensively studied for its potential therapeutic effects, particularly in cancer treatment.

TIMER is another valuable resource used alongside cBioPortal to investigate the relationship between genetic alterations and immune cells within the tumor

microenvironment. This platform enables researchers to understand how gene-expression changes, such as those influenced by Quercetin, affect the infiltration and activity of immune cells like T cells and macrophages. In Quercetin-related studies, TIMER can demonstrate how genetic alterations might modulate immune responses, potentially enhancing the compound's efficacy by promoting an immune-stimulatory environment. Study findings revealed that genes such as *MYC, CXCL10, CXCL11*, and *E2F1* showed differential expressions across tumor types and were associated with immune cell abundance [125]. Analysis using cBioPortal also revealed that these genes exhibited significant CNVs and were linked to patient survival. TIMER was used in conjugation with Quercetin to examine the relationship between immune infiltration and various factors, including gene expression, clinical outcomes, somatic mutations, and alterations in somatic copy number.

However, there are certain limitations in the use of TIMER. For instance, the results generated by TIMER may not be fully applicable to all cancer types or animal models due to the limited tumor populations included in its database, which restricts the generalizability of the findings in every clinical context. Additionally, the accuracy of TIMER's results is highly dependent on the quality and quantity of gene-expression data in the underlying database. Variability in data samples, such as differences in quality and quantity from diverse sources, can affect the precision of immune cell infiltration estimates provided by the TIMER algorithm. By using TIMER and cBioPortal, researchers were able to identify connections between genetic alterations and immune responses, offering deeper insights into the molecular mechanisms underlying breast cancer and uncovering potential new therapeutic targets [125]. Overall, the combined use of cBioPortal and TIMER in analyzing Quercetin's genetic alterations highlights the significance of these tools in cancer research. They enhance our understanding of the genetic landscape influenced by natural compounds, thereby supporting the development of more precise and targeted cancer therapies.

Survival Analysis

Survival analysis refers to the ability of a factor to predict the progression or outcome of a disease or medical condition. In early-stage breast cancer patients, survival analysis of a gene is conducted to determine whether that gene can serve as an indicator for predicting patient survival [126]. Bioinformatics approaches used in survival analysis aid in forecasting patient outcomes based on genetic and clinical data. Some of the commonly used omics-based databases or web servers for this analysis include Kaplan-Meier Plotter, Breast Cancer Gene-Expression Miner (bc-GenExMiner), Gene-Expression Profiling Interactive Analysis (GEPIA), and OncoLnc (Table 1).

Kaplan-Meier Plotter is the most widely used tool for examining survival analysis in relation to various single natural compounds and natural products derived from plants or TCMs, as well as their ability to enhance the efficacy of conventional drugs when used in combination for breast cancer treatment and their effects on different breast cancer subtypes (Supplementary Table 1). This web server enables researchers to generate survival plots based on gene-expression levels. It allows comparison of patient survival across varying expression levels of genes and assists in identifying prognostic biomarkers [127].

Kaplan-Meier Plotter is primarily used to assess the survival analysis of gene expression in cancer patients, based on survival data such as overall survival and relapse-free survival. It is employed to examine the relationship between specific gene-expression levels and clinical outcomes following treatment. Some of the compounds studied using this web server include Tangeretin and tamoxifen, which were evaluated for their combined effect on breast cancer patients. Findings indicated that this combination could improve survival in TNBC patients [93]. Survival analysis using Kaplan-Meier Plotter also demonstrated that borneol, when combined with tamoxifen, improved survival rates in patients with the luminal A subtype, suggesting the potential of borneol as a co-chemotherapeutic agent in treating luminal A breast cancer [22]. Additionally, Kaplan-Meier Plotter showed that the combination of Honokiol and trastuzumab enhanced survival in patients with the HER2+ subtype of breast cancer (48). Several other natural products, such as schisandrin, Hesperetin, Curcumin, *Cremastra appendiculata*, [10]-Gingerol, Cardiac glycosides from *Vernonia amygdalina*, Xiaoyaosan have also been studied using Kaplan-Meier Plotter to evaluate their potential to improve survival in patients with specific breast cancer subtypes [11, 28, 29, 39, 117, 128, 129]. While the platforms offers Kaplan-Meier graphs to visualize the relationship between gene expression and survival, as well as insights into multigene interaction, the results must be confirmed with other studies or additional data as this platform only uses limited public data.

The Breast Cancer Gene-Expression Miner (bc-GenExMiner) also provides access to gene-expression data specific to breast cancer, enabling survival analysis based on gene-expression profiles. Data generated from this web server is often used to identify genes with potential as therapeutic targets. The Breast Cancer Gene-Expression Miner is designed to explore the expression of specific genes across various breast cancer datasets and correlate them with clinical outcomes. This platform focuses on analyzing gene-expression based on the molecular subtype and clinical status of the patient [130]. One compound studied using this web server was anhydroicaritin, which was analyzed to determine its relationship with gene expression in breast cancer [42]. Although it offers tools for Kaplan-Meier survival analysis and multigene analysis, the reliability of the results depends on the quality and relevance of the data available on the platform. Therefore, further validation through clinical studies is necessary to confirm the findings.

GEPIA provides a tool for gene-expression analysis utilizing data from TCGA and GTEx. It enables users to visualize differences in gene expression across various cancer subtypes and examine their relationship with prognosis. GEPIA is also a bioinformatics web server used to assess gene expression and perform survival analysis

on RNA-seq data derived from TCGA. It facilitates the exploration of correlations between gene expression and clinical outcomes [131]. Ten studies have employed this web server to conduct survival analysis involving various single natural compounds and natural products derived from plants or TCM in both unspecified breast cancer subtypes and TNBC subtypes (Supplementary Table 1). CKI, a TCM, has been evaluated using GEPIA, showing that it may enhance survival in breast cancer patients [19, 46]. Cannabidiol analogs, *Flos daturae,* Syringin, and *Salvia miltiorrhiza* were also analyzed using a combination of GEPIA and Km-plotter to enrich the survival data obtained [25, 132, 133]. Although GEPIA provides gene-expression analysis across various cancers, including breast cancer from TCGA and the Genotype-Tissue Expression (GTEx) project, the data may not always capture patient variability outside of these datasets.

OncoLnc is another web server that offers insights into the relationship between gene expression and survival, integrating TCGA data to support in-depth analyses [134]. One study employed Oncloc in the survival analysis of breast cancer patients, specifically examining the combination of PGV-1 and Piperine [135]. While it allows exploration the gene expression-survival relationships and gene interactions, offering Kaplan-Meier plots, heatmaps, and expression visualizations, the findings still require further validation, as the conclusions depend solely on TCGA data, which may not fully represent broader patient populations.

Survival analysis plays a vital role in breast cancer management by offering critical insights into prognosis and treatment response. By using various omics-based databases or web servers such as Kaplan-Meier Plotter and GEPIA, researchers can assess the effects of individual natural compounds and plant-derived or TCM. These tools also help evaluate how such compounds, when combined with conventional therapies, influence gene expression and clinical outcomes, aiding in the identification of potential prognostic biomarkers. The data generated from these analyses may contribute to the development of more personalized and effective treatments for breast cancer patients.

Conclusions and Future Direction

The use of omics-based databases, web servers, and software in bioinformatics approaches is essential for identifying target genes in breast cancer treatment. These tools enable in-depth and large-scale data analysis, facilitating the study of gene-expression patterns, mutations, predicting gene-target interaction predictions, and genetic changes relevant to specific diseases. This supports rational drug design prior to laboratory validation. Advances in sequencing technologies and sophisticated analytical algorithms have led to more accurate and in-depth data interpretation, accelerating the discovery of novel biomarkers and the identification of more specific drug targets. Additionally, network simulation models contribute to target gene discovery by predicting the impact of new therapies on complex biological networks.

Bioinformatics and omics-based approaches have proven effective in identifying potential target genes for breast cancer therapy using anticancer agents from single natural compounds and natural products derived from plants or TCMs. The integration of bioinformatics methods with various omics-based databases, web servers, or software facilitates a deeper understanding of the molecular mechanisms underlying therapeutic responses and the enhancement of conventional drug efficacy when used in combination. The application of data mining, GO and KEGG pathway analyses, PPI network and hub gene analysis, genetic alteration, and survival analysis enables a more comprehensive examination of gene functions and interactions. This review outlines the features, advantages, and limitations of various omics-based databases, web servers, and software used in the search for potential target genes in breast cancer therapy. Such knowledge aids researchers in selecting the most suitable database for their investigations, accelerates the discovery of anticancer agents for breast cancer treatment, and enables validation in preclinical models, both *in vitro* and *in vivo*, to ensure clinical relevance.

Author Contribution Statement

Conceptualization: II, AH; Data curation: II, AKR, DM, WRW, MHW, BWS, IH, AA, SRN; Supervision: DDPP, AH; Writing - original draft: II; Funding acquisition: AH; Writing - review and editing: II, AKR, IMBKY, AH.

Acknowledgements

General

II expresses her gratitude to the Beasiswa Unggulan 2023 program for its support during the Master's studies in Pharmaceutical Sciences at the Faculty of Pharmacy, Universitas Gadjah Mada. The author also extends her thanks to Mrs. Ririn Widarti for administrative assistance and to Badan Penerbit dan Publikasi Universitas Gadjah Mada for their support in writing.

Funding Statement

The authors gratefully acknowledge financial support of Penelitian Tesis Magister from the Ministry of Education's DRTPM BIMA Funding, under contract number 048/E5/PG.02.00.PL/2024, dated June 11, 2024.

Scientific Approval and Thesis Acknowledgment

This study is not part of a thesis but fulfills one of the requirements for the completion of the II Master's degree in Pharmaceutical Sciences at Universitas Gadjah Mada. The research was conducted under faculty supervision and was reviewed and approved as part of the university's academic requirements.

Ethical Declaration

This study is exempt from ethical review and approval, as it is a narrative review based on publicly available data from previously published research.

Availability of Data and Materials

All data used in this study are publicly available and

derived from previously published research. Any datasets supporting the findings can be accessed through the references cited within the manuscript.

Study Registration

As a narrative review, this study does not involve primary data collection, clinical trials, or meta-analyses, and thus does not require formal registration in a research database.

Conflict of Interest

The authors declare that there are no conflicts of interest.

References

1. Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. Breast. 2022;66:15-23. https://doi.org/10.1016/j.breast.2022.08.010.
2. Borunda EO, Patricia AN, Lucero EA, Francisco OGV, Claudia ARV. Breast Cancer [Internet]. Brisbane (AU): Exon Publications; 2022 [Cited 2024 Oct 2024]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK583808/
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209-49. https://doi.org/10.3322/caac.21660.
4. Khan SU, Jan S, Fatima K, Wani A, Malik F. Future Directions and Challenges in Overcoming Drug Resistance in Cancer. In: Khan SU, Malik F, editors. Drug Resistance in Cancer: Mechanisms and Strategies. Singapore: Springer Nature Singapore; 2024. p. 351–72.
5. De Matteis A, Nuzzo F, D'Aiuto G, Labonia V, Landi G, Rossi E, et al. Docetaxel plus epidoxorubicin as neoadjuvant treatment in patients with large operable or locally advanced carcinoma of the breast: A single-center, phase II study. Cancer. 2002;94(4):895–901. https://doi.org/10.1002/cncr.20335.abs
6. Ji X, Lu Y, Tian H, Meng X, Wei M, Cho WC. Chemoresistance mechanisms of breast cancer and their countermeasures. Biomed Pharmacother. 2019;114:108800. https://doi.org/10.1016/j.biopha.2019.108800.
7. Alam MS, Sultana A, Reza MS, Amanullah M, Kabir SR, Mollah MNH. Integrated bioinformatics and statistical approaches to explore molecular biomarkers for breast cancer diagnosis, prognosis and therapies. PLoS One. 2022;17(5):e0268967. https://doi.org/10.1371/journal.pone.0268967.
8. Chen R, Ma S, Qiao H, Su F, Wang L, Guan QL. Identification of target genes and prognostic evaluation for colorectal cancer using integrated bioinformatics analysis. All Life. 2022;15(1):160–73. https://doi.org/10.1080/26895293.2022.2026825
9. Skolastika S, Hanif N, Ikawati M, Hermawan A. Comprehensive computational analysis of honokiol targets for cell cycle inhibition and immunotherapy in metastatic breast cancer stem cells. Evid Based Complement Alternat Med. 2022;2022:4172531. https://doi.org/10.1155/2022/4172531.
10. Hermawan A, Putri H, Hanif N, Fatimah N, Prasetio HH. Identification of potential target genes of honokiol in overcoming breast cancer resistance to tamoxifen. Front Oncol. 2022;12:1019025. https://doi.org/10.3389/fonc.2022.1019025
11. Hermawan A, Satria D, Hasibuan PAZ, Huda F, Tafrihani AS, Fatimah N, et al. Identification of potential target genes of cardiac glycosides from Vernonia amygdalina Delile in HER2+ breast cancer cells. S Afr J Bot. 2024;164:401 18. https://doi.org/https://doi.org/10.1016/j.sajb.2023.12.002
12. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210. https://doi.org/10.1186/s13643-016-0384-4
13. Suneetha L, Marsakatla P, Suneetha RS, Suneetha S. Bioinformatics as a Tool to Identify Infectious Disease Pathogen Peptide Sequences as Targets for Antibody Engineering. InAntibody Engineering. IntechOpen;2018.
14. Rahayu P, I Gede IS, Suryani, Arie S, Achmad R, I Gede MD, et al. BUKU AJAR DATA MINING [Internet]. Jambi: PT. Sonpedia Publishing Indonesia; 2024 [cited: 2024 Oct 12]. Available from: https://www.researchgate.net/publication/377415198
15. Roberto Cesar MO, German LB, Paola Patricia AC, Eugenia AR, Elisa Clementina OM, Jose CO, et al. Method Based on Data Mining Techniques for Breast Cancer Recurrence Analysis. In: Tan Y, Shi Y, Tuba M, editors. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer; 2020. p. 584–96.
16. Oskouei RJ, Kor NM, Maleki SA. Data mining and medical world: breast cancers' diagnosis, treatment, prognosis and challenges. Am J Cancer Res. 2017;7(3):610-27.
17. Clough E, Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Res. 2024;52(D1):D138–44. https://doi.org/10.1093/nar/gkad965
18. Jia L, Ling Y, Li K, Zhang L, Wang Y, Kang H. A 10-gene signature for predicting the response to neoadjuvant trastuzumab therapy in her2-positive breast cancer. Clin Breast Cancer. 2021;21(6):e654-e64. https://doi.org/10.1016/j.clbc.2021.04.010
19. Liu S, Hu X, Fan X, Jin R, Yang W, Geng Y, et al. A Bioinformatics Research on Novel Mechanism of Compound Kushen Injection for Treating Breast Cancer by Network Pharmacology and Molecular Docking Verification. Evid Based Complement Alternat Med. 2020;2020(1):2758640. https://doi.org/10.1155/2020/2758640
20. Yang Q, Yang G, Wu Y, Zhang L, Song Z, Yang D. Bioinformatics analysis and validation of genes related to paclitaxel's anti-breast cancer effect through immunogenic cell death. Heliyon. 2024;10(7):e28409. https://doi.org/10.1016/j.heliyon.2024.e28409
21. Mirzaei M, Sheikholeslami SA, Jalili A, Bereimipour A, Sharbati S, Kaveh V, et al. Investigating the molecular mechanisms of Tamoxifen on the EMT pathway among patients with breast cancer. J Med Life. 2022;15(6):835–44. https://doi.org/10.25122/jml-2022-0085
22. Ali Z, Hermawan A. Bioinformatics Analysis Uncovers the Importance of RTK-RAS-PI3K/Akt Regulation by Borneol in Overcoming Breast Cancer Resistance to Tamoxifen. Indones J Pharm. 2022;33(1). https://doi.org/10.22146/ijp.2346
23. Hermawan A, Putri H. Identification of potential gene associated with berberine in overcoming tamoxifen resistance by functional network analysis. J Appl Pharm Sci. 2020;10(7):9–18. https://doi.org/10.7324/JAPS.2020.10702
24. Azzahra SNA, Hanif N, Hermawan A. MDM2 is a Potential Target Gene of Glycyrrhizic Acid for Circumventing Breast Cancer Resistance to Tamoxifen: Integrative Bioinformatics Analysis. Asian Pac J Cancer Prev. 2022;23(7):2341–50. https://doi.org/10.31557/APJCP.2022.23.7.2341
25. Xiao Y, Ma M, Li Y, Xiao Y. Polygenic Regulation by Flos

Daturae in the Treatment of Breast Cancer: A Study based on Network Pharmacology and Bioinformatics. Lett Drug Des Discov. 2023;20(6):649–61. https://doi.org/10.2174/1570180820666230214104234
26. Zhuo Z, Zhang D, Lu W, Wu X, Cui Y, Zhang W, et al. Reversal of tamoxifen resistance by artemisinin in er+ breast cancer: Bioinformatics analysis and experimental validation. Oncol Res. 2024;32(6):1093-107. https://doi.org/10.32604/or.2024.047257.
27. Qiao Z, Xing Y, Zhang Q, Tang Y, Feng R, Pang W. Tamoxifen resistance-related ceRNA network for breast cancer. Front Cell Dev Biol. 2022;10. https://doi.org/10.3389/fcell.2022.1023079
28. Chen Z, Lu P, Li M, Zhang Q, He T, Gan L. Curcumin suppresses metastasis of triple-negative breast cancer cells by modulating emt signaling pathways: An integrated study of bioinformatics analysis. Medicine (Baltimore). 2024;103(8):e37264. https://doi.org/10.1097/md.0000000000037264
29. Huang P, Zhou P, Liang Y, Wu J, Wu G, Xu R, et al. Exploring the molecular targets and mechanisms of [10]-gingerol for treating triple-negative breast cancer using bioinformatics approaches, molecular docking, and in vivo experiments. Transl Cancer Res. 2021;10(11):4680-93. https://doi.org/10.21037/tcr-21-1138.
30. Chou CW, Huang YM, Chang YJ, Huang CY, Hung CS. Identified the novel resistant biomarkers for taxane-based therapy for triple-negative breast cancer. Int J Med Sci. 2021;18(12):2521-31. https://doi.org/10.7150/ijms.59177.
31. Zhang Y, Ma X, Li H, Zhuang J, Feng F, Liu L, et al. Identifying the effect of ursolic acid against triple-negative breast cancer: Coupling network pharmacology with experiments verification. Front Pharmacol. 2021;12:685773. https://doi.org/10.3389/fphar.2021.685773.
32. Wei S, Zhang Y, Ma X, Yao Y, Zhou Q, Zhang W, et al. Mat as a promising therapeutic strategy against triple-negative breast cancer via inhibiting pi3k/akt pathway. Scientific Reports. 2023;13(1):12351. https://doi.org/10.1038/s41598-023-39655-9
33. Zhao M, Fu L, Xu P, Wang T, Li P. Network pharmacology and experimental validation to explore the effect and mechanism of kanglaite injection against triple-negative breast cancer. Drug Des Devel Ther. 2023;17:901-17. https://doi.org/10.2147/dddt.S397969.
34. Yang B, Wang N, Wang S, Li X, Zheng Y, Li M, et al. Network-pharmacology-based identification of caveolin-1 as a key target of oldenlandia diffusa to suppress breast cancer metastasis. Biomed Pharmacother. 2019;112:108607. https://doi.org/10.1016/j.biopha.2019.108607
35. Han H, Qian C, Zong G, Liu H, Wang F, Tao R, et al. Systemic pharmacological verification of salvia miltiorrhiza-ginseng chinese herb pair in inhibiting spontaneous breast cancer metastasis. Biomed Pharmacother. 2022;156:113897. https://doi.org/10.1016/j.biopha.2022.113897
36. Ge X, Zhang D, Song S, Mi Y, Shen Y, Jiang Q, et al. Usp18 reduces paclitaxol sensitivity of triple-negative breast cancer via autophagy. Biochem Biophys Res Commun. 2022;599:120-6. https://doi.org/10.1016/j.bbrc.2022.02.048
37. Daina A, Michielin O, Zoete V. Swisstargetprediction: Updated data and new features for efficient prediction of protein targets of small molecules. Nucleic Acids Res. 2019;47(W1):W357-w64. https://doi.org/10.1093/nar/gkz382
38. Lu Y, Bi J, Li F, Wang G, Zhu J, Jin J, et al. Differential gene analysis of trastuzumab in breast cancer based on network pharmacology and medical images. Front Physiol. 2022;13:942049. https://doi.org/10.3389/fphys.2022.942049
39. Chen L, Ren LQ, Liu Z, Liu X, Tu H, Huang XY. Bioinformatics and in vitro experiments reveal the mechanism of schisandrin a against mda-mb-231 cells. Bioengineered. 2021;12(1):7678-93. https://doi.org/10.1080/21655979.2021.1982307
40. Mariana Kustiawan P, Siregar K, Syaifie PH, Zein Muttaqin F, Ibadillah D, Miftah Jauhar M, et al. Uncovering the anti-breast cancer activity potential of east kalimantan propolis by in vitro and bioinformatics analysis. Heliyon. 2024;10(13):e33636. https://doi.org/10.1016/j.heliyon.2024.e33636
41. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinformatics. 2016;54(1):1.30.1-1.30.33. https://doi.org/10.1002/cpbi.5
42. Li F, Shi Y, Yang X, Luo Z, Zhang G, Yu K, et al. Anhydroicaritin inhibits emt in breast cancer by enhancing gpx1 expression: A research based on sequencing technologies and bioinformatics analysis. Front Cell Dev Biol. 2021;9:764481. https://doi.org/10.3389/fcell.2021.764481
43. Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, Chen X, et al. New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLoS Comput Biol. 2019;15(3):e1006701. https://doi.org/10.1371/journal.pcbi.1006701
44. Lawal B, Wu AT, Chen CH, T AG, Wu SY. Identification of infg/stat1/notch3 as γ-mangostin's potential targets for overcoming doxorubicin resistance and reducing cancer-associated fibroblasts in triple-negative breast cancer. Biomed Pharmacother. 2023;163:114800. https://doi.org/10.1016/j.biopha.2023.114800
45. Yu H, Hu K, Zhang T, Ren H. Identification of target genes related to sulfasalazine in triple-negative breast cancer through network pharmacology. Med Sci Monit. 2020;26:e926550. https://doi.org/10.12659/msm.926550
46. Ye Y, Zhang B, Liang Q, Wang D, Bai F, Li Y, et al. Exploring the pharmacological mechanism of compound kushen injection in the treatment of breast cancer using in vitro experiments: Coupling network pharmacology with geo database. Front Oncol. 2022;12:946758. https://doi.org/10.3389/fonc.2022.946758
47. Kim S. Exploring chemical information in pubchem. Curr Protoc. 2021;1(8):e217. https://doi.org/10.1002/cpz1.217
48. Putra IMR, Lestari IA, Fatimah N, Hanif N, Ujiantari NSO, Putri DDP, et al. Bioinformatics and in vitro study reveal erα as the potential target gene of honokiol to enhance trastuzumab sensitivity in her2+ trastuzumab-resistant breast cancer cells. Comput Biol Chem. 2024;111:108084. https://doi.org/10.1016/j.compbiolchem.2024.108084
49. Szklarczyk D, Santos A, Von Mering C, Jensen LJ, Bork P, Kuhn M. STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic Acids Res. 2016;44(D1):D380–4. https://doi.org/10.1093/nar/gkv1277
50. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 2020;48(D1):D845–55. https://doi.org/10.1093/nar/gkz1021
51. 1. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online mendelian inheritance in man (omim), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005;33(Database issue):D514-7. https://doi.org/10.1093/nar/gki033
52. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, et al. DrugBank: a knowledgebase for drugs,

drug actions and drug targets. Nucleic Acids Res. 2008;36(suppl_1):D901–6. https://doi.org/10.1093/nar/gkm958
53. Liu X, Ouyang S, Yu B, Liu Y, Huang K, Gong J, et al. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. Nucleic Acids Res. 2010;38(suppl_2):W609–14. https://doi.org/10.1093/nar/gkq300
54. Wang Z, Liang L, Yin Z, Lin J. Improving chemical similarity ensemble approach in target prediction. J Cheminform. 2016;8:20. https://doi.org/10.1186/s13321-016-0130-x
55. Chen X, Ji ZL, Chen YZ. TTD: Therapeutic Target Database. Nucleic Acids Res. 2002;30(1):412–5. https://doi.org/10.1093/nar/30.1.412
56. Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. Bindingdb: A web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res. 2007;35(Database issue):D198-201. https://doi.org/10.1093/nar/gkl999
57. Yao ZJ, Dong J, Che YJ, Zhu MF, Wen M, Wang NN, et al. TargetNet: a web service for predicting potential drug–target interaction profiling via multi-target SAR models. J Comput Aided Mol Des. 2016;30(5):413–24. https://doi.org/10.1007/s10822-016-9915-2
58. Coker EA, Mitsopoulos C, Tym JE, Komianou A, Kannas C, Di Micco P, et al. Cansar: Update to the cancer translational research and drug discovery knowledgebase. Nucleic Acids Res. 2019;47(D1):D917-d22. https://doi.org/10.1093/nar/gky1129
59. Sangkuhl K, Berlin DS, Altman RB, Klein TE. Pharmgkb: Understanding the effects of individual genetic variants. Drug Metab Rev. 2008;40(4):539-51. https://doi.org/10.1080/03602530802413338
60. Davis AP, Grondin CJ, Johnson RJ, Sciaky D, King BL, McMorran R, et al. The comparative toxicogenomics database: Update 2017. Nucleic Acids Res. 2017;45(D1):D972-d8. https://doi.org/10.1093/nar/gkw838
61. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the xena platform. Nat Biotechnol. 2020;38(6):675-8. https://doi.org/10.1038/s41587-020-0546-8
62. Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, et al. Ualcan: A portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia. 2017;19(8):649-58. https://doi.org/10.1016/j.neo.2017.05.002
63. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 2023;51(D1):D638–46. https://doi.org/10.1093/nar/gkac1000
64. Vivek-Ananth RP, Mohanraj K, Sahoo AK, Samal A. Imppat 2.0: An enhanced and expanded phytochemical atlas of indian medicinal plants. ACS Omega. 2023;8(9):8827-45. https://doi.org/10.1021/acsomega.3c00156
65. Gao K, Liu L, Lei S, Li Z, Huo P, Wang Z, et al. Herb 2.0: An updated database integrating clinical and experimental evidence for traditional chinese medicine. Nucleic Acids Res. 2025;53(D1):D1404-d14. https://doi.org/10.1093/nar/gkae1037
66. Daina A, Michielin O, Zoete V. Swissadme: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 2017;7:42717. https://doi.org/10.1038/srep42717
67. Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, et al. The ChEMBL bioactivity database: an update. Nucleic Acids Res. 2014;42(D1):D1083–90. https://doi.org/10.1093/nar/gkt1031
68. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, 3rd, et al. Next-generation characterization of the cancer cell line encyclopedia. Nature. 2019;569(7757):503-8. https://doi.org/10.1038/s41586-019-1186-3
69. Monga M, Sausville EA. Developmental therapeutics program at the nci: Molecular target and drug discovery process. Leukemia. 2002;16(4):520-6. https://doi.org/10.1038/sj.leu.2402464.
70. Knox C, Wilson M, Klinger CM, Franklin M, Oler E, Wilson A, et al. DrugBank 6.0: the DrugBank Knowledgebase for 2024. Nucleic Acids Res. 2024;52(D1):D1265–75. https://doi.org/10.1093/nar/gkad976
71. Consortium TU. UniProt: the Universal Protein Knowledgebase in 2025. Nucleic Acids Res. 2024;53(D1):1010. https://doi.org/10.1093/nar/gkae1010
72. Afendi F, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, et al. KNApSAcK Family Databases: Integrated Metabolite-Plant Species Databases for Multifaceted Plant Research. Plant Cell Physiol. 2011;53(2):e1. https://doi.org/10.1093/pcp/pcr165
73. Dunkel M, Günther S, Ahmed J, Wittig B, Preissner R. Superpred: Drug classification and target prediction. Nucleic Acids Res. 2008;36(Web Server issue):W55-9. https://doi.org/10.1093/nar/gkn307
74. Espe S. MalaCards: the human disease database. J Med Libr Assoc. 2018;106(1):140.
75. Yan D, Zheng G, Wang C, Chen Z, Mao T, Gao J, et al. HIT 2.0: an enhanced platform for Herbal Ingredients' Targets. Nucleic Acids Res. 2022;50(D1):D1238–43. https://doi.org/10.1093/nar/gkab1011
76. Gong J, Cai C, Liu X, Ku X, Jiang H, Gao D, et al. ChemMapper: a versatile web server for exploring pharmacology and chemical structure association based on molecular 3D similarity method. Bioinformatics. 2013;29(14):1827–9. https://doi.org/10.1093/bioinformatics/btt270
77. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50. https://doi.org/10.1073/pnas.0506580102
78. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, et al. ArrayExpress update – from bulk to single-cell expression data. Nucleic Acids Res. 2019;47(D1):D711–5. https://doi.org/10.1093/nar/gky964
79. Hosen SMZ, Junaid M, Alam MS, Rubayed M, Dash R, Akter R, et al. Greenmolbd: Nature derived bioactive molecules' database. Med Chem. 2022;18(6):724-33. https://doi.org/10.2174/1573406418666211129103458
80. Freeberg MA, Fromont LA, D'Altri T, Romero AF, Ciges JI, Jene A, et al. The European Genome-phenome Archive in 2021. Nucleic Acids Res. 2022;50(D1):D980–7. https://doi.org/10.1093/nar/gkab1059
81. Wu Y, Zhang F, Yang K, Fang S, Bu D, Li H, et al. SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. Nucleic Acids Res. 2019;47(D1):D1110–7. https://doi.org/10.1093/nar/gky1021
82. Yanqiong Z, Xin L, Yulong S, Tong C, Zhijian X, Ping W, et al. Etcm v2.0: An update with comprehensive resource and rich annotations for traditional chinese medicine. Acta Pharmaceutica Sinica B. 2023;13(6):2559-71. https://doi.org/10.1016/j.apsb.2023.03.012

83. Lv C, Wu X, Wang X, Su J, Zeng H, Zhao J, et al. The gene expression profiles in response to 102 traditional chinese medicine (tcm) components: A general template for research on tcms. Sci Rep. 2017;7(1):352. https://doi.org/10.1038/s41598-017-00535-8
84. Kong X, Liu C, Zhang Z, Cheng M, Mei Z, Li X, et al. BATMAN-TCM 2.0: an enhanced integrative database for known and predicted interactions between traditional Chinese medicine ingredients and target proteins. Nucleic Acids Res. 2024;52(D1):D1110–20. https://doi.org/10.1093/nar/gkad926
85. Zhu F, Xu D. Predicting gene signature in breast cancer patients with multiple machine learning models. Discov Oncol. 2024;15(1):516. https://doi.org/10.1007/s12672-024-01386-2.
86. Roncaglia P, Martone ME, Hill DP, Berardini TZ, Foulger RE, Imam FT, et al. The Gene Ontology (GO) Cellular Component Ontology: Integration with SAO (Subcellular Anatomy Ontology) and other recent developments. J Biomed Semantics. 2013;4(1):20. https://doi.org/10.1186/2041-1480-4-20.
87. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. Kegg: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017;45(D1):D353-d61. https://doi.org/10.1093/nar/gkw1092
88. Ibadurrahman W, Hanif N, Hermawan A. Functional network analysis of p85 and pi3k as potential gene targets and mechanism of oleanolic acid in overcoming breast cancer resistance to tamoxifen. J Genet Eng Biotechnol. 2022;20(1):66. https://doi.org/10.1186/s43141-022-00341-4.
89. Wibowo MA, Nugroho EP, Hermawan A. Genomic Understanding Reveals the Important Role of FGFR2 as Paeoniflorin Target for Circumventing Breast Cancer Resistance to Tamoxifen. Asian Pac J Cancer Prev. 2021;22(12):3949–58. https://doi.org/10.31557/APJCP.2021.22.12.3949
90. Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022;50(W1):W216–21. https://doi.org/10.1093/nar/gkac194
91. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. Nucleic Acids Res. 2019;47(W1):W199–205. https://doi.org/10.1093/nar/gkz401
92. Ge SX, Jung D, Yao R. Shinygo: A graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36(8):2628-9. https://doi.org/10.1093/bioinformatics/btz931
93. Hermawan A, Putri H, Ikawati M. Bioinformatic analysis reveals the molecular targets of tangeretin in overcoming the resistance of breast cancer to tamoxifen. Gene Rep. 2020;21:100884. https://doi.org/10.1016/j.genrep.2020.100884
94. Hermawan A, Putri H, Hanif N, Ikawati M. Integrative Bioinformatics Study of Tangeretin Potential Targets for Preventing Metastatic Breast Cancer. Evid Based Complement Alternat Med. 2021;2021:1–15. https://doi.org/10.1155/2021/2234554
95. Lestari IA, Putra IMR, Fatimah N, Ujiantari NSO, Putri DDP, Hermawan A. Characterization of Potential Target Genes of Borneol in Increasing Trastuzumab Sensitivity in HER2+ Trastuzumab-Resistant Breast Cancer: Bioinformatics and In Vitro Studies. Asian Pac J Cancer Prev. 2024;25(5):1623–34. https://doi.org/10.31557/APJCP.2024.25.5.1623
96. Hermawan A, Putri H, Utomo RY. Comprehensive bioinformatics study reveals targets and molecular mechanism of hesperetin in overcoming breast cancer chemoresistance. Mol Divers. 2020;24(4):933–47. https://doi.org/10.1007/s11030-019-10003-2
97. Hermawan A, Putri H. Targets and molecular mechanisms of a citrus flavonoid, hesperidin, against luminal breast cancer cells: an integrative bioinformatics analysis. Asian Pac J Trop Biomed. 2019;9(12):531. https://doi.org/10.4103/2221-1691.271727
98. Hermawan A, Ikawati M, Jenie RI, Khumaira A, Putri H, Nurhayati IP, et al. Identification of potential therapeutic target of naringenin in breast cancer stem cells inhibition by bioinformatics and in vitro studies. Saudi Pharm J. 2021;29(1):12–26. https://doi.org/10.1016/j.jsps.2020.12.002
99. Hermawan A, Ikawati M, Putri DDP, Fatimah N, Prasetio HH. Nobiletin Inhibits Breast Cancer Stem Cell by Regulating the Cell Cycle: A Comprehensive Bioinformatics Analysis and In Vitro Experiments. Nutr Cancer. 2024;76(7):638–55. https://doi.org/10.1080/01635581.2024.2348217
100. Adrian, Lubis MF, Syahputra RA, Astyka R, Sumaiyah S, Yudha Harahap MA, et al. The Potential Effect of Aporphine Alkaloids from Nelumbo Nucifera Gaertn. As Anti-Breast Cancer Based on Network Pharmacology and Molecular Docking. Int J App Pharm. 2024;280–7. https://doi.org/10.22159/ijap.2024v16i1.49171
101. Joseph N, Festus MT, Fred W, Esther NM. Effects of tea catechin extracts from bb35 and purple (trfk 306) tea clones on the gene expression of egfr, app, bcl2, dnmt, casp3, hif1a, gadd45b and psmb5 genes involved in triple negative breast cancer diseases: In silico and in vitro study. Inform Med Unlocked. 2024;46:101469. https://doi.org/10.1016/j.imu.2024.101469
102. Consortium GO. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 2004;32(suppl_1):D258–61. https://doi.org/10.1093/nar/gkh036
103. Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics. 2015;31(4):608–9. https://doi.org/10.1093/bioinformatics/btu684
104. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587–92. https://doi.org/10.1093/nar/gkac963
105. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523. https://doi.org/10.1038/s41467-019-09234-6
106. Lo Surdo P, Iannuccelli M, Contino S, Castagnoli L, Licata L, Cesareni G, et al. Signor 3.0, the signaling network open resource 3.0: 2022 update. Nucleic Acids Res. 2023;51(D1):D631-d7. https://doi.org/10.1093/nar/gkac883
107. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set knowledge discovery with enrichr. Curr Protoc. 2021;1(3):e90. https://doi.org/10.1002/cpz1.90.
108. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. Cluego: A cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25(8):1091-3. https://doi.org/10.1093/bioinformatics/btp101
109. Bindea G, Galon J, Mlecnik B. Cluepedia cytoscape plugin: Pathway insights using integrated experimental and in silico data. Bioinformatics. 2013;29(5):661-3. https://doi.org/10.1093/bioinformatics/btt019
110. Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, et

al. Srplot: A free online platform for data visualization and graphing. PLoS One. 2023;18(11):e0294236. https://doi.org/10.1371/journal.pone.0294236
111. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the panther classification system. Nat Protoc. 2013;8(8):1551-66. https://doi.org/10.1038/nprot.2013.092
112. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, et al. The harmonizome: A collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford). 2016;2016. https://doi.org/10.1093/database/baw100.
113. Hasan M, Kumar N, Majeed A, Ahmad A, Mukhtar S. Protein-protein interaction network analysis using networkx. Methods Mol Biol. 2023;2690:457-67. https://doi.org/10.1007/978-1-0716-3327-4_35.
114. Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi AA. Protein-protein interaction networks (ppi) and complex diseases. Gastroenterol Hepatol Bed Bench. 2014;7(1):17-31.
115. Kim M, Park J, Bouhaddou M, Kim K, Rojc A, Modak M, et al. A protein interaction landscape of breast cancer. Science. 2021;374(6563):eabf3066. https://doi.org/10.1126/science.abf3066
116. Doncheva NT, Morris JH, Holze H, Kirsch R, Nastou KC, Cuesta-Astroz Y, et al. Cytoscape stringapp 2.0: Analysis and visualization of heterogeneous biological networks. J Proteome Res. 2023;22(2):637-46. https://doi.org/10.1021/acs.jproteome.2c00651.
117. Hermawan A, Ikawati M, Khumaira A, Putri H, Jenie RI, Angraini SM, et al. Bioinformatics and in vitro studies reveal the importance of p53, pparg and notch signaling pathway in inhibition of breast cancer stem cells by hesperetin. Adv Pharm Bull. 2021;11(2):351-60. https://doi.org/10.34172/apb.2021.033.
118. Sheybatzadeh K, Moshtaghie SAA, Shahanipour K, Golab F. Integrative bioinformatics analysis reveals potential target genes and pten signaling in breast cancer and effect of zingiber officinale (ginger) and allium sativum (garlic) extract on it. Asian Pac J Cancer Prev. 2024;25(3):893-908. https://doi.org/10.31557/apjcp.2024.25.3.893
119. Zheng L, Jiang H, Li R, Song L, Chen R, Dong H. The pharmacological mechanisms of xiaochaihutang in treating breast cancer based on network pharmacology. Contrast Media Mol Imaging. 2022;2022:3900636. https://doi.org/10.1155/2022/3900636
120. Zhang Z, Zhang Z, Song J, Wu W, Chen Y, Li J, et al. Prognostic model construction and target identification of si-wu-tang against breast cancer. Heliyon. 2023;9(10):e20709. https://doi.org/10.1016/j.heliyon.2023.e20709.
121. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The genemania prediction server: Biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38(Web Server issue):W214-20. https://doi.org/10.1093/nar/gkq537.
122. Unberath P, Mahlmeister L, Reimer N, Busch H, Boerries M, Christoph J. Searching of clinical trials made easier in cbioportal using patients' genetic and clinical profiles. Appl Clin Inform. 2022;13(2):363-9. https://doi.org/10.1055/s-0042-1743560.
123. Hermawan A, Putri H. Integrative bioinformatics analysis reveals potential target genes and tnfα signaling inhibition by brazilin in metastatic breast cancer cells. Asian Pac J Cancer Prev. 2020;21(9):2751-62. https://doi.org/10.31557/apjcp.2020.21.9.2751
124. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, et al. Timer: A web server for comprehensive analysis of tumor-infiltrating immune cells. Cancer Res. 2017;77(21):e108-e10. https://doi.org/10.1158/0008-5472.Can-17-0307
125. 1. Ma Y, Peng Z, Pan R, Zhu Z, Meng X, Hu H, et al. The bioinformatics analysis of quercetin in octagonal lotus for the screening of breast cancer myc, cxcl10, cxcl11, and e2f1. Int J Immunopathol Pharmacol. 2021;35:20587384211040903. https://doi.org/10.1177/20587384211040903
126. Allison KH. Prognostic and predictive parameters in breast pathology: A pathologist's primer. Mod Pathol. 2021;34(Suppl 1):94-106. https://doi.org/10.1038/s41379-020-00704-7
127. Lánczky A, Győrffy B. Web-based survival analysis tool tailored for medical research (kmplot): Development and implementation. J Med Internet Res. 2021;23(7):e27633. https://doi.org/10.2196/27633
128. Chen G. Molecular basis of breast cancer with comorbid depression and the mechanistic insights of xiaoyaosan in treating breast cancer-associated depression. Medicine (Baltimore). 2023;102(38):e35157. https://doi.org/10.1097/md.0000000000035157
129. Zhang L, Yang K, Wang M, Zeng L, Sun E, Zhang F, et al. Exploring the mechanism of cremastra appendiculata (suanpanqi) against breast cancer by network pharmacology and molecular docking. Comput Biol Chem. 2021;94:107396. https://doi.org/10.1016/j.compbiolchem.2020.107396
130. Jézéquel P, Gouraud W, Ben Azzouz F, Guérin-Charbonnel C, Juin PP, Lasla H, et al. Bc-genexminer 4.5: New mining module computes breast cancer differential gene expression analyses. Database (Oxford). 2021;2021. https://doi.org/10.1093/database/baab007.
131. Tang Z, Kang B, Li C, Chen T, Zhang Z. Gepia2: An enhanced web server for large-scale expression profiling and interactive analysis. Nucleic Acids Res. 2019;47(W1):W556-w60. https://doi.org/10.1093/nar/gkz430
132. C Contreras-Puentes N, Alviz-Amador A, Zabaleta-Guzman JA, Pineda Aleman R, Tarón Dunoyer A. Identification of genes hub associated with triple-negative breast cancer and cannabidiol analogs potential inhibitory agents: An in-silico study. Asian Pac J Cancer Prev. 2024;25(5):1649-61. https://doi.org/10.31557/apjcp.2024.25.5.1649
133. Wang F, Yuan C, Liu B, Yang YF, Wu HZ. Syringin exerts anti-breast cancer effects through pi3k-akt and egfr-ras-raf pathways. J Transl Med. 2022;20(1):310. https://doi.org/10.1186/s12967-022-03504-6
134. Anaya J. OncoLnc: linking TCGA survival data to mRNAs, miRNAs, and lncRNAs. PeerJ Computer Science. 2016;2:e67.
135. Endah E, Wulandari F, Putri Y, Jenie RI, Meiyanto E. Piperine increases pentagamavunon-1 anti-cancer activity on 4t1 breast cancer through mitotic catastrophe mechanism and senescence with sharing targeting on mitotic regulatory proteins. Iran J Pharm Res. 2022;21(1):e123820. https://doi.org/10.5812/ijpr.123820
136. Li D, Fan H, Dong J, Sun C, Su Y, Liu J, et al. Based on batman-tcm to explore the molecular mechanism of xihuang pill regulating immune function to treat breast precancerous lesions. Breast Cancer (Dove Med Press). 2021;13:725-42. https://doi.org/10.2147/bctt.S339607
137. Novitasari D, Muntafiah L, Sari NF, Meiyanto E, Hermawan A. Ethanolic extract of sappan wood (Caesalpinia sappan L.) inhibits MCF-7 and MCF-7/HER2 mammospheres' formation: An in vitro and bioinformatic study. Indones J Biotechnol. 2021;26(3):133–41.
138. Hermawan A, Putri H, Utomo R. Exploration of targets and molecular mechanisms of cinnamaldehyde in overcoming fulvestrant-resistant breast cancer: A bioinformatics study.

Netw Model Anal Health Inform Bioinform. 2021;10. https://doi.org/10.1007/s13721-021-00303-9.
139. Luo L, Chen Y, Ma Q, Huang Y, Hong T, Shu K, et al. Exploring the mechanism of an active ingredient of ginger, dihydrocapsaicin, on triple negative breast cancer based on network pharmacology and in vitro experiments. Oncol Lett. 2023;25(5):195. https://doi.org/10.3892/ol.2023.13781
140. Hermawan A, Khumaira A, Ikawati M, Putri H, Jenie RI, Angraini SM, et al. Identification of key genes of hesperidin in inhibition of breast cancer stem cells by functional network analysis. Comput Biol Chem. 2021;90:107427. https://doi.org/10.1016/j.compbiolchem.2020.107427.
141. Manivannan HP, Veeraraghavan VP, Francis AP. Identification of molecular targets of trigonelline for treating breast cancer through network pharmacology and bioinformatics-based prediction. Mol Divers. 2024;28(6):3835-57. https://doi.org/10.1007/s11030-023-10780-x
142. Huang B, Wen G, Li R, Wu M, Zou Z. Integrated network pharmacology, bioinformatics, and molecular docking to explore the mechanisms of berberine regulating autophagy in breast cancer. Medicine (Baltimore). 2023;102(36):e35070. https://doi.org/10.1097/md.0000000000035070.
143. Zeng C, Lin M, Jin Y, Zhang J. Identification of key genes associated with brain metastasis from breast cancer: A bioinformatics analysis. Med Sci Monit. 2022;28:e935071. https://doi.org/10.12659/msm.935071.
144. Qian P, Mu XT, Su B, Gao L, Zhang DF. Identification of the anti-breast cancer targets of triterpenoids in liquidambaris fructus and the hints for its traditional applications. BMC Complement Med Ther. 2020;20(1):369. https://doi.org/10.1186/s12906-020-03143-8
145. Xu C, Sun L, Wang H, Sun J, Feng Y, Wang X, et al. Identifying the mechanism of polysaccharopeptide against breast cancer based on network pharmacology and experimental verification. BMC Cancer. 2024;24(1):726. https://doi.org/10.1186/s12885-024-12494-1.
146. Li H, Tang X, Sun Z, Qu Z, Zou X. Integrating bioinformatics and experimental models to investigate the mechanism of the chelidonine-induced mitotic catastrophe via the akt/foxo3/foxm1 axis in breast cancer cells. Biomol Biomed. 2024;24(3):560-74. https://doi.org/10.17305/bb.2023.9665
147. Srinivasan M, Gangurde A, Chandane AY, Tagalpallewar A, Pawar A, Baheti AM. Integrating network pharmacology and in silico analysis deciphers withaferin-a's anti-breast cancer potential via hedgehog pathway and target network interplay. Brief Bioinform. 2024;25(2). https://doi.org/10.1093/bib/bbae032
148. Hermawan A, Putri H. Integrative bioinformatics analysis reveals possible target and mechanism of ellipticine against breast cancer stem cells. Indones J Pharm. 2021;32. https://doi.org/10.22146/ijp.1126.
149. Xie YI, Tang C, Qin JP, Gu HQ, Wang ZW, Liu Q. Molecular docking technology and network pharmacology based on rhapontici radix-cremastrae pseudobulbus drug pair in treating breast cancer. Eur Rev Med Pharmacol Sci. 2023;27(21):10204-12. https://doi.org/10.26355/eurrev_202311_34295.
150. Chen L, Yang A, Li Y, Liu X, Jiang W, Hu K. Molecular mechanism of oroxyli semen against triple-negative breast cancer verified by bioinformatics and in vitro experiments. Medicine (United States). 2023;102(37):E34835. https://doi.org/10.1097/MD.0000000000034835
151. Long S, Yuan C, Wang Y, Zhang J, Li G. Network Pharmacology Analysis of Damnacanthus indicus C.F.Gaertn in Gene-Phenotype. Evid Based Complement Alternat Med. 2019;2019:1368371. https://doi.org/10.1155/2019/1368371
152. Sheng J, Cheng J, Chu W, Dong M, Jiang K. Network pharmacology and experimental validation to reveal the pharmacological mechanisms of Qizhu prescription for treating breast cancer. J Tradit Chin Med Sci. 2024;11(3):303–15. https://doi.org/10.1016/j.jtcms.2024.06.006
153. Bhagya N. Network pharmacology based investigation on the mechanism of tetrandrine against breast cancer. Phytomedicine Plus. 2023 Feb 1;3(1):100381.
Available from: https://www.sciencedirect.com/science/article/pii/S2667031322001609
154. Lakshmi SG, Kamaraj M, Nithya TG, Chidambaranathan N, Pushpalatha GGL, Santhosh P, et al. Network pharmacology integrated with molecular docking reveals the anticancer mechanism of Jasminum sambac Linn. essential oil against human breast cancer and experimental validation by in vitro and in vivo studies. Appl Biochem Biotechnol. 2023;196(1):350–81. https://doi.org/10.1007/s12010-023-04481-2
155. Alshehri FF, Alshehri ZS. Network pharmacology-based screening of active constituents of Avicennia marina and their clinical biochemistry related mechanism against breast cancer. J Biomol Struct Dyn. 2023;42(9):4506–21. https://doi.org/10.1080/07391102.2023.2220801
156. Li Y, Li R, Zeng Z, Li S, Luo S, Wu J, et al. Prediction of the mechanisms of Xiaoai Jiedu Recipe in the treatment of breast cancer: A comprehensive approach study with experimental validation. J Ethnopharmacol. 2020;252:112603. https://doi.org/10.1016/j.jep.2020.112603
157. Hermansyah D, Putra A, Munir D, Lelo A, Amalina ND, Alif I. Synergistic Effect of Curcuma longa Extract in Combination with Phyllanthus niruri Extract in Regulating Annexin A2, Epidermal Growth Factor Receptor, Matrix Metalloproteinases, and Pyruvate Kinase M1/2 Signaling Pathway on Breast Cancer Stem Cell. Open Access Maced J Med Sci. 2021;9(A):271–85. https://doi.org/10.3889/oamjms.2021.5941
158. Karabi Z, Moradian F, Kheirabadi M. The effect of lactoferrin on ulk1 and atg13 genes expression in breast cancer cell line mcf7 and bioinformatics studies of protein interaction between lactoferrin and the autophagy initiation complex. Cell Biochem Biophys. 2022;80(4):795-806. https://doi.org/10.1007/s12013-022-01097-x
159. Sun M, Lv F, Qin C, Du D, Li W, Liu S. The potential mechanism of liujunzi decoction in the treatment of breast cancer based on network pharmacology and molecular docking technology. Curr Pharm Des. 2024;30(9):702-26. https://doi.org/10.2174/0113816128289900240219104854
160. Basavarajappa GM, Rehman A, Shiroorkar PN, Sreeharsha N, Anwer MK, Aloufi B. Therapeutic effects of crataegus monogyna inhibitors against breast cancer. Front Pharmacol. 2023;14:1187079. https://doi.org/10.3389/fphar.2023.1187079
161. Poustforoosh A, Faramarz S, Negahdaripour M, Tüzün B, Hashemipour H. Tracing the pathways and mechanisms involved in the anti-breast cancer activity of glycyrrhizin using bioinformatics tools and computational methods. J Biomol Struct Dyn. 2024;42(2):819-33. https://doi.org/10.1080/07391102.2023.2196347.
162. Lee JH, Kim J, Kim HS, Kang YJ. Unraveling connective tissue growth factor as a therapeutic target and assessing kahweol as a potential drug candidate in triple-negative breast cancer treatment. Int J Mol Sci. 2023;24(22):16307. https://doi.org/10.3390/ijms242216307