Identifying Lymph Node Metastasis-Related Factors in Breast Cancer Using Differential Modular and Mutational Structural Analysis

Xingyi Liu1 · Bin Yang1 · Xinpeng Huang1 · Wenying Yan1,2 · Yujuan Zhang3 · Guang Hu1,2

Received: 25 January 2023 / Revised: 13 April 2023 / Accepted: 13 April 2023 / Published online: 28 April 2023
© International Association of Scientists in the Interdisciplinary Areas 2023

Abstract
Complex diseases are generally caused by disorders of biological networks and/or mutations in multiple genes. Comparisons of network topologies between different disease states can highlight key factors in their dynamic processes. Here, we propose a differential modular analysis approach that integrates protein–protein interactions with gene expression profiles for modular analysis, and introduces inter-modular edges and date hubs to identify the “core network module” that quantifies the significant phenotypic variation. Then, based on this core network module, key factors, including functional protein–protein interactions, pathways, and driver mutations, are predicted by the topological–functional connection score and structural modeling. We applied this approach to analyze the lymph node metastasis (LNM) process in breast cancer. The functional enrichment analysis showed that both inter-modular edges and date hubs play important roles in cancer metastasis and invasion, and in metastasis hallmarks. The structural mutation analysis suggested that the LNM of breast cancer may be the outcome of the dysfunction of rearranged during transfection (RET) proto-oncogene-related interactions and the non-canonical calcium signaling pathway via an allosteric mutation of RET. We believe that the proposed method can provide new insights into disease progression such as cancer metastasis.

Xingyi Liu and Bin Yang contributed equally to this work.

Wenying Yan
wyyan@suda.edu.cn

Yujuan Zhang
zhangyujuan@suda.edu.cn

Guang Hu
huguang@suda.edu.cn

1 Center for Systems Biology, Department of Bioinformatics, School of Biology and Basic Medical Sciences, Soochow University, Suzhou 215123, Jiangsu, China
2 Jiangsu Province Engineering Research Center of Precision Diagnostics and Therapeutics Development, Suzhou 215123, Jiangsu, China
3 Experimental Center of Suzhou Medical College, Soochow University, Suzhou 215123, Jiangsu, China
Graphical Abstract

Keywords  Network module · Date hubs · Cancer dynamics · Allosteric mutation

Abbreviations

| Acronym | Description                              |
|---------|------------------------------------------|
| LNM     | Lymph node metastasis                    |
| RET     | Rearranged during transfection           |
| PPIN    | Protein–protein interaction network      |
| EGFR    | Epidermal growth factor receptor         |
| BRCA    | Breast cancer                            |
| DEGs    | Differentially expressed genes           |
| aPCC    | Average Pearson’s correlation coefficient |
| CPL     | Characteristic path length               |
| TFC     | Topological–functional connection        |
| TCGA    | The Cancer Genome Atlas                  |
| KEGG    | Kyoto Encyclopedia of Genes and Genomes  |
| GSEA    | Gene set enrichment analysis             |

COSMIC    Catalogue of Somatic Mutations in Cancer
PDB       Protein Data Bank
PRISM     Protein Interactions by Structural Matching

1 Introduction

Metastasis is the hallmark of cancer and is responsible for the greatest number of cancer-related deaths [1, 2], constituting the primary cause of death for >90% of patients with cancer. The activation of cancer metastasis has three distinguishing features: location-dependence, environmental interaction, and a dynamic selection process. For example,
the metastasis of breast cancer mainly occurs through the lymphatic system [3]. Therefore, understanding the biological process and mechanism of LNM in breast cancer will help guide the treatment of breast cancer and improve the prognosis of patients [4]. Although several studies have focused on the identification of disease markers in metastatic breast cancer from the perspective of genome-wide expression profiles [5, 6] and comparative analysis of PPI networks [7], the molecular understanding of LNM in breast cancer is still very poor.

Benefiting from the advances of network science and high-throughput biomedical technologies, studying biological systems from network biology has attracted much attention in recent years. Networks have long been central to our understanding of biological systems, in the form of linkage maps among genotypes, phenotypes, and the corresponding environmental factors [8]. With the tremendous increase in human protein interaction data, the protein–protein interaction network (PPIN) approach is commonly used to understand the molecular mechanisms of cancers [9]. However, PPINs merely provide a static snapshot of the molecular interactions within a tissue, whereas biological systems are highly dynamic. Thus, differential network analysis can be used to study the dynamic properties of networks related to cancer metastasis and to highlight network changes between conditions [10]. Differential network methods developed to date differ in the entities and measures that they compare [11]. Node-based methods focus on differences in node-related measures, such as node connectivity [12]. Interaction-based methods focus on differences in the context-specific weights associated with each interaction [13]. Furthermore, integrating co-expression data with differential network analysis can reveal the dynamic context of gene expression profiles [14, 15]. Such differential co-expression networks are useful tools to identify changes in response to an external perturbation, such as mutations predisposed to cancer progression, and to identify changes in the activity of gene expression regulators or signaling.

Three-dimensional protein structural data at the molecular level are pivotal for successful precision medicine. Such data are crucial not only for discovering drugs that act to block the active site of the target mutant protein but also for clarifying to the patient and the clinician how the mutations harbored by the patient work [16]. In oncological research, structure-based methods can be highly beneficial for addressing the diversity of cancer hallmarks [17]. Recently, a structure–function-based approach was shown to improve the prediction of drug sensitivity in epidermal growth factor receptor (EGFR)-mutant non-small cell lung cancer [18]. Actually, linking structural or mutation information with PPI network or biological pathway information is useful for predicting the genotype–phenotype relationship, providing insight into signaling mechanisms, helping to understand the mechanism of disease-related mutations, and helping in drug discovery [19]. Such an approach has been used for the detailed analysis of cancer and the cancer metastasis-related PPI binding interface [20, 21], and to delineate the mechanism of oncogenic mutations and single nucleotide polymorphism mutations in inflammation and cancer [22].

In this paper, based on the idea that network modules serve as a more robust indicator of cancer prognosis [23, 24], we propose a differential modular analysis approach to identify key network modules correlated with LNM in breast cancer (Fig. 1). First, weighted PPI networks for non-LNM and LNM were constructed by incorporating human interactome and gene expression data. Then, based on network modular analysis, “inter-modular edges” and “date hubs” [25] were introduced to detect the altered modularity of PPI networks, which may correspond to the key dynamic region in LNM. Finally, we evaluated the importance and potential application of the core network module by mutational structural analysis at both the edge and node levels. We hope that this study provides a novel perspective for the analysis of mutation effects to facilitate network-guided precision medicine.

2 Materials and Methods

2.1 Data Collection and Pre-processing

RNA-seq data on breast cancer (BRCA) from The Cancer Genome Atlas (TCGA) with clinical information was retrieved from UCSC XENA (https://xena.ucsc.edu/). According to the extent of the lymph node metastases (LNM) of the clinical information, BRCA patients were divided into N0 (non-LNM) and N+ (LNM) groups and missing information was discarded. Complete classification information is provided in Table 1. The value of expression was converted to \( \log_2(TPM + 1) \). We selected the genes with the highest median absolute deviation of 75% in the expression profile screening mean, at least larger than 0.01. Data analysis was performed to analyze the differences between the N+ and N0 groups using the package “limma” in R, and \( P \) value < 0.01 was set as the cut-off to screen for differentially expressed genes (DEGs).

To verify the prognostic prediction performance of LNM, Kaplan–Meier survival analysis and the log-rank test were performed to identify the prognostic significance of LNM between the N0 and N+ samples. Kaplan–Meier survival curves and log-rank tests were executed using the R packages “survival” and “survminer”. The mutation data of the corresponding sample using the MuTect2 pipeline were downloaded to compare the tumors to a pool of normal samples to find somatic variations. Somatic mutations were
Fig. 1 Overview of our approach. First, two weighted PPI networks for both LNM and non-LNM in breast cancer were constructed, whose topology is based on the STRING database and DEGs of LNM and non-LNM, weighted by the r-z transformation of co-expression data. Second, module analysis was performed for the two weighted PPI networks, and inter/intra-modular edges and date/party hubs, as well as KEGG, Reactome, and cancer hallmark enrichment analysis were introduced to compare the topology and biological functions of different modules. Third, the core network module for breast LNM was further characterized by assessing TFC scores for edges [26], structural modeling, and mutation mapping, and GSEA enrichment analysis. Fourth, three key factors were predicted: allosteric mutations, key PPIs, and key pathways for LNM in breast cancer.
analyzed with the R package “maftools” and visualized in a waterfall plot \[27\].

2.2 Differential Interaction Networks Construction

Two differentially weighted PPI networks under specific conditions (non-LNM and LNM) were built with the following steps. First, the protein–protein interaction network for \textit{Homo sapiens} was retrieved from the STRING database \[28\], filtering interactions by a combined score $\geq 0.4$ (medium confidence in STRING database). Second, DEGs between N$^0$ and N$^+$ groups were used to construct the topological architectures of the two target networks. Third, the Pearson’s correlation coefficients (PCCs) of expression level of a pair of interacting nodes were calculated in N$^0$ and N$^+$ groups, which were calculated by (PCCs) of expression level of a pair of interacting nodes were calculated in N$^0$ and N$^+$ groups, which were calculated by

\begin{equation}
r = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}},
\end{equation}

\begin{equation}
z = \frac{1}{2} \ln \left( \frac{1 + r}{1 - r} \right),
\end{equation}

where $X_i$ denotes the sample gene expressed value indexed $i$, analogously for $Y_i$, and $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ (the gene expression mean value), analogously for $\bar{Y}$. The absolute value of $z$ is the weight of the interaction between $i$ and $j$.

2.3 Differential Modular Analysis

2.3.1 Module Detection

Integrated co-expression networks were then clustered using the multi-level modularity optimization algorithm \[30\]. The method is based on the modularity measure and a hierarchical approach, and the resolution parameter is set to 1. Here, only modules larger than ten nodes were considered. Module detection was performed with the R package “igraph” and represented using Cytoscape \[31\]. Next, we calculated the similarity between two modules using the Jaccard similarity coefficient, defined as

\begin{equation}
J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|},
\end{equation}

where $A$ and $B$ represent different node sets, belonging to two modules.

2.3.2 Edge Knockout Experiment

To investigate which type interactions were more critical in the topology of the network prior to LNM, edge knockout experiments of the interactions were conducted on N$^0$ weighted network. Two common topological measures were calculated \[32\]: the betweenness centrality, which measures the information flow through networks; and the characteristic path length (CPL), which is the average of the shortest path between all nodes in a network, defined as:

\begin{equation}
\text{CPL} = \frac{\sum \sum L_{ij}}{N(N - 1)},
\end{equation}

where $L_{ij}$ denotes the weighted shortest path length between nodes $i$ and $j$.

The change of the two topological measures was applied to systematically assess the robustness of the network by removing the equivalent number of different edges.

We randomly selected several interactions of a group type, removed them from the network, and used “igraph” in the R language package to calculate the topological parameters of the network formed by the remaining interactions. Finally, we calculated the average value of the topological parameters of the nodes in the network. This process counts as one removal and measures how the removed edges affect the rest of the network via topological parameters. To describe the effect of removing a certain type of interaction more accurately on the network, the random sampling was repeated 200 times.
2.3.3 Date and Party Hubs

The average Pearson’s correlation coefficient (aPCC) was calculated for each interaction in the PPI network based on the co-expression of two interacting genes. Then, two types of hub genes were defined as date hubs and party hubs with low and high aPCC values, respectively. Party hub genes bind with most of other genes simultaneously, while date hub genes interact other genes at different times and locations to organize the biological process [23, 25]. Random sampling of the PCC was used to ascertain that the observed edges were nonrandom. A number of interactions were randomly selected with the observed edges several times in the non-LNM and LNM networks.

2.3.4 Topological–Functional Connection

The topological–functional connection (TFC) was calculated for the prioritization of PPIs. TFC integrates the edge betweenness and the gene ontology (GO) semantic similarity [26] to define a new edge measure:

\[ \text{TFC} = \sum_{n=1}^{N} \frac{T_n^* + F_n}{|T_n^* + F_n - 2|} \times 100, \quad (5) \]

\[ T_n^* = \frac{T_n \text{Min}_T}{\text{Max}_T - \text{Min}_T}, \quad (6) \]

where \( N \) represents the number of interactions, and \( T_n \) and \( F_n \) represent the edge betweenness and GO semantic similarity, respectively, of interaction \( n \). Thus, the TFC score is proposed to identify key protein interactions by integrating network topology and biological characteristics.

2.4 Enrichment Analysis

Four types of functional enrichment analysis were used in our work. To identify the significant biological pathways of each module, we used the R package “clusterProfiler” [33] to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment [34]. Terms with a \( P \) value < 0.01 were considered significant. Gene set enrichment analysis (GSEA) was performed to investigate the particular pathway across the whole expression profile, using the MSigDB database [35] and the R package “GSVA” [36]. Genes were sorted according to the logFC in the results of the \( N^0 \) and \( N^+ \) group difference analysis.

Unlike common Reactome enrichment analysis based on gene sets, we performed another pathway enrichment based on interaction (or edge) annotations from the Reactome database [37]. The annotations were inferred between all protein components of a complex. The annotation information only retains the information regarding the interactions between proteins for enrichment analysis, which includes 21,327 interactions annotated to 1633 Reactome pathways. We used a hypergeometric test to detect whether the input interactions were significantly enriched in the pathway, formulated as follows:

\[ P = \Pr(X \geq k) = 1 - \sum_{i=0}^{k-1} \binom{M}{i} \binom{N-M}{n-i}, \quad (7) \]

where \( N \) is the total number of edges with annotations present in the database, \( n \) denotes the number of edges with annotations under study, \( M \) is the total number of interactions related to the Reactome pathway, and \( k \) denotes the number of interactions related to the Reactome pathway in the set of interactions under study.

To determine whether the gene set corresponded to cancer hallmarks, the hallmark annotations from the Catalogue of Somatic Mutations in Cancer (COSMIC) data resource were used [38]. Both of these enrichments were performed using the R package “clusterProfiler”. Terms with a \( P \) value < 0.05 were considered significant.

2.5 Structural Modeling of Protein–Protein Interactions

Protein structures were obtained from the Protein Data Bank (PDB) by deleting small molecules, specifically, PDB id: 6NEC for RET, PDB id: 3PP0 for ERBB2, PDB id: 6OP9 for ERBB3, and PDB id: 2PSQ for FGFR2. We used PRISM [39, 40] (Protein Interactions by Structural Matching) to predict the structures of protein–protein interactions (PPIs) and the effect of the mutations on interactions. Predicted protein complexes were ranked by FiberDock [41] according to their energies, and complexes with the lowest binding energy were selected to evaluate the effects of mutations on PPIs. Thus, the effect of a mutation on PPI is defined as:

\[ \Delta \Delta G = \Delta \Delta G_{mt} - \Delta \Delta G_{wt}, \quad (8) \]

where \( \Delta \Delta G_{mt} \) and \( \Delta \Delta G_{wt} \) are the lowest binding energies of the mutant and wild-type complexes, respectively. \( \Delta \Delta G \) is the binding energy change caused by a single mutation.

2.6 Allosteric Effects of Mutations

The structure-based statistical mechanical model of allostery was used to obtain a direct estimate of the allosteric effects caused by the single mutation, using AlloSigMA [42]. In the approach, two types of mutations were defined: UP-mutation, which models the situation of an actual mutation to a bulky residue with over-stabilizing
effects on the local contact network; conversely, DOWN-mutation models the destabilization of the residue’s contact network similarly to Ala/Gly-like mutations. The allosteric free energy of the residue quantifies the strength and sign of allosteric communication associated with the mutation, where positive and negative signs correspond to a kinetic increase (local destabilization) and decrease (local stabilization), respectively.

3 Results

3.1 \(N^0\) and \(N^+\) PPI Networks Present Different Modular Structures

According to the data in Table 1, 1579 differentially expressed genes (DEGs) were obtained, including 1299 up-regulated and 280 down-regulated genes in \(N^+\) compared with \(N^0\). A volcano plot of the DEGs is shown in Supplementary Fig. S1a, in which \(MISP, BLVRA, RNF223, KRT8, PTK6, PQLC3, KRT18, BCL2L14, ETV6,\) and \(LMO4\) were the 10 most significantly differentially expressed genes. In addition, a Kaplan–Meier model for \(N^0\) and \(N^+\) was constructed, and we observed that the survival status of the lymph node metastasis group was significantly worse than that of the non-lymph node metastasis group, with a log rank test \(P\) value < 0.0001 (Supplementary Fig. S1b). This result also confirmed that lymph node metastasis could be an independent predictor of survival in breast cancer.

Based on these DEGs, the corresponding topology of the LNM-related PPI network was generated, with 1516 nodes and 8286 edges. The absolute value of the Pearson coefficient between the expressions in the \(N^0\) and \(N^+\) groups was introduced as network weights, and then two different weighted PPI networks for LNM-related PPI networks were finally constructed, defined as the \(N^0\) and \(N^+\) PPI networks. As such, these two networks could preliminarily reflect the dynamic nature of the LNM process. Next, the modular detection algorithm was used to discover the sub-network structures and functions of the PPI networks. From a global perspective, the \(N^0\) and \(N^+\) PPI networks can be separated into 15 and 17 modules (Fig. 2a and Supplementary Fig. S1c), respectively. Almost all modules tend to rewire, and smaller modules can be obtained in the \(N^+\) PPI network. A heat map of module similarity further quantifies this result; that is, the topological properties of only three modules are preserved (green asterisks), including purine metabolism, nucleotide excision repair, and spliceosome, as highlighted in blue font. Furthermore, some statistically significant pathways in the \(N^0\) modular network were not significant in the \(N^+\) state (highlighted in red font). Among these pathways, we found that most corresponded to signaling transformation and cancer metastasis. In particular, the imbalance of the calcium signaling pathway in breast cancer will lead to the migration, invasion, proliferation, tumorigenicty, or metastasis of cancer cells. The expression of some key proteins in breast cancer is closely related to its progression. For example, oncogenic receptor tyrosine kinase \(ERBB2\) (also called \(HER2\)) is overexpressed in approximately 20% of breast cancers, resulting in ligand-independent dimerization and activation. The preliminary modular analysis shows that \(ERBB2-ERBB3, ERBB2-RET,\) and \(ERBB2-RTK6\) interactions are involved in the same module of the \(N^0\) network, but between different modules of the \(N^+\) network (Fig. 2d). Due to the importance of \(ERBB2\) in breast cancer, the change in \(ERBB2\)-related interactions during LNM may be useful as an indicator of breast cancer metastasis.

The comparison of the network modules of \(N^0\) and \(N^+\) PPI networks demonstrated that most biological functions are conserved during LNM, although some module evaluation was also detected. Accordingly, further study of these different modules will help us understand the LNM mechanism and facilitate the prediction of key modules, interactions, and genes involved in cancer metastasis.

3.2 Inter-modular Edges Correspond to Cancer Metastasis and Network Signaling

To further characterize how modular structures change during LNM, we classified the edges or PPIs involved in dynamic changes into two types: (1) inter-modular edges, whose interacting nodes are located within the same module in the \(N^0\) PPI network, but within different modules in the \(N^+\) PPI network, or vice versa; and (2) intra-modular edges, in which the two interacting nodes are always located within the same module or between different modules. In all, 1546 inter-modular edges and 6740 intra-modular edges were identified in the LNM-related network dynamic process. Topological and functional analyses of these edges were then performed in terms of edge knockout experiments, as well as Reactome pathway enrichment analysis based on edge annotation.

In the edge knockout experiment, we calculated the average betweenness and average CPL of each remaining network by systematically removing two types of edges randomly based on the \(N^0\) network. We randomly removed 20 inter-modular edges and intra-modular edges respectively, and repeated the process 200 times. We then
Fig. 2 Overview of modular structures and functions of LNM-related PPI networks. a Modular structure of the N₀ network, where nodes are colored according to different modules. b Assignment flow to KEGG pathways from modules in the N₀ to N⁺ networks, shown in a Sankey diagram. Each module is annotated by the three most significant pathways with colored beads, or by all related pathways if the enriched pathway number is fewer than three. Shared pathways by the N₀ and N⁺ modular networks are marked by green asterisks, and pathways colored by red and blue indicate cancer metastasis-related biological pathways. c Example of ERBB2-related interactions in the modular structures between the N₀ and N⁺ networks. Dashed red lines correspond to the observed interactions that are located between different modules of the N⁺ network and inside a module of the N₀ network.
gradually increased the number of knockdown interactions with a gradient of 20. As there were fewer inter-modular edges than intra-modular edges, we removed a maximum of 1546 of the two interactions. The average betweenness of the interaction network decreased more quickly by removing inter- rather than intra-modular edges (Fig. 3a), while the average shortest path increased more quickly by removing inter- rather than intra-modular edges (Fig. 3b). Betweenness measures the information flow through networks, with high betweenness indicating high biological signaling ability. CPL indicates the network entropy, which means that a biological network system with a higher CPL is more chaotic. Both topological measures showed that their values were more sensitive to inter-modular edges, indicating that inter-modular edges contributed more to maintaining the global connectivity of the network, and played key roles in network signaling.

To identify specific biological functions, pathway enrichment of these two types of edges was performed based on “edge annotation” in the Reactome database. As shown in Fig. 3c, inter-modular edges were significantly enriched with several cancer emergence- and metastasis-related pathways. For example, VEGFA-VEGFR2 pathway regulates breast cancer angiogenesis, which is key for metastasis [43].

![Fig. 3](image_url)

**Fig. 3** Topological and functional analysis of inter/intra-modular edges. **a** Network betweenness as a function of removing equivalent numbers of inter-modular and intra-modular edges. **b** The CPL of the network as a function of removing equivalent numbers of inter-modular and intra-modular edges. Significantly enriched Reactome pathways of edges in the **c** inter-modular and **d** intra-modular groups. **“Edge Ratio”** represents the number of interactions in the input that are annotated to be in a certain pathway divided by the total number of interactions annotated to be in that pathway in the database.
RET signaling also a potential driver of breast cancer metastasis [44]. The disruption of Ptk6 significantly delays and reduces ERBB2-induced breast gland tumor formation and metastasis [45]. Autocrine PDGFR signaling promotes breast cancer metastasis [46]. The enrichment of interactions within these pathways also has a direct or indirect relationship with breast cancer metastasis, such as ERBB2-ETBB3 [47], ERBB2-PTK6 [45], and PDGFB-PDGFRB [48]. On the other hand, significant Reactome pathways of the intra-modular edges are mainly involved in transport, metabolism, protein modification, growth and development, and signal communication, such as Endosomal Sorting Complex Required For Transport (ESCRT), Translocation of SLC2A4 (GLUT4) to the plasma membrane, VEGFR2-mediated vascular permeability, collagen biosynthesis, and modifying enzymes (Fig. 3d). Together, the intermodular edges contributed to the major events implicated in the network signaling pathway and served as attractive targets for inhibiting LNM [49].

3.3 Date Hubs Reveal the Invasion and Metastasis Hallmarks

Topological analysis of the PPI networks revealed that most of the proteins were connected to relatively few, highly connected proteins, termed hub proteins. Hubs in PPI networks have been classified into party and date hubs based on the co-expression of the interacting proteins [1]. Whereas date hubs display low co-expression with their partners, party hubs have high co-expression. It has been proposed that date hubs were global connectors, whereas party hubs were local coordinators. Date hubs tended to have transient interactions due to their low average co-expression correlation with their interaction partners, which not only played a role in connecting biological modules to each other but also showed more dynamic properties.

Here, we defined nodes with a degree > 20 in PPI networks as hub nodes. The average Pearson’s correlation coefficients (PCCs) of the edges related to these hub nodes in the N⁰ network and the N⁺ network were calculated. Then, the date hubs and party hubs were defined as having lower PCCs and higher PCCs by the median of these values. To investigate the functions of these types of hubs, we first compared the relationship between hub proteins and inter/ intra-modular edges. A boxplot showed that date hubs were significantly more involved in inter-modular interactions than party hubs (Fig. 4a). In robustness tests of the N⁰ and N⁺ networks (Fig. 4b and c), the average PCC of inter-modular edges interacting with date hubs was significantly lower than the average PCC of interactions generated in random sampling (658 samples with 100,000 random times). In addition, compared with party hubs, date hubs showed obviously higher degrees and betweenness, such as HSP90AA1, ERBB2, and VEGFA in the top 10 inter-modular edge-enriched pathways (Supplementary Fig. S2a), and MAPK3 and HSP90AA1 in the top 10 intramodular edge-enriched pathways (Supplementary Fig. S2b).

Gene expression changes in cancer cells are related to a limited set of special characteristics, often termed cancer hallmarks [50]. We further assessed whether date hubs and party hubs inferred different clinical outcomes. Cancer hallmark enrichment analysis of these hub genes was performed, based on the COSMIC manual annotation of hallmark identification. The hypergeometric distribution was used to infer the significance of the hub genes in cancer hallmarks, with all genes as the background to compute the P value. Overall, date hubs were more highly enriched cancer hallmarks than party hubs, suggesting that date hubs may play a more important role in driving the occurrence and progression of cancer (Fig. 4d). Among the ten cancer hallmarks, only cell-replicative immortality was not enriched for date hubs. The most significant cancer hallmarks for date hubs were invasion and metastasis, followed by proliferative signaling and escaping programmed cell death. In contrast, the party hubs showed that “genome instability and mutations” is the most significant cancer hallmark. Specifically, there are 11 date hubs enriched in the “invasion and metastasis” hallmark: GATA3, DDB2, FGFR1, RAC1, ERBB3, ERBB2, GATA2, RET, CDH1, NF2, and FGFR2. However, there are only four party hubs enriched in the “invasion and metastasis” hallmark: FOXA1, CUX1, PDGFA, and PRKARIA. The above analysis recapitulated the roles of date hubs and party hubs in terms of hallmark enrichments. This suggests that date hubs are more relevant in the cancer context, especially regarding invasion and metastasis.

3.4 Core Network Module of Lymph Node Metastasis in Breast Cancer

Both date hubs and inter-modular edges describe how protein–protein interaction (PPI) networks change, but date hubs introduce invariance in the dynamic evolution of the network and drive cooperative inter-modular interactions. Accordingly, the “core network module” was constructed by connecting date hubs with inter-modular edges. We suggest that this core network module contains highly dynamic regions that reorganize to drive or respond to lymph node metastasis in breast cancer (Fig. 5a). To evaluate the correlation between this core module and the LNM process, four types of analysis were performed: (i) we determined the number of nodes in the core network related to breast cancer progression or metastasis, (ii) we calculated the TFC score to detect key interactions, (iii) we performed pathway enrichment analysis, and (iv) we performed mutation analysis.
For the breast cancer LNM-related core module, most genes (68 of 76) are reported to be associated with breast cancer-related genes, while almost half of the genes (36/76) are related to lymph node metastasis in breast cancer, according to the PubMed database. Additionally, a new score, named TFC, was defined as an edge parameter that was obtained by integrating edge betweenness and the GO semantic similarity of interactions. Here, we ranked the importance of each interaction according to the TFC score.

The KEGG enrichment pathways of the top 20 PPIs are shown in Table S1. Among them, PI3K-Akt, MAPK, ErbB, and Calcium signaling pathways were found to be related to invasion and LNM in breast cancer. As listed in Table 2, among the top six interactions (FGFR2-RET, RET-VEGFA, AP2A1-SYT1, AP2A1-SYT1, ERBB3-RET, and ERBB2-RET), four are rearranged during transfection (RET)-related, including interactions with two metastatic breast cancer hallmark genes (ERBB2 and ERBB3) (Fig. 5b). This interesting finding suggested a key role of RET, as it may be involved in key interactions that need further investigation.

Furthermore, KEGG pathway enrichment analysis was performed on the core network module. We found that five genes (VEGFA, FGFR2, RET, ERBB3, and ERBB2) involved in RET-related interactions were enriched in the calcium signaling pathway (P value = 0.0109). To verify whether these genes could affect breast LNM through the calcium signaling pathway, we used the GSEA algorithm to evaluate the status of the calcium signaling pathway in the whole expression profile. The results showed that these genes tended to be up-regulated in the calcium signaling pathway, with a P value of 0.0089 (Fig. 5c).

Next, we investigated the genomic mutational signatures, which provided clues for the TNM staging of breast cancer. The somatic mutations in all samples were compared between date hubs and party hubs. Figure 5c and Supplementary Fig. S3a show the mutation spectrums of highly mutated date-hub and party-hub genes in breast tumor samples. Overall, the mutation frequency of date hubs (39.3%) was higher than that of party hubs (23%). Except for the two most highly mutated genes (CDH1 and GATA3; > 10%), among the date hubs, three more genes (ERBB2, ERBB3, and CTCF) had relatively high mutation frequencies. In addition, FGFR2 is a low frequency mutated gene, but the important role of its mutations in breast cancer has been reported [51].

Fig. 4 Distribution of date and party hubs and their corresponding KEGG pathways. a Statistical comparison of the numbers of intermodular edges consisting of date hubs and party hubs. Sampling tests of inter-modular edges in the b N0 network and c N+ network. Solid red lines correspond to the average weight values of inter-modular edges. d Cancer hallmark enrichment analysis for date hubs and party hubs. The blue and red dotted lines represent the threshold of −log10 (0.05) calculated by the hypergeometric distribution. The number of genes is marked on the histogram.
Accordingly, the above analysis suggested the importance of the sub-network, consisting of RET and its connected genes, as shown in Supplementary Fig. S3b. First, the sub-network contained four of the six most important interactions with the highest TFC scores. Second, the sub-network contained all five calcium signaling pathway-related genes. Third, seven of ten genes in the sub-network had important mutation information, including the core gene of RET. By focusing on the sub-network, we further suggested that RET-ERBB2, RET-ERBB3, and RET-FGFR2 interactions are three key interactions for breast cancer LNM.

3.5 Structure-Based Assessment of the Effect of Mutations on RET Interactions

The occurrence and development of cancer is not only related to changes in expression levels but also to somatic mutations in the coding regions of key genes and their interaction partners. The dynamic assembly of protein complexes is a central mechanism of many cell signaling pathways, which may be regulated by their mutations. To characterize the effect of mutations on RET interactions, structural models of RET-ERBB2, RET-ERBB3, and RET-FGFR2 complexes were first constructed. Using PRISM,

---

**Fig. 5** Molecular characteristics of the “core network module”. a The core network module consists of 76 date hubs and 147 inter-molecular edges. Edges are weighted by TFC scores, and the top 20 edges with the highest TFC scores are denoted in purple. Date hubs with the highest mutation frequency are denoted as yellow nodes. b Polar area diagram of network edge prioritization by the TFC scores in the core network module. Asterisks indicate RET-involved interactions. c GSEA revealed that the genes of the calcium signaling pathway were enriched across the whole expression profile, and the enrichment status of the calcium signaling pathway was not affected by the supervisor selection of differential genes. d Mutated date hubs (rows, top 30) are ordered by mutation rate; samples (columns) are arranged to emphasize mutual exclusivity among mutations. The row on the right shows the mutation percentage, and the top histogram shows the overall number of mutations. The color coding indicates the mutation type.
we found that all three interactions can form stable protein complexes, with binding energies of $-44.88$, $-51.97$, and $-46.33$ kcal/mol for the RET-FGFR2, RET-ERBB2, and RET-ERBB3 protein complexes, respectively (Table 3). The structural models for the three protein complexes and their key interfacial residues are shown in Supplementary Fig. S4. We then mapped missense mutations to the obtained structural models. It can be observed that only the mutation of the D769 in ERBB2 is located at the RET-ERBB2 interface. The distribution of other mutations is relatively scattered over the whole structure.

Next, we analyzed the effects of mutations on the stability of protein–protein interactions by calculating the binding affinity. From the TCGA sample, we found three mutations in RET: the tyrosine domain V778I mutation; and two kinase domain mutations, L846I and L963V. As shown in Table 3, the change in the binding energy ($\Delta \Delta G$) of the three mutations (V778I, L846I, and L963V) on the RET-ERBB2, RET-ERBB3, and RET-FGFR2 complexes shows that their binding energy may become more stable or decrease slightly. Similar results were also found for mutations of ERBB2, ERBB3, and FGFR2 (Table S2), and even for the D769H(Y) mutation located at the RET-ERBB2 interface. In conclusion, the structural modeling, in silico mutagenesis, and comparison of the predicted binding energies revealed that the mutations did not disturb protein–protein interactions, suggesting other kinds of mechanisms.

In addition, we used a structure-based statistical mechanical model implemented in AlloSigMA [42] to obtain a direct estimate of the allosteric effects across protein–protein interactions caused by these missense mutations. As shown in Supplementary Figs. S5–S8, except for V778I in RET, almost all other mutations have no allosteric effects on their protein partners. We observed that V778I in RET showed allosteric effects on all its partners, including ERBB2, ERBB3, and FGFR2 (Fig. 6a). The V778I mutation is a large amino acid, while in AlloSigMA, it is defined as an UP mutation. The allosteric energy profile showed that the energy of residues near V778 was negative and relatively low, which demonstrated the greater local stability of this region. However, for the interaction partners of RET, the allosteric free energy of residues was generally positive and relatively high, indicating that the interaction partners became unstable above the V778I mutation in RET.

In detail, we predicted that FGFR2 had an obvious peak (0.99 kcal/mol) at residue P582 (Fig. 6b), and ERBB2 had a local increase in kinetics at residues E719, I740 to V746, D873, E874, H878, I886, L891 to L895, I926, E930 to R940, and I961 (Fig. 6c). In addition, we observed a distal effect on the ERBB3 activation loop (P842-K853) and residues G713 to S715, where the contact network became unstable.

### Table 2
The 20 interactions with the highest TFC scores in the core network module

| Proteins | Proteins | TFCs |
|----------|----------|------|
| FGFR2    | RET      | 173.0058 |
| RET      | VEGFA    | 157.5393 |
| AP2A1    | SYT1     | 149.0898 |
| AP2A1    | AREG     | 143.0134 |
| ERBB3    | RET      | 134.7705 |
| ERBB2    | RET      | 117.165 |
| AREG     | ERBB3    | 101.4081 |
| AREG     | FGFR2    | 100.4558 |
| CDKN1A   | DDB2     | 90.53797 |
| PSM7     | UB1      | 87.0691 |
| HSP90AA1 | PSM7     | 82.14306 |
| MAPK3    | RET      | 82.12968 |
| NCAM1    | RET      | 69.61411 |
| MMP9     | VEGFA    | 68.23644 |
| CDH1     | RET      | 65.6743 |
| SYT1     | VAMP2    | 65.67153 |
| HSP90AA1 | RET      | 63.11788 |
| AREG     | GATA3    | 62.60333 |
| PXN      | RET      | 58.36475 |
| STXBP1   | SYT1     | 56.96024 |

### Table 3
Effects of RET mutations on the stability of protein–protein interactions

| Complex    | $\Delta \Delta G_{\text{mut}}$ (kcal/mol) | RET Mutation | $\Delta \Delta G_{\text{mut}}$ (kcal/mol) | $\Delta \Delta G$ (kcal/mol) |
|------------|------------------------------------------|--------------|------------------------------------------|-----------------------------|
| RET-FGFR2  | $-44.88$                                 | L846I        | $-45.18$                                | $-0.3$                     |
|            |                                          | L963V        | $-44.44$                                | $0.44$                     |
|            |                                          | V778I        | $-44.25$                                | $0.63$                     |
| RET-ERBB2  | $-51.97$                                 | L846I        | $-60.36$                                | $-8.39$                    |
|            |                                          | L963V        | $-55.49$                                | $-3.52$                    |
|            |                                          | V778I        | $-51.97$                                | $0$                        |
| RET-ERBB3  | $-46.33$                                 | L846I        | $-55.8$                                 | $-9.47$                    |
|            |                                          | L963V        | $-41.3$                                 | $5.03$                     |
|            |                                          | V778I        | $-55.81$                                | $-9.48$                    |

$\text{wt}$ represents the wild-type protein complex, $\text{mt}$ represents the mutant protein complex.

[42] Springer
Fig. 6 Allosteric effects caused by V778I in RET. a Allosteric energy profiles for V778I in RET in three protein complexes, predicted by AlloSigMA. Mapping of allosteric energy on 3D structures of the b RET-FGFR2, c RET-ERBB2, and d RET-ERBB3 interactions.

Overall, our results showed that the V778I mutation in the RET protein had an unstable effect on the contact network in its interacting partners. This may affect the interaction through allosteric communication between RET and its interacting partners.

4 Discussion and Conclusion

Cancer metastasis is, therefore, an evolving disease and the combined outcome of cells that metastasize and a series of microenvironmental factors that they interact with, collude, or surmount. Although each instance of metastasis could be unique, the quest is to find commonalities that could be targeted therapeutically. The complexity of metastasis through its chronological progression, and its manifestation at various biological scales, calls for a systems approach to understand metastasis mechanistically. In this work, we proposed a method that combines differential modular analysis with mutational structural analysis to study the dynamic process and molecular mechanism of cancer metastasis. By applying this method to study the LNM of breast cancer, we identified some key factors, including the core network module, key PPIs, and a potential allosteric mutation.

From the methodology, we combined gene co-expression data, PPI networks and structures, and genetic variation together to understand disease progression at both the systems and molecular levels. First, we constructed two PPI networks with the same topology but weighted with different gene co-expression data. Then, different modular structures were obtained by modular analysis and the two PPI networks associated with different disease states were compared. The third and fourth steps represent the main idea of our method, that inter-modular edges and date hubs obtained from different modular analysis afforded more functional contexts for disease progression. The network module associated with disease progression was constructed by connecting date hubs with inter-modular edges. Based on the core network module, subsequent analysis included ranking all the interactions by TFC scores and mutation analysis.
Finally, the structures of PPIs were modeled and mutations with important functions were predicted.

In summary, the novelty of our method was two fold. First, differential network analysis methods were normally node- and edge-based, or they involved a comparison of the global topology. Our method here was module-based, which could help in discovering shared or changed functional patterns. The differential modular analysis lies somewhere between node/edge analysis and global topological differential network analysis, shedding more light on the underlying mechanisms of biological systems. Second, the proposed method was novel insofar as it linked allosteric mutation with PPI network analysis to characterize the differentially interacting modules. As such, it could identify key PPIs and active mutations during LNM. This idea has been used previously to study the role of Bcl-2 proteins in breast cancer [52].

In breast cancer, the activation mutation in ERBB2 is a well-known oncogenic driver. The interaction of ERBB2 with other protein partners (or dimerization) activates various oncogenic signaling pathways related to breast cancer metastasis, including the Smad2/3, RAS/RAF/MAPK, PKC, and PI3K-Akt signaling pathways [53]. Recent computational structural modeling with biochemical and cell biological analyses suggests that ERBB2 mutations need to mutate ERBB3, and then promote oncogenesis and invasion of breast cancer via PI3K pathway activation [54].

However, we identified that calcium signaling pathways could be considered an ERBB2 participation noncanonical pathway related to LNM in breast cancer. As a second messenger, the intracellular calcium ion (Ca^{2+}) plays direct and robust roles in many biological processes. Several studies have reported that calcium signaling pathways are essential to cancer progression. In particular, calcium signaling pathways regulate key processes, from inflammation to apoptosis, that are involved in breast cancer tumorigenesis [55], metastasis [56], and resistance to chemotherapy.

The current standard of care for the treatment of metastatic ERBB2 breast cancer is the combination of seven inhibitors: trastuzumab, pertuzumab, trastuzumab emtansine, trastuzumab deruxtecan, lapatinib, neratinib, and tucatinib [57]. However, drug resistance is common and remains a major unresolved clinical problem. Targeting the interaction between ERBB2 and a number of non-canonical RTKs (i.e. EGFR, ERBB3, and ERBB4) has emerged as a promising therapeutic method that overcomes drug resistance in treating breast cancer metastasis [58].

Here, we considered that breast cancer LNM may be the outcome of allosteric driver mutations in RET [59]. As one of the RTK members, RET is a target for several kinds of human cancer, such as thyroid, breast, and colorectal carcinoma. Many RET missense mutations are known to be causally associated with breast cancer, including extracellular domain mutations C611R, C620F, L633V, C634R, C634F, and T636M, and the kinase domain mutation M918T [60]. A structural understanding of the mutational roles of RET in different interactions suggested V778I as an allosteric mutation, revealing potential targets to prevent LNM in breast cancer from overcoming resistance to ERBB2.

Altogether, the analysis characterized a differential module representing significant changes in their interaction patterns during LNM in breast cancer, and their functional roles in terms of enrichment analyses and mutation analysis based on the structural level. In addition, the differential module is potentially valuable not only for understanding LNM but also for developing effective diagnosis, prognosis, and treatment strategies.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12539-023-00568-w.

Funding This work was supported by the National Natural Science Foundation of China (32271292, 31872723), a project funded by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions, the China Postdoctoral Science Foundation (2016M590495), and the Jiangsu Planned Projects for Postdoctoral Research Funds (1601168C).

Data Availability The code for this work is available at https://github.com/CSB-SUDA/DMA.

Declarations Conflict of Interest The authors declare no competing financial interest.

References
1. Welch DR, Hurst DR (2019) Defining the hallmarks of metastasis. Can Res 79(12):3011–3027. https://doi.org/10.1158/0008-5472.CAN-19-0458
2. Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y (2020) Molecular principles of metastasis: a hallmark of cancer revisited. Signal Transduct Target Ther 5(1):28. https://doi.org/10.1038/s41392-020-0134-x
3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71(3):209–249. https://doi.org/10.3322/caac.21660
4. Dihge L, Vallon-Christersson J, Hegardt C, Saal LH, Hakkinen J, Larsson C, Ehinger A, Loman N, Malmberg M, Bendahl PO, Borg A, Staaf J, Ryden L (2019) Prediction of lymph node metastasis in breast cancer by gene expression and clinicopathological models: development and validation within a population-based cohort. Clin Cancer Res 25(21):6368–6381. https://doi.org/10.1158/1078-0432.CCR-19-0075
5. Angus L, Smid M, Wilting SM, van Riel J, Van Hoeck A, Nguyen L, Nik-Zainal S, Steenbruggen TG, Tjan-Heijnen VCG, Labots M, van Riel J, Bloemendal HJ, Steeghs N, Lolkema MP,
Voest EE, van de Werken HJG, Jager A, Cuppen E, Sleijfer S, Martens JW (2019) The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. Nat Genet 51(10):1450–1458. https://doi.org/10.1038/s41588-019-0507-7

6. Khan A, Rehman Z, Hashmi HF, Khan AA, Junaid M, Sayaf AM, Ali SS, Hassan FU, Heng W, Wei DQ (2020) An integrated systems biology and network-based approaches to identify novel biomarkers in breast cancer cell lines using gene expression data. Interdiscip Sci 12(2):155–168. https://doi.org/10.1007/s12539-020-00360-0

7. Hozhibari H, GhasemiDehkohnehs RS, Razavi SM, Razavi SM, Salarian F, Rasooli A, Azami J, GhasemiShiran M, Kardan Z, Farrokhzad N, MikaeliNamini A, Salari A (2022) Comparative analysis of protein-protein interaction networks in metastatic breast cancer. PLoS ONE 17(1):e0260584. https://doi.org/10.1371/journal.pone.0260584

8. Liu C, Ma YF, Zhao J, Nussinov R, Zhang YC, Cheng FX, Zhang ZK (2020) Computational network biology: data, models, and applications. Phys Rev 846:1–66. https://doi.org/10.1016/j.physrep.2019.12.004

9. Yan WY, Liu XY, Wang YB, Han SQ, Wang F, Liu X, Xiao F, Hu G (2020) Identifying drug targets in pancreatic ductal adenocarcinoma through machine learning, analyzing biomolecular networks, and structural modeling. Front Pharmacol. https://doi.org/10.3389/fphar.2020.00534

10. Basha O, Argov CM, Arzty R, Zoabi Y, Hekselman I, Alfandari L, Chalifa-Caspi V, Yeger-Lotem E (2020) Differential network analysis of multiple human tissue interactomes highlights tissue-selective processes and genetic disorder genes. Bioinformatics 36(9):2821–2828. https://doi.org/10.1093/bioinformatics/btaa034

11. Lichtblau Y, Zimmermann K, Haldemann B, Lenze D, Hummel M, Lerer U (2017) Comparative assessment of differential network analysis methodis. Brief Bioinform 18(5):837–850. https://doi.org/10.1093/bib/bbw061

12. Goenawan IH, Bryan K, Lynn DJ (2016) DyNet: visualization and analysis of dynamic molecular interaction networks. Bioinformatics 32(17):2713–2715. https://doi.org/10.1093/bioinformatics/btw187

13. Guldadan G, Turanli B, Beklen H, Sinha R, Arga KY (2020) Pan-cancer mapping of differential protein-protein interactions. Sci Rep 10(1):3272. https://doi.org/10.1038/s41598-020-60127-x

14. Savino A, Provero P, Poli V (2020) Differential co-expression analyses allow the identification of critical signalling pathways altered during tumour transformation and progression. Int J Mol Sci. https://doi.org/10.3390/ijms21249461

15. Hu B, Chang X, Liu X (2019) Predicting functional modules of liver cancer based on differential network analysis. Interdiscip Sci 11(4):636–644. https://doi.org/10.1007/s12539-018-0314-3

16. Nussinov R, Jang H, Nir G, Tsai CJ, Cheng F (2022) Open structural data in precision medicine. Annu Rev Biomed Data Sci 5:95–117. https://doi.org/10.1146/annurev-biodata-122220-012951

17. Gorostiza Gonzalez M, Janssen APA, IJzerman AP, Heitman LH, van Westen GJP (2022) Oncological drug discovery: AI meets structure-based computational research. Drug Discover Today 27(6):1661–1670. https://doi.org/10.1016/j.drudis.2022.03.005

18. Robichaux JP, Le X, Vijayan RSK, Hicks JK, Hecke S, Elamir YY, Lin HY, Udagawa H, Skoulidis F, Tran H, Varghese S, He J, Zhang F, Nilsson MB, Hu L, Poteete A, Rinsurongkawong W, Zhang X, Ren C, Liu X, Hong L, Zhang J, Diao L, Madison R, Schrock AB, Saam J, Raymond V, Fang B, Wang J, Ha MJ, Cross JB, Gray JE, Heymach JV (2021) Structure-based classification predicts drug response in EGFR-mutant NSCLC. Nature 597(7878):732–737. https://doi.org/10.1038/s41586-021-03898-1

19. Wu H, Chen Z, Wu Y, Zhang H, Liu Q (2022) Integrating protein-protein interaction networks and somatic mutation data to detect driver modules in pan-cancer. Interdiscip Sci 14(1):151–167. https://doi.org/10.1007/s12539-021-00475-y

20. Kar G, Gursoy A, Keskin O (2009) Human cancer protein-protein interaction network: a structural perspective. PLoS Comput Biol 5(12):e1000601. https://doi.org/10.1371/journal.pcbi.1000601

21. Engin HB, Guney E, Keskin O, Oliva B, Gursoy A (2013) Integrating structure to protein-protein interaction networks that drive metastasis to brain and lung in breast cancer. PLoS ONE. https://doi.org/10.1371/journal.pone.0081035

22. AcunerOzbabacan SE, Gursoy A, Nussinov R, Keskin O (2014) The structural pathway of interleukin 1 (IL-1) initiated signaling reveals mechanisms of oncogenic mutations and SNPs in inflammation and cancer. PLoS Comput Biol 10(2):e1003470. https://doi.org/10.1371/journal.pcbi.1003470

23. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wanla J (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 27(2):199–204. https://doi.org/10.1038/nbt.1522

24. Chuang HY, Lee E, Liu YT, Lee D, Ideker T (2007) Network-based classification of breast cancer metastasis. Mol Syst Biol 3:140. https://doi.org/10.1038/msb4100180

25. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, Vidal M (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430(6995):88–93. https://doi.org/10.1038/nature02555

26. Wang F, Han S, Yang J, Yan W, Hu G (2021) Knowledge-guided “Community Network” analysis reveals the functional modules and candidate targets in non-small-cell lung cancer. Cells. https://doi.org/10.3390/cells10020402

27. Mayakonda A, Lin DC, Assenov Y, Plass C, Koehler HP (2018) Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 28(11):1747–1756. https://doi.org/10.1101/gr.239244.118

28. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffe N, Huyenen MA, Bork P (2005) STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic acids Res 33(Database issue):D433–D437. https://doi.org/10.1093/nar/gki005

29. Hotelling H (1953) New light on the correlation coefficient and its transforms. J R Stat Soc B 15(2):193–232. https://doi.org/10.1111/j.2517-6161.1953.tb00135.x

30. Subelj L, Bajec M (2011) Unfolding communities in large complex networks: combining defensive and offensive label propagation for core extraction. Phys Rev 83(3–2):036103. https://doi.org/10.1103/PhysRevE.83.036103

31. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504. https://doi.org/10.1101/gr.123930

32. Hu G, Yan W, Zhou J, Shen B (2014) Residue interaction network analysis of Droopna and a DNA clamp. J Theor Biol 348:55–64. https://doi.org/10.1016/j.jtbi.2014.01.023

33. Yu G, Wang LG, Han Y, He QY (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16(5):284–287. https://doi.org/10.1089/omi.2011.0118

34. Ogata H, Goto S, Kato F, Fujibuchi W, Bono H, Kanehisa M (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 27(1):29–34. https://doi.org/10.1038/nbt.29

35. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P (2015) The Molecular Signatures Database (MSigDB)
hallmark gene set collection. Cell Syst 1(6):417–425. https://doi.org/10.1016/j.cels.2015.12.004

36. Hanzelmann S, Castelo R, Guinney J (2013) GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14:7. https://doi.org/10.1186/1471-2105-14-7

37. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidirooulos K, Cook J, Gillespie M, Hav R, Loney F, May B, Milacic M, Rothfels K, Sevilla C, Shamovsky V, Shorser S, Varasai T, Weiser J, Wu G, Stein L, Hermjakob H, D’Eustachio P (2020) The reactome pathway knowledgebase. Nucleic Acids Res 48(D1):D498–D503. https://doi.org/10.1093/nar/gka1031

38. Forbes S, Clements J, Dawson E, Bamford S, Webb T, Dogan A, Flanagan A, Teague J, Wooster R, Futreal PA, Stratton MR (2006) Cosmic 2005. Br J Cancer 94(2):318–322. https://doi.org/10.1038/sj.bjc.6602928

39. Zhang Q, Lu S, Li T, Zhang Y, Zeng H, Qian X, Bi J, Lin Y (2021) PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes. Nucleic Acids Res 42(Web Server issue):W285–W289. https://doi.org/10.1093/nar/gku397

40. Tuncbag N, Gursoy A, Nussinov R, Keskin O, Gursoy A (2014) PRISM: a web server for prediction of evolutionary and structural similarities at interfaces using PRISM. Nat Protoc 9(9):1341–1354. https://doi.org/10.1038/nprot.2011.367

41. Mashiach E, Nussinov R, Wolfson HJ (2010) FiberDock: flexible induced-fit backbone refinement in molecular docking. Proteins 78(6):1503–1519. https://doi.org/10.1002/prot.22668

42. Guanera E, Tan ZW, Zheng Z, Berezovsky IN (2017) AlloSigMA: allosteric signaling and mutation analysis server. Bioinformatics 33(24):3996–3998. https://doi.org/10.1093/bioinformatics/btx430

43. Zhang Q, Lu S, Li T, Yu L, Zhang Y, Zeng H, Qian X, Bi J, Lin Y (2019) ACE2 inhibits breast cancer angiogenesis via suppressing the VEGFα/VEGFR2/ERK pathway. J Exp Clin Cancer Res 38(1):173. https://doi.org/10.1186/s13046-019-1156-5

44. Pecar G, Liu S, Hoo DA, Atkinson JM, Oesterreich S, Lee AV (2023) SET signaling in breast cancer therapeutic resistance and metastasis. Breast cancer Res 25(1):26. https://doi.org/10.1186/s13058-023-01622-7

45. Peng M, Ball-Kell SM, Tyner AL (2015) Protein tyrosine kinase 6 promotes ERBB2-induced mammary gland tumorigenesis in the mouse. Cell Death Dis 6(8):e1846. https://doi.org/10.1038/cddis.2015.210

46. Jechlinger M, Sommer A, Morrigl R, Seither P, Kraut N, Capodiecci P, Donovan M, Cordon-Cardo C, Beug H, Grunert S (2006) Autocrine PDGFR signaling promotes mammary cancer metastasis. J Clin Invest 116(6):1561–1570. https://doi.org/10.1172/JCI24652

47. Xue C, Liang F, Mahmood R, Vuolo M, Wycuff J, Qian H, Tsai KL, Kim M, Locker J, Zhang ZY, Segall JE (2006) ErbB3-dependent motility and intravasation in breast cancer metastasis. Can Res 66(3):1418–1426. https://doi.org/10.1158/0008-5472.CAN-05-0550

48. Jitariu AA, Raica M, Cimpean AM, Suciuc SC (2018) The role of PDGF-B/PDGFR-BETA axis in the normal development and carcinogenesis of the breast. Crit Rev Oncol Hematol 131:46–52. https://doi.org/10.1016/j.critrevonc.2018.08.002

49. Zhou HB, Lei YJ, Padera TP (2021) Progression of metastasis through lymphatic system. Cells. https://doi.org/10.3390/cells10030972

50. Hanahan D (2022) Hallmarks of cancer: new dimensions. Cancer Discov 12(1):31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

51. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, Wang J, Yu K, Chatterjee N, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover RN, Thomas G, Chanock SJ (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 39(7):870–874. https://doi.org/10.1038/ng2075

52. Konig SM, Rissler V, Terkelsen T, Lambrughi M, Papaleo E (2019) Alterations of the interactome of Bcl-2 proteins in breast cancer at the transcriptional, mutational and structural level. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.1007485

53. Ghanri MA, Raza A, Hayat U, Atif N, Iqbal HMM, Bilal M (2022) Mechanistic insights expatiating the biological role and regulatory implications of estrogen and HER2 in breast cancer metastasis. Biochim Biophys Acta Acta 1866(5):130113. https://doi.org/10.1016/j.bbadis.2022.130113

54. Hanker AB, Brown BP, Meiler J, Marin A, Jayanthan HS, Ye D, Lin CC, Akamatsu H, Lee KM, Chatterjee S, Sudhan DR, Servetto A, Brewer MR, Koch JP, Sheehan H, He J, Lalani AS, Arteaga CL (2021) Co-occurring gain-of-function mutations in HER2 and HER3 modulate HER2/HER3 activation, oncogenesis, and HER2 inhibitor sensitivity. Cancer Cell 39(8):1099–1114. https://doi.org/10.1016/j.ccell.2021.06.001. (e8)

55. Makena MR, Rao R (2020) Subtype specific targeting of calcium signaling in breast cancer. Cell Calcium 85:102109. https://doi.org/10.1016/j.ceca.2020.102109

56. Lee D, Hong HJ (2021) Ca2+ signaling as the untact mode during signaling in metastatic breast cancer. Cancers. https://doi.org/10.3390/Cancers13061473

57. Zhang Y (2021) The root cause of drug resistance in HER2-positive breast cancer and the therapeutic approaches to overcoming the resistance. Pharmacol Therapeutics 218:107677. https://doi.org/10.1016/j.pharmthera.2020.107677

58. Kennedy SP, Han JZR, Portman N, Nobis M, Hastings JF, Murphy KJ, Latham SL, Cadell AL, Miladinovic D, Marriett GR, O’Donnell YEL, Shearer RF, Williams JT, Muzoz AG, Cox TR, Watkins DN, Saunders DN, Timpson P, Lim E, Kolch W, Croucher DR (2019) Targeting promiscuous heterodimerization of HER3 modulate HER2/HER3 activation, oncogenesis, and HER2 inhibitor sensitivity. Cancer Cell 39(8):1099–1114. https://doi.org/10.1016/j.cccel.2021.06.001. (e8)

59. Nussinov R, Zhang M, Maloney R, Liu Y, Tsai CJ, Jang H (2022) Allostery: allosteric cancer drivers and innovative allosteric drugs. J Mol Biol. https://doi.org/10.1016/j.jmb.2022.167569

60. Paratla BS, Chung JH, Williams CB, Yilmazel B, Petrosky W, Lin CC, Akamatsu H, Lee KM, Chatterjee S, Sudhan DR, Servetto A, Brewer MR, Koch JP, Sheehan H, He J, Lalani AS, Arteaga CL (2021) Co-occurring gain-of-function mutations in HER2 and HER3 modulate HER2/HER3 activation, oncogenesis, and HER2 inhibitor sensitivity. Cancer Cell 39(8):1099–1114. https://doi.org/10.1016/j.ccell.2021.06.001. (e8)

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.