

Multi-Omic Data Interpretation to Repurpose Subtype Specific Drug Candidates for Breast Cancer

Beste TURANLI^{5, 2, 6}, Kubra Karagoz⁷, Gholamreza Bidkhori^{2, 4}, Raghu Sinha⁸, Michael L. Gatza⁷, Mathias Uhlen⁹, Adil Mardinoglu^{2, 3, 4*}, Kazim Yalcin Arga^{1*}

¹Department of Bioengineering, Marmara University, Turkey, ²Science for Life Laboratory (SciLifeLab), Sweden, ³Department of Chemical and Biological Engineering, Chalmers University of Technology, Sweden, ⁴Dental Institute, King's College London, United Kingdom, ⁵Department of Bioengineering, Faculty of Engineering, Marmara University, Turkey, ⁶Department of Bioengineering, Istanbul Medeniyet University, Turkey, ⁷Department of Radiation Oncology, Rutgers Cancer Institute of New Jersey, United States, ⁸Department of Biochemistry and Molecular Biology, Pennsylvania State University, United States, ⁹Science for Life Laboratory, Royal Institute of Technology, Sweden

Submitted to Journal:
Frontiers in Genetics

Specialty Section:
Bioinformatics and Computational Biology

ISSN:
1664-8021

Article type:
Original Research Article

Received on:
23 Nov 2018

Accepted on:
17 Apr 2019

Provisional PDF published on:
17 Apr 2019

Frontiers website link:
www.frontiersin.org

Citation:
Turanli B, Karagoz K, Bidkhori G, Sinha R, Gatza ML, Uhlen M, Mardinoglu A and Arga K(2018) Multi-Omic Data Interpretation to Repurpose Subtype Specific Drug Candidates for Breast Cancer. *Front. Genet.* 10:420. doi:10.3389/fgene.2019.00420

Copyright statement:
© 2019 Turanli, Karagoz, Bidkhori, Sinha, Gatza, Uhlen, Mardinoglu and Arga. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

This Provisional PDF corresponds to the article as it appeared upon acceptance, after peer-review. Fully formatted PDF and full text (HTML) versions will be made available soon.

Provisional

Multi-Omic Data Interpretation to Repurpose Subtype Specific Drug Candidates for Breast Cancer

Running title: Breast Cancer Subtype Specific Drugs

Beste Turanli^{1,2,3,§}, Kubra Karagoz^{4,§}, Gholamreza Bidkhori^{2,7}, Raghu Sinha⁵, Michael L. Gatz⁴, **Mathias Uhlen²**, Adil Mardinoglu^{2,6,7,*}, Kazim Yalcin Arga^{1,*}

1 Department of Bioengineering, Marmara University, Istanbul, Turkey

2 Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, Sweden

3 Department of Bioengineering, Istanbul Medeniyet University, Istanbul, Turkey

4 Department of Radiation Oncology, Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey, USA

5 Department of Biochemistry and Molecular Biology, Penn State College of Medicine, Hershey, Pennsylvania, USA

6 Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

7 Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom

§ Both authors contributed equally to this work.

* Correspondence:

Kazim Yalcin Arga, Ph.D.

kazim.arga@marmara.edu.tr

Adil Mardinoglu, Ph.D.

adilm@scilifelab.se

Abstract

Triple-negative breast cancer (TNBC), which is largely synonymous with the basal-like molecular subtype, is the 5th leading cause of cancer deaths for women in the USA. The overall prognosis for TNBC patients remains poor given that few treatment options exist; including targeted therapies (not FDA approved), and multi-agent chemotherapy as standard-of-care treatment. TNBC like other complex diseases is governed by the perturbations of the complex interaction networks thereby elucidating the underlying molecular mechanisms of this disease in the context of network principles, which have the potential to identify targets for drug development. Here, we present an integrated “omics” approach based on the use of transcriptome and interactome data to identify dynamic/active protein-protein interaction networks in TNBC patients. We have identified three highly connected modules, EED, DHX9 and AURKA, which are extremely activated in TNBC tumors compared to both normal tissues and other breast cancer subtypes. Based on the functional analyses, we propose that these modules are potential drivers of proliferation and, as such, should be considered candidate molecular targets for drug development or drug repositioning in TNBC. Consistent with this argument, we repurposed steroids, anti-inflammatory agents, anti-infective agents, cardiovascular agents for patients with basal-like breast cancer. Finally, we have performed essential metabolite analysis on personalized genome-scale metabolic models and found that metabolites such as sphingosine-1-phosphate and cholesterol-sulfate have utmost importance in TNBC tumor growth.

Keywords: breast cancer, drug repositioning, non-cancer therapeutics, repurposing, basal subtype, personalized metabolic models

Introduction

Breast cancer is the most commonly diagnoses and second leading cause of cancer-related deaths in women in the United States with an estimated 268,600 new cases and 41,760 deaths in 2019 (Siegel et al., 2019). Although overall survival has significantly improved over the past several decades owing in part to advances in early diagnostic techniques and an increasing understanding of the underlying biological basis of the disease, which has led to improved treatment strategies. On a molecular level, breast cancer can be defined as five predominant molecular subtypes including the luminal A (LumA), luminal B (LumB) and Normal-like (NL) subtypes which are predominantly estrogen receptor (ER) and progesterone receptor (PR) positive; the HER2 Enriched subtype (HER2E) subtype; and basal-like tumors which are largely synonymous with Triple Negative Breast cancer (TNBC) and are ER/PR/HER2 negative. The considerable differences among these molecular subtypes are a consequence of dramatically altered genomic and proteomic profiles which manifest as changes in activated signaling networks (Gatza et al., 2014) and manifest as differences in risk factors, incidence, age, prognosis and response to treatment. Therefore, there is a clear need to develop reliable biomarkers and to identify potential drug targets in each molecular and clinical subtype (Perou et al., 2000; Curtis et al., 2012; Weigman et al., 2012; Gatza et al., 2014; Ciriello et al., 2015 and Mertins et al., 2016).

Basal-like breast cancers disproportionately affect younger women and women of African American decent. This subtype, which is highly concordant with TNBC, accounts for ~15-20% of diagnosed breast tumors but more than 1-in-4 breast cancer related deaths each year. This is, due in part, to the lack of effective therapeutic options for TNBC patients aside from multi-agent chemotherapy, which remains the standard-of-care treatment despite a limited and varied response among patients and the related toxic side-effects (Solzak et al., 2017). In this context, we and others, have proposed that systems level analyses can assist in revealing the underlying molecular mechanism of the diseases, discovery of biomarkers for specific subtypes, identification of subtype specific drug targets and reposition of drugs that can be used in effective treatment of patients (Mardinoglu et al., 2018; Mardinoglu and Nielsen, 2015; Turanli et al., 2018).

Publicly available “omics” datasets including The Cancer Genome Atlas (TCGA) (Ciriello et al., 2015), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) (Curtis et al., 2012) and the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (NCI-

CPTAC) (Mertins et al., 2016) enhance our understanding of the subtype specific molecular mechanisms of breast cancer. Moreover, integrative and comparative analysis of “omics” data together with network modeling provided a comprehensive platform for the drug repositioning and multi-target drug design (Kibble et al., 2015; Turanli et al., 2017; Vitali et al., 2016). A number of studies also combined genomic, transcriptomic, proteomic data with protein-protein interaction networks (PPINs) and identified putative druggable candidates in breast cancer by analyzing topological features of the reconstructed networks (Karagoz et al., 2015; Li et al., 2018; Liu et al., 2017; Nuncia-Cantarero et al., 2018). These bioinformatics pipelines have their own power through decreasing the number of candidate therapeutic targets/drugs and proposing potential treatment strategies for subsets of breast cancer patients.

The overall prognosis for patients with basal-like breast cancer remains poor and there is an urgent need to identify molecular targets to develop effective therapeutic strategies. To take advantage of the extensive publicly available “omics” data, we integrated transcriptome with interactome data and calculated network entropy for each protein-protein interactions (PPIs) to identify the dynamic states in basal-like breast cancer. Our analyses identified modules as systems biomarkers at gene expression level and these networks were confirmed at the proteomic level. Importantly, functional annotation and analysis of module activity scores demonstrated that these modules were subtype specific. Using these models essential metabolites and drug candidates were identified within the context of basal-like specific modules. Collectively, these analyses suggest that the proposed strategy incorporating multi-omics analyses of human breast tumors has the capacity to define novel signaling networks and link these features to existing therapeutic opportunities.

Methods

Data Collection

Throughout the study, we integrated multi-omics data including genomics, transcriptomics and proteomics using network analysis (Table 1). TCGA data were obtained from (<https://gdac.broadinstitute.org/>), METABRIC and CPTAC data were collected from supplementary files of these studies. At transcriptomic level, gene expressions were obtained from two major initiatives presenting RNA-Seq data from the TCGA study and microarray data from the METABRIC study. Normalized gene expression values for 179 basal and 852 non-basal like breast cancer samples ($n = 1031$) from TCGA, and 331 basal and 1665 non-basal samples from the METABRIC project ($n =$

1992) were used in integrative analysis. At the protein level, two different sources were used, i) expression data of 160 basal and 777 non-basal like samples ($n = 937$) in TCGA, using Reverse Phase Protein Array (RPPA)- based analysis of 226 proteins, and ii) expression data of 19 basal and 58 non-basal like samples ($n = 77$) from CPTAC which performed comprehensive mass-spectrometry methods including around 10,000 proteins (Mertins et al., 2016).

[Table 1]

RNA sequencing data from TCGA ($n = 1031$) were used as a discovery set whereas, microarray data from METABRIC and proteomic data from TCGA and/or CPTAC were used as independent validation data sets in the study (Table 1).

Differential Interactome

To obtain a differential view of human interactome between two different phenotypes, and to identify PPIs that are up- or down-regulated in each phenotype relative to the other one, we used the gene expression profiles of interacting protein pairs and recruited the differential interactome analysis as previously described (Ayyildiz et al., 2017). For this purpose, normalized gene expression profiles from TCGA (179 basal-like and 852 non-basal like samples) were categorized into three levels: high (1), moderate (0), and low (-1) expression levels according to comparison of each gene expression with the average expression within each sample. The probability distributions for any possible co-expression profile of gene pairs (encoding proteins interacting with each other) were estimated, and the uncertainty of determining whether or not a PPI is encountered in a phenotype was estimated through an entropy formulation. In order to define possible PPIs, we used the high confidence human PPIs (Karagoz et al., 2016), comprising 147,923 interactions among 13,213 proteins. Karagoz and coworkers assembled and integrated physical protein-protein interactions of Homo sapiens from six publicly available databases including BioGRID (Chatr-Aryamontri et al., 2015), DIP (Salwinski, 2004), IntAct (Orchard et al., 2014), HIPPIE (Schaefer et al., 2012), HomoMINT (Persico et al., 2005) and HPRD (Prasad et al., 2009). Then, PPIs analyzed the differential view of human interactome between the basal and non-basal subtypes of breast cancer; P-value <0.05 was considered statistically significant for these analyses.

Differentially expressed genes and proteins

Both differentially expressed genes (DEGs) between 179 basal and 852 non-basal samples in TCGA cohort, and differentially expressed proteins (DEPs) between 19 basal and 61 non-basal samples in CPTAC cohort were identified by using the Significance Analysis of Microarrays (SAM) method

implemented in R software (Gómez-Pozo et al., 2017; Hu et al., 2016; Tusher et al., 2001). False Discovery Rate (FDR), adjusted p-value was set at p-value < 0.05, and fold changes >1 between basal-like and non-basal samples were considered as up-regulated differentially expressed genes and proteins in basal tumors.

Module extraction from basal specific networks

Basal subtype specific PPI networks were constructed by using the differential interactome from basal-like tumors. The interactions associated with proteins corresponding to DEGs that are up-regulated in basal-like tumors were identified and used to construct up-regulated PPI networks specific to basal-like breast cancer. The networks were visualized by using Cytoscape software (version 3.4.0) (Lopes et al., 2011). The topological analysis of the networks was performed via CytoNCA plugin of Cytoscape (version 2.1) (Tang et al., 2015). Two different topological metrics, degree, which is defined by the number of adjacent nodes of a node in the network, and betweenness centrality, which characterizes nodes by how often they occur on the shortest path between two other nodes in the network, were simultaneously employed to define hub nodes. Hub nodes with higher degree or betweenness values were reported to have significant roles in cellular signal trafficking and could be potential candidate biomarkers or drug targets. Modules were identified as highly connected subnetworks within up-regulated networks. Gene expression data from METABRIC were used for validation of the gene expression modules in basal-like breast cancer.

Functional annotation

Functional enrichment analysis associated with the three protein-protein interaction modules were analyzed using QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity).

Module Activity

In order to convert the identified EED, AURKA, and DHX9 modules to gene expression signatures that can be used to quantify pathway activity in a given sample from independent datasets, the module was converted to a gene list and the mean expression of unweighted gene list was used to calculate a pathway score. For these studies, a score was calculated for each sample in the TCGA (discovery) and METABRIC cohort (validation). Analysis of variance (ANOVA) tests were used to quantify differences between the EED-module, DHX9-module and AURKA-module activity scores between breast cancer subtypes in each dataset. A Student's t-test was used to evaluate levels

of EED, DHX9e and AURKA signature scores between adjacent normal breast tissue and basal-like tumors. To infer the functional roles of these modules, a panel of 270 experimentally derived gene expression signatures that predict activation of various oncogenic signaling pathways, was performed by integrating gene expression data as described previously (Gatza et al., 2014). To identify the association of the modules with oncogenic pathways, a Spearman's rank correlation was used between oncogenic pathway activity scores and EED, DHX9 and AURKA activity scores.

Module specific drug repositioning

To identify small molecules that can potentially reverse gene expression of basal-like tumors, we utilized the Library of Integrated Network-based Cellular Signatures (LINCS) - L1000 data which includes gene expression data from ~ 50 human cell line in response to ~20,000 compounds (Campillos et al., 2008). We queried basal-like specific module genes which are all up-regulated and down-regulated DEGs (Fold Change<0.2) signatures as input. We used the L1000CDS2 (Duan et al., 2016) search engine, which contains 30,000 significant signatures that were processed from the LINCS L1000 data, to identify small molecule signatures associated with each module. The identified drugs were ranked based on their scores and the top 50 were acquired for each query. Drugs were checked through literature review and publicly available datasets such as CTD (Davis et al., 2017) and KEGG DRUG (Kanehisa et al., 2012) to identify those that were previously investigated within the context of breast cancer.

Subtype specific essential metabolites

We next acquired 917 personalized genome scale metabolic models (GEMs) of breast cancer patients (Uhlen et al., 2017). We analyzed each patient GEM to identify essential metabolites for tumor growth by removing the reactions in which the metabolite functions as substrate regardless of compartmentalization (Bidkhori et al., 2018). Next, we categorized personalized models based on clinical information to create subtype-specific patient metabolic models and found the percentage of subtype representation of each metabolite. A Fisher exact test was applied to identify statistically significant difference between basal-like and non-basal-like (*i.e.* all other tumors) for each metabolite. Significant difference between subtypes was determined based on a P-value of <0.05.

Results

Basal-like Subtype Specific PPI Elucidation via Differential Interactome

Cancer cells are characterized by increase in network entropy comprising high uncertainty, pathway redundancy and promiscuous signaling resulting from intra-sample heterogeneity. Recently, a differential interactome network analysis were presented to show the uncertainties of PPIs in ovarian cancer (Ayyildiz et al., 2017). In this study, we employed differential interactome algorithm utilizing the entropy concept using a comprehensive gene expression data and human PPI network to reveal the heterogeneity among the breast cancer subtypes (*i.e.* basal-like vs. non-basal-like). To do so, we categorized the expression of each gene and for each patient using 179 basal and 852 non-basal-like samples from TCGA into three classes as -1, 0, 1. These classes were then integrated with a high confident PPI network (Karagoz et al., 2016) and the frequency of PPIs estimated for both basal-like and non-basal-like tumors. Using a 95% confidence interval ($p < 0.05$), significant values < 0.2 and > 0.8 as well as corresponding H values < 0.7 were calculated for each class. As a result, 3,002 interactions among 1,652 proteins were considered significant across the entire dataset. These analyses identified 2,291 interactions among 1,391 proteins as being significantly activated in basal-like tumors whereas 712 interactions among 612 proteins were identified as significant in non-basal-like samples; 351 proteins were common across both subgroups of tumors (Table S1).

Since low entropy presents low uncertainty, low redundancy and deterministic signaling resulting with homogeneity in the population, we next focused on the basal-like subtype to identify low entropy interactions ($H < 0.1$). These analyses identified the EED protein network which is defined by 82 interactions within the group of 98 proteins. Importantly, the lowest entropy profile of the EED centroid network only identified an interaction with one protein (CTCF) in non-basal-like tumors. We further identified a sub-set of proteins, excluding 351 common signatures evident in both basal-like and non-basal-like tumors to identify a basal-like subtype specific network (Table S2). All differential interactome networks and basal-like subtype specific networks were delimited regarding up-regulated genes in the basal-like tumors through 2-class SAM analysis (Tusher et al., 2001)(Table S3). Through the integration of SAM analysis and the above detailed differential interactome framework, we identified three significant modules: EED centroid module, covering relatively low entropy PPIs (Figure 1A); the DHX9 centroid module, covering mixed of low and high entropy PPIs (Figure 1B); and the AURKA centroid module, covering relatively high entropy PPIs (Figure 1C).

Further analyses of the EED, DHX9 and AURKA modules determined that genes included in EED-module play roles in cyclins and cell cycle regulation ($p= 6.1e-19$), cell cycle: G1/S checkpoint regulation ($p= 3.5e-18$), regulation of cellular mechanics by calpain protease ($p= 1.6e-11$), aryl hydrocarbon receptor signaling ($p= 4.3e-11$), apoptosis signaling ($p= 7.0e-10$), TWEAK signaling ($p= 1.8e-09$), and GADD45 signaling ($p= 4.3e-9$). In contrast, the genes in DHX9-module contribute to mTOR signaling ($p= 4.1e-06$), regulation of eIF4 and p70S6K signaling ($p= 7.9e-06$), EIF2 signaling ($p= 7.2e-05$), Inflammasome pathway ($p= 1.4e-04$), assembly of RNA Polymerase I Complex ($p= 1.1e-03$), DNA double strand break repair ($p= 1.8e-03$) and cell cycle ($p= 3.5e-03$) while the genes associated with the AURKA-module are involved in DNA damaged-induced 14-3-3A signaling ($p= 1.8e-10$), mitotic roles of Polo like kinase ($p= 2.1e-09$), role of CHK proteins in cell cycle checkpoint control ($p= 6.0e-08$), ATM signaling ($p= 9.3e-07$) and mismatch repair ($p= 3.1e-06$), role of BRCA1 in DNA damage response ($p= 1.3e-05$), and cell cycle ($p= 9.8e-05$). These data suggest that each module represent a unique aspect of basal-like breast cancer signaling. Some of these pathways such as TWEAK signaling, apoptosis signaling, mTOR signaling, ATM signaling showed that the chemotherapy targeted pathways are also activated in basal-like tumors in which chemotherapy is the front-line treatment option, nowadays (Figure S1).

Proteomic analysis of basal specific modules

We next reconstructed PPI networks using transcriptome data and validated our findings at proteomic level by leveraging orthogonal genomic and proteomic data from the TCGA and CPTAC projects. Transcriptome data from 937 sample was compared to RPPA analysis of the same samples to assess the relationship between each network at the 226 proteins and phosphoproteins from TCGA. Likewise the gene expression data from a subset of 77 of these samples was used to examine the relationship between each module and 10,062 proteins and phosphoproteins using mass spectrometry-derived proteomic data from the CPTAC project. First, we used CPTAC proteome data to compare each gene to its corresponding protein across all basal-like tumors and assessed correlation for those pairs. Overall, 52.6%-64.5% of the mRNA-protein pairs showed statistically significant positive Spearman correlations ($P < 0.05$) when changes in mRNA abundance were compared to changes in relative protein abundance. These proteins in basal-like samples are shown in darker colors in Figure 1A-C. Then, we identified differentially expressed proteins between basal-like and non-basal-like samples by using both RPPA and CPTAC data. Although RPPA data has limited number of proteins, we identified

several up-regulated proteins including CCNE1, RAF1, SRC, CDK1, EGFR, MYC, MYH9, PCNA associated with the EED-module. Similarly, NDGR1 and CCNB1 were associated with the DHX9 and AURKA modules, respectively. We also analyzed differentially expressed proteins between basal-like and non-basal-like tumors by using CPTAC data which is more comprehensive than RPPA data and it covered 69.4% - 56.4% of the module genes and 29.4% - 36.4% of these proteins were identified as being up-regulated in basal-like tumors (Table S3).

Modules as basal specific signatures

In order to quantitatively assess the activity of each modular in each patient sample, we next generated a gene expression signature on the basis of median expression of each gene in the module. This strategy was used to calculate a module score for each sample in the TCGA (discovery set) and METABRIC (validation set) datasets. We then quantitatively evaluated the differences in the module activities across breast cancer subtypes by an ANOVA test. These analyses demonstrated that EED ($P = 1.13e-244$), DHX9 ($P = 2.4e-236$) and AURKA ($P = 2.05e-175$) activity was highest in basal-like tumors in the TCGA cohort (Figure 2A-C); these findings were validated by analysis of module activity in the METABRIC cohort (Figure 2D-F). Finally, we determined that the EED ($P=1.06e-96$), DHX9 ($P=2.44e-85$), and AURKA modules ($P=6.61e-127$) were expressed at significantly higher levels in basal-like tumors compared to adjacent normal tissue (Figure 3A-C).

Functionality of basal specific modules

We examined the functional roles of these modules by exploring the correlations with a series of previously published gene expression signatures which are capable of measuring oncogene or tumor suppressor pathway activity, aspects of the tumor microenvironment and other tumor characteristics. We identified pathway activities, which were positively (or negatively) correlated with module activities using a Spearman rank correlation to assess the relationship between pathway activity and the EED, DHX9 or AURKA module activity scores. As expected, our data recapitulated known characteristics of basal-like tumors including low hormone receptor signaling and high expression of proliferation pathway activity and demonstrated the relationship between these characteristics and the expression of each module (i.e., EED, DHX9 and AURKA). Moreover, these modules were associated with multiple indicators of proliferation including, RB_LOSS, RB_LOH and bMYB highly correlated with these module activities as well as RAS, PIK3CA, β -catenin, MYC and HER1_Cluster 1, HER1_Cluster 2 and HER1_Cluster 3 signatures (Figure 4A). Consistent results were obtained using the METABRIC data (Figure 4B). Importantly, we also confirmed the ability of the transcriptomic

module signatures to assess the functional roles of EED, DHX9 and AURKA modules by exploring relationships between the module signature scores and protein expression. Analysis of RPPA data from basal-like samples confirmed that these tumors with high module scores have significantly higher levels of CHK1, CHK2, CDK1, Cyclin B1, Cyclin E1, FOXM1 and PCNA protein expression consistent with their role in cell cycle regulation and proliferation (Figure 4C).

Drug repositioning based on basal subtype specific modules

As discussed above, the EED, DHX9 and AURKA modules were converted to gene expression signatures on the basis of up-regulated genes specific to each module; as would be expected down-regulated genes (Fold Change<0.2) were common for all modules. We asked the question of whether each module/signature identified potential therapeutic opportunities. To do so, we queried each gene signatures separately against the LINCS database L1000CDS2 (Duan et al., 2016) in order to identify concordant and discordant patterns of gene expression between each module and gene expression profiles associates with drug-induced and/or disease expression. Drugs that resulted in a gene expression profile that was negatively correlated with each module were identified and selected as potential candidate compounds that had the potential to reverse the activity of each module network that was associated with basal-like tumors (Figure S2). Since we have demonstrated specificity of the modules to basal-like tumors, we may also propose that our candidate drugs are specifically targeting basal-like tumors.

After removing the duplicated drugs from query results, we found that EED and AURKA modules were associated with 41 candidate compounds while DHX9 was associated with 31 candidate small molecules. Networks comprising drug candidates and modules were found to have 114 interaction between three modules and 80 drugs (Figure 5A). The 80 identified drugs were categorized as molecular inhibitors (23%), anti-neoplastic agents (15%), heterocyclic compounds (10%), anti-infective agents (6%) or steroids (6%). Moreover, a number of the drugs specific to each module (as well as some common candidates) were also identified in each drug category (Figure 5B). There are at least 19 approved, 24 investigational and 6 experimental drugs listed in DrugBank (version 5.1.1), however there are perturbagens used in L1000 platform without detailed information (Table S4).

Nine of the drugs including selumetinib, trametinib and several other investigational drugs were common to each of the three modules. Consistent with our results, selumetinib as MEK inhibitor was

reported to suppresses cell proliferation, migration and trigger apoptosis, following G1 arrest in TNBC cells (Zhou et al., 2016). Furthermore, the MEK inhibitor, trametinib is also a therapy of significant interest for the treatment of TNBC since TNBC cell lines have been shown to be especially sensitive to this drug (Davis et al., 2014; Jing et al., 2012). Finally, we noted some overlap between drugs associated with each module. For instance, the three common drugs (i.e., wortmannin, mestanolone, NVP-TAE684) are associated with both the EED and AURKA modules while twelve drugs (i.e. radicicol, lapatinib, alvocidib, zileuton, geldanamycin, exemestane) are consistent between the EED and DHX9 module (Table S4). Intriguingly, ten of our candidate drugs were previously associated with the breast cancer based on at least one of the sources including CTD, KEGG Drug, Clinical Trials and scientific literature (Table 2).

[Table 2]

Since EED module has the lowest entropy level between PPIs, we focused on 17 drug candidates which are only related to EED module in addition to common drugs. Three of these drugs are anti-neoplastic agents and five of them are unknown however, others belonged to steroids (BRD-A94793051, Oxymetholone, Testosterone propionate), PLK inhibitor (BI-2536), heterocyclic compounds (BRD-K17953061, GDC-0980, TG101348), cardiovascular agents (BRD-K52080565, S-2500), anti-inflammatory (oxaprozin) and anti-infective agents (5-fluorocytosine).

Essential metabolites and anti-metabolites as drug candidates

GEMs reconstructed for different cancer tissues have been used for characterization of metabolic modifications; disease stratification and determination of drug targets using essential genes or metabolites (Agren et al., 2012; Bidkhori et al., 2018; Folger et al., 2011). To address this question, we first identified a panel of 917 personalized genome scale metabolic models (GEMs) derived from breast cancer patients (Uhlen et al., 2017). We then categorized each GEMs based on clinical information to create subtype-specific patient metabolic models. These models were then used to identify subtype-specific metabolites essential for tumor growth. After categorization of BCS, percentage of abundance for each essential metabolite was calculated. Significant alteration between the abundance of basal-like and non-basal BCS were determined based on FDR adjusted P-value threshold ($P\text{-adj} < 0.05$) (Table S5). These analyses identified 27 essential metabolites (Table S6); eleven were significantly enriched in basal-like tumors while the remaining 16 were enriched in non-basal-like samples. Further analyses

determined that the essential metabolites that are expressed at higher levels in basal-like tumors were associated with steroid metabolism, biotin metabolism, nucleotide metabolism, sphingolipid metabolism and transport. Conversely, the identified metabolites down-regulated in basal-like samples were involved in beta-alanine metabolism, arginine and proline metabolism, cysteine and methionine metabolism, and carnitine shuttle (Figure 6).

Discussion

The dynamics of cells are regulated by PPIs and properties of networks such as entropy provide information about the current state of the network. Given that cancer cells are reported to have an increase in network entropy, several previous studies have integrated gene expression data with PPI network information to compute the energetic state of cancer cells by calculating entropy (Rietman et al., 2016; Teschendorff et al., 2015 and West et al., 2012). Likewise, a number of studies have used a network-based entropy approach to identify disease specific PPIs as biomarker candidates, proliferative and prognostic markers in lung and breast cancer, as well as to demonstrate the association between network entropy and tumor initiation, progression, and anticancer drug responses (Ayyildiz et al., 2017; Lecca and Re, 2015; Varadan and Anastassiou, 2006; Xiong et al., 2010; Banerji et al., 2013 and Cheng et al., 2016).

The current study employed a novel multi-omics-based approach to integrate genomic, proteomic and metabolomic tumor data. Our analyses of mRNA expression data identified three highly connected modules which are centered on the activation of the EED, DHX9 and AURKA signaling networks. These data demonstrated that each module is highly activated in basal-like tumors compared to non-basal-like tumors as well as adjacent normal tissues. Importantly, by analyzing proteome data, our results confirmed the correlation between the expression of genes and proteins that comprise each identified module. By analyzing the association between module expression and oncogenic signaling using a panel of more than 250 gene expression signatures, we were able to assess the functional relationship of these modules with known oncogenic and signaling features. Our results demonstrated the correlation between EED, DHX9 and AURKA module activity and proliferative oncogenic pathways including RAS, PI3K and Rb/E2F signaling in basal-like tumors. Consistent with these results, CHK1, CHK2, CDK1, Cyclin B1, Cyclin E1, and PCNA protein expression levels were identified higher in tumors with high module scores. Through integrated

analyses, we identified candidate drugs to target three modules by drug repositioning. Utilizing multiple omics data including genome, transcriptome, and interactome, we repurposed 519 agents for breast cancer by incorporating data from the LINCS project (Duan et al., 2016) into our analyses. In another drug repositioning study, five of the identified repurposed candidate agents showed superior therapeutic indices compared to doxorubicin in *in vitro* assays in basal sub-type cell line (SUM149) in addition to luminal cell line (MCF7) (Chen et al., 2016). Moreover, Lee et al. (2016) developed an integrative approach for drug repositioning using the expression signature, chemical structure, target signatures and LINCS data. They applied this strategy to identify candidate anti-cancer drugs for breast cancer (Lee et al., 2016). Although there are previous computational drug-repositioning efforts that utilized LINCS as mentioned, the methodologies are focused on breast cancer regardless of disease heterogeneity and subtype information.

In addition, our analyses identified subtype-specific metabolites, including several specific to basal-like tumors, which may provide opportunity to design anti-metabolite drugs for breast cancer. Results in essential metabolite analysis emphasized sphingolipids and steroid metabolism for basal-like breast cancer. Sphingolipid levels in breast cancer tissue are generally higher than normal breast tissue and bioactive sphingolipids, such as sphingosine-1-phosphate (S1P) has many cellular functions like cell proliferation, migration, survival, immune cell trafficking, and angiogenesis which are related to cancer progression and metastasis (Nagahashi et al., 2016). However, sphingosine and S1P were recently highlighted as important for signaling mechanisms in metastatic TNBC and its targeted therapy (Maiti et al., 2017). A recent lipidomics profiling of TNBC tumors also supported sphingolipids as potential prognostic markers and associated enzymes as candidate therapeutic targets (Purwaha et al., 2018) in parallel to our results.

TNBC was associated with expression pattern of 2-pore domain potassium (K2p) channels which enable background leak of potassium (K^+). Differential expression on K2p-channels may be suggested as a novel molecular marker related to potassium levels in basal like BCS (Dookeran et al., 2017). In another study, expression of calcium-activated potassium (SK4) channels were also associated with TNBC and cellular functions such as proliferation, migration, apoptosis, and EMT processes (Zhang et al., 2016).

Breast cancer is known as one of the malignancies in which steroid hormones drive cellular proliferation (Capper et al., 2017). As steroid metabolism associated metabolite, cholesterol sulfate, is quantitatively the most important known sterol sulfate in human plasma and may play a role in cell

adhesion, differentiation and signal transduction (Strott and Higashi, 2003). Given that current standard-of-care therapy for TNBC is largely limited to multi-agent cytotoxic chemotherapy, the potential of incorporating identified repurposed drugs and/or targeting identified modules and/or metabolites represents a potential therapeutic opportunity for a subset of patients with limited treatment options.

Given these data, we would propose that the strategy outlined here can be used to repurposed drugs in order to identify novel candidate compounds or drugs to be utilized in not only monotherapy but also in combination therapy for the treatment of TNBC. Consistent with this argument, a number of the candidate drugs identified by our analyses have been incorporated in ongoing clinical trials. For instance, TNBC patients who received pre-operative sequential epirubicin and cyclophosphamide followed by docetaxel were found to have a significant increase in pathological complete response (PCR) (Warm et al., 2010). Although a great number of pre-clinical trials will be necessary to support the *in silico* modeling detailed in the current study prior to initiation of clinical trials, a large number of identified candidates have significant *in vitro* and *in vivo* support to indicate that these represent potential therapeutic opportunities. For instance, drugs inhibiting cyclin-dependent kinases (CDKs), including the CDK9 inhibitor alvociclib have been reported to be effective against TNBC (Ocana and Pandiella, 2015).

Erlotinib also showed anti-tumor effect on TNBC in a xenograft model (Ueno and Zhang, 2011). Likewise, targeting the MET and EGFR receptors, which regulate RAS/ERK and PI3K/AKT signaling, resulted in improved treatment compared to monotherapy (Linklater et al., 2016).

The current study has defined a novel approach to identify breast cancer subtype-specific network modules via a network entropy-based approach. This strategy can be used for both the identification of potentially novel signaling networks but also to identify subtype-specific therapeutic opportunities through drug repositioning. Importantly, we demonstrate that this approach can be used to link signaling networks with and subtype-specific essential metabolites which represents additional therapeutic opportunities. As such, the current studies have the potential enhancing the impact of existing therapeutics or multi-agent therapeutic strategies by identifying novel drug/target networks in the context of breast cancer and in breast cancer subtypes. On a broader scale, this strategy is largely applicable to all cancer and disease type/subtypes where multi-platform genomic, proteomic,

and metabolomic data exists and thus represents a potential strategy to define novel signaling networks unique to each disease and identify disease/subtype-specific therapeutic strategies.

Author Contributions Statement

BT and KK designed the study. GB performed essential metabolite analysis. BT and KK performed the all other analyses. BT and KK wrote the manuscript. MLG, RS, AM, MU and KYA supervised the work and contributed to the manuscript during the progress of the work. All authors reviewed and approved the final manuscript.

Funding

This work was supported by Knut and Alice Wallenberg Foundation and Marmara University Scientific Projects Committee (BAPKO) in the context of the project FEN-C-DRP-250816-0417. R00CA166228 from the National Cancer Institute of the National Institutes of Health and V2016-013 from the V Foundation for Cancer Research to MLG and DHFS-18PPC-024 from the New Jersey Commission for Cancer Research to KK. We also thank the TUBITAK BİDEB 2211 National Doctoral Fellowship Program provided to BT.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Agren, R., Bordel, S., Mardinoglu, A., Pornputtapong, N., Nookaew, I., and Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* doi:10.1371/journal.pcbi.1002518.
- Ayyildiz, D., Gov, E., Sinha, R., and Arga, K. Y. (2017). Ovarian Cancer Differential Interactome and Network Entropy Analysis Reveal New Candidate Biomarkers. *Omi. A J. Integr. Biol.* 21, 285–294. doi:10.1089/omi.2017.0010.
- Bidkhorı, G., Benfeitas, R., Elmas, E., Kararoudi, M. N., Arif, M., Uhlen, M., et al. (2018). Metabolic network-based identification and prioritization of anticancer targets based on expression data in hepatocellular carcinoma. *Front. Physiol.* 9. doi:10.3389/fphys.2018.00916.
- Campillos, M., Kuhn, M., Gavin, A. C., Jensen, L. J., and Bork, P. (2008). Drug target identification using side-effect similarity. *Science (80-.)*. doi:10.1126/science.1158140.
- Capper, C. P., Rae, J. M., and Auchus, R. J. (2017). The metabolism, analysis, and targeting of steroid hormones in breast and prostate cancer. 7, 149–164. doi:10.1007/s12672-016-0259-0.The.
- Catania, C., De Pas, T. M., Pelosi, G., Manzotti, M., Adamoli, L., Nolè, F., et al. (2006). Erlotinib-induced breast cancer regression. *Ann. Pharmacother.* doi:10.1345/aph.1H252.
- Chatr-Aryamontri, A., Breitkreutz, B. J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., et al.

- 499 (2015). The BioGRID interaction database: 2015 update. *Nucleic Acids Res.*
500 doi:10.1093/nar/gku1204.
- 501 Chen, H.-R., Sherr, D. H., Hu, Z., DeLisi, C., Jin, G., Fu, C., et al. (2016). A network based approach
502 to drug repositioning identifies plausible candidates for breast cancer and prostate cancer. *J Natl*
503 *Compr Canc Netw* 8, 1–21. doi:10.1016/j.jacc.2007.01.076.White.
- 504 Ciriello, G., Gatza, M. L., Beck, A. H., Wilkerson, M. D., Rhie, S. K., Pastore, A., et al. (2015).
505 Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell*.
506 doi:10.1016/j.cell.2015.09.033.
- 507 Curtis, C., Shah, S. P., Chin, S. F., Turashvili, G., Rueda, O. M., Dunning, M. J., et al. (2012). The
508 genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups.
509 *Nature*. doi:10.1038/nature10983.
- 510 Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., King, B. L., McMorran, R., et al. (2017). The
511 Comparative Toxicogenomics Database: Update 2017. *Nucleic Acids Res.* 45, D972–D978.
512 doi:10.1093/nar/gkw838.
- 513 Davis, S. L., Eckhardt, S. G., Tentler, J. J., and Diamond, J. R. (2014). Triple-negative breast cancer:
514 Bridging the gap from cancer genomics to predictive biomarkers. *Ther. Adv. Med. Oncol*.
515 doi:10.1177/1758834013519843.
- 516 Dookeran, K. A., Zhang, W., Stayner, L., and Argos, M. (2017). Associations of two-pore domain
517 potassium channels and triple negative breast cancer subtype in the Cancer Genome Atlas:
518 Systematic evaluation of gene expression and methylation. *BMC Res. Notes* 10.
519 doi:10.1186/s13104-017-2777-4.
- 520 Duan, Q., Reid, S. P., Clark, N. R., Wang, Z., Fernandez, N. F., Rouillard, A. D., et al. (2016).
521 L1000CDS2: LINCS L1000 characteristic direction signatures search engine. *npj Syst. Biol.*
522 *Appl.* doi:10.1038/npjbsa.2016.15.
- 523 Finn, R. S., Martin, M., Rugo, H. S., Jones, S., Im, S.-A., Gelmon, K., et al. (2016). Palbociclib and
524 Letrozole in Advanced Breast Cancer. *N. Engl. J. Med.* doi:10.1056/NEJMoa1607303.
- 525 Folger, O., Jerby, L., Frezza, C., Gottlieb, E., Rupp, E., and Shlomi, T. (2011). Predicting selective
526 drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* 7, 1–10.
527 doi:10.1038/msb.2011.35.
- 528 Gámez-Pozo, A., Trilla-Fuertes, L., Berges-Soria, J., Selevsek, N., López-Vacas, R., Díaz-Almirón,
529 M., et al. (2017). Functional proteomics outlines the complexity of breast cancer molecular
530 subtypes. *Sci. Rep.* doi:10.1038/s41598-017-10493-w.
- 531 Gatza, M. L., Silva, G. O., Parker, J. S., Fan, C., and Perou, C. M. (2014). An integrated genomics
532 approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nat. Genet.*
533 doi:10.1038/ng.3073.
- 534 Giampaglia, M., Chiuri, V. E., Tinelli, A., De Laurentiis, M., Silvestris, N., and Lorusso, V. (2010).
535 Lapatinib in breast cancer: Clinical experiences and future perspectives. *Cancer Treat. Rev.*

- doi:10.1016/S0305-7372(10)70024-4.
- Goss, P. E., Ingle, J. N., Pritchard, K. I., Ellis, M. J., Sledge, G. W., Budd, G. T., et al. (2013). Exemestane versus anastrozole in postmenopausal women with early breast cancer: NCIC CTG MA.27 - A randomized controlled phase III trial. *J. Clin. Oncol.* doi:10.1200/JCO.2012.44.7805.
- Gschwantler-Kaulich, D., Grunt, T. W., Muhr, D., Wagner, R., Kölbl, H., and Singer, C. F. (2016). HER specific TKIs exert their antineoplastic effects on breast cancer cell lines through the involvement of STAT5 and JNK. *PLoS One* 11, 1–14. doi:10.1371/journal.pone.0146311.
- Hu, J., Ye, F., Cui, M., Lee, P., Wei, C., Hao, Y., et al. (2016). Protein profiling of bladder urothelial cell carcinoma. *PLoS One* 11, 1–16. doi:10.1371/journal.pone.0161922.
- Jing, J., Greshock, J., Holbrook, J. D., Gilmartin, A., Zhang, X., McNeil, E., et al. (2012). Comprehensive Predictive Biomarker Analysis for MEK Inhibitor GSK1120212. *Mol. Cancer Ther.* doi:10.1158/1535-7163.MCT-11-0505.
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, 109–114. doi:10.1093/nar/gkr988.
- Karagoz, K., Sevimoglu, T., and Arga, K. Y. (2016). Integration of multiple biological features yields high confidence human protein interactome. *J. Theor. Biol.* 403, 85–96. doi:10.1016/j.jtbi.2016.05.020.
- Karagoz, K., Sinha, R., and Arga, K. Y. (2015). Triple negative breast cancer: a multi-omics network discovery strategy for candidate targets and driving pathways. *OMICS* 19, 115–30. doi:10.1089/omi.2014.0135.
- Kibble, M., Saarinen, N., Tang, J., Wennerberg, K., Mäkelä, S., and Aittokallio, T. (2015). Network pharmacology applications to map the unexplored target space and therapeutic potential of natural products. *Nat. Prod. Rep.* doi:10.1039/c5np00005j.
- Lecca, P., and Re, A. (2015). Detecting modules in biological networks by edge weight clustering and entropy significance. *Front. Genet.* doi:10.3389/fgene.2015.00265.
- Lee, H., Kang, S., and Kim, W. (2016). Drug repositioning for cancer therapy based on large-scale drug-induced transcriptional signatures. *PLoS One* 11, 1–17. doi:10.1371/journal.pone.0150460.
- Li, C., Luo, L., Wei, S., and Wang, X. (2018). Identification of the potential crucial genes in invasive ductal carcinoma using bioinformatics analysis. *Oncotarget* 9, 6800–6813. doi:10.18632/oncotarget.23239.
- Li, J., Li, F., Wang, H., Wang, X., Jiang, Y., and Li, D. (2012). Wortmannin reduces metastasis and angiogenesis of human breast cancer cells via nuclear factor- κ B-dependent matrix metalloproteinase-9 and interleukin-8 pathways. *J. Int. Med. Res.* doi:10.1177/147323001204000305.
- Linklater, E. S., Tovar, E. A., Essenburg, C. J., Turner, L., Madaj, Z., Winn, M. E., et al. (2016).

- 573 Targeting MET and EGFR crosstalk signaling in triple-negative breast cancers. *Oncotarget*.
574 doi:10.18632/oncotarget.12065.
- 575 Liu, Y., Yin, X., Zhong, J., Guan, N., Luo, Z., Min, L., et al. (2017). Systematic identification and
576 assessment of therapeutic targets for breast cancer based on genome-wide RNA interference
577 transcriptomes. *Genes (Basel)*. doi:10.3390/genes8030086.
- 578 Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q., Bader, G. D., et al. (2011). Cytoscape
579 Web: An interactive web-based network browser. in *Bioinformatics*
580 doi:10.1093/bioinformatics/btq430.
- 581 Maiti, A., Takabe, K., and Hait, N. C. (2017). Metastatic triple-negative breast cancer is dependent
582 on SphKs/S1P signaling for growth and survival. 32, 85–92. doi:10.1016/j.cellsig.2017.01.021.
- 583 Mardinoglu, A., Boren, J., Smith, U., Uhlen, M., and Nielsen, J. (2018). Systems biology in
584 hepatology: approaches and applications. *Nat. Rev. Gastroenterol. Hepatol.* 15, 365–377.
585 doi:10.1038/s41575-018-0007-8.
- 586 Mardinoglu, A., and Nielsen, J. (2015). New paradigms for metabolic modeling of human cells.
587 *Curr. Opin. Biotechnol.* 34, 91–97. doi:10.1016/j.copbio.2014.12.013.
- 588 Mertins, P., Mani, D. R., Ruggles, K. V., Gillette, M. A., Clauser, K. R., Wang, P., et al. (2016).
589 Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*.
590 doi:10.1038/nature18003.
- 591 Murphy, C. G., and Dickler, M. N. (2015). The Role of CDK4/6 Inhibition in Breast Cancer.
592 *Oncologist* 20, 483–490. doi:10.1634/theoncologist.2014-0443.
- 593 Nagahashi, M., Tsuchida, J., Moro, K., Hasegawa, M., Tatsuda, K., Woelfel, I. A., et al. (2016). High
594 levels of sphingolipids in human breast cancer. *J. Surg. Res.* 204, 435–444.
595 doi:10.1016/j.jss.2016.05.022.
- 596 Nuncia-Cantarero, M., Martinez-Canales, S., Andrés-Pretel, F., Santpere, G., Ocaña, A., and Galan-
597 Moya, E. M. (2018). Functional transcriptomic annotation and protein–protein interaction
598 network analysis identify NEK2, BIRC5, and TOP2A as potential targets in obese patients with
599 luminal A breast cancer. *Breast Cancer Res. Treat.* doi:10.1007/s10549-017-4652-3.
- 600 Ocana, A., and Pandiella, A. (2015). Targeting oncogenic vulnerabilities in triple negative breast
601 cancer: biological bases and ongoing clinical studies. *Oncotarget*.
602 doi:10.18632/oncotarget.14731.
- 603 Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., et al. (2014). The
604 MIntAct project - IntAct as a common curation platform for 11 molecular interaction databases.
605 *Nucleic Acids Res.* doi:10.1093/nar/gkt1115.
- 606 Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C., et al. (2000).
607 Molecular portraits of human breast tumours. *Nature*. doi:10.1038/35021093.
- 608 Persico, M., Ceol, A., Gavrila, C., Hoffman, R., Florio, A., and Cesareni, G. (2005). HomoMINT: An

- 609 inferred human network based on orthology mapping of protein interactions discovered in
610 model organisms. *BMC Bioinformatics*. doi:10.1186/1471-2105-6-S1-S21.
- 611 Prasad, T. S. K., Kandasamy, K., and Pandey, A. (2009). “Human Protein Reference Database and
612 Human Proteinpedia as Discovery Tools for Systems Biology,” in doi:10.1007/978-1-60761-
613 232-2_6.
- 614 Purwaha, P., Gu, F., Piyarathna, D. W. B., Rajendiran, T., Ravindran, A., Omilian, A. R., et al.
615 (2018). Unbiased lipidomic profiling of triple-negative breast cancer tissues reveals the
616 association of sphingomyelin levels with patient disease-free survival. *Metabolites* 8, 1–14.
617 doi:10.3390/metabo8030041.
- 618 Rietman, E. A., Platig, J., Tuszynski, J. A., and Lakka Klement, G. (2016). Thermodynamic
619 measures of cancer: Gibbs free energy and entropy of protein–protein interactions. *J. Biol. Phys.*
620 42, 339–350. doi:10.1007/s10867-016-9410-y.
- 621 Salwinski, L. (2004). The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res.*
622 doi:10.1093/nar/gkh086.
- 623 Schaefer, M. H., Fontaine, J. F., Vinayagam, A., Porras, P., Wanker, E. E., and Andrade-Navarro, M.
624 A. (2012). Hippie: Integrating protein interaction networks with experiment based quality
625 scores. *PLoS One*. doi:10.1371/journal.pone.0031826.
- 626 Siegel, R. L., Miller, K. D., and Jemal, A. (2018). Cancer statistics, 2018. *CA. Cancer J. Clin.* 68, 7–
627 30. doi:10.3322/caac.21442.
- 628 Solzak, J. P., Atale, R. V., Hancock, B. A., Sinn, A. L., Pollok, K. E., Jones, D. R., et al. (2017). Dual
629 PI3K and Wnt pathway inhibition is a synergistic combination against triple negative breast
630 cancer. *npj Breast Cancer*. doi:10.1038/s41523-017-0016-8.
- 631 Strott, C. A., and Higashi, Y. (2003). Cholesterol sulfate in human physiology. *J. Lipid Res.* 44,
632 1268–1278. doi:10.1194/jlr.R300005-JLR200.
- 633 Tang, Y., Li, M., Wang, J., Pan, Y., and Wu, F. X. (2015). CytoNCA: A cytoscape plugin for
634 centrality analysis and evaluation of protein interaction networks. *BioSystems*.
635 doi:10.1016/j.biosystems.2014.11.005.
- 636 Teschendorff, A. E., Banerji, C. R. S., Severini, S., Kuehn, R., and Sollich, P. (2015). Increased
637 signaling entropy in cancer requires the scale-free property of protein interaction networks. *Sci.*
638 *Rep.* doi:10.1038/srep09646.
- 639 Turanli, B., Grøtli, M., Boren, J., Nielsen, J., Uhlen, M., Arga, K. Y., et al. (2018). Drug
640 Repositioning for Effective Prostate Cancer Treatment. 9, 1–20. doi:10.3389/fphys.2018.00500.
- 641 Turanli, B., Gulfidan, G., and Arga, K. Y. (2017). Transcriptomic-Guided Drug Repositioning
642 Supported by a New Bioinformatics Search Tool: geneXpharma. *Omi. A J. Integr. Biol.* 21,
643 584–591. doi:10.1089/omi.2017.0127.
- 644 Tusher, V. G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the
645 ionizing radiation response. *Proc. Natl. Acad. Sci.* doi:10.1073/pnas.091062498.

- 646 Ueno, N. T., and Zhang, D. (2011). Targeting EGFR in triple negative breast cancer. *J. Cancer* 2,
647 324–328. doi:10.7150/jca.2.324.
- 648 Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., et al. (2017). A pathology
649 atlas of the human cancer transcriptome. *Science* (80-.). 357. doi:10.1126/science.aan2507.
- 650 Varadan, V., and Anastassiou, D. (2006). Inference of disease-related molecular logic from systems-
651 based microarray analysis. *PLoS Comput. Biol.* doi:10.1371/journal.pcbi.0020068.
- 652 Vitali, F., Cohen, L. D., Demartini, A., Amato, A., Eterno, V., Zambelli, A., et al. (2016). A network-
653 based data integration approach to support drug repurposing and multi-Target therapies in triple
654 negative breast cancer. *PLoS One*. doi:10.1371/journal.pone.0162407.
- 655 Warm, M., Kates, R., Große-Onnebrink, E. M., Stoff-Khalili, M., Hoopmann, M., Mallmann, P., et
656 al. (2010). Impact of tumor biology, particularly triple-negative status, on response to pre-
657 operative sequential, dose-dense epirubicin, cyclophosphamide followed by docetaxel in breast
658 cancer. *Anticancer Res.*
- 659 West, J., Bianconi, G., Severini, S., and Teschendorff, A. E. (2012). Differential network entropy
660 reveals cancer system hallmarks. *Sci. Rep.* doi:10.1038/srep00802.
- 661 Xiong, J., Liu, J., Rayner, S., Li, Y., and Chen, S. (2010). Protein-protein interaction reveals
662 synergistic discrimination of cancer phenotype. *Cancer Inform.*
- 663 Zhang, P., Yang, X., Yin, Q., Yi, J., Shen, W., Zhao, L., et al. (2016). Inhibition of SK4 potassium
664 channels suppresses cell proliferation, migration and the epithelial-mesenchymal transition in
665 triple-negative breast cancer cells. *PLoS One* 11, 1–17. doi:10.1371/journal.pone.0154471.
- 666 Zhang, Y. G., Du, Q., Fang, W. G., Jin, M. L., and Tian, X. X. (2008). Tyrphostin AG1478 supresses
667 proliferation and invasion of human breast cancer cells. *Int. J. Oncol.*
668 doi:10.3892/ijo_00000045.
- 669 Zhou, Y., Lin, S., Tseng, K. F., Han, K., Wang, Y., Gan, Z. hua, et al. (2016). Selumetinib suppresses
670 cell proliferation, migration and trigger apoptosis, G1 arrest in triple-negative breast cancer
671 cells. *BMC Cancer*. doi:10.1186/s12885-016-2773-4.

672

673 **Table 1.** Validation and discovery sets used in this study

Data type	Data portal	‘Omic’ level	Number of basal samples	Number of non-basal samples	Set type
Gene expression levels	TCGA	Transcriptomic	179	852	Discovery

Gene expression levels	METABRIC	Transcriptomic	331	1655	Validation
Protein expression levels	CPTAC	Proteomic	19	58	Validation
Protein expression levels	TCGA	Proteomic	160	777	Validation

674

675 **Table 2.** Various drug candidates that already associated with breast cancer via different sources

Drug Name	Literature Evidence	CTD	KEGG Drug	Clinical Trials
epirubicin	(Warm et al., 2010)	✓	✓	NCT00176488
erlotinib	(Catania et al., 2006)			NCT01650506
lapatinib	(Giampaglia et al., 2010)	✓		NCT00694252
exemestane	(Goss et al., 2013)	✓	✓	NCT00810797
wortmannin	(Li et al., 2012)	✓		
alvocidib	(Murphy and Dickler, 2015)	✓		NCT00039455
tyrphostin ag 1478	(Zhang et al., 2008)	✓		
canertinib	(Gschwantler-Kaulich et al., 2016)	✓		NCT00051051
danazol	(Coombes et al., 1983)	✓		
palbociclib	(Finn et al., 2016)	✓	✓	NCT02513394

676

677 **Figure Legends**

678 **Figure 1.** Basal like breast cancer specific highly connected protein-protein interaction modules. (A)
679 EED module, (B) DHX9 module, (C) AURKA module. Darker nodes indicate the statistically
680 significant positive correlations between mRNA and protein pairs. **Thicker edges indicate lowest**
681 **entropy levels between interacting pairs.**

Figure 2. The pattern of basal like breast cancer specific modules activity across breast cancer subtypes. (A-C) EED, DHX9 and AURKA modules are highly activated in basal like tumors by using TCGA cohort-discovery set. (D-F) EED, DHX9 and AURKA modules are highly activated in basal like tumors by using METABRIC cohort-validation set.

Figure 3. The activity levels of basal like breast cancer specific modules in normal and basal like tumors. (A) EED module, (B) DHX9 module, (C) AURKA module.

Figure 4. The functional analysis of basal like breast cancer specific modules (A) The activity of oncogenic pathways correlated with module activities in TCGA cohort-discovery set. (B) The activity of oncogenic pathways correlated with module activities in METABRIC cohort-validation set. (C) High module activities characterized by high expression of cell cycle proteins.

Figure 5. Drug repositioning for basal like breast cancer specific modules (A) Module-drug networks (B) Drug categories of module specific drugs and common drugs among modules

Figure 6. Significant essential metabolite differences between non-basal and basal like breast cancer specific personalized metabolic models and their associated pathways.

Supplementary File Legends

Figure S1. Functional enrichment results of the genes involved in each basal-like module using Ingenuity Pathway Analysis (IPA).

Figure S2. The gene signatures of three modules separately on L1000CDS2 for elucidating the differences and similarities between drug-induced expression profiles and disease expression. Drugs were ranked for each module and we elected drugs that showed negatively correlated action mechanisms with the module gene signatures to reverse disease gene expression.

Table S1. Non-basal and basal-like subtype specific PPI elucidation via differential interactome

Table S2. Three modules for only in basal-like subtype specific networks

Table S3. Statistical values of differential expressed genes and proteins

Table S4. Information of repurposed module specific and common drug signatures

Table S5. Essential metabolite and personalized model matrix and breast cancer categorization of personalized GEMs

Table S6. Significant essential metabolites between non-basal and basal-like breast cancer

Figure 01.JPEG

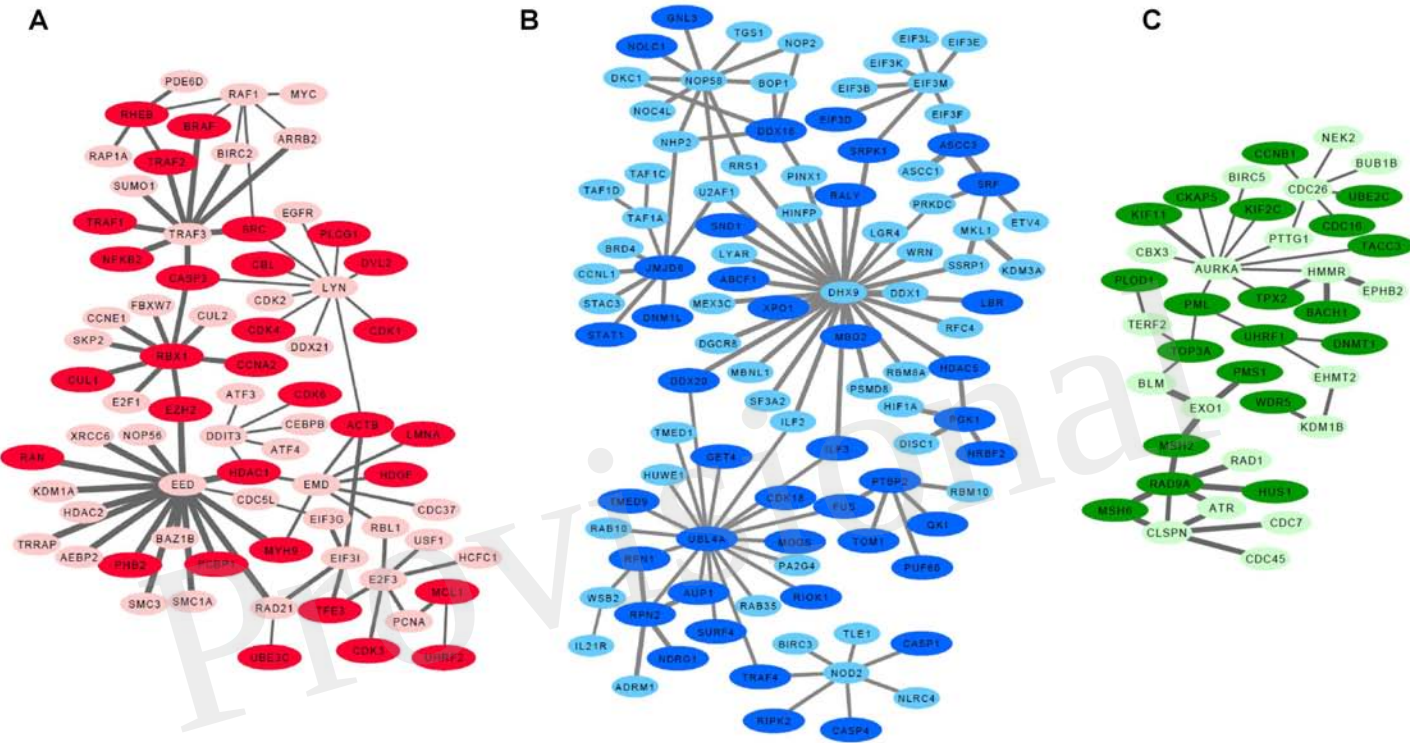


Figure 02.JPEG

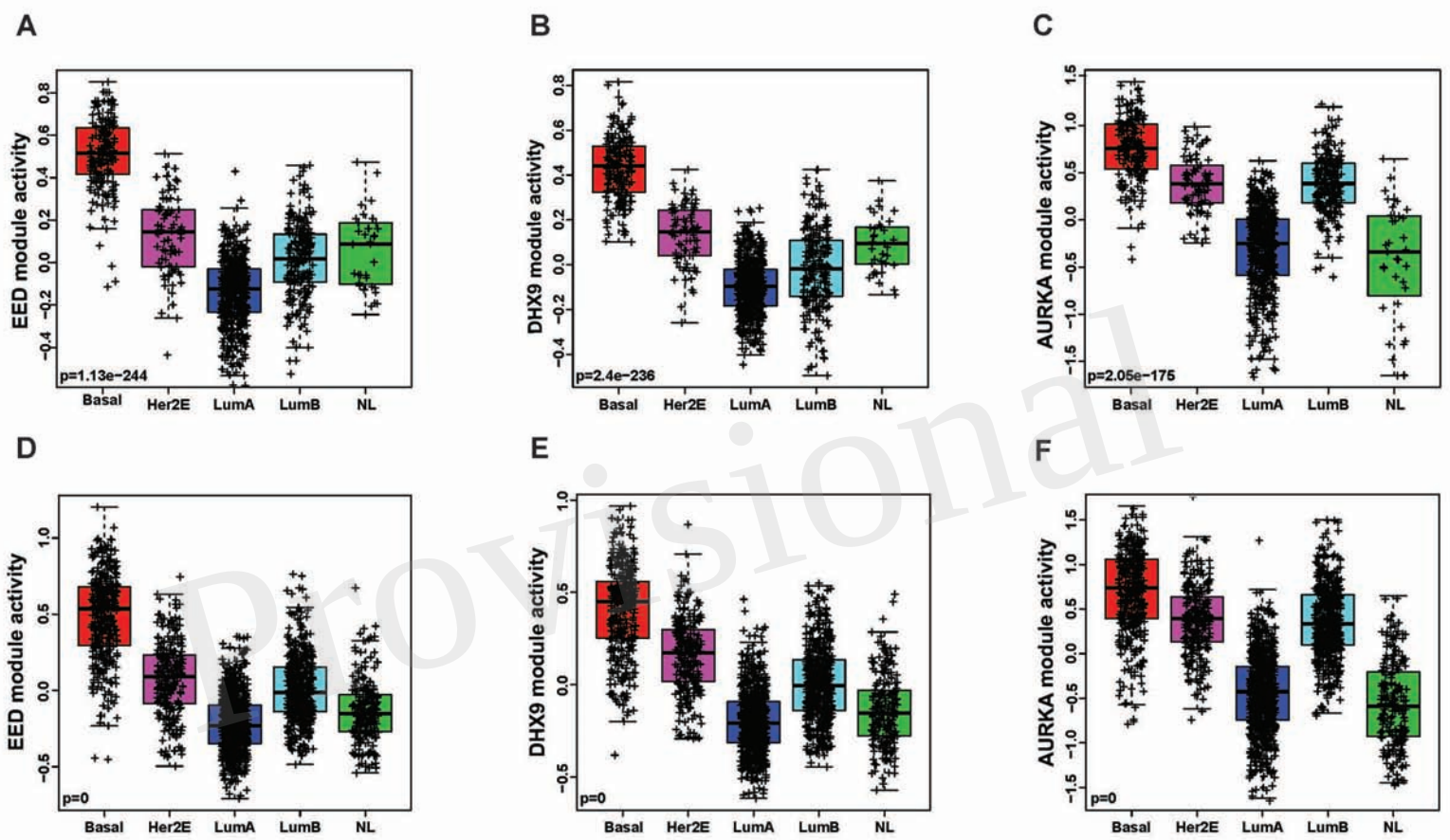


Figure 03.JPEG

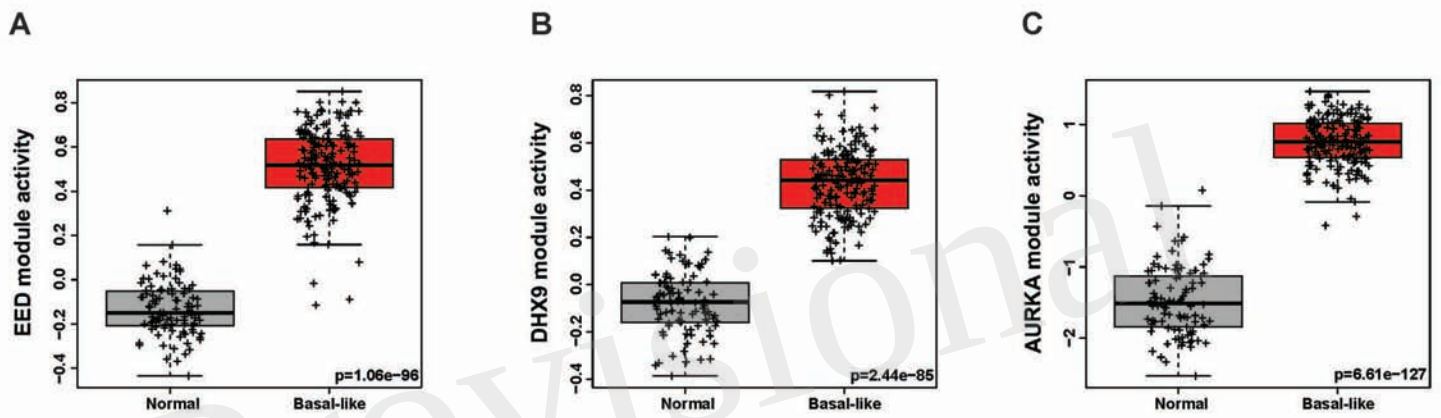


Figure 04.JPEG

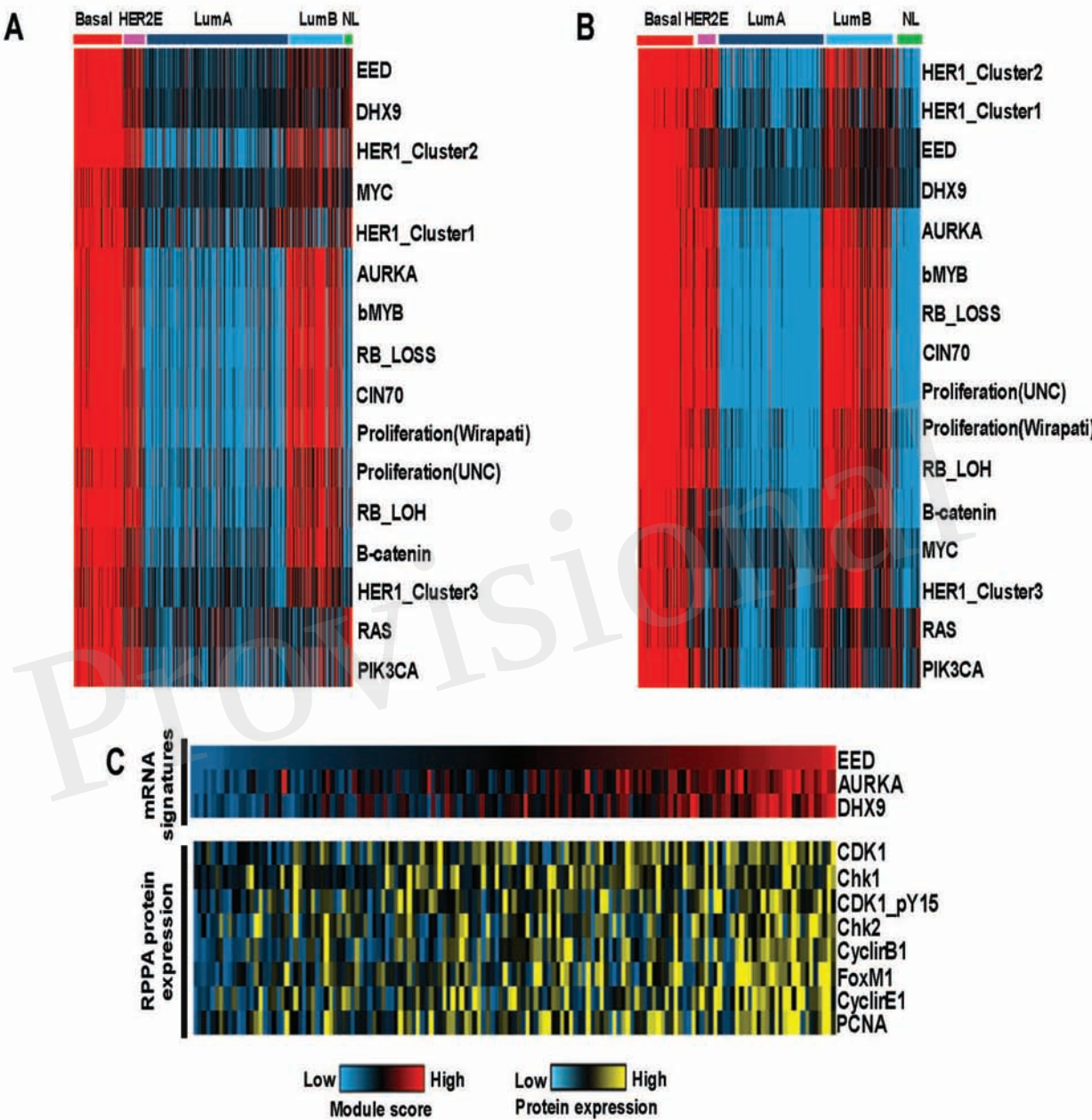


Figure 05.JPEG

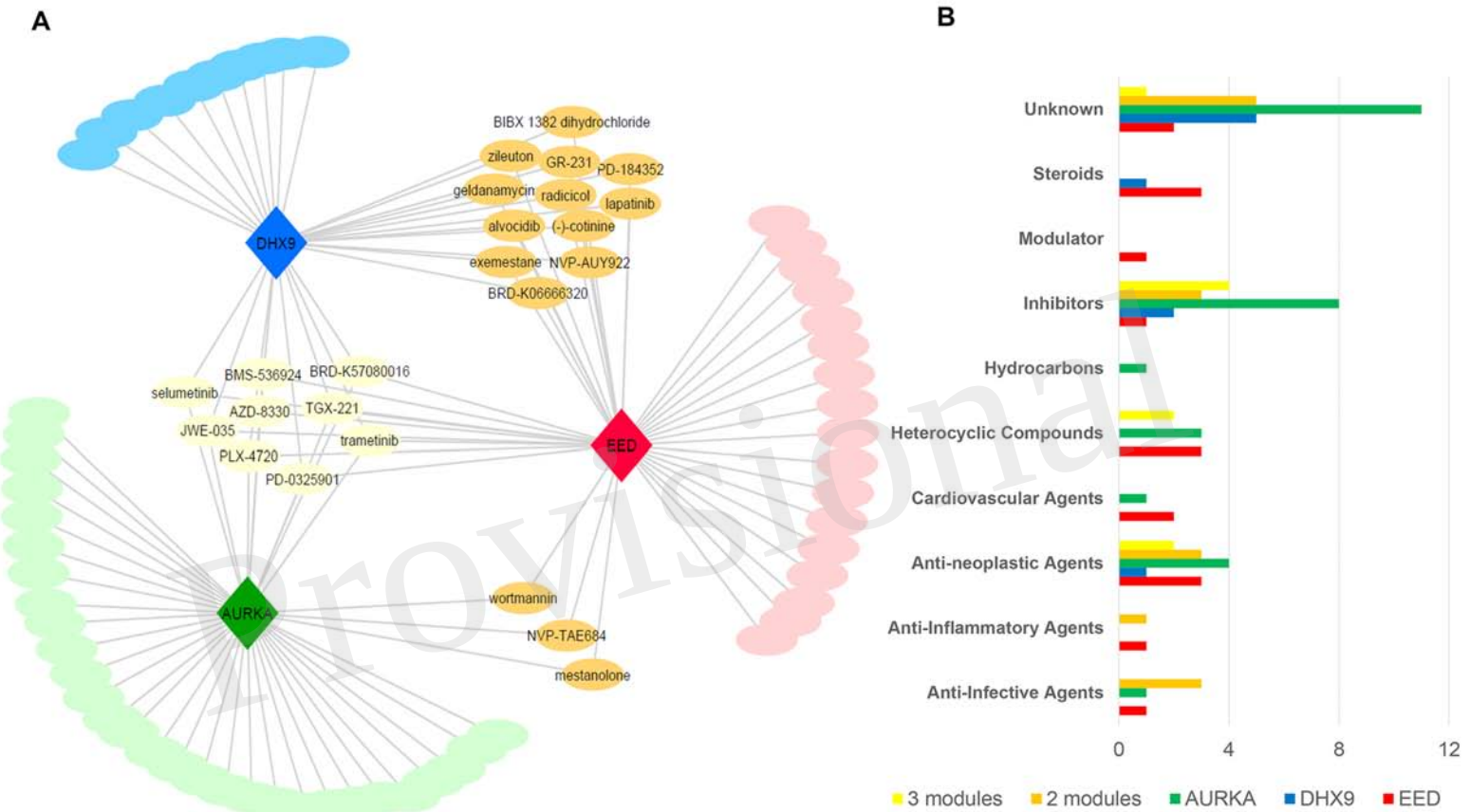


Figure 06.JPEG

