Bioinformatics methods for the discovery of network signatures towards understanding of underlying molecular mechanisms and investigation of candidate drugs

Marilena Bourdakou*

National and Kapodistrian University of Athens,
Department of Informatics and Telecommunications
kmarlen1988@gmail.com

Abstract. Systemic approaches are essential in the discovery of disease-specific genes, offering a different perspective and new tools on the analysis of several types of molecular relationships, such as gene co-expression or protein-protein interactions. However, due to lack of experimental information, this analysis is not fully applicable. The aim of this study is to reveal the multi-potent contribution of statistical network inference methods in highlighting significant genes and interactions. We have investigated the ability of statistical co-expression networks to highlight and prioritize genes for breast cancer in terms of: (i) classification efficiency, (ii) gene network pattern conservation, (iii) indication of involved molecular mechanisms and (iv) systems level momentum to drug repurposing pipelines. We have found that statistical network inference methods are advantageous in gene prioritization, are capable to contribute to meaningful network signature discovery, give insights regarding the disease-related mechanisms and boost drug discovery pipelines from a systems point of view.

KEYWORDS: network inference methods, gene expression data, co-expression networks, molecular mechanisms, drug repurposing

1 Introduction

Breast cancer is a major public health problem, since it remains the most frequently diagnosed cancer and ranked second as a cause of death in women population. Outbreaks are increasing in most countries, despite current efforts have been made to avoid the disease [1]. This happens because breast cancer is a complex disease with many contributing factors affecting the progress of the disease. Despite the fact that many studies have been conducted, neither the exact etiology of the breast cancer, nor

* Dissertation Advisor: Emmanouil Sagkriotis, Associate Professor.

the mechanisms behind the heterogeneity from patient to patient are known. For this, the diagnosis and the treatment of breast cancer remain a both challenging and fascinating task [2].

With the rapid development of genome-wide gene expression profiling methodologies, many bioinformatics data analysis pipelines have been developed to identify breast cancer related genes and discover gene signatures for prognosis and treatment prediction. However, since breast cancer is a complex disease, it should be determined not only by individual genes, but also by the coordinated effect of numerous genes [3]. The information behind gene interaction networks is of great importance due to the fact that all cellular functions are regulated by gene patterns, where the presence or absence of an interaction may cause the emergence of a disease.

Network analysis and graph theory support the study of interactions among relatively large number of genes in order to conclude to large lists of statistically significant genes [4]. Several bioinformatics tools prioritize genes by combining gene expression data with the protein-protein interaction (PPI) network through a random walk approach to enrich the candidate genes and finally re-rank them. The majority of these methods necessitate prior knowledge to re-rank genes accordingly. However, due to the absence of functional characterizations for a significant number of genes, these approaches are not fully applicable [5]. Genome-wide association studies (GWAS) have recognized DNA variants that are related to common complex diseases but for many of these studies, functional associations between genes and diseases are unknown [6].

In order to overcome this hurdle, several network inference methods have been adopted to construct statistical co-expression networks, based on gene expression data. These network inference approaches identify groups of genes that are highly correlated in expression levels to multiple samples according to a variety of correlation functions and algorithms [7].

In this study, we investigate the ability of statistical co-expression networks to highlight and prioritize significant genes at four different breast cancer molecular subtypes, including Luminal A, Luminal B, HER2 and Triple Negative as well as at four different disease stages (I-IV) in terms of: (i) classification efficiency, (ii) gene subnetwork conservation, (iii) involved molecular mechanisms investigation and (iv) potential boost to drug repurposing pipelines.

Specifically, we have used mRNA gene expression microarray data concerning Breast Invasive Carcinoma, retrieved from The Cancer Genome Atlas – TCGA (http://gdac.broadinstitute.org/runs/stddata__latest/samples_report/BRCA.html), to reconstruct 17 different networks (twelve based on mathematical correlation and six based on the literature) of the top differentially expressed genes. Using a mathematical function that combines gene expression data with custom networks, we prioritized genes based on each network. Furthermore, in order to investigate the quality of each prioritized gene list, we elucidated the impact of each one over sample discrimination, by applying a hold out validation scheme using the TCGA data as training set and a number of Breast cancer datasets from the transcriptional data repository Gene Expression Omnibus GEO (http://www.ncbi.nlm.nih.gov/geo/) as test sets. Using the network inference method that performed the highest classification score, we constructed co-expression networks for all datasets (train and test sets) to

find the most significant gene-gene links that recur in all networks. With the proposed pipeline, we concluded to breast cancer specific network patterns per subtype and stage. Analyzing each pattern we concluded in specific mechanisms per subtype and stage related to cellular community (cell communication, focal adhesion), signaling (in terms of extracellular matrix and cytokine receptor interactions), cell growth and death (cell cycle), immune system (including complement and coagulation cascades and toll like receptor signaling pathway), endocrine system (ppar and adipocytokine signaling pathway), carbohydrate, lipid and amino acid metabolism (glycolysis/gluconeogenesis, fatty acid and glycerolipid metabolism, bile acid biosynthesis, as well as tyrosine, phenylalanine, glycine, serine, threonine metabolism) and xenobiotics biodegradation and metabolism (3 chloroacylic acid and 1,2 methylnaphthalene degradation, metabolism of xenobiotics by cytochrome p450). Interestingly, all the derived network patterns include genes found in breast cancer specific regions of significant somatic copy number alterations (SCNA) [8]. Finally, the genes from the conserved network patterns were used in a drug repurposing pipeline, revealing drugs that have the potential to suppress breast cancer specifically for each molecular subtype and stage of the disease.

2 Methods

2.1 Datasets and preprocessing

Reference Set: TCGA mRNA (microarray) gene expression data for Breast Invasive Carcinoma cases are obtained from Firehose (http://gdac.broadinstitute.org/). From a total 587 samples (526 primary solid tumor samples and 61 primary solid normal samples - 17.814 genes), we have selected a subset of tumor data containing information regarding breast cancer staging, HER2, ER and PR status with their corresponding normal samples. Concerning staging, selection of stages I, II, III and IV was performed based on the clinical records accompanying each sample, while for the case of subtyping, the selection was performed as followed: (i) Luminal A for ER+ and/or PR+ , HER2- , (ii) Luminal B for ER+ and/or PR+ , HER2+ , (iii) HER2 for ER- , PR- , HER2+ and (iv) Triple Negative for ER-, PR-, HER2-. The eight distinct TCGA dataset were statistically analyzed with the LIMMA R package in order to select the Differentially Expressed Genes (DEGs) in breast cancer samples compared with the normal ones. The top 1000 genes of each sub-dataset with p-value < 0.01 and q-value < 0.01, sorted based on their log Fold Change absolute value, were used as the reference sets in our analysis.

Validation Sets: We searched in Gene Expression Omnibus accessed on 19 November 2015 using queries, in order to find microarray datasets for each breast cancer stage and subtype. Finally we concluded in 7 independent datasets from which, one contain clinical feature from both stages and subtypes.

2.2 Network Reconstruction

We have examined 3 major categories of statistical network inference methods: (i) Mutual Information-based methods, (ii) Correlation-based methods and (iii) Tree-

based methods. Also, we utilized Biological information-based network methods and one ensemble scheme using all statistical network inference methods. More specifically, we have used 11 network inference methods to reconstruct gene co-expression networks for each dataset including the top 1000 DEGs from the TCGA dataset. All the selected methods are implemented in R packages. Six mutual information based methods are used (Aracne.a, Aracne.m, CLR, MRNET, MRNETB and C3NET), four correlation based (Lasso, Adaptive Lasso, GeneNET and WGCNA) and one tree based – Genie3. Furthermore, we have used the Cytoscape platform and more specifically the GeneMania plug-in to reconstruct a gene network using biological information. The GeneMANIA algorithm inside the homonymous plugin obtains information from a combination of potentially heterogeneous sources. This plug-in uses a large data set unifying functional networks comprising approximately 800 networks for 6 organisms including Homo sapiens. Using the Homo sapiens network we constructed a sub – network for the top 1000 DEGs from the TCGA dataset merging 5 Network types (Co-expression, co-localization, physical interaction, genetic interaction and pathway). We also used the manually curated human signaling network [9] based on the literature since 2005 (Version 6). The signaling network contains more than 6,000 proteins and 63,000 relations from different data sources including BioCarta, CST Signaling pathways, Pathway Interaction database (PID), iHOP, and many review papers on cell signaling. The signaling network comprised of three different relations (activation, inhibition and physical interactions). This network was used not only as a whole network (all relations), but was further divided into three sub-networks based on the different relation types.

Finally, we have created a union unique gene list based on the different top 100 re-ranked gene lists from the eleven statistical network inference methods. Based on the highest frequency of the appearance, the minimum mean rank and the minimum coefficient of variation across all statistical network inference methods we selected the top 100 genes.

2.3 Gene re-ranking using underlying networks

In order to investigate the influence of the reconstructed 17 gene networks (12 statistically and 5 biologically inferred) on gene prioritization, we applied a method that allows for a custom network selection combining the log fold change absolute values with the selected underlying network in order to re-rank the initial DEGs. The basic idea of the method is the reconciliation of the gene expression values taking into account an underlying gene network. This approach is available as part of the Biorithm software in the Network Reconciliation package [10].

2.4 Scoring the ranked gene lists

Each method is scored according to the maximum achieved mean classification accuracy across datasets, modified by two multiplicative weights: w_n that is related to the number of genes required for the maximum accuracy and w_{cv} that is related to the coefficient of variation (CV) of the classification accuracy along the first 100 genes.

Finally, we calculated the average score of each method across the stages and the subtypes.

3 Results

3.1 Evaluation of gene re-ranking through a classification scheme

The top 1000 re-ranked gene lists for each subtype and stage, along with the initially ranked list, gave us a total number of 18 ranked gene lists. In order to evaluate each list, we elucidated the impact of the top 100 genes from each list over sample discrimination, by applying a hold out validation scheme. More precisely, we employed a Support Vector Machine (SVM) – based classification scheme using the e1071 R package through sequential gene selection of the first 100 genes, using as Train set the expression values of each top 100 gene list from the reference set (TCGA) and as Test sets the expression values of the same top 100 genes from a number of independent GEO datasets (discovery sets) available for each subtype and stage. We followed the same procedure for each top 100 gene lists and we calculated the mean classification accuracy from the discovery datasets in a sequential gene selection manner. Figures 1 and 2 show the box plots of the mean classification accuracies of the top 100 sequential genes for each network approach using the Page Rank reconciling method for each stage and subtype. We observe that, in most cases the median classification performances of the top 100 gene lists from network inference methods are either better or equivalent compared to the median performance of the initial gene list.

Each ranking method is scored according to the maximum achieved mean classification accuracy across datasets, evaluated by a score (see Methods). The maximum average score for breast cancer stages and subtypes was achieved by Genenet network inference method and MRNETB, respectively. For this reason we adopted them for the rest of our analysis. It is worth mentioning that the selected statistical network inference methods achieved a higher or equivalent score compared to the initial ranking in most cases.

3.2 Deriving a common Network Pattern

We applied the Genenet and MRNETB network inference methods to reconstruct gene co-expression networks for each of the available dataset for each stage and subtype. In order to highlight any common gene network pattern, we found the common edges across all datasets. We performed a dynamic filtering to keep only the highly weighted gene - gene links. Finally, we came up with 205 genes-nodes and 216 edges for Stage I, 561 genes-nodes and 896 edges for Stage II, 289 nodes and 380 edges for Stage III and 132 genes-nodes and 169 edges for Stage IV. As far as subtypes are concerned, we came up with 196 genes-nodes and 872 edges for Triple Negative, 201 genes-nodes and 272 edges for Luminal A, 155 genes-nodes and 305 edges for Luminal B and 544 genes-nodes and 573 edges for HER2.

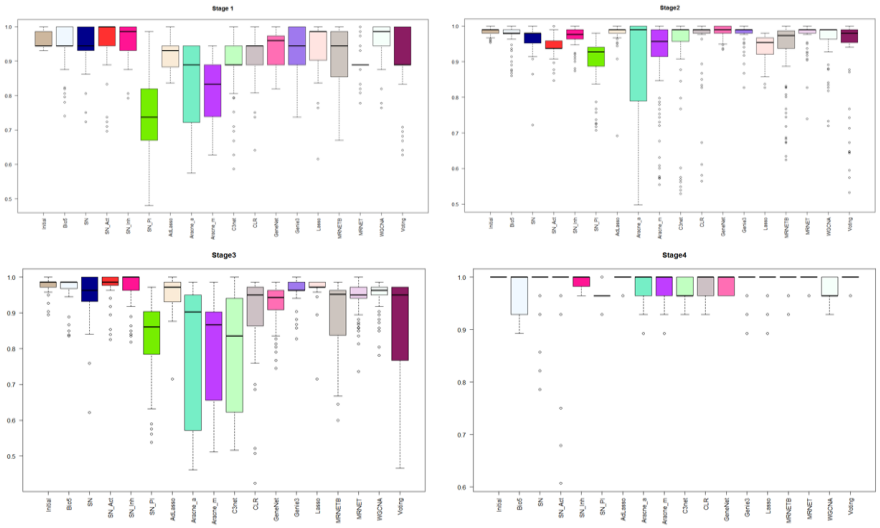

Figure 1: Box plots of the mean accuracy rates of the top 100 sequential genes from all ranked and re-ranked gene lists in combination with PageRank reconciling method, using hold out validation with train set the TCGA expression values and test set the expression values from GEO independent datasets for breast cancer stages.

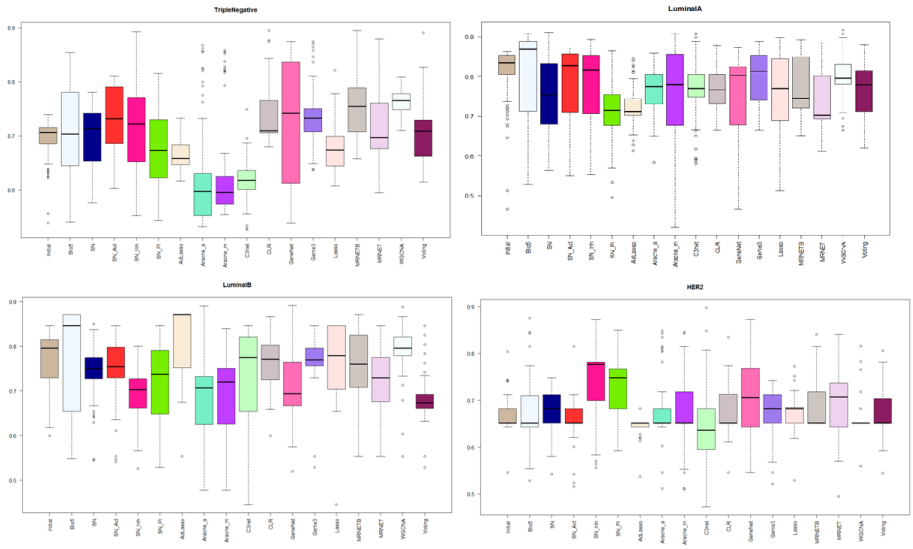

Figure 2: Box plots of the mean accuracy rates of the top 100 sequential genes from all ranked and re-ranked gene lists in combination with PageRank reconciling method, using hold out validation with train set the TCGA expression values and test set the expression values from GEO independent datasets for breast cancer subtypes.

3.3 Network inference, underlying mechanisms

We used the Enrichr web-based software application in order to find the underlying significant biological pathways derived from genes of each network pattern. Common and exclusive mechanisms of each stage and subtype were further investigated.

Following pathway analysis of our findings for the case of Staging, we have found four exclusive stage-related pathways including phenylalanine metabolism for Stage II, peroxisome proliferator-activated (PPAR) signaling pathway and glycolysis and gluconeogenesis for Stage III and toll like receptor signaling pathway for Stage IV.

For the case of Luminal A, Luminal B, HER2 and TN subtypes, we have found seven exclusive subtype-related pathways, including glycine serine and threonine metabolism pathway for Luminal B, glycerolipid metabolism, fatty acid metabolism, complement and coagulation cascades and bladder cancer for HER2 and small cell lung cancer and metabolism of xenobiotics by cytochrome p450 for TN.

3.4 Network inference and drug repurposing

The network patterns were further processed in order to investigate their contribution regarding the discovery of potential drugs for breast cancer subtypes and stages. Actually, genes that constitute the common network patterns from each subtype and stage were divided into up and down regulated, based on their Fold Change from the initial statistical analysis of the TCGA reference sets. The up and down regulated genes formed disease signatures that were queried in a well-established drug repurposing pipeline. Namely, LINCS-L1000 (http://www.lincscloud.org/) is the advanced version of cMap [11] with significantly increased number of drug treatments, cell types and gene signatures based on L1000 high throughput technology. We used the LINCS-L1000 detailed report and we collected the top 20 drugs for each gene list with the most negative enrichment scores. The negative score suggests that the drugs are considered to be inhibitors. We then derived a list of 80 drugs regarding the stages (20 drugs per stage) and 80 drugs regarding the subtypes (20 drugs per subtype). DrugBank database, as well as ChemSpider tool was used to find their chemical structures. The resulted drug lists (names and structures) were further evaluated via ChemBioServer [12], a web application for searching, filtering and comparing drug structures. More specifically, we compared each top 20 drug list from LINCS with 25 known FDA-approved Breast cancer therapeutic drugs. Hierarchical clustering using tanimoto similarity (Soergel distance) was applied to each of the top 20 drug list from LINCS and the 25 known FDA-approved Breast cancer therapeutic drugs. In synopsis, the unique drugs for the breast cancer stages were 63 and for the breast cancer subtypes 58, as we have located common drugs across them.

To further examine the resulted drugs, we constructed a super network that combines each of the top 20 drugs extracted from our analysis with the 25 FDA approved breast cancer drugs, with their target genes and finally with the respective common network pattern. We used the DrugBank database in order to find the target

genes of all drugs from LINCS and the 25 FDA approved Breast Cancer drugs. GeneMANIA plug-in was applied to identify which genes from each pattern were physically interacting with the target genes. Our goal was to understand the correlations between drugs, drug targets and conserved co-expressed genes from a network-based view, in order to outline small paths that are of great importance in breast cancer stages and subtypes. Each network consists of four sub-networks, two drug – drug similarity networks, a drug – target network and a drug target – common pattern genes co-expression network, as shown in Figures 3-4. In Figures 3-4, the yellow cycles represent each top 20 drug list from LINCS and the green cycles the 25 FDA Breast cancer Drugs. Edges between the two cycles represent their structural similarity. As much thicker is the edge, the greater the similarity between the drugs. Only edges with similarity greater than 0.5 are presented. Grey cycles (Figures 3-4) depict the target genes. As we described above, we found the corresponding target genes of the total drugs by means of the DrugBank database. Drug- target associations are represented with red dots. Purple ellipses typify top 100 genes from each common network pattern. Blue edges represent physical interactions between target genes and genes from each common network pattern.

As shown in Figure 3, one drug out of 25 FDA approved Breast cancer drugs, Gemcitabine, was proposed as repurposed drug by the LINCS for breast cancer stage I. Furthermore, Gemcitabine is quite similar (tanimoto similarity greater than 80%) with Clofarabine and Kinetin-riboside (repurposed drugs from LINCS). Clofarabine is also an anti-cancer, antineoplastic chemotherapy drug and is classified as an antimetabolite. Moreover, Vinblastine – Breast Cancer drug was found to be greater than 60% structurally similar with Sepantronium bromide (repurposed drug from LINCS), which is a small-molecule proapoptotic agent with potential antineoplastic activity. Vinblastine has three target genes TUBA1A, TUBB and JUN. The latter was found to physically interact with three genes (ATF3, FOS and EGR1) of the breast cancer stage I network pattern. As shown in Figure 4, Idarubicin (repurposed drug from LINCS) was also found to be 85% structurally similar with Doxorubicin and Epirubicin and they are all topoisomerase 2 inhibitors (TOP2A). Super Networks were constructed and analyzed for each stage and breast cancer subtype.

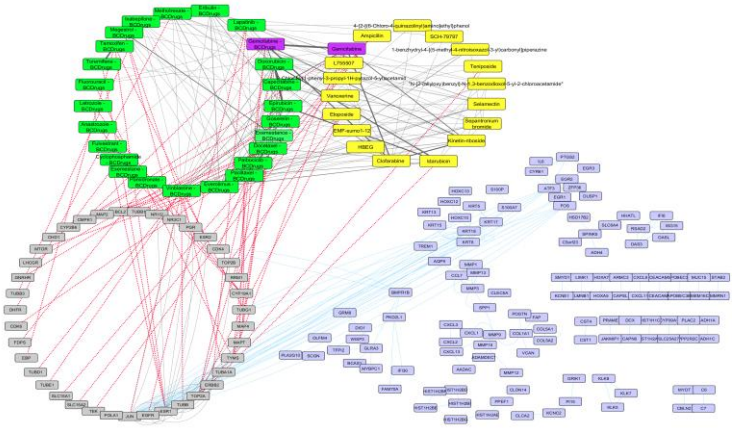

Figure 3. Super Network for breast cancer Stage I- consists of 4 sub-networks: 1) two drug – drug networks: with yellow cycle are represented the 21 drugs from LINCS and with green cycle the 20 therapeutic breast cancer drugs 2) drug – target network: grey round rectangles represent the target genes of all drugs (red dots edges) and 3) target - pattern genes network: physical interactions (blue edges) between target genes and genes from the network pattern (purple ellipses). One out of the 25 FDA approved Breast cancer drugs (Gemcitabine), was found in the top 20 drug list from LINCS from breast cancer stage I (dark magenta).

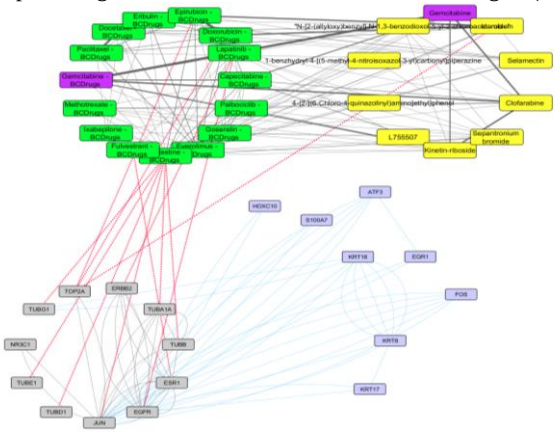

Figure 4. Highlighted target genes that physically interact with genes from the breast cancer stage I common network pattern and their corresponding repurposed drugs from LINCS, along with their structurally similar Breast cancer drugs.

4 Discussion

In the present work, we used eleven network inference methods and one ensemble scheme to reconstruct gene co-expression networks in order to examine their contribution in identifying significant genes and gene-gene links related to different breast cancer stages and subtypes. During this assessment, we demonstrated that, in

most cases of breast cancer stages and subtypes, the statistical co-expression networks produce either similar or more enriched lists with significant genes (in terms of maximum classification accuracy achieved) for each breast cancer stage and subtype than the conventional statistical approach or the networks based solely on the biological information extracted from the literature. Actually, the dominance of statistical networks is profound in the analysis of breast cancer subtypes, whereas in the case of stage analysis, the simple statistical method (Initial) and the signaling network based on inhibition (SN_I) give slightly better (almost equivalent) scoring than statistical networks.

Furthermore, our analysis concluded to eight network patterns, four for the stages (I, II, III and IV) and four for the subtypes (Triple Negative, Luminal A, Luminal B and HER2). Additionally, we further analyzed the gene patterns, in order to investigate potential mechanisms and drugs for breast carcinomas staging and subtypes. As described in the previous section, we have found four exclusive stage-related pathways. Peroxisome proliferator-activated (PPAR) signaling pathway has been implicated in the pathology of numerous diseases including obesity, diabetes, atherosclerosis, and cancer. More specifically, PPAR signaling pathway has been reported as a possible important predictor of breast cancer response to neoadjuvant chemotherapy [13]. Five dehydrogenase (ADH) isoenzymes and aldehyde dehydrogenases (ALDH) genes from the breast cancer Stage III network pattern were involved in the glycolysis and gluconeogenesis pathway. It has been reported that patients with advanced breast cancer had changes in the activity of activity of ADH isoenzymes and ALDH [14]. Furthermore, from the breast cancer Stage IV pattern we have found an exclusive pathway - toll like receptor signaling pathway for which it is well known that supports tumor cell growth in vitro and in vivo [15]. For the case of breast cancer subtypes, we have found seven exclusive subtype-related pathways. Hyperactivation Glycine serine and threonine metabolism pathway drives to oncogenesis and recent developments support that this pathway may provide novel opportunities for drug development and biomarker identification of human cancers [16]. Moreover, from the Triple Negative pattern we found the metabolism of xenobiotics by cytochrome p450 pathway. Cytochromes P450 (CYPs) play a pivotal role in cancer formation and cancer treatment as they participate in the inactivation and activation of anticancer drugs [17].

Most of the specific mechanisms per subtype and stage are related to cellular community, signaling, cell growth and death, immune and endocrine systems, carbohydrate, lipid and amino acid metabolism as well as xenobiotics biodegradation and metabolism. Furthermore, all the derived network patterns include genes found in breast cancer specific regions of significant somatic copy number alterations (SCNA) [8]. These results are fully aligned to the up-to-date recognized cancer hallmarks related to cell growth, metabolism, immune system, inflammation and genome duplication [18].

The resulted network patterns were also analyzed by means of LINCS drug reposition pipeline. Two out of 25 therapeutic FDA approved breast cancer drugs (Gemcitabine and Palbociclib) were also found as repurposed drugs from LINCS. In Stage I, two repurposed drugs Clofarabine and Kinetin-riboside were found to be

structurally similar to Gemcitabine. Clofarabine seems to have potential efficacy in epigenetic therapy of solid tumours, especially at early stages of carcinogenesis [19]. For Stage II, Cladribine (repurposed drug) was found to be structurally similar with Triciribine (repurposed drug) and Gemcitabine and Capecitabine Breast cancer drugs. In clinical trial (June, 2015) triciribine phosphate, when given together with paclitaxel, doxorubicin hydrochloride, and cyclophosphamide, works in treating patients with stage IIB-IV breast cancer (https://clinicaltrials.gov).

Moreover, in Stage III Ruxolitinib and Pyrvinium-pamoate repurposed drugs from LINCS have been found as structurally similar with Letrozole and Vinblastine Breast cancer drugs respectively. An ongoing clinical trial (October, 2015) compares the overall survival of women with advanced (Stage III) or metastatic (Stage IV) HER2-negative breast cancer who receive treatment with Capecitabine in combination with Ruxolitinib versus those who receive treatment with Capecitabine alone (https://clinicaltrials.gov). Irinotecan has been examined in a clinical trial in Phase II in order to find its objective response rate in patients with metastatic breast cancer (Stage IV) (https://clinicaltrials.gov).

In case of repurposed drugs for breast cancer subtypes, we have found that Etoposide and Teniposide (repurposed drugs) as structurally similar with two Breast cancer drugs Epirubicin and Doxorubicin in Triple Negative subtype. These four drugs are topoisomerase ii inhibitors (TOP2A) and Etoposide has been found as effective drug in Chinese women with heavily pretreated metastatic breast cancer [20].

In Luminal A, the target genes of Vorinostat physically interact with two genes (RUNX1T1 and SMYD1) from the Luminal A pattern. It has been reported that Vorinostat in combination with Tamoxifen may treat patients with hormone therapy-resistant breast cancer [21]. In Luminal B, F10 and EGFR genes from Luminal B pattern are also target genes of Menadione (repurposed drug from LINCS) and Lapatinib Breast cancer drug. Menadione has been examined on its antiproliferative action on breast cancer cells50. Finally in HER2 subtype, Palbociclib is also a Breast cancer drug that was found from the drug repurposing analysis of HER2 pattern. It has quite similar structure with WZ-4002 repurposed drug which is a novel, mutant inhibitor of EGFR.

Finally, the action of the remaining mechanisms and drugs found from LINCS may be further investigated since they have been derived from significantly relevant genes related to breast cancer stages and subtypes.

References

1. Howell, A. *et al.* Risk determination and prevention of breast cancer. Breast Cancer Res 16, 446, doi:10.1186/s13058-014-0446-2 (2014).
2. Hutchinson, L. Breast cancer: challenges, controversies, breakthroughs. Nat Rev Clin Oncol 7, 669-670, doi:10.1038/nrclinonc.2010.192 (2010).
3. Zhang, J. et al. Weighted frequent gene co-expression network mining to identify genes involved in genome stability. PLoS Comput Biol 8, e1002656, doi:10.1371/journal.pcbi.1002656 (2012).

4 Cheng, F. et al. Prediction of drug-target interactions and drug repositioning via network-based inference. PLoS Comput Biol 8, e1002503, doi:10.1371/journal.pcbi.1002503 (2012).
5 Chen, J., Aronow, B. J. & Jegga, A. G. Disease candidate gene identification and prioritization using protein interaction networks. BMC Bioinformatics 10, 73, doi:10.1186/1471-2105-10-73 (2009).
6 Nayak, R. R., Kearns, M., Spielman, R. S. & Cheung, V. G. Coexpression network based on natural variation in human gene expression reveals gene interactions and functions. Genome Res 19, 1953-1962, doi:10.1101/gr.097600.109 (2009).
7 Emmert-Streib, F., Glazko, G. V., Altay, G. & de Matos Simoes, R. Statistical inference and reverse engineering of gene regulatory networks from observational expression data. Front Genet 3, 8, doi:10.3389/fgene.2012.00008 (2012).
8 Zack, T. I. et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet 45, 1134-1140, doi:10.1038/ng.2760 (2013).
9 Cui, Q. et al. A map of human cancer signaling. Molecular systems biology 3, 152, doi:10.1038/msb4100200 (2007).
10 Poirel, C. L. et al. Reconciling differential gene expression data with molecular interaction networks. Bioinformatics 29, 622-629, doi:10.1093/bioinformatics/btt007 (2013).
11 Lamb, J. et al. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. Science 313, 1929-1935, doi:10.1126/science.1132939 (2006).
12 Athanasiadis, E., Cournia, Z. & Spyrou, G. ChemBioServer: a web-based pipeline for filtering, clustering and visualization of chemical compounds used in drug discovery. Bioinformatics 28, 3002-3003, doi:10.1093/bioinformatics/bts551 (2012).
13 Chen, Y. Z. et al. PPAR signaling pathway may be an important predictor of breast cancer response to neoadjuvant chemotherapy. Cancer chemotherapy and pharmacology 70, 637-644, doi:10.1007/s00280-012-1949-0 (2012).
14 Jelski, W., Chrostek, L., Markiewicz, W. & Szmitkowski, M. Activity of alcohol dehydrogenase (ADH) isoenzymes and aldehyde dehydrogenase (ALDH) in the sera of patients with breast cancer. Journal of clinical laboratory analysis 20, 105-108, doi:10.1002/jcla.20109 (2006).
15 Ahmed, A., Redmond, H. P. & Wang, J. H. Links between Toll-like receptor 4 and breast cancer. Oncoimmunology 2, e22945, doi:10.4161/onci.22945 (2013).
16 Amelio, I., Cutruzzola, F., Antonov, A., Agostini, M. & Melino, G. Serine and glycine metabolism in cancer. Trends in biochemical sciences 39, 191-198, doi:10.1016/j.tibs.2014.02.004 (2014).
17 Rodriguez-Antona, C. & Ingelman-Sundberg, M. Cytochrome P450 pharmacogenetics and cancer. Oncogene 25, 1679-1691, doi:10.1038/sj.onc.1209377 (2006).
18 Wang, E. et al. Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. Seminars in cancer biology 30, 4-12, doi:10.1016/j.semcancer.2014.04.002 (2015).
19 Lubecka-Pietruszewska, K. et al. Clofarabine, a novel adenosine analogue, reactivates DNA methylation-silenced tumour suppressor genes and inhibits cell growth in breast cancer cells. Eur J Pharmacol 723, 276-287, doi:10.1016/j.ejphar.2013.11.021 (2014).
20 Yuan, P. et al. Oral etoposide monotherapy is effective for metastatic breast cancer with heavy prior therapy. Chin Med J (Engl) 125, 775-779 (2012).
21 Munster, P. N. et al. A phase II study of the histone deacetylase inhibitor vorinostat combined with tamoxifen for the treatment of patients with hormone therapy-resistant breast cancer. British journal of cancer 104, 1828-1835, doi:10.1038/bjc.2011.156 (2011).