frontiers in Physiology

ORIGINAL RESEARCH
published: 22 November 2016
doi: 10.3389/fphys.2016.00568

Transcriptional Network Architecture of Breast Cancer Molecular Subtypes

Guillermo de Anda-Jáuregui[1], *Tadeo E. Velázquez-Caldelas*[1], *Jesús Espinal-Enríquez*[1,2] *and Enrique Hernández-Lemus*[1,2]*

[1] *Computational Genomics, National Institute of Genomic Medicine, Mexico City, Mexico,* [2] *Complejidad en Biología de Sistemas, Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico*

Breast cancer heterogeneity is evident at the clinical, histological and molecular level. High throughput technologies allowed the identification of intrinsic subtypes that capture transcriptional differences among tumors. A remaining question is whether said differences are associated to a particular transcriptional program which involves different connections between the same molecules. In other words, whether particular transcriptional network architectures can be linked to specific phenotypes. In this work we infer, construct and analyze transcriptional networks from whole-genome gene expression microarrays, by using an information theory approach. We use 493 samples of primary breast cancer tissue classified in four molecular subtypes: Luminal A, Luminal B, Basal and HER2-enriched. For comparison, a network for non-tumoral mammary tissue (61 samples) is also inferred and analyzed. Transcriptional networks present particular architectures in each breast cancer subtype as well as in the non-tumor breast tissue. We find substantial differences between the non-tumor network and those networks inferred from cancer samples, in both structure and gene composition. More importantly, we find specific network architectural features associated to each breast cancer subtype. Based on breast cancer networks' centrality, we identify genes previously associated to the disease, either, generally (i.e., CNR2) or to a particular subtype (such as LCK). Similarly, we identify LUZP4, a gene barely explored in breast cancer, playing a role in transcriptional networks with subtype-specific relevance. With this approach we observe architectural differences between cancer and non-cancer at network level, as well as differences between cancer subtype networks which might be associated with breast cancer heterogeneity. The centrality measures of these networks allow us to identify genes with potential biomedical implications to breast cancer.

Keywords: gene regulatory networks, breast cancer, molecular subtypes, network topology, clinical genomics

OPEN ACCESS

Edited by:
Argyris Papantonis,
University of Cologne, Germany

Reviewed by:
Guanglong Jiang,
Indiana University School of Medicine, USA
Alessandro Giuliani,
Istituto Superiore di Sanità, Italy

***Correspondence:**
Enrique Hernández-Lemus
ehernandez@inmegen.gob.mx

Specialty section:
This article was submitted to Systems Biology, a section of the journal Frontiers in Physiology

Received: *20 June 2016*
Accepted: *08 November 2016*
Published: *22 November 2016*

Citation:
de Anda-Jáuregui G, Velázquez-Caldelas TE, Espinal-Enríquez J and Hernández-Lemus E (2016) Transcriptional Network Architecture of Breast Cancer Molecular Subtypes. Front. Physiol. 7:568. doi: 10.3389/fphys.2016.00568

BACKGROUND

Breast cancer is a heterogeneous disease. The identification of *molecular subtypes* (Perou et al., 2000) was a major breakthrough in order to categorize this heterogeneity, made possible by the emergence of whole-genome microarray technology (Rueda, 2014). Molecular subtypes associate expression of certain markers to phenotypical differences in cancer, pointing to different prognosis as well as distinct therapies (Liu et al., 2014).

The identification of breast cancer molecular subtypes is a paramount example of the impact of high throughput technologies in the study of cancer. These technologies are able to provide

even deeper biological understanding, when analyzed by an integrative, systemic framework. In this regard, *network theory* has emerged as a major tool to achieve this goal (Newman, 2010; Chen et al., 2015) under a systems biology view of matters.

A network is a mathematical construct composed by a set of *nodes* or vertices, and a set of *links* that represent a relation between them. A *biological network*, is a network where nodes represent any kind of biological molecules: genes, transcripts, proteins, metabolites, etc., and links represent physical or chemical interactions between those molecules (Jeong et al., 2000; Hasty et al., 2001; Jeong et al., 2001; Thattai and Van Oudenaarden, 2001; Lee et al., 2002; Maslov and Sneppen, 2002; Davidson and Levin, 2005; Guimera and Amaral, 2005; Levine and Davidson, 2005; Davidson and Erwin, 2006). With gene expression microarray technologies, it is factible to construct **transcriptional networks** where nodes are transcribed genes, and links represent a correlation between expression values of said genes, which point to a *possible* interaction between them at the transcriptional level (Tovar et al., 2015).

Many correlation measures have been implemented in order to construct biologically meaningful transcriptional interaction networks based on the inference of statistical dependency (Friedman et al., 2000; Gardner et al., 2003; Giuliani et al., 2004; Wang et al., 2005; Cowell, 2006; Nielsen and Jensen, 2009). This is especially fitting in view of the isomorphism existing between a network structure and a correlation matrix whose elements are the strength of the interaction between the intervening nodes. We must notice, however, that this correlation structure is usually given in the presence of an accompanying variance structure among gene expression levels (Giuliani et al., 2004). It has long been proven that the best estimator of statistic dependency is *mutual information (MI)* (Basso et al., 2005; Margolin et al., 2006; Hernández-Lemus and Rangel-Escareño, 2011; de Matos Simoes and Emmert-Streib, 2012; Hernández-Lemus and Siqueiros-García, 2013). These statistically inferred networks provide a deeper level of biological understanding in two main directions: giving support to previously identified biological observations, and giving new insights regarding novel biological interactions.

Network structural properties have been related to features in the biological context (Kitano, 2002; Albert, 2005; Serrano, 2007; Hakes et al., 2008). Therefore, analyzing these properties in transcriptional networks may provides us better understanding of the underlying biological phenomena. Global network metrics often provide information regarding the system as a whole; while local parameters provide information regarding the relevance of particular nodes (Barabasi and Oltvai, 2004; Newman, 2010; Barabási et al., 2011; Biane et al., 2016; Robinson and Nielsen, 2016). The transcriptional network approach has proven be useful to unveil transcriptional regulation in cancer (Carro et al., 2010; House et al., 2010; Pe'er and Hacohen, 2011; Madhamshettiwar et al., 2012) and in particular in breast cancer (Van De Vijver et al., 2002; Lim et al., 2009; Cicatiello et al., 2010; Gu et al., 2010; Tovar et al., 2015; Castro et al., 2016).

Transcriptional networks are representations of the regulatory programs behind phenotypes. Given the intrinsic heterogeneity of breast cancer molecular subtypes, a fundamental question which remains unsolved is whether the transcriptional architecture of these subtypes is different. To answer this we constructed transcriptional networks for breast cancer molecular subtypes based on mutual information of genome-wide gene expression. We compared them to a network of healthy mammary tissue.

We identified differences in network architecture between phenotypes. We observed major differences between the cancer subtype networks and the non-tumor network. Particular architectural features were associated to the different molecular subtypes. We find that in these networks, the connectivity of particular genes may indicate differences of their role in the transcriptional program of each subtype. Identifying such differences may be key to understand how the specific transcriptional program shapes a particular phenotype. This in turn, will enhance our insight on the nature of molecular subtypes, with basic and clinical implications.

METHODS

Sample Data and Classification

Four hundred and ninty three-microarray expression profiles for breast cancer samples and 61 microarray expression profiles corresponding to healthy breast tissue were collected from several experimental datasets that are available on the Gene Expression Omnibus site (Edgar et al., 2002). We used microarray data from GSE 4922 (Ivshina et al., 2006), 1456 (Pawitan et al., 2005), 7390 (Desmedt et al., 2007), 1561 (Farmer et al., 2005), 2603 (Minn et al., 2005), 2990 (Sotiriou et al., 2006), GSE 9574 (Tripathi et al., 2008), GSE 15852 (Pau Ni et al., 2010), GSE 6883 (Liu et al., 2007) and 3494 (Miller et al., 2005). All experiments were performed following protocol GPL96, using total mRNA on the Affymetrix HGU133A microarray platform. This platform contains probes for 18,400 transcripts and variants. Raw microarray data was processed following a pipeline for *Robust Multi-array Average* (Irizarry et al., 2003), previously implemented in our workgroup (Baca-López et al., 2012; Tovar et al., 2015). Breast cancer samples were classified using the well-validated `PAM50` algorithm (Parker et al., 2009).

Since all samples were downloaded from properly documented public databases, no ethics committee approval was required. All raw data is available in the NCBI-GEO database, with accession keys and references as stated above to guarantee full data availability.

Sample Comparability

Comparability is a key issue during analysis from microarrays, in particular when dealing with data coming from different sources (laboratories, technicians, etc.). Biases may exist even when the same protocols have been followed (Grass, 2009). Chen et al. (2011) tested six different algorithms to eliminate this so-called *batch effect* and found that the best results were obtained by using the empirical Bayesian assessment methods, such as **ComBat** (Johnson et al., 2007). The strategy followed here consisted in preprocessing all arrays with the **frma** algorithm (McCall et al., 2010), and using summarization with robust weighted average with no background correction, we split the datasets into cases/controls, and then applied **ComBat** to both datasets

separately. After that, we re-joined the two resulting datasets and re-normalized them together with the **cyclic loess algorithm** (Ballman et al., 2004), in such way that both conditions belong now to the *same dynamic range*.

In order to assess the extent of this effect within our samples, so that we could remove the corresponding bias as accurately as possible, we resort to Principal Variance Component Analysis (PVCA) which is an algorithm combining the advantages of the principal component analysis (reduction of dimensionality) with the statistical reliability of the analysis of variance (Grass, 2009). After bias reduction, a PVCA analysis corroborated that such a confounding or batch effect almost disappeared. (Further information can be found in **Supplementary Material 1**).

Transcriptional Network Inference

Gene regulatory network inference from experimental data involves the solution of an inverse problem (also called a deconvolution) which consists in unveiling the interactions (edges or links) from the properties of observables such as gene expression levels. Inferring a network implies the uncovering of the statistical dependencies within a Joint Probability Distribution (JPD). A usual way to do this is by quantifying the new information content that arise when we compare the full JPD to a series of successive independent approximations. In practice doing this is rather difficult because one is faced with large numbers of variables with a strong nonlinear behavior. Mutual Information (*MI*) is a measure from information theory that is able to deal with these issues since it is model independent, non-parametric and capable of capturing non-linear dependencies (Hernández-Lemus and Siqueiros-García, 2013).

Transcriptional network inference based on *MI* has been successfully employed (Hernández-Lemus and Siqueiros-García, 2013; Khosravi et al., 2015; Rodriguez-Barrueco et al., 2015). ARACNE (Margolin et al., 2006) is one of many algorithms used to calculate MI based on gene expression. ARACNE algorithm presents a relatively fast and reliable implementation of the inference of gene regulatory network from gene expression data. The method works as follows: a normalized gene expression matrix (i.e., an N by M matrix containing the gene expression levels of N genes in M samples) is used as input. The algorithm then calculates the *empirical* marginal probability distribution for the expression levels for all genes (i), as well as the *empirical* joint probability distribution for all the gene-couples (i,j) by approximating them by using Gaussian kernels. With these probabilities, a value of the *MI* between any two genes is calculated.

The method associates a *MI* value to each significance value (*p*-value) based on permutation analysis, as a function of the sample size. Therefore, a *MI* threshold (MI_0) can be defined. For every pair of genes with $MI_{i,j} > MI_0$ an interaction of weight $MI_{i,j}$ is reported. Pairs of genes with $MI_{i,j} < MI_0$ are considered to be non-interacting. Correlation analysis is made in the presence of *range restriction*, i.e., a sufficient amount of variance to allow for the detection of correlation structure. Hence a system observed at different scales (as given by the cut-off values) will give rise to different solutions. As it has long been known, in biological systems there is no *preferential scale of observation* a fact that makes scaling analysis a relevant approach to the quantitative analysis in biology (Giuliani et al., 2004).

Using the previously described expression data, we inferred transcriptional networks for each of the following molecular subtypes of breast cancer: Luminal A, Luminal B, Basal and HER2-enriched, and one for the non-tumor breast tissue phenotype. The construction of our networks proceeded as follows:

1. MI was calculated for every pair of (non-self) probesets in the microarray platform, using the ARACNE algorithm.
2. Those interactions ranked highest by Mutual Information values were kept.
3. Probesets were mapped to HUGO gene symbols. Those probesets that did not correspond to a gene symbol were discarded.

Network Analysis and Comparison

In order to compare the network structure of each phenotype, we analyzed each network by calculating the following metrics: the number of nodes and edges; the node degree (number of nodes connected to a specific node); connected components (a subset of nodes connected among them and not connected to the rest of nodes in the graph); and clustering coefficient (the number of existing node triplets over the total number of triplets) (Luce and Perry, 1949; Watts and Strogatz, 1998). Network connectivity degree is perhaps the most obvious *centrality measure*, i.e., it is an indicator of the relevance of a particular node to the large scale structure of the network. Genes that are more connected (that is, have a higher degree) are the ones that partake in most interactions, thus connecting a largest number of biological processes. Hence, highest degree genes are known to be quite important for the establishment of phenotypes.

The number of connected components (or islands) in a complex biological networks is a simple yet important indicator about the way in which different parts of the network (subnetworks) work together. A small number, or even a single connected component means that most (or all of the) interactions in the network have impact at a *global* level, whereas a higher number of connected components may imply a certain degree of *modularity* in which interactions within a subnetwork are somehow *autonomous* from interactions in other subnetworks.

The clustering coefficient in these networks is also indicative of the modularity and connectivity patterns at a lower (more local) scale than that of the number of connected components. Higher values of the average clustering coefficient can be related to greater redundancy and robustness in biological networks. Network topological analysis and visualization was performed by using Cytoscape v.3.0 (Shannon et al., 2003).

As previously mentioned, a network is defined by a set of nodes and a set of links between said nodes. Particularly in these inferred networks, nodes are representing genes and links correspond to potential transcriptional interactions. Similarity in these sets among phenotypes points to similarity in transcriptional programs. To compare these sets we used the

Jaccard index J, this is obtained by dividing the size of the intersection over the size of the union of two sets. This was done for gene sets and link sets.

Jaccard indexes are used as measures of *similarity* between two sets: the closer J is to 1, the more similar the sets; Conversely, $J = 0$ implies completely dissimilar sets. Indeed, Jaccard indexes are actual probability measures. Here, Jaccard indexes are calculated for the sets of nodes (genes) of different networks, to see to what degree transcriptional networks representing different phenotypes (breast cancer subtypes) share genes, regardless of their particular interactions. On the other hand, Jaccard indexes calculated for the sets of interactions or links in those networks, reveal to what extent different transcriptional regulatory programmes share, not only groups of genes but also connection patterns among those genes.

Since biological functions depend not only in sets of specific molecules, but also in the interaction patterns among them, the joint consideration of similarities (and dissimilarities also) between gene sets (lists) and interaction sets may broaden our scope as to what are the differences and commonalities of breast cancer subtypes in terms of *biological features* (García-Campos et al., 2015).

Network Threshold Assessment

As already mentioned the cut-off value in the *MI* distribution will affect the membership of particular interactions as well as the structure of the inferred networks. In some sense, the choice of this threshold is indeed related to *feature selection*. The choice of cut-off value to construct *meaningful* (at the light of the feature selection procedure) networks is an open problem in contemporary research in biology. This is to say that, a particular cut-off choice depends on what kinds of features are to be selected (Giuliani et al., 2004; Censi et al., 2011). In order to avoid unnecessary biases, we decided to test different *MI* cut-off values compliant with quite general topological structure constraints of the underlying networks. Network size, for instance, is one of the most important constraints: extremely small and stringent networks will not capture the essential biological information, whilst extremely large, low confidence networks will present a larger number of false positive interactions and are much harder to analyze in order to unveil biological function. We assessed the threshold influence by calculating network metrics for different cut-off values, restricting our analysis to those networks which have a node-to-link ratio around 0.1, as this ratio value is characteristic of biological complex networks (Albert and Barabási, 2002; Barabasi and Oltvai, 2004; Barabási et al., 2011).

RESULTS

We inferred transcriptional networks for each breast cancer molecular subtype and for non-tumor mammary tissue. Network approaches are highly relevant for the understanding of the connection between sample state-variability and gene expression patterns, essential to elucidate the role that such expression patterns play in the establishment of cellular phenotypes. In this work we decided to pursue this by calculating mutual information correlation measures rather that parametric correlation coefficients due to the higher generality of the former.

Based on the evaluation of network threshold (**Supplementary Material 2**), we will present the results for networks constructed from the 10,000 interactions ranked with highest *MI* values, which are also, as previously mentioned, those with the highest statistical significance ($p\text{-}value < 10^{-10}$). This cut-off value lies in the identified range (between the 10^3 and 10^5 highest ranked interactions) in which the node-to-link ratio is consistent with the expected values for biological complex networks.

Transcriptional Networks Show Different Architectures

A graphical representation of all phenotype networks constructed with our approach can be observed in **Figure 1**, starting with the 4 breast cancer subtypes (**Figures 1A–D**), and the non-cancer mammary tissue network (**Figure 1E**). In this representation, nodes correspond to genes and links are a representation of the *MI* value. We will now show the main differences in the transcriptional network gene composition, highlighting the differences both between non-tumor and tumor networks as well as between tumor subtypes.

Figure 1 shows the differences in transcriptional network architectures among phenotypes. These differences in structure can be thought of as a representation of potential differences in the transcriptional regulatory program. As it has been discussed previously (Censi et al., 2011), in systems under stress (high correlation, and high variance) (Gorban et al., 2010) changes in the general correlation structure become visible as changes in the associated networks. The most evident differences in network structure are seen between the tumor and the non-tumor networks: the non-tumor network is dominated by a giant component; meanwhile, the subtype networks are formed by several coexisting components of different sizes. In this regard, we can expect that changes in the gene correlations between the normal and disease conditions may appear, at the local—pairwise correlations—, modular—or functional—and global—at the genome-wide transcriptional network—levels (Censi et al., 2011). It is also evident that while more similar among themselves, each breast cancer subtype has a unique network architecture. This is reflected in the network metrics reported in **Table 1**.

Non-tumor Transcriptional Network Architecture

The transcriptional network for non-tumor tissue is shown in **Figure 1E**. This network is dominated by a giant component which contains 913 genes and 9823 links between them; in other words, about 90% of the genes in the network belong to this larger component, with the rest scattered in small islands of less than ten genes. The network's clustering coefficient (CC) is 0.3722, which indicates a sparse network structure. This particular network architecture, a *sparse network mostly comprised by a giant component*, was found to be unique to the non-tumor phenotype.

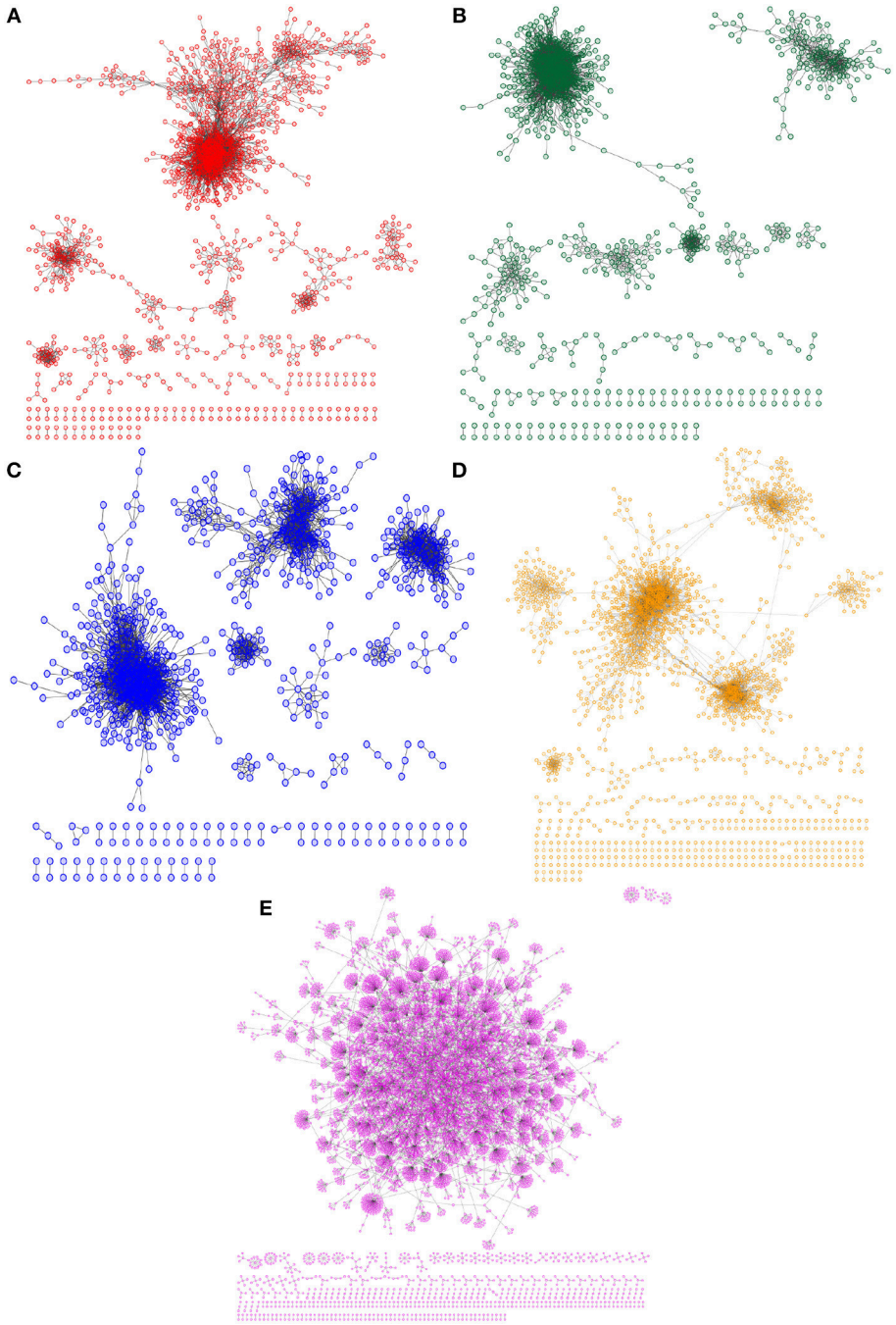

FIGURE 1 | Network architectures for breast cancer subtypes and non-tumor breast tissue. In each panel, the transcriptional network structure of each breast tumor subtype is shown: **(A)** Luminal A (red nodes); **(B)** Luminal B (green); **(C)** Basal (blue) **(D)** and HER2-enriched molecular subtype (orange). **(E)** Shows the transcriptional architecture for non-tumor breast tissue. Please notice the that **(A–D)** show networks with a large component and multiple medium-sized components, while **(E)** presents a network dominated by a single giant component, followed by small components.

Breast Cancer Transcriptional Networks Show Different Topologies between Molecular Subtypes

The transcriptional networks for each molecular subtype are shown in **Figures 1A–D**. This graphical representation makes immediately evident the architectural differences between networks. Luminal A subtype (**Figure 1A**) shows a network with one larger component, four other mid-sized components with a small number of connected genes, and several small islands of less than 10 genes. This network contains 1451 annotated genes, with 9,941 links between them, and a CC of 0.3658.

TABLE 1 | Network metrics.

Parameter	Luminal A	Luminal B	Basal	HER2-enriched	Non-tumor
Nodes	1451	1018	1046	2100	1027
Edges	9941	9898	9966	9856	9894
Connected components	88	70	56	162	46
Clustering coefficient	0.3658	0.4044	0.4114	0.3586	0.3722
Min *P*-value	5.09×10^{-25}	1.34×10^{-23}	2.2×10^{-21}	1.95×10^{-12}	9.06×10^{-19}

The Luminal B subtype network (**Figure 1B**) shows one larger component, followed by four other smaller components, and several genes scattered in small islands of less than 20 genes. This network has a total of 1018 genes, with 9898 links between them, and a CC of 0.4044.

In the Basal subtype network (**Figure 1C**), we observe again a larger component, but in this network, the next two components in size are in the same order of magnitude, and smaller sized islands ranging from thirty to two nodes in size. This network has 1046 nodes and 9966 links between them, and a CC of 0.4114.

Finally, the network for the HER2-enriched subtype (1d) has an architecture dominated by a larger component, which is composed by several clusters linked together by few genes acting as bridges; the rest of the network is composed from islands ranging in size from 40 to 2 nodes. This network has the largest number of individual genes (2100) and 9856 links between them, with a CC of 0.3586.

It is important to notice that the disruption of connected components in the cancer networks is a phenomenon that it is very likely intertwined with the presence of stress induced correlations (Gorban et al., 2010), as such one need to be cautious as to assign a degree of importance to each of these two features on the issue of phenotypic differences on subtype-associated networks.

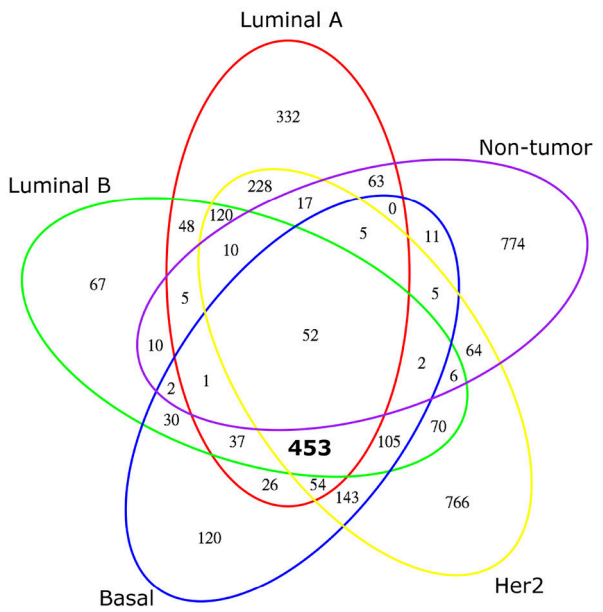

FIGURE 2 | Venn diagram of network node composition. A five-set Venn diagram of the nodes that form each transcriptional network: Luminal A is outlined in red, Luminal B in green, Basal in blue, HER2-enriched in yellow and non-tumor breast tissue in purple. Numbers in graph represent how many genes belong to each subset. Notice that the intersection of the four breast cancer molecular subtypes is large (453, marked in bold); meanwhile the number of shared genes between tumor and non-tumor networks is much smaller, with only 52 genes shared by all networks.

Network Composition Analysis

Networks were constructed following restrictions that makes them comparable in size. However, it was found that each phenotype is has a unique network composition of both genes and links between them. Network composition constitutes a different level of analysis in the multi-scale characterization of biological systems (Giuliani et al., 2004). The actual molecular make-up of gene regulatory networks extends its influence not only to the upper level—the topological network structure itself—but also to the somewhat lower scale level of description given by the molecular pathways and particular biochemical processes behind physiological functions and phenotypes (Censi et al., 2011), for this reason, in what follows we will present a detailed description of the molecular composition for phenotype (i.e., subtype)-specific networks.

Transcriptional Networks Reveal Different Gene Compositions

Each transcriptional network inferred contains a particular set of genes. Each of these sets is not completely dissimilar to another, as there are intersections between them. **Figure 2** shows a Venn diagram representation of the overlap between gene sets for each transcriptional network. In this figure, it can be seen that the intersection of genes in all breast cancer and non-cancer transcriptional networks is relatively small, containing only 52 genes. Meanwhile, the intersection of genes in all breast cancer networks, excluding those shared with the non-cancer network, contains 453 genes, a "breast cancer core set" which we will discuss later. Finally, notice how the number of non-shared, "exclusive" genes for each breast cancer transcriptional network is different.

The differences between network genesets can also be seen in **Table 2**, containing *J* for each network pair. This table shows varying levels of similarity between molecular subtypes: we can see how the Luminal A and Basal subtypes are the most different in gene composition, and interestingly, the most similar in gene composition are the Luminal B and Basal subtypes.

TABLE 2 | Jaccard indexes for each pair of transcriptional network node sets.

	Luminal A	Luminal B	Basal	HER2-enriched	Non-tumor
Luminal A	–	0.4165	0.3360	0.3595	0.0658
Luminal B	0.4165	–	0.4935	0.3556	0.0450
Basal	0.3360	0.4935	–	0.3519	0.0391
HER2-enriched	0.3595	0.3556	0.3519	–	0.0543

TABLE 3 | Jaccard indexes for each pair of transcriptional network link sets.

	Luminal A	Luminal B	Basal	HER2	Non-tumor
Luminal A	–	0.2181	0.1461	0.0956	0.0024
Luminal B	0.2181	–	0.2254	0.1176	0.0014
Basal	0.1461	0.2254	–	0.1189	0.0014
HER2	0.0956	0.1176	0.1189	–	0.0014

Even more striking is the fact that similarity between any cancer network nodeset and the non-tumor network nodeset is about an order of magnitude smaller than any J value for any cancer network pair, indicating that non-tumor network vastly differs from any breast cancer network; with 774 unique genes, the non-tumor transcriptional network is 90% unique.

Transcriptional Networks Reveal Different Gene-Interaction Compositions

Transcriptional networks show even more differences in their link composition. **Table 3** shows J values indicating the similarity in link sets between networks. Again, it is interesting to note that the highest similarity in links is between the Luminal B and the Basal networks, followed by the similarity between the Luminal A and Luminal B networks. Again, the linkset similarity between non-tumor and any tumor network, is about two orders of magnitude smaller than between any two tumor networks.

The differences (and similarities) in link sets are determinant for phenotype definition. Even though gene composition describes *which molecules participate* in the transcriptional landscape, the *potential transcriptional interactions* that exist between each other, are the essence of the regulatory program.

A Core Set of Genes Is Shared among Breast Cancer Network Subtypes

As we have shown before, both gene and link composition define distinct network architectures for each transcriptional network, reflecting the heterogeneous nature of cancer. Particularly, different connections between the same genes may be involved in the development of features specific to each subtype. This can be evident, for instance, in the connections between the previously described *breast cancer core set* of genes shared by all breast cancer networks. **Figure 3** shows the subgraphs of each breast cancer network containing only these core genes. These highlight how, even when taking into account the same genes, the interactions between them vary across the molecular subtype landscape. As an example, take the isolated component which contains the IRF8 transcription factor, present in all subnetworks; notice how the degree of IRF8 is different in all subtypes: ranging from only three neighbors in the Luminal A subtype, to 21 neighbors in the HER2-enriched subtype.

Node Degree As a Measure of Relative Influence of Genes in the Transcriptional Context

We have focused on global network properties which define breast cancer network structures. In any network, each node has individual topological properties related to the global structure. For instance node degree, the number of neighbors of a particular node, is a measure of its centrality in the network context. For our purposes, we can think of node degree in breast cancer transcriptional networks as a measure of the influence which a particular gene may have in the regulatory program. In this sense, transcriptional network architecture not only gives global information about phenotypes, but also insights on the role of particular genes.

Given the differences in network structures, it can be expected that even the same genes may have different influences in each specific molecular subtype; at the same time, it is not unreasonable to think that some genes will play an important role in all molecular subtypes. As a proof of context, we have selected three genes which cover these scenarios: CNR2, LCK, and LUZP4. Those genes were selected on the basis of their being highly central in at least one of the cancer subtypes and being present in all the subtypes' networks, so as to have a rational means for comparison of the different effects they may have in different biological networks structures. Subnetworks containing these genes and their first neighbors for each breast cancer molecular subtype can be seen in **Figure 4**.

DISCUSSION

Deconvoluting transcriptional network architecture is a step toward uncovering differences between health and disease, as well as between different manifestations of disease. Doing this is relevant for two main lines of research: basic biological knowledge and clinical therapeutical interventions. On the one hand, unveiling the hidden relationship between system's level transcriptional programmes—as sketched by gene regulatory networks—and molecular and physiological phenotypes—given by the different breast cancer subtypes—is of great importance on the way to understand to a deeper level the elusive genotype/phenotype relationships (Barabási et al., 2011; Censi et al., 2011; Ahmad et al., 2012). On the other hand, detailed and specific knowledge about the differences and similarities of phenotype-specific gene regulatory networks, and their underlying biochemical processes as represented by metabolic and signaling pathways, may pave the way to find specific pharmacological targets useful in the clinical and therapeutical management of different breast cancer subtypes.

In this work, by using a computational approach based on information theory, we inferred transcriptional networks from

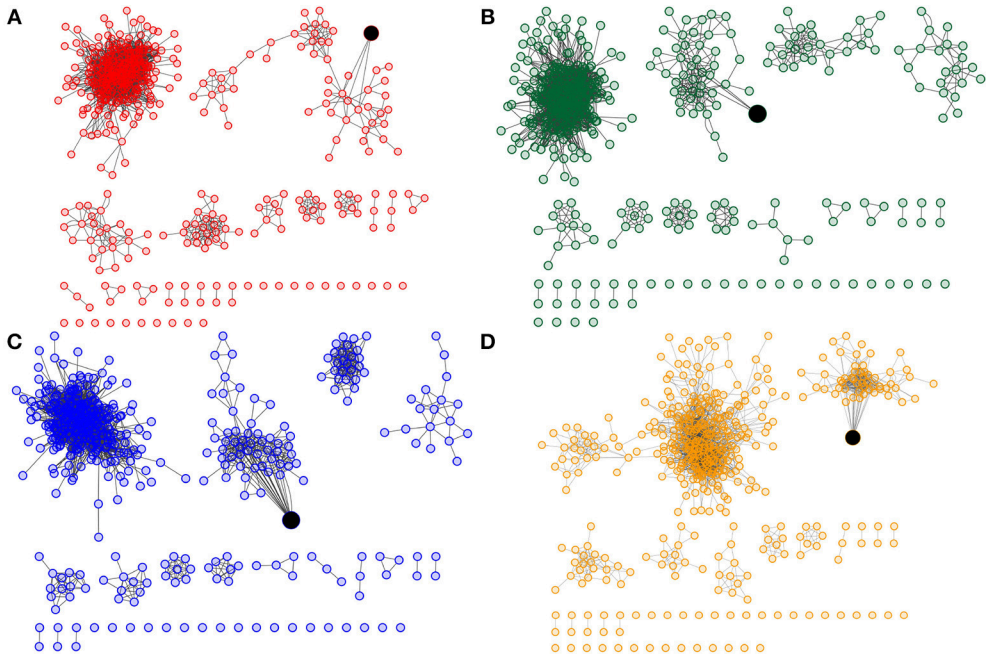

FIGURE 3 | Subnetworks of *core genes* shared by all breast cancer networks have different connections. (A–D) Show the 453 core genes for Luminal A, Luminal B, Basal and HER2-enriched subtypes. All subnetworks have a component with IRF8 transcription factor (shown in black).

genome-wide gene expression microarrays of the following breast cancer subtypes: Luminal A, Luminal B, Basal, and HER2-enriched; these were contrasted with comparable non-tumor breast tissue networks. We showed that the architecture of these transcriptional networks is unique to each phenotype, and emphasize the major differences between non-tumor and cancer networks, as well as the differences between different subtypes of breast cancer. We propose that these transcriptional networks reflect the transcriptional programs behind each phenotype, which is involved in the physiopathological differences observed in breast cancer. To our knowledge, this is the first time that transcriptional network structure has been characterized at any level of description for these clinically important breast cancer subtypes.

Breast cancer heterogeneity is a widely acknowledged fact, with molecular, histological, and clinical manifestations. In this work, we have shown breast cancer heterogeneity manifested in transcriptional network architectures; since molecular subtypes are defined by differences in gene expression profiles (see **Supplementary Material 2**), we could expect that transcriptional networks, derived from such data, would reflect this. It is well-known that network structure is intimately linked to functionality in biological networks (Kitano, 2002; Albert, 2005; Serrano, 2007; Hakes et al., 2008). In this case, we propose that our inferred networks are a representation of an underlying transcriptional program associated to each of the studied phenotypes.

We expected a drastic difference between non-tumor and tumor *networks*, just as there is a drastic difference between non-tumor and tumor *tissue*. Indeed, we found a marked difference in structure between breast cancer networks and a non-cancer network; this is in agreement with the evident differences in gene expression profiles between these two states (see **Supplementary Material 2**). The non-tumor transcriptional network is dominated by a giant component, while the networks for each of the breast cancer subtypes studied present a larger number of disconnected components. This suggests the existence of generalized transcriptional communication in healthy cells, which is lost and supplanted with a fractured, more autonomous regulation in different cancer manifestations. We consider that more research is needed in order to find mechanistic causes to this phenomenon.

The differences between cancer and non-cancer networks are also seen in the genetic composition of these networks, suggesting differences in the importance of genes in the regulatory programs of health and disease. There are only 52 genes which are part of both the non-tumor and all the cancer networks. Most of these are located in small "islands" in the network, where they interact with few other genes, rendering their impact to the global network topology negligible. Only 5 out of these 52 genes belong to the larger network components. These genes seem to be associated to general biological functions:

- ASCL3 encodes a transcription factor highly involved in determination of cell fate and the development and differentiation of numerous tissues (Yoshida et al., 2001; Jonsson et al., 2004).
- DNAJ4 encodes a highly conserved heat shock protein which serves a chaperone (Walker et al., 2010; Hageman et al., 2011).
- NCR1 encodes the natural cytotoxicity triggering receptor recognizes a broad spectrum of ligands in natural killer cells

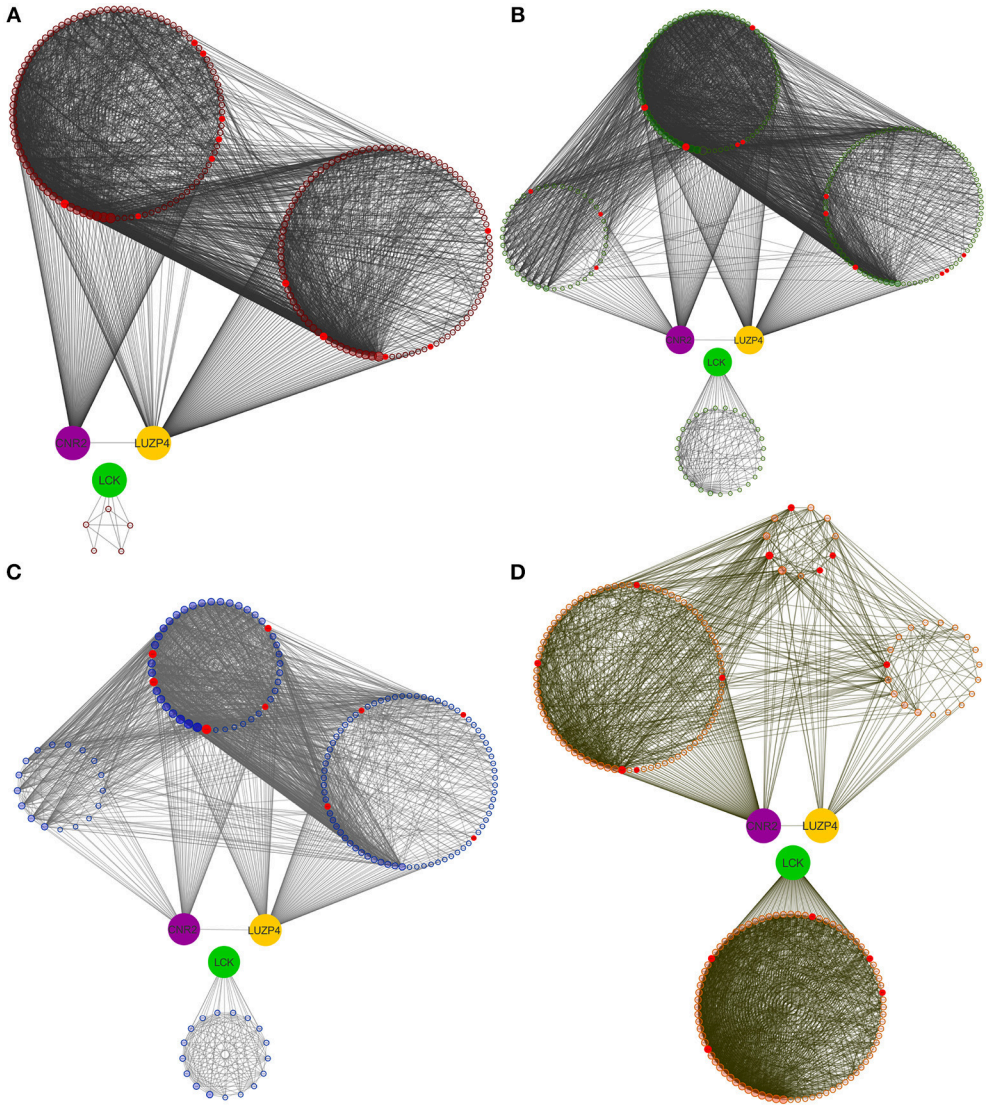

FIGURE 4 | First neighbors of CNR2, LCK, and LUZP4 genes. Genes that are connected to CNR2 (violet), LCK (green) and LUZP4 (yellow) are shown. **(A–D)** Show the same order than **Figures 1**, **2**. It can be seen that CNR2 is highly connected in all subtypes; LCK gene has a high degree relevance in HER2-enriched subtype. Finally, LUZP4 gene has different degree relevance depending on the molecular subtype. It is worth noticing that CNR2 and LUZP4 share first neighbors, but not with LCK.

(Kruse et al., 2013), and has been observed to participate regulating several functions in those cells (Pembroke et al., 2014; Fu et al., 2015; Tanimine et al., 2016).

- SLN encodes sarcolipin, a Ca^{2+}-ATPase located at the sarcoplasmic reticulum, catalyzes the ATP-dependent transport of calcium ion from the cytosol into the sarcoplasmic reticulum in muscle cells (Fajardo et al., 2013; Gorski et al., 2013; Espinoza-Fonseca et al., 2015).
- BMP15 encodes a protein associated to oocyte maturation, and follicular development (Fenwick et al., 2013; Persani et al., 2014; Sutton-McDowall et al., 2015).

Breast cancer heterogeneity, as reflected by molecular subtypes, is mostly related to the cellular composition of the normal mammary tissue: the mammary epithelium has an inner layer composed by luminal cells, and an outer layer formed by basal cells (Skibinski and Kuperwasser, 2015); the luminal and basal subtypes show phenotypical similarities to these cells. HER2-enriched breast cancer is mostly determined by an overexpression of the ERBB2 receptor, regardless of whether its cell exhibits luminal or basal characteristics. With this in mind, we expected transcriptional networks to reflect this. It would be expected to find common features in transcriptional networks for the two luminal subtypes, and also to think that these networks would be somewhat different from the basal network, with the HER2-enriched network perhaps exhibiting commonalities with all networks. Our results show the expected similarity between

the luminal A and B transcriptional networks. However, it was surprising to find that the highest similarity in both nodes and links was not between them, but rather between the luminal B and Basal networks. Meanwhile, the HER2-enriched subtype network showed was found to have the most unique architecture, as well as being the most different of all breast cancer networks. It is worth noting that we were able to recover a genetic network composition that captures the differences between normal breast and breast cancer tissue, as well as the differences between the various molecular subtypes, at the gene expression level (see **Supplementary Material 2**).

We believe that these results suggest that the oncogenic processes behind breast cancer originate unique transcriptional programs which drive each molecular subtype. The similarities of the luminal B subtype network to both the luminal A and the Basal networks are in line with the idea of a common breast cancer progenitor cell for these subtypes; at the same time, the nature of the HER2-enriched network could be indicative of a different molecular origin. (Skibinski and Kuperwasser, 2015).

The existence of a "core set" of genes which are shared among all breast cancer subtype networks was noteworthy. More interesting was the fact that even though all subtypes share these genes, the wiring structure of these is not the same in each phenotype, which again suggests that differences in the regulatory program are driven not only by the genes that participate, but by the relationships between them. The transcriptional networks define a landscape in which different elements may be playing distinct roles. Therefore, our breast cancer transcriptional networks not only provide us with global network features, but are also useful in order to identify the influence that any given gene may have in the transcriptional program. In order to illustrate how the same genes may have different roles in different breast cancer manifestations, we selected, based on their degree in each breast cancer network, the CNR2, LCK, and LUZP4 genes.

CNR2 is a gene highly connected in all breast cancer networks. CNR2 codifies the cannabinoid receptor 2 (CB2). This receptor is associated to immunomodulation and related processes by endocannabinoids (Munro et al., 1993). CB2 alterations have been found in different types of cancer (Guida et al., 2010; Jha et al., 2012; Pisanti et al., 2013), including breast cancer (Nasser et al., 2011; Pérez-Gómez et al., 2015; Sophocleous et al., 2015). Our analysis identifies CNR2 as one of the top 30 highest degree nodes in our breast cancer subtype transcriptional networks. Based on this, CNR2 might be pointed as an important gene in the general breast cancer transcriptional architecture. Cannabinoid receptors have been previously proposed as pharmacological targets for cancer (Chakravarti et al., 2014; Velasco et al., 2016), including breast cancer (Qamri et al., 2009; Morales et al., 2015). Our transcriptional network findings suggest that, if therapeutic benefits to this type of treatment are found, they may be of use to all types of breast cancer.

LCK gene was identified as a key player in HER2-enriched subtype while having a low degree in the other molecular subtypes network. LCK codifies the LCK proto-oncogene, Src family tyrosine kinase, a protein involved in signal transduction. LCK has been found expressed in breast cancer (Köster et al., 1990). Furthermore, a role in breast cancer progression and angiogenesis has been identified (Chakraborty et al., 2006). Interestingly enough, there are reports of a LCK-associated molecular signature with prognostic utility in HER2-enriched breast tumors (Rody et al., 2009). The result presented here reinforces the functional implication of LCK in the context of HER2-enriched breast cancer and emphasizes the necessity of further, focused studies.

LUZP4 was identified as a high degree gene in all breast cancer networks, except in the HER2-enriched. LUZP4 codifies a leucine zipper protein. This protein has not been extensively explored, however, recently, this leucine zipper has been identified as involved in mRNA exporting in cancer cells (Viphakone et al., 2015). Our results indicate that this gene may be an important player in Luminal A, Luminal B, or Basal. Furthermore, unlike the two previously discussed genes, LUZP4 has not been extensively studied in the context of breast cancer. This gene is an example of the type of new biological information that can be recovered from existing data only with the use of network-based approaches. We believe that further experimental exploration of this molecule may be of interest in the future.

As it has been largely discussed (Margolin et al., 2006; Baca-López et al., 2012), the probabilistic inference of gene regulatory networks is fundamental (and in some cases almost mandatory) to unveil complex transcriptional interactions that would be otherwise extremely difficult to notice. However, as it has also been noticed (Margolin et al., 2006), reverse engineering methods have also a number of limitations. In the particular case of transcriptional networks inferred via Mutual Information calculations, one important aspect is that regulatory interactions may include not only canonical transcription factor to target interactions, but also a number of indirect and more complex relationships. In this sense, MI-inferred networks (such as the ones discussed here) provide us with information about statistical dependency in the transcriptional profiles (Hernández-Lemus et al., 2009; Baca-López et al., 2012; Creixell et al., 2015).

A major open question in the field of network reconstruction is still how to establish a proper cut-off value to include a link in a network (Serrano et al., 2009). We decided to make this choice based on two constraints: on one hand, to include links with a high statistical confidence, as measured with the associated *p*-value provided by ARACNE, and on the other one, to obtain a network that belongs to the *complex network regime*, as described by Albert and Barabási (2002), which exhibit a node-to-link ratio in which *nodes* $<<$ *links*.

It is evident that the threshold choice will modify network parameters. We identified that, in an interval surrounding our selected node-to-link ratio (0.1), network parameters are generally stable (**Supplementary Material 3**). More importantly, even though the actual numeric values may be modified, is important to notice that the behaviors such as similarity between networks (**Supplementary Material 4**) and the relative influence of genes, based on their node degree (**Supplementary Material 5**) are preserved in this interval. Importantly, this shows robustness of the transcriptional regulatory program, while at the same time allowing us to observe structural differences between phenotypes along the

interval. It is worth noticing that the network structures obtained with our methodology are quite different to networks generated at random (see **Supplementary Material 6**).

We have shown how a network approach can be useful to understand the heterogeneity of breast cancer. We were able to infer and compare the transcriptional programs of breast cancer molecular subtypes, and contrast them with that of healthy mammary tissue. We showed how this paradigm can help to identify novel roles of molecules in different manifestations of breast cancer, with potential biomedical implications.

AUTHOR CONTRIBUTIONS

GdA devised the experiment, implemented the methodologies and provided this manuscript's first draft. TV contributed to the discussion and visualization of results and helped in the revision of the manuscript. JE contributed to the analysis of results, discussed biological implications, and helped in the writing of the manuscript. EH proposed the experiment, participated in its design and coordination and helped in the writing of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by CONACYT (grant no.179431/2012), as well as by federal funding from the National Institute of Genomic Medicine (Mexico). Additional support has been granted by the National Laboratory of Complexity Sciences (grant no. 232647/2014 CONACYT).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fphys.2016.00568/full#supplementary-material

Supplementary Material 1 | Normalization of microarrays with batch effect correction and information on subtype composition of each microarray experiment.

Supplementary Material 2 | Principal component analysis visualizations of gene expression.

Supplementary Material 3 | A table containing the results of the network analysis performed for each phenotype at different cut-off values.

Supplementary Material 4 | A table containing the node and edge Jaccard indexes *J* between each pair of phenotypes at different cut-off values.

Supplementary Material 5 | A zipped file containing lists with the 100 highest ranked genes by degree for each phenotype at different cut-off values.

Supplementary Material 6 | Null model for network connectivity: nodes from transcriptional networks were reconnected according to the Erdős-Rényi model.

REFERENCES

Ahmad, F. K., Deris, S., and Othman, N. (2012). The inference of breast cancer metastasis through gene regulatory networks. *J. Biomed. Inform.* 45, 350–362. doi: 10.1016/j.jbi.2011.11.015

Albert, R. (2005). Scale-free networks in cell biology. *J. Cell Sci.* 118, 4947–4957. doi: 10.1242/jcs.02714

Albert, R., and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97. doi: 10.1103/RevModPhys.74.47

Baca-López, K., Mayorga, M., Hidalgo-Miranda, A., Gutiérrez-Nájera, N., and Hernández-Lemus, E. (2012). The role of master regulators in the metabolic/transcriptional coupling in breast carcinomas. *PLoS ONE* 7:e42678. doi: 10.1371/journal.pone.0042678

Ballman, K. V., Grill, D. E., Oberg, A. L., and Therneau, T. M. (2004). Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* 20, 2778–2786. doi: 10.1093/bioinformatics/bth327

Barabási, A.-L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68. doi: 10.1038/nrg2918

Barabasi, A.-L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272

Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human b cells. *Nat. Genet.* 37, 382–390. doi: 10.1038/ng1532

Biane, C., Delaplace, F., and Klaudel, H. (2016). Networks and games for precision medicine. *Biosystems* 150, 52–60. doi: 10.1016/j.biosystems.2016.08.006

Carro, M. S., Lim, W. K., Alvarez, M. J., Bollo, R. J., Zhao, X., Snyder, E. Y., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. *Nature* 463, 318–325. doi: 10.1038/nature08712

Castro, M. A., de Santiago, I., Campbell, T. M., Vaughn, C., Hickey, T. E., Ross, E., et al. (2016). Regulators of genetic risk of breast cancer identified by integrative network analysis. *Nat. Genet.* 48, 12–21. doi: 10.1038/ng.3458

Censi, F., Giuliani, A., Bartolini, P., and Calcagnini, G. (2011). A multiscale graph theoretical approach to gene regulation networks: a case study in atrial fibrillation. *IEEE Trans. Biomed. Eng.* 58, 2943–2946. doi: 10.1109/TBME.2011.2150747

Chakraborty, G., Rangaswami, H., Jain, S., and Kundu, G. C. (2006). Hypoxia regulates cross-talk between syk and lck leading to breast cancer progression and angiogenesis. *J. Biol. Chem.* 281, 11322–11331. doi: 10.1074/jbc.M512546200

Chakravarti, B., Ravi, J., and Ganju, R. K. (2014). Cannabinoids as therapeutic agents in cancer: current status and future implications. *Oncotarget* 5:5852. doi: 10.18632/oncotarget.2233

Chen, C., Grennan, K., Badner, J., Zhang, D., Gershon, E., Jin, L., et al. (2011). Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. *PLoS ONE* 6:e17238. doi: 10.1371/journal.pone.0017238

Cicatiello, L., Mutarelli, M., Grober, O. M., Paris, O., Ferraro, L., Ravo, M., et al. (2010). Estrogen receptor α controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microrrnas. *Am. J. Pathol.* 176, 2113–2130. doi: 10.2353/ajpath.2010.090837

Cowell, R. G. (2006). *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. New York, NY: Springer Science & Business Media.

Creixell, P., Reimand, J., Haider, S., Wu, G., Shibata, T., Vazquez, M., et al. (2015). Pathway and network analysis of cancer genomes. *Nat. Methods* 12:615. doi: 10.1038/nmeth.3440

Davidson, E., and Levin, M. (2005). Gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4935–4935. doi: 10.1073/pnas.0502024102

Davidson, E. H., and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* 311, 796–800. doi: 10.1126/science.1113832

de Matos Simoes, R., and Emmert-Streib, F. (2012). Bagging statistical network inference from large-scale gene expression data. *PLoS ONE* 7:e33624. doi: 10.1371/journal.pone.0033624

Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., et al. (2007). Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the transbig multicenter independent validation series. *Clin. Cancer Res.* 13, 3207–3214. doi: 10.1158/1078-0432.CCR-06-2765

Edgar, R., Domrachev, M., and Lash, A. E. (2002). Gene expression omnibus: ncbi gene expression and hybridization array data repository. *Nucl. Acids Res.* 30, 207–210. doi: 10.1093/nar/30.1.207

Espinoza-Fonseca, L. M., Autry, J. M., and Thomas, D. D. (2015). Sarcolipin and phospholamban inhibit the calcium pump by populating a similar metal ion-free intermediate state. *Biochem. Biophys. Res. Commun.* 463, 37–41. doi: 10.1016/j.bbrc.2015.05.012

Fajardo, V. A., Bombardier, E., Vigna, C., Devji, T., Bloemberg, D., Gamu, D., et al. (2013). Co-expression of serca isoforms, phospholamban and sarcolipin in human skeletal muscle fibers. *PLoS ONE* 8:e84304. doi: 10.1371/journal.pone.0084304

Farmer, P., Bonnefoi, H., Becette, V., Tubiana-Hulin, M., Fumoleau, P., Larsimont, D., et al. (2005). Identification of molecular apocrine breast tumours by microarray analysis. *Breast Cancer Res.* 7(Suppl. 2):P2–P11. doi: 10.1186/bcr1122

Fenwick, M. A., Mora, J. M., Mansour, Y. T., Baithun, C., Franks, S., and Hardy, K. (2013). Investigations of tgf-β signaling in preantral follicles of female mice reveal differential roles for bone morphogenetic protein 15. *Endocrinology* 154, 3423–3436. doi: 10.1210/en.2012-2251

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000). Using bayesian networks to analyze expression data. *J. Comput. Biol.* 7, 601–620. doi: 10.1089/106652700750050961

Fu, R., Liu, H., Zhang, J., Liu, C., Ding, S., Li, L., et al. (2015). Expression of nk-activating receptor-nkp46/ncr1 on nk cells in patients with severe aplastic anemia. *Clin. Lab.* 61, 1221–1229. doi: 10.7754/Clin.Lab.2015.150130

García-Campos, M. A., Espinal-Enríquez, J., and Hernández-Lemus, E. (2015). Pathway analysis: state of the art. *Front. Physiol.* 6:383. doi: 10.3389/fphys.2015.00383

Gardner, T. S., di Bernardo, D., Lorenz, D., and Collins, J. J. (2003). Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301, 102–105. doi: 10.1126/science.1081900

Giuliani, A., Zbilut, J. P., Conti, F., Manetti, C., and Miccheli, A. (2004). Invariant features of metabolic networks: a data analysis application on scaling properties of biochemical pathways. *Physica A* 337, 157–170. doi: 10.1016/j.physa.2004.01.053

Gorban, A. N., Smirnova, E. V., and Tyukina, T. A. (2010). Correlations, risk and crisis: from physiology to finance. *Physica A* 389, 3193–3217. doi: 10.1016/j.physa.2010.03.035

Gorski, P. A., Glaves, J. P., Vangheluwe, P., and Young, H. S. (2013). Sarco(endo)plasmic reticulum calcium atpase (serca) inhibition by sarcolipin is encoded in its luminal tail. *J. Biol. Chem.* 288, 8456–8467. doi: 10.1074/jbc.M112.446161

Grass, P. (2009). "Experimental design," in *Batch Effects and Noise in Microarray Experiments*, chapter 3, ed A. Scherer (Hoboken, NJ: John Wiley & Sons, Ltd.), 19–31. doi: 10.1002/9780470685983.ch3

Gu, F., Hsu, H.-K., Hsu, P.-Y., Wu, J., Ma, Y., Parvin, J., et al. (2010). Inference of hierarchical regulatory network of estrogen-dependent breast cancer through chip-based data. *BMC Syst. Biol.* 4:170. doi: 10.1186/1752-0509-4-170

Chen, G., Wang, X., and Li, X. (2015). *Fundamentals of Complex Networks: Models, Structures and Dynamics*. San Francisco, CA: Blackwell Publ.

Guida, M., Ligresti, A., De Filippis, D., D'Amico, A., Petrosino, S., Cipriano, M., et al. (2010). The levels of the endocannabinoid receptor cb2 and its ligand 2-arachidonoylglycerol are elevated in endometrial carcinoma. *Endocrinology* 151, 921–928. doi: 10.1210/en.2009-0883

Guimerà, R., and Amaral, L. A. (2005). Functional cartography of complex metabolic networks. *Nature* 433, 895–900. doi: 10.1038/nature03288

Hageman, J., van Waarde, M. A., Zylicz, A., Walerych, D., and Kampinga, H. H. (2011). The diverse members of the mammalian hsp70 machine show distinct chaperone-like activities. *Biochem. J.* 435, 127–142. doi: 10.1042/BJ20101247

Hakes, L., Pinney, J. W., Robertson, D. L., and Lovell, S. C. (2008). Protein-protein interaction networks and biologywhat's the connection? *Nat. Biotechnol.* 26, 69–72. doi: 10.1038/nbt0108-69

Hasty, J., McMillen, D., Isaacs, F., and Collins, J. J. (2001). Computational studies of gene regulatory networks: in numero molecular biology. *Nat. Rev. Genet.* 2, 268–279. doi: 10.1038/35066056

Hernández-Lemus, E., and Rangel-Escareño, C. (2011). "The role of information theory in gene regulatory network inference," in *Information Theory: New Research*, eds P. Deloumeaux and J. D. Gorzalka (New York, NY: Nova Publishing Group), 109–144.

Hernández-Lemus, E., and Siqueiros-García, J. M. (2013). Information theoretical methods for complex network structure reconstruction. *Complex Adapt. Syst. Model.* 1, 1–22. doi: 10.1186/2194-3206-1-8

Hernández-Lemus, E., Velázquez-Fernández, D., Estrada-Gil, J. K., Silva-Zolezzi, I., Herrera-Hernández, M. F., and Jiménez-Sánchez, G. (2009). Information theoretical methods to deconvolute genetic regulatory networks applied to thyroid neoplasms. *Physica A* 388, 5057–5069. doi: 10.1016/j.physa.2009.08.013

House, C. D., Vaske, C. J., Schwartz, A. M., Obias, V., Frank, B., Luu, T., et al. (2010). Voltage-gated na+ channel scn5a is a key regulator of a gene transcriptional network that controls colon cancer invasion. *Cancer Res.* 70, 6957–6967. doi: 10.1158/0008-5472.CAN-10-1169

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., et al. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264. doi: 10.1093/biostatistics/4.2.249

Ivshina, A. V., George, J., Senko, O., Mow, B., Putti, T. C., Smeds, J., et al. (2006). Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res.* 66, 10292–10301. doi: 10.1158/0008-5472.CAN-05-4414

Jeong, H., Mason, S. P., Barabási, A.-L., and Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* 411, 41–42. doi: 10.1038/35075138

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabási, A.-L. (2000). The large-scale organization of metabolic networks. *Nature* 407, 651–654. doi: 10.1038/35036627

Jha, S., Jones, V., Burridge, K., and Mukhopadhyay, S. (2012). Cb2 receptor-mediated regulation of prostate cancer cell migration: involvement of rhoa and stress fiber formation. *FASEB J.* 26(Suppl.):782.11

Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127. doi: 10.1093/biostatistics/kxj037

Jonsson, M., Björntorp Mark, E., Brantsing, C., Brandner, J. M., Lindahl, A., and Asp, J. (2004). Hash4, a novel human achaete-scute homologue found in fetal skin. *Genomics* 84, 859–866. doi: 10.1016/j.ygeno.2004.07.004

Khosravi, P., Gazestani, V., Law, B., Bader, G., and Sadeghi, M. (2015). "Comparative analysis of co-expression networks reveals molecular changes during the cancer progression," in *World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015* (Toronto, ON: Springer), 1481–1487. doi: 10.1007/978-3-319-19387-8_360

Kitano, H. (2002). Systems biology: a brief overview. *Science* 295, 1662–1664. doi: 10.1126/science.1069492

Köster, A., Landgraf, S., Leipold, A., Sachse, R., Gebhart, E., Tulusan, A. H., et al. (1990). Expression of oncogenes in human breast cancer specimens. *Anticancer Res.* 11, 193–201.

Kruse, P. H., Matta, J., Ugolini, S., and Vivier, E. (2013). Natural cytotoxicity receptors and their ligands. *Immunol. Cell Biol.* 92, 221–229. doi: 10.1038/icb.2013.98

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., et al. (2002). Transcriptional regulatory networks in saccharomyces cerevisiae. *Science* 298, 799–804. doi: 10.1126/science.1075090

Levine, M., and Davidson, E. H. (2005). Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4936–4942. doi: 10.1073/pnas.0408031102

Lim, W. K., Lyashenko, E., and Califano, A. (2009). "Master regulators used as breast cancer metastasis classifier," in *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (Santa Clara, CA: NIH Public Access), 504.

Liu, R., Wang, X., Chen, G. Y., Dalerba, P., Gurney, A., Hoey, T., et al. (2007). The prognostic role of a gene signature from tumorigenic breast-cancer cells. *New Engl. J. Med.* 356, 217–226. doi: 10.1056/NEJMoa063994

Liu, Z., Zhang, X.-S., and Zhang, S. (2014). Breast tumor subgroups reveal diverse clinical prognostic power. *Sci. Rep.* 4:4002. doi: 10.1038/srep04002

Luce, R. D., and Perry, A. D. (1949). A method of matrix analysis of group structure. *Psychometrika* 14, 95–116. doi: 10.1007/BF02289146

Madhamshettiwar, P. B., Maetschke, S. R., Davis, M. J., Reverter, A., and Ragan, M. A. (2012). Gene regulatory network inference: evaluation and application to ovarian cancer allows the prioritization of drug targets. *Genome Med.* 4, 1–16. doi: 10.1186/gm340

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., et al. (2006). Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7(Suppl. 1):S7. doi: 10.1186/1471-2105-7-S1-S7

Maslov, S., and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science* 296, 910–913. doi: 10.1126/science.1065103

McCall, M. N., Bolstad, B. M., and Irizarry, R. A. (2010). Frozen robust multiarray analysis (fRMA). *Biostatistics* 11, 242–253. doi: 10.1093/biostatistics/kxp059

Miller, L. D., Smeds, J., George, J., Vega, V. B., Vergara, L., Ploner, A., et al. (2005). An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13550–13555. doi: 10.1073/pnas.0506230102

Minn, A. J., Gupta, G. P., Siegel, P. M., Bos, P. D., Shu, W., Giri, D. D., et al. (2005). Genes that mediate breast cancer metastasis to lung. *Nature* 436, 518–524. doi: 10.1038/nature03799

Morales, P., Blasco-Benito, S., Andradas, C., Gómez-Cañas, M., Flores, J. M., Goya, P., et al. (2015). Selective, nontoxic cb2 cannabinoid o-quinone with *in vivo* activity against triple-negative breast cancer. *J. Med. Chem.* 58, 2256–2264. doi: 10.1021/acs.jmedchem.5b00078

Munro, S., Thomas, K. L., and Abu-Shaar, M. (1993). Molecular characterization of a peripheral receptor for cannabinoids. *Nature* 365, 61–65. doi: 10.1038/365061a0

Nasser, M. W., Qamri, Z., Deol, Y. S., Smith, D., Shilo, K., Zou, X., et al. (2011). Crosstalk between chemokine receptor cxcr4 and cannabinoid receptor cb 2 in modulating breast cancer growth and invasion. *PLoS ONE* 6:e23901. doi: 10.1371/journal.pone.0023901

Newman, M. (2010). *Networks: An Introduction*. Oxford: Oxford University Press.

Nielsen, T. D., and Jensen, F. V. (2009). *Bayesian Networks and Decision Graphs*. New York, NY: Springer Science & Business Media.

Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* 27, 1160–1167. doi: 10.1200/JCO.2008.18.1370

Pau Ni, I. B., Zakaria, Z., Muhammad, R., Abdullah, N., Ibrahim, N., Aina Emran, N., et al. (2010). Gene expression patterns distinguish breast carcinomas from normal breast tissues: the malaysian context. *Pathol. Res. Pract.* 206, 223–228. doi: 10.1016/j.prp.2009.11.006

Pawitan, Y., Bjöhle, J., Amler, L., Borg, A.-L., Egyhazi, S., Hall, P., et al. (2005). Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res.* 7:R953. doi: 10.1186/bcr1325

Pe'er, D., and Hacohen, N. (2011). Principles and strategies for developing network models in cancer. *Cell* 144, 864–873. doi: 10.1016/j.cell.2011.03.001

Pembroke, T., Christian, A., Jones, E., Hills, R. K., Wang, E. C., Gallimore, A. M., et al. (2014). The paradox of nkp46+ natural killer cells: drivers of severe hepatitis c virus-induced pathology but *in-vivo* resistance to interferon α treatment. *Gut* 63, 515–524. doi: 10.1136/gutjnl-2013-304472

Pérez-Gómez, E., Andradas, C., Blasco-Benito, S., Caffarel, M. M., García-Taboada, E., Villa-Morales, M., et al. (2015). Role of cannabinoid receptor cb2 in her2 pro-oncogenic signaling in breast cancer. *J. Nat. Cancer Inst.* 107:djv077. doi: 10.1093/jnci/djv077

Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747–752. doi: 10.1038/35021093

Persani, L., Rossetti, R., Di Pasquale, E., Cacciatore, C., and Fabre, S. (2014). The fundamental role of bone morphogenetic protein 15 in ovarian function and its involvement in female fertility disorders. *Hum. Reprod. Update* 20, 869–883. doi: 10.1093/humupd/dmu036

Pisanti, S., Picardi, P., D'Alessandro, A., Laezza, C., and Bifulco, M. (2013). The endocannabinoid signaling system in cancer. *Trends Pharmacol. Sci.* 34, 273–282. doi: 10.1016/j.tips.2013.03.003

Qamri, Z., Preet, A., Nasser, M. W., Bass, C. E., Leone, G., Barsky, S. H., et al. (2009). Synthetic cannabinoid receptor agonists inhibit tumor growth and metastasis of breast cancer. *Mol. Cancer Ther.* 8, 3117–3129. doi: 10.1158/1535-7163.MCT-09-0448

Robinson, J. L., and Nielsen, J. (2016). Integrative analysis of human omics data using biomolecular networks. *Mol. Biosyst.* 12, 2953–2964. doi: 10.1039/c6mb00476h

Rodriguez-Barrueco, R., Yu, J., Saucedo-Cuevas, L. P., Olivan, M., Llobet-Navas, D., Putcha, P., et al. (2015). Inhibition of the autocrine il-6–jak2–stat3–calprotectin axis as targeted therapy for hr-/her2+ breast cancers. *Genes Dev.* 29, 1631–1648. doi: 10.1101/gad.262642.115

Rody, A., Holtrich, U., Pusztai, L., Liedtke, C., Gaetje, R., Ruckhaeberle, E., et al. (2009). T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and her2-positive breast cancers. *Breast Cancer Res.* 11:R15. doi: 10.1186/bcr2234

Rueda, L. (2014). *Microarray Image and Data Analysis: Theory and Practice*. Londres: CRC Press.

Serrano, L. (2007). Synthetic biology: promises and challenges. *Mol. Syst. Biol.* 3:158. doi: 10.1038/msb4100202

Serrano, M. A., Boguna, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6483–6488. doi: 10.1073/pnas.0808904106

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Skibinski, A., and Kuperwasser, C. (2015). The origin of breast tumor heterogeneity. *Oncogene* 34, 5309–5316. doi: 10.1038/onc.2014.475

Sophocleous, A., Marino, S., Logan, J. G., Mollat, P., Ralston, S. H., and Idris, A. I. (2015). Bone cell-autonomous contribution of type 2 cannabinoid receptor to breast cancer-induced osteolysis. *J. Biol. Chem.* 290, 22049–22060. doi: 10.1074/jbc.M115.649608

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., et al. (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J. Nat. Cancer Inst.* 98, 262–272. doi: 10.1093/jnci/djj052

Sutton-McDowall, M. L., Purdey, M., Brown, H. M., Abell, A. D., Mottershead, D. G., Cetica, P. D., et al. (2015). Redox and anti-oxidant state within cattle oocytes following *in vitro* maturation with bone morphogenetic protein 15 and follicle stimulating hormone. *Mol. Reprod. Dev.* 82, 281–294. doi: 10.1002/mrd.22470

Tanimine, N., Tanaka, Y., Abe, T., Piao, J., Chayama, K., and Ohdan, H. (2016). Functional behavior of nkp46-positive intrahepatic natural killer cells against hepatitis c virus reinfection after liver transplantation. *Transplantation* 100, 355–364. doi: 10.1097/TP.0000000000001031

Thattai, M., and Van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* 98, 8614–8619. doi: 10.1073/pnas.151588598

Tovar, H., García-Herrera, R., Espinal-Enríquez, J., and Hernández-Lemus, E. (2015). Transcriptional master regulator analysis in breast cancer genetic networks. *Comput. Biol. Chem.* 59, 67–77. doi: 10.1016/j.compbiolchem.2015.08.007

Tripathi, A., King, C., de la Morenas, A., Perry, V. K., Burke, B., Antoine, G. A., et al. (2008). Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. *Int. J. Cancer* 122, 1557–1566. doi: 10.1002/ijc.23267

Van De Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *New Engl. J. Med.* 347, 1999–2009. doi: 10.1056/NEJMoa021967

Velasco, G., Hernández-Tiedra, S., Dávila, D., and Lorente, M. (2016). The use of cannabinoids as anticancer agents. *Prog. Neuro Psychopharmacol. Biol. Psychiatry* 64, 259–266. doi: 10.1016/j.pnpbp.2015.05.010

Viphakone, N., Cumberbatch, M. G., Livingstone, M. J., Heath, P. R., Dickman, M. J., Catto, J. W., et al. (2015). Luzp4 defines a new mrna export pathway in cancer cells. *Nucl. Acids Res.* 43, 2353–2366. doi: 10.1093/nar/gkv070

Walker, V. E., Wong, M. J., Atanasiu, R., Hantouche, C., Young, J. C., and Shrier, A. (2010). Hsp40 chaperones promote degradation of the herg potassium channel. *J. Biol. Chem.* 285, 3319–3329. doi: 10.1074/jbc.M109.024000

Wang, W., Cherry, J. M., Nochomovitz, Y., Jolly, E., Botstein, D., and Li, H. (2005). Inference of combinatorial regulation in yeast transcriptional networks: a case study of sporulation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1998–2003. doi: 10.1073/pnas.0405537102

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature* 393, 440–442. doi: 10.1038/30918

Yoshida, S., Ohbo, K., Takakura, A., Takebayashi, H., Okada, T., Abe, K., et al. (2001). Sgn1, a basic helix-loop-helix transcription factor delineates the salivary gland duct cell lineage in mice. *Dev. Biol.* 240, 517–530. doi: 10.1006/dbio.2001.0473

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.

Copyright © 2016 de Anda-Jáuregui, Velázquez-Caldelas, Espinal-Enríquez and Hernández-Lemus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or 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.