



The Hierarchical Modular Structure of HER2+ Breast Cancer Network

Sergio Antonio Alcalá-Corona^{1,2†}, Jesús Espinal-Enríquez^{1,2†},
Guillermo de Anda-Jáuregui¹ and Enrique Hernández-Lemus^{1,2*}

¹ Computational Genomics, National Institute of Genomic Medicine, Mexico City, Mexico, ² Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Ciudad de México, México

OPEN ACCESS

Edited by:

Matteo Barberis,
University of Amsterdam, Netherlands

Reviewed by:

Adil Mardinoglu,
Chalmers University of Technology,
Sweden

Monika Heiner,
Brandenburg University of
Technology Cottbus-Senftenberg,
Germany

*Correspondence:

Enrique Hernández-Lemus
ehernandez@inmegen.gob.mx

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 22 March 2018

Accepted: 19 September 2018

Published: 11 October 2018

Citation:

Alcalá-Corona SA, Espinal-Enríquez J,
de Anda-Jáuregui G and
Hernández-Lemus E (2018) The
Hierarchical Modular Structure of
HER2+ Breast Cancer Network.
Front. Physiol. 9:1423.
doi: 10.3389/fphys.2018.01423

HER2-enriched breast cancer is a complex disease characterized by the overexpression of the ERBB2 amplicon. While the effects of this genomic aberration on the pathology have been studied, genome-wide deregulation patterns in this subtype of cancer are also observed. A novel approach to the study of this malignant neoplasm is the use of transcriptional networks. These networks generally exhibit modular structures, which in turn may be associated to biological processes. This modular regulation of biological functions may also exhibit a hierarchical structure, with deeper levels of modular organization accounting for more specific functional regulation. In this work, we identified the most probable (maximum likelihood) model of the hierarchical modular structure of the HER2-enriched transcriptional network as reconstructed from gene expression data, and analyzed the statistical associations of modules and submodules to biological functions. We found modular structures, independent from direct ERBB2 amplicon regulation, involved in different biological functions such as signaling, immunity, and cellular morphology. Higher resolution submodules were identified in more specific functions, such as micro-RNA regulation and the activation of viral-like immune response. We propose the approach presented here as one that may help to unveil mechanisms involved in the development of the pathology.

Keywords: modular networks, breast cancer, gene regulatory networks, transcription, genetic, signaling pathways, molecular subtypes

1. INTRODUCTION

Breast cancer is the malignant neoplasm with the highest incidence and mortality among women worldwide (Ferlay et al., 2014). Breast cancer is a heterogeneous disease, and this poses a challenge for its treatment: a multitude of clinical, physiological and survival outcomes, all affect the choice of therapeutic options (Polyak, 2011; Network, 2012). The heterogeneity of breast cancer can be traced down to the subcellular level, which includes changes in the transcriptional programs through mutations, alterations of epigenetic regulation, chromosomal aberrations, among others.

HER2-enriched (HER2+) breast cancer is a paradigmatic example of the role of alterations in the chromosomal structure throughout the development of cancer. HER2+ breast cancer is characterized by the overexpression of the *HER2* receptor, encoded in the *ERBB2* gene located on Chromosome 17: Amplification of the Chr17q12 locus leads to the overexpression of the receptor that can be identified through immunohistochemical and transcriptomic approaches (Perou et al., 2000; Burstein, 2005). This *HER2 amplicon* also includes genes such as *STARD3*, *GRB7*, *PGAP3*, *TOP2*, *MED1*, *THRA*, *RARA*, *IGFPB4*, *CCR7*, *KRT20*, *KRT19*, and *GAST*.

While the effects that genes in this region have in the HER2+ breast cancer phenotype have been extensively explored, high-throughput technologies that allow genome-wide studies open the possibility for further exploration of the genomic landscape of this pathology. Major breakthroughs in the study of genomic cancer landscapes have been accomplished through the use of network theory. For instance, relationships between gene expression levels can be modeled as transcriptional networks. These networks have proven to be useful to associate different biological features of interest to a particular phenotype (de Anda-Jáuregui et al., 2016; Espinal-Enríquez et al., 2017).

A major topological feature of transcriptional networks is the fact that these have *modular* structures. Modules in networks, also known in some contexts as *communities*, are structural sub-units (subnetworks) broadly defined as subsets of tightly interconnected nodes so that the density of *within-connections* is higher than that of *between-connections* (Girvan and Newman, 2002; Porter et al., 2009; Fortunato, 2010). One challenge of the community structure finding algorithms is that there is no consensus in the appropriate within/between ratio. One measure of the efficacy of different algorithms with variable ratios is given by the LFR benchmark (Lancichinetti et al., 2009). In some cases, modules in transcriptional networks are associated to biological processes and functions observed in a given phenotype (Cantini et al., 2015). Thus, modules in a gene transcriptional network may capture some functional aspects underlying the phenomenology.

In this work, we explored the modular structure in the HER2+ breast cancer transcriptional network. We previously inferred said network (de Anda-Jáuregui et al., 2015), using a Mutual Information (MI)- based algorithm, from gene expression patterns measured using microarrays. Then, we analyzed the modularity of this network at three different levels: connected components (islands), modules in the largest component, and submodules in the largest modules. Over-representation analysis (ORA) was used to identify modular structures associated with functional categories as defined in Gene Ontology (GO).

Our analysis identified a transcriptional network with an explicit modular structure at each level of modularity resolution explored. We found associations between some of these modules and biological features of interest. The highest level of modular organization, that of connected components, captures the transcriptional relationships associated with the HER2 amplicon. Modules inside connected components were identified linked to biological processes, including Extracellular Matrix (ECM) organization, signaling, and immune response. Finally, at the finest level of modular organization, we found groups of genes associated to more specific processes, such as viral response, plasma membrane organization, and micro-RNA regulation. These results show the usefulness of understanding the transcriptional architecture of disease in order to dissect the mechanisms behind the appearance of a pathological phenotype.

2. METHODS

We used the *hierarchical map equation* (Rosvall and Bergstrom, 2011) to find nested submodules into the *HER2+ breast cancer network*, and then, using the approach of Alcalá-Corona et al. (2016), we identified whether those submodules were associated to a particular biological function. This section is divided as follows: network inference, differential expression analysis, modularity and submodularity detection and the hierarchical map equation, and functional analysis. A graphical representation of this methodology can be observed in **Figure 1**. It is worth mentioning that all the code used in this work is publicly available in our Github site https://github.com/CSB-IG/BioNetworkInference_ModularAnalysis.

2.1. Network Inference

The network architecture of breast cancer molecular subtypes has been previously analyzed (de Anda-Jáuregui et al., 2016). There, network inference was carried out by using data on 493 microarray expression profiles for breast cancer samples processed on the Affymetrix HGU133A platform. Mutual information calculations were performed by means of the ARACNe algorithm (Margolin et al., 2006). PAM50-subtyped gene expression datasets were obtained as in de Anda-Jáuregui et al. (2015). From PAM50 algorithm we conserved the HER2+ subtypes only. In this work, we built upon such transcriptional network structure to carry out posterior analyses. As a result of this network inference, nodes represent genes in the transcriptional space and edges is the statistical dependence between two genes, which is a robust measure of the degree of co-expression existing in any couple of genes, and edges the statistical dependence between two genes, which is a robust measure of the degree of co-expression existing in any pair of genes.

2.2. Differential Expression Analysis

Independent gene-based linear models were adjusted using *limma* R package (R Core Team, 2013) to find differential expressed genes (DEGs) in the tumor samples compared to the control ones (61 samples in the same array, as described in de Anda-Jáuregui et al., 2016) using (1):

$$y_{ij} = \mu + \alpha_i + \varepsilon_j \quad (1)$$

where y_{ij} , is the \log_2 (normalized gene expression); μ , is the global mean; α_i is the i th experimental condition (normal or tumor) and $\varepsilon_j \sim N(0, \sigma)$ is the random error term of the j th sample. We also used a hypothesis tests based on empirical Bayes moderation of the standard errors toward a common value, in order to obtain p -values which were adjusted to control multiple comparisons using the False Discovery Rate (FDR) (Benjamini and Hochberg, 1995). We then defined a gene as differentially expressed if it had a $FDR < 1 \times 10^{-5}$ and a $|\log_2(\text{fold change})| > 1$. As it is observed, **Figure 1B** reflects how genes are grouped in a first modular structure, based on their differential expression, overexpressed genes have more overexpressed neighbors. Analogously, underexpressed genes are clustered with other underexpressed genes.

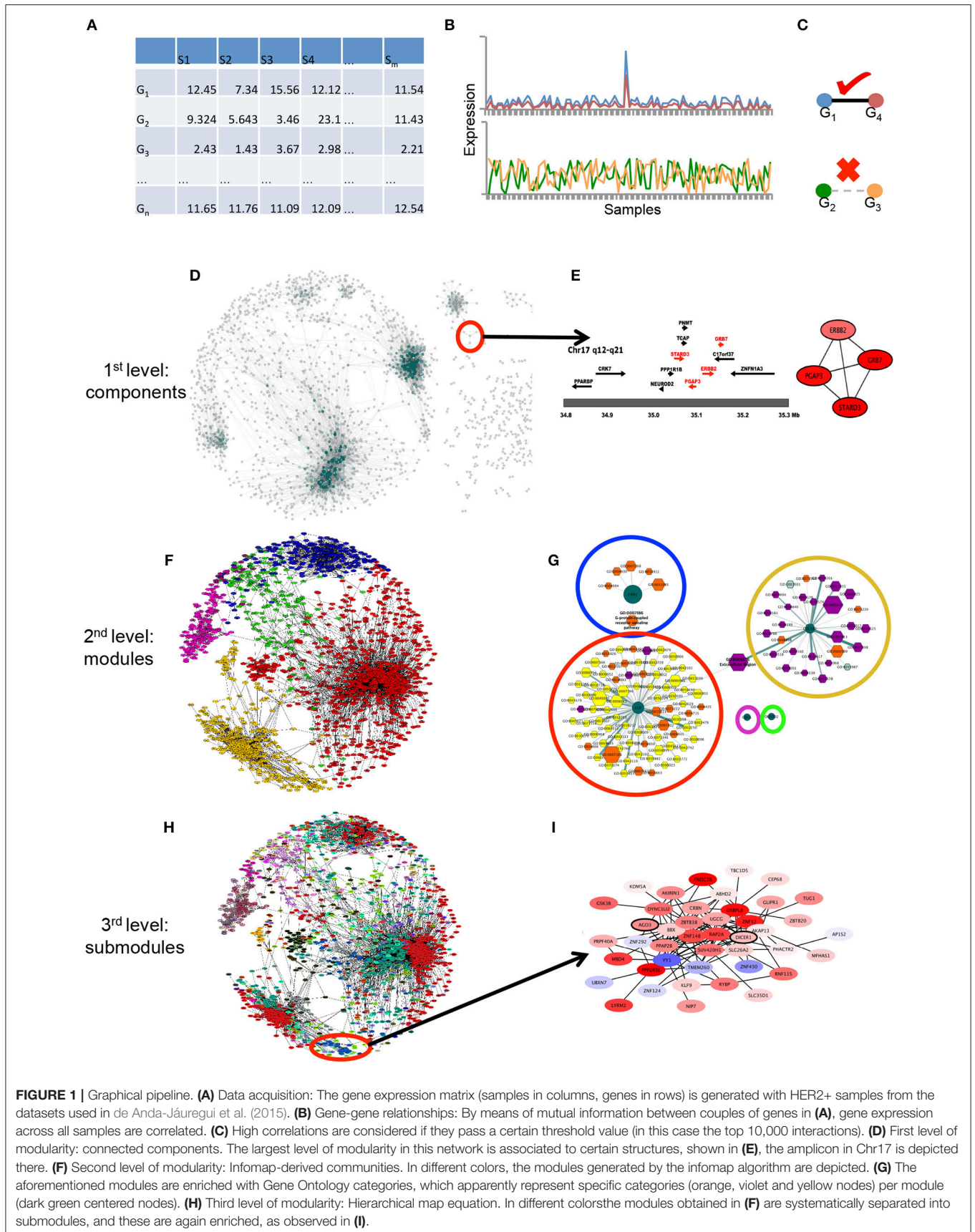


FIGURE 1 | Graphical pipeline. **(A)** Data acquisition: The gene expression matrix (samples in columns, genes in rows) is generated with HER2+ samples from the datasets used in de Anda-Jáuregui et al. (2015). **(B)** Gene-gene relationships: By means of mutual information between couples of genes in **(A)**, gene expression across all samples are correlated. **(C)** High correlations are considered if they pass a certain threshold value (in this case the top 10,000 interactions). **(D)** First level of modularity: connected components. The largest level of modularity in this network is associated to certain structures, shown in **(E)**, the amplicon in Chr17 is depicted there. **(F)** Second level of modularity: Infomap-derived communities. In different colors, the modules generated by the infomap algorithm are depicted. **(G)** The aforementioned modules are enriched with Gene Ontology categories, which apparently represent specific categories (orange, violet and yellow nodes) per module (dark green centered nodes). **(H)** Third level of modularity: Hierarchical map equation. In different colors the modules obtained in **(F)** are systematically separated into submodules, and these are again enriched, as observed in **(I)**.

2.3. Modularity and Submodularity Detection and the Hierarchical Map Equation

In order to find the hierarchical modular structure in the network, several methods to obtain modules from a given network have been developed (Girvan and Newman, 2002; Clauset et al., 2004; Adamcsek et al., 2006; Alves, 2007; Fortunato and Barthelemy, 2007; Arenas et al., 2008; Amini et al., 2013 for an in-depth revision, see Fortunato, 2010). Additionally, several of these algorithms have been used to find modules into gene regulatory networks (Dey and Meyer, 2015; Binder et al., 2016; Bonsang-Kitzis et al., 2016; Feng et al., 2016; Miecznikowski et al., 2016). For this work, we decide to use the generalization of the *map equation*, developed by Rosvall and Bergstrom (Rosvall and Bergstrom, 2011), as this is one of the best performing network partition algorithms as revealed by several stringent benchmark tests (Lancichinetti et al., 2009). In general, the hierarchical *map equation* method allows for multiple description levels of the movements of a random walker within the modules and allows an arbitrary number of such movements inside modules, submodules, subsubmodules, and so on, to the finest level. This additional description allows exploiting the fact that fine-level modules are themselves organized into larger modules. Modules and submodules are labeled according to the gene with the highest PageRank, a centrality measure that takes into account the cumulative weight of adjacent links (Brin and Page, 1998).

Graph partition by means of random walk-based methods (such as the method used here), generally speaking, optimize information flow within partitions and over the whole network. To do so, most of these algorithms perform constrained optimization over an entropic probability measure. In the present case, this is done by minimum description length optimization and bootstrapped permutation over an ensemble of encoded trajectories to generate an optimum partition. The outcome of the method is a maximum entropy partition. Hence, there is no need to evaluate different partitions, since there is no better partition compliant with the data than the one obtained. Infomap is one of the flagship of non-overlapping community detection methods. It has shown to be robust in the most stringent benchmarks (such as the LFR, Q-maximization and Spectral matrix-based tests), its computational complexity is linear in time, i.e., $O(N)$. These facts have been already tested and are well established within the complex networks research community (Fortunato, 2010; Fortunato and Hric, 2016).

2.4. Functional Analyses

Once the modules of the network have been detected, we explored whether the genes in these modules, arisen from the connections in the network (co-regulation model), are associated to a particular biological function. This is achieved by performing an *Over Representation Analysis (ORA)* based on the **hypergeometric test** over a category of genes whose function is known or annotated in a database. For this study, we use the *Gene Ontology Consortium* database **GO** (Ashburner et al., 2000). This *hypergeometric test* represents a *null model* to calculate how probable is it that a set of randomly chosen genes k belongs

to a category (biological function) annotated in said database. Thus, a *p-value* is associated with this test, so the lower the *p-value* the lower the probability that the set k of genes belongs randomly to the category, and therefore represents a statistical trust about the particular gene set over the whole category. Additional analyses included the identification of differentially expressed genes (DEG) using the **limma** R package (Ritchie et al., 2015), and their functional characterization using *Ingenuity Pathway Analysis (IPA)* (Krämer et al., 2014).

2.5. Validation

In order to provide validation of our biological findings on an independent experimental dataset, we used 89 HER2+ samples from the METABRIC database. METABRIC (Molecular Taxonomy of Breast International Consortium) is a collection of clinically annotated primary fresh-frozen breast cancer specimens from tumor banks from the UK and Canada (Curtis et al., 2012), with transcriptomics data measured with the Illumina HT-12 v3 microarray platform.

Due to the differences in coverage between the Affymetrix 133A and the Illumina HT-12 v3 microarrays, we decided to discard for the downstream analysis genes that could not be simultaneously measured in both platforms. We used this filtered METABRIC expression matrix to infer the full set of Mutual Information values for all gene pairs using ARACNE. Then, this network was pruned by keeping the 10,000 gene pairs with the highest MI values, which formed our transcriptional network. The hierarchical modular structure of this transcriptional network was analyzed using the Infomap algorithm, and the resulting modules and submodules were inspected for functional associations following the same criteria previously used for the discovery network.

3. RESULTS AND DISCUSSION

Complex networks are mathematical models that represent the intricate interrelationship structure in complex systems. In the case of biological networks (in particular, gene regulatory networks), there is a search for modular partitions that may reflect a semi-mechanistic (ideally, a fully mechanistic) structure in which particular modules (and submodules) are responsible to carry out certain biological functions.

Conceptually, one may think of modules within a regulatory network that perform some functions in a way that it is not completely independent of the whole genome regulation program but is, to a certain extent, autonomous. This is thought to be so, since biological functions are often robust in their control and hence, resilient to potentially harmful systemic damage. Control theory has proven that one of the easiest way to combine robustness, with relative autonomy and global control is modularization.

The graph partition algorithm used here (InfoMap/MapEquation) is based on the consideration of an ensemble of random walkers performing stochastic trajectories over the network, a coding procedure is performed on the trajectories generating an ensemble of travel codes that are then subject to minimum description length optimization. The

whole process is repeated in a large number of bootstrapped permutations that aside from the minimum description length optimization (and the fact that random walk trajectories are subject to the central limit theorem) ensures that one has the most probable modular partition of the network. Since this modular partition is subject to stringent optimization, we may refer to as the modular partition of the network, because even if it is not unique, is by far the most likely to happen.

3.1. The Hierarchical Structure of HER2+ Network

Transcriptional networks associated with specific molecular subtypes of breast cancer have characteristic structures (de Anda-Jáuregui et al., 2016). In the case of the HER2+ molecular subtype, we can observe a network with 2,100 nodes and 9,856 edges; this network is integrated by a giant component, along with several (161) smaller connected components or islands (Figure 2). The basic topological parameters of this network are shown in Table 1.

Breast cancer transcriptional networks also exhibit modular structures, as it has been previously reported (Alcalá-Corona et al., 2017). The HER2+ molecular subtype is no exception.

The first level of modularity observed in this network is the aforementioned distribution of nodes into connected components. As it is shown in Figure 1, the bulk of connections in the network are part of the giant component. Moreover 1,649 out of the total number of nodes are part of it. It is in the giant component where we can analyze the modular hierarchical structure of HER2+ breast cancer network.

In Figure 3 we show different visualizations of the giant component, highlighting different properties: In panel A, we show this largest component using a spring-embedded layout. In panel B, we color the nodes according to differential expression (ranging from blue to red in terms of under and over expression);

TABLE 1 | Network metrics.

Parameter	Value
Nodes	2,100
Edges	9,856
Avg. node degree	4,583
Char. path length	5,917
Connected components	162

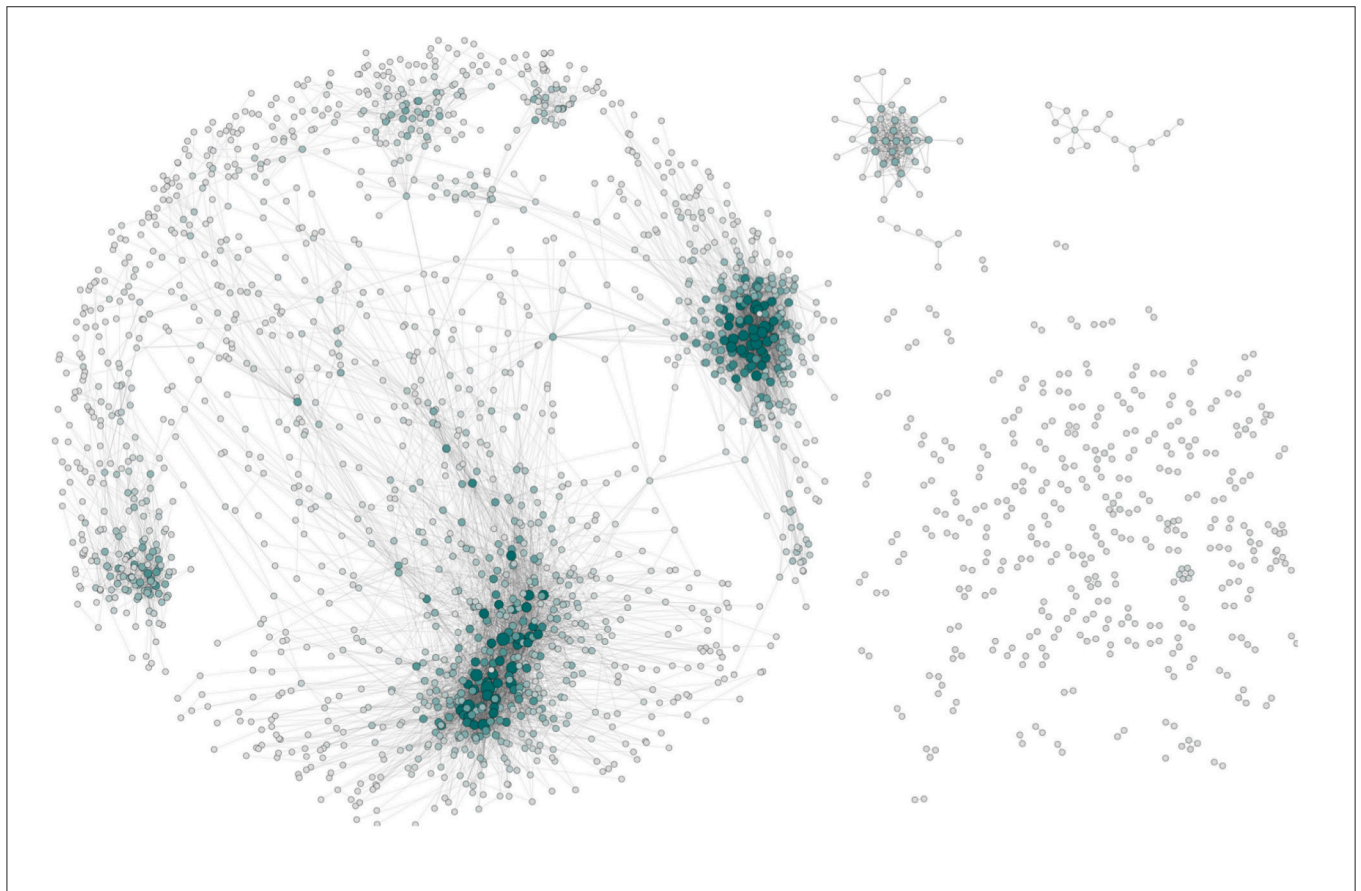
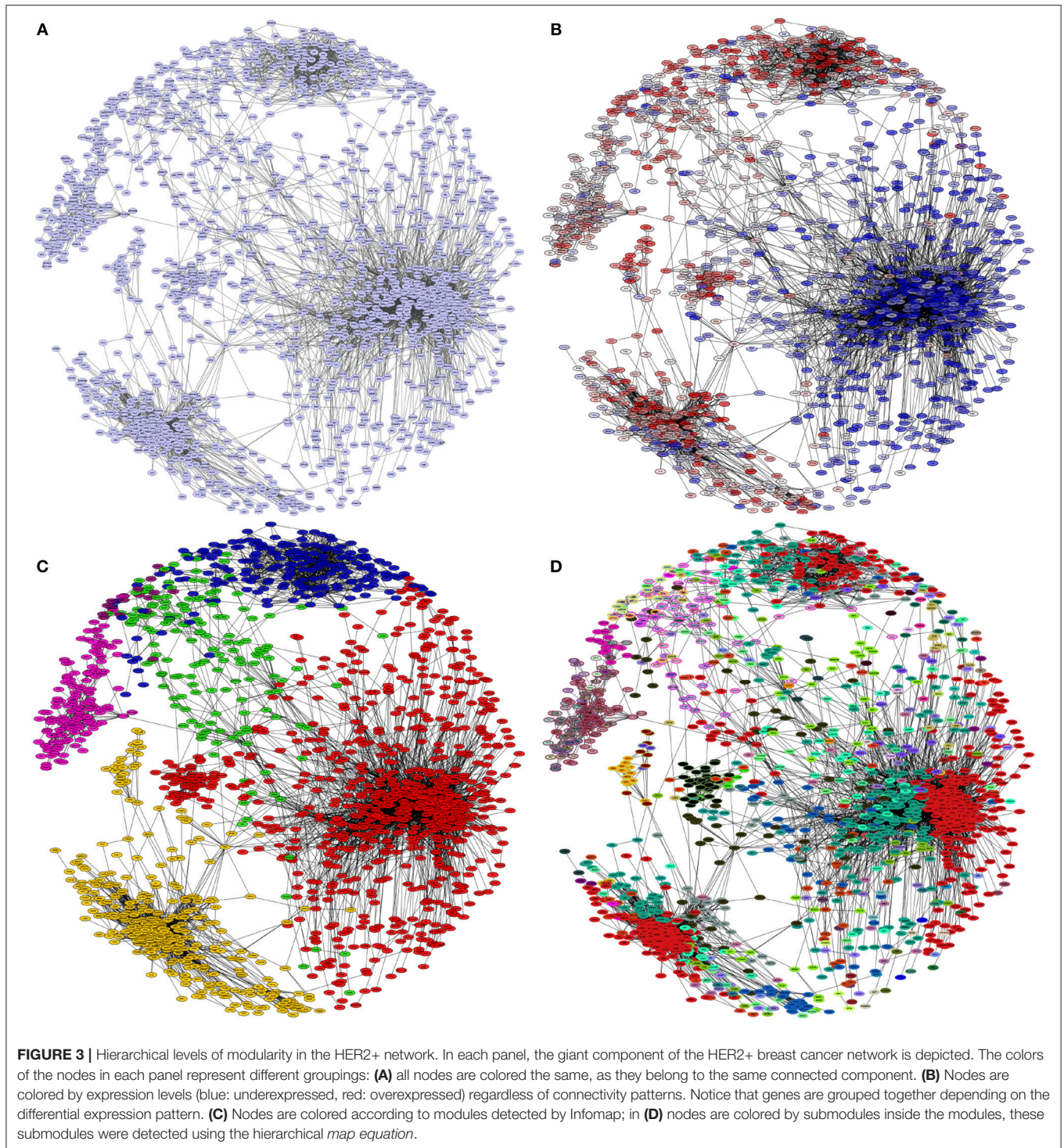


FIGURE 2 | The network architecture of Her2+ breast cancer. The visualization shows the connected components (islands) of this network separately. In this representation nodes are colored and sized according to their node degree (i.e., the number of neighbors connected to a gene). The layout and arrangement of the network showcases the giant component (Left) and several small islands (Right).



nodes with similar expression values are arranged together and have a higher density of connections among them, which is in accordance to the fact that connections in this network arise from common expression patterns. This in turn leads to panels C and D, that show modules and submodules, respectively, based on the connectivity patterns exhibited in the network, as identified

using the Infomap algorithm based on the hierarchical *map equation*.

Previously, we have shown that modular structures in transcriptional networks are usually associated with biological features (Alcalá-Corona et al., 2016, 2017). Here we show that new insights on the biological features of HER2+ breast cancer

may be revealed by exploring the functionality related to different layers of modularity.

3.2. The HER2 Amplicon Genes Are Modularly Isolated in the Transcriptional Network

As we have previously mentioned, the nodes that integrate the HER2+ transcriptional network are distributed in 162 different connected components. The vast majority of nodes belong to a single giant component, whereas the rest of connected components are much smaller: the next component in size has only 39 nodes, and 120 components are composed by two nodes each. At this level of modularity, it is hard to systematically analyze functionality of connected components, as it is unlikely to identify biological functions and processes that can be associated with two or three molecules alone.

However, an important exception is the case of a component integrated by four genes that belong to the HER2 amplicon: ERBB2, GRB7, PGAP3, and STARD3. These four genes are closely located at the core of the HER2 amplicon, which is the single, most important hallmark of this particular subtype of cancer (Kauraniemi and Kallioniemi, 2006). Their fully connected pattern (clique), is indicative of a close coexpression exhibited by these four genes, which may be related to their genomic proximity in the amplified region (Figure 4). In previous work (Espinal-Enríquez et al., 2017) we have shown that transcriptional networks of breast cancer recover connections between genes belonging to the same chromosome, which are related to a loss of trans regulation associated with breast cancer. In the case of the network analyzed in this work, it appears that the more statistically significant connections that are recovered for the ERBB2 gene are to its closest genomic neighbors.

3.3. Functional Role of Modules in HER2 Transcriptional Network

To analyze the modular structure of the giant component and their functionality is important to have a more detailed vision of these genes in the disease. This giant component has a modular structure composed of six modules. In Figure 5 we see that

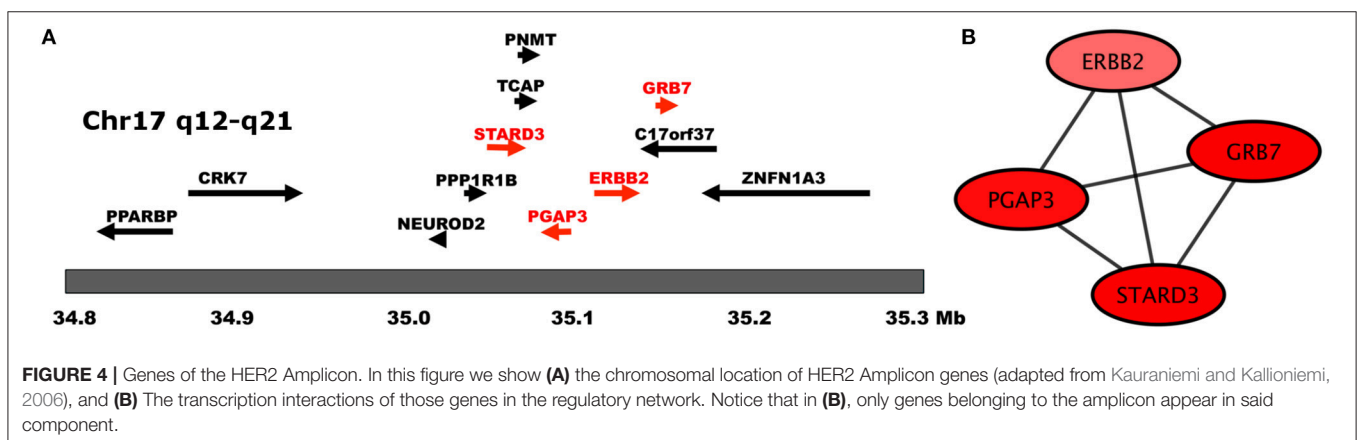
the functionality associated with these different modules can be clearly divided along functional themes, indicating a functional compartmentalization reflected in the transcriptional program.

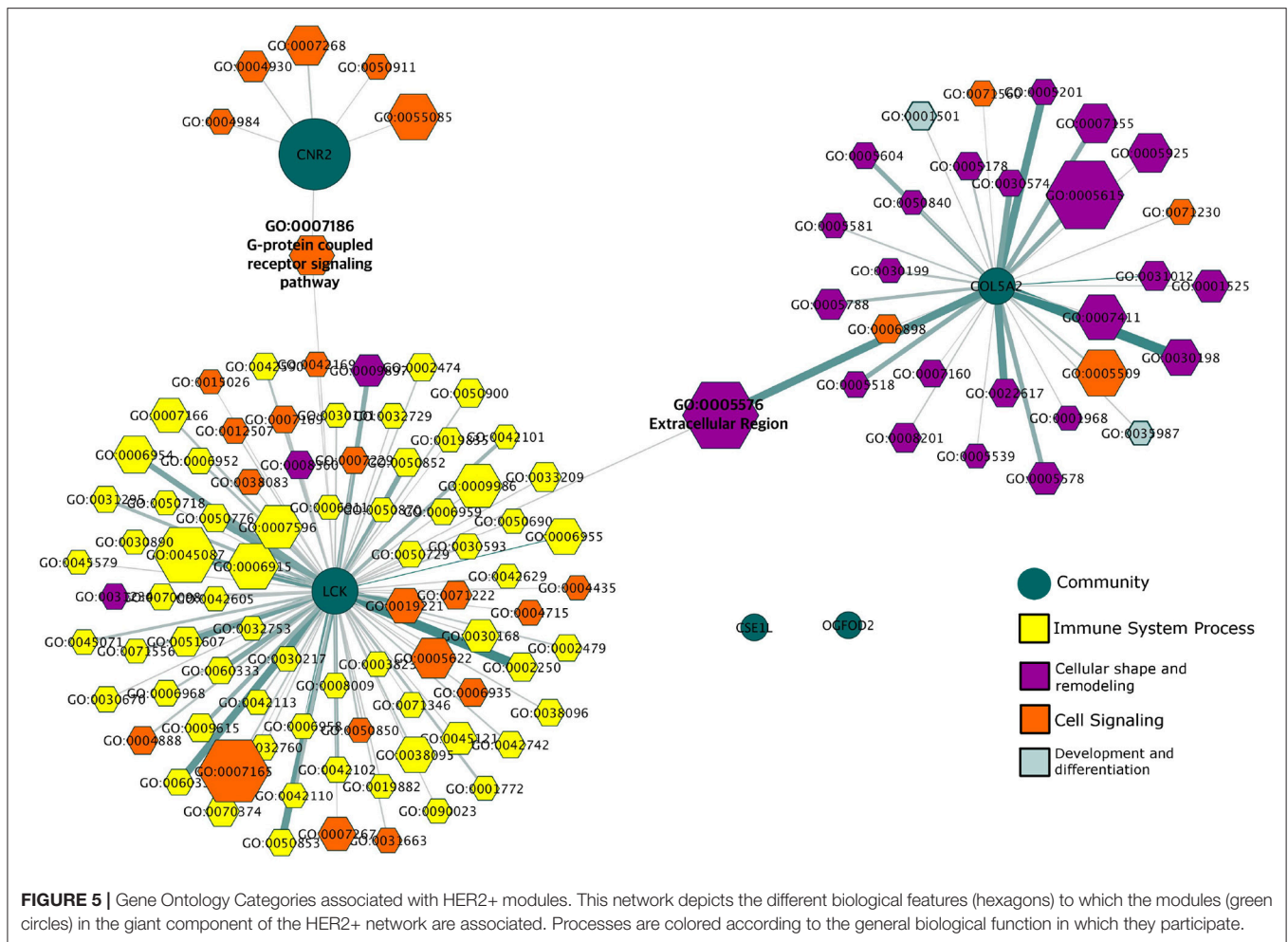
We can observe that the three largest modules (labeled CNR2, LCK, and COL5A2 based on the genes with the highest PageRank in each module) are related to processes of signaling and transport, immunity, and extracellular matrix organization respectively. We can also observe the existence of two processes that connect these communities at a functional level: GO:007186 (G-protein coupled receptor signaling pathway) which is associated with both the CNR2 and the LCK modules; and GO:0005576 (Extracellular region) which is associated with both the LCK and COL5A2 modules. In this regard, it is important to recall that the hierarchic *map equation* recovers non-overlapping modules: therefore, the fact that these processes can be associated with two distinct modules indicates that the regulation of those processes involves two different sets of genes.

3.3.1. The CNR2 Submodules Are Associated With the Regulation of Signaling and Micro-RNA Assembly

The largest module in the giant component of the HER2+ network has 763 genes with 4,574 links between them. The gene with the highest PageRank in this community is the CNR2 gene, which codifies for the cannabinoid receptor type 2 (CB2). This CNR2 module is composed by 49 submodules. Additionally, two of these modules: GPATCH4 and ZBTB38, have sub-submodular structures themselves (with 4 and 5 sub-submodules, respectively). As mentioned before, the CNR2 module is associated with signaling and transport processes. Interestingly, few of their submodules show enrichment (each submodule contains genes that participate in a specific function): these are the CNR2, PGLYRP4, and ZBTB38 submodules.

The aforementioned modules may be associated with specific cellular components, with the first two modules being involved in the cell membrane, while the last one correlates to the cytosolic region. Furthermore, the ZBTB38 submodule (defined by the gene that encodes zinc finger and BTB domain containing 38) is also associated with protein binding processes; taken into account together, this may be indicative that the ZBTB38 submodule





is more related to intracellular signaling than to membrane receptor-dependent signaling.

When the expression levels of these genes are taken into account (as seen in **Figure 6A**) we may observe a clear separation in this submodule, where only the ZBTB38 submodule is overexpressed, while the rest of the submodules in the CNR2 module are underexpressed. Inside the ZBTB38 we can also find two crucial genes for micro-RNA assembly and regulation: DICER1 and AGO (**Figure 6B**). These two are the main components in the RISC complex (Daugaard and Hansen, 2017). It has been shown that micro-RNA regulation is crucial for the induction of both epithelial-to-mesenchymal transition and mesenchymal-to-epithelial processes in breast cancer Drago-García et al. (2017). Hence a submodule related to micro-RNA-gene regulation may be relevant to identify novel therapeutic options.

3.3.2. The Submodules of COL5A2 Are Involved in Extracellular Matrix Organization

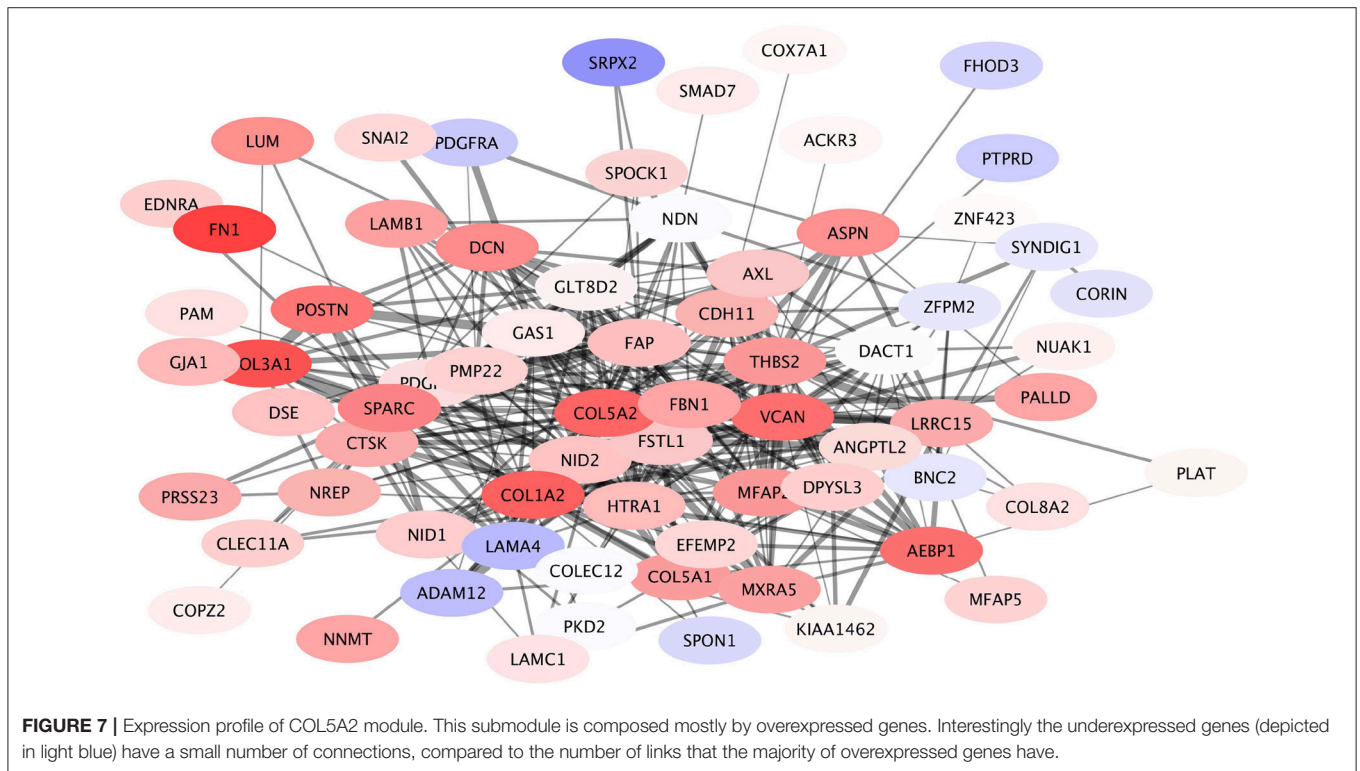
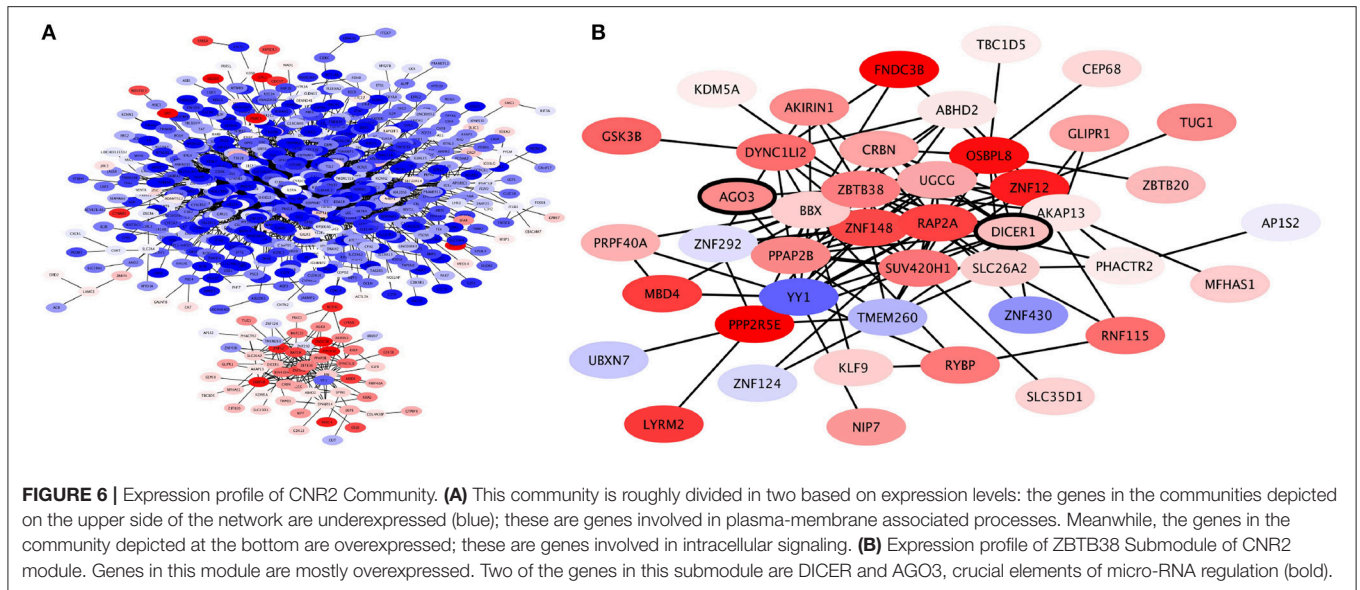
A module containing 169 nodes and 789 edges was identified in the giant component, containing the COL5A2 gene as the node with the highest PageRank. This COL5A2 module was associated

with Extracellular Matrix (ECM) organization. This module in turn is composed by 18 submodules, with COL5A2 being the defining gene of the largest one. This COL5A2 submodule again is associated with ECM organization, collagen organization and cell adhesion. The COL5A2 gene codifies for the Collagen 5A2 protein, a key participant in shaping the ECM. Other genes in this subcommunity that are widely known to play important roles in ECM organization include genes such as LUM, fibronectin, VCAN and members of the collagen family beyond COL5A2, among others.

In **Figure 7** the expression patterns of these genes can be seen. It shows that overall the members of this submodule are overexpressed. This is a clear indicative that in HER2+ breast cancer subtype, remodeling of extracellular matrix is a key participant in shaping the phenotype, it is fundamental for invasiveness, migration, Epithelial-to-mesenchymal transition (EMT) and other processes ubiquitous in cancer.

3.3.3. The Role of the Submodules of LCK in Immune Response

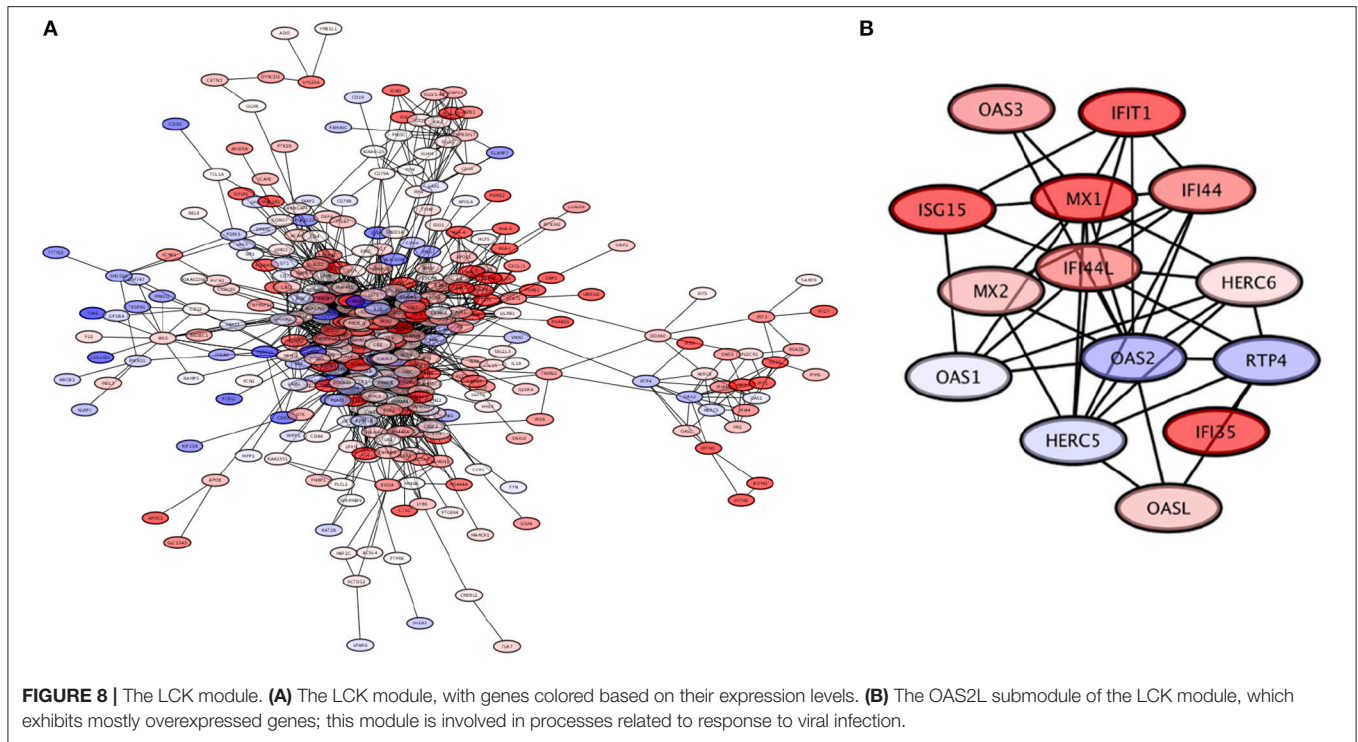
The third largest module of the giant component is composed by 371 genes with 3,011 connections among them, with the



highest *Page-Ranked* one being the LCK gene (which encodes the lymphocyte-specific protein tyrosine kinase). This module is subdivided into 27 submodules, with one of them (the OAS2L module) in turn having three sub-submodules. In **Figure 8A**, a visualization of this module is provided.

The LCK module is associated to the functioning of the immune system. In turn, five submodules of LCK: LCK, OAS2, PSMB9, SLAMF8, and TNFRSF17 show enrichment, again related to functions of the immune system. The OAS2L

submodule in particular called our attention. At the level of submodular structure, this module can be associated with processes related to the response to viral infections: “defense response to virus,” “type I interferon signaling pathway,” “response to virus,” and “negative regulation of viral genome replication.” When the three sub-submodules of OAS2L (DDX60, IFITM1, OAS2) are analyzed independently, they in turn are associated with seven processes that are also involved in viral infection response: this includes the



forementioned “defense response to virus,” “type I interferon signaling pathway,” “response to virus,” and “negative regulation of viral genome replication,” as well as new emerging processes “response to interferon-gamma,” “negative regulation of viral entry into host cell,” and “interferon-gamma-mediated signaling pathway.”

These results are complemented by the findings of analyzing the gene expression data using Ingenuity Pathway Analysis. The results of such analysis are shown in **Figure 9**. In particular, we highlight the appearance of the “infectious diseases” category; importantly, notice that the response detected is associated only to viral (as opposed to, for instance, bacterial) infection. The expression patterns of the OAS2L submodule (as seen in **Figure 8B**), leads us to argue that an activation of immunity with mechanisms used for the response to viral infections is found in HER2+ breast cancer subtype.

3.4. Validation

In the validation network we were able to recover several of the previously discussed findings regarding modular structure and biological functional associations (see **Supplementary Material**).

By applying the same methodology as in our discovery approach, we were able to recover an independent component that is composed by genes in the HER2 amplicon. In the discovery network, this component contains STARD3, GRB7, PGAP3 and ERBB2, whereas in the validation network it was composed by, GRB7, PGAP3, MIEN1, TCAP and ERBB2. MIEN1 and TCAP are also part of the amplicon. In both cases the important functional finding is the tight co-expression of the genes in the amplicon, and how these are completely isolated

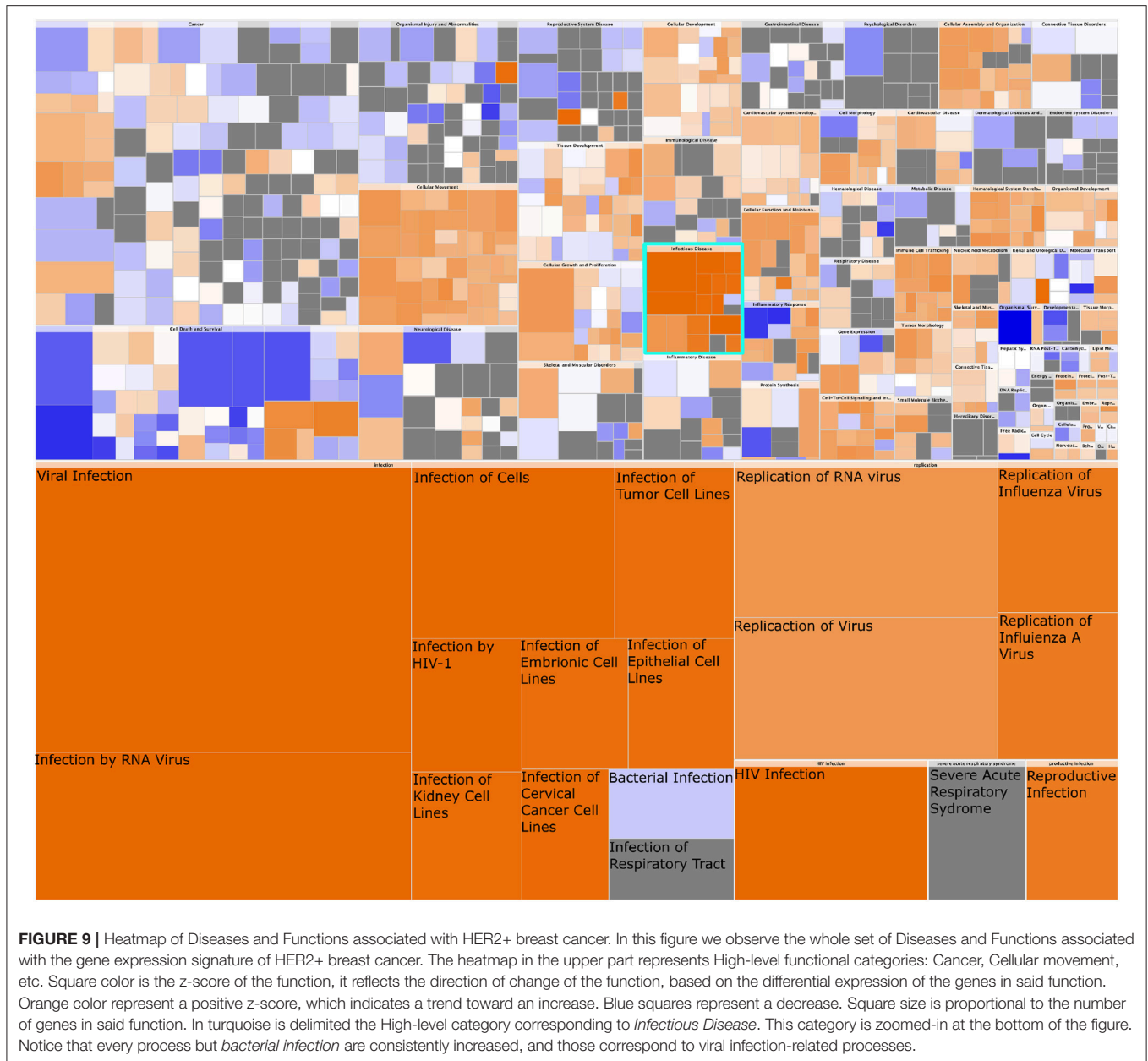
from other elements of the transcriptional network, including the largest connected component.

A largest connected component also appears in the validation network. As previously mentioned, this component is completely disconnected from HER2 amplicon regulation, suggesting independent mechanisms of regulation. We were also able to recover modular and sub-modular structure in this largest connected component with similar functional associations to those found in the discovery network. Some of these functional enrichments are described in the following lines:

We recovered a module containing several collagen-encoding genes, analogous to the COL5A2 module in the discovery network. Furthermore, this module was significantly associated to Extracellular Matrix processes, as well as other structural functions, concordant with the significant functional associations found in the discovery network.

We were also able to find a module functionally associated to immune system processes, as observed in the LCK module in the discovery network; in both cases, these were the largest modules. Furthermore, this validation immune module contain modular substructures, including a small group (25 genes), including OAS1, 2, and 3, as well as interferon alpha-inducing protein family members. This submodule was statistically associated to viral response processes, just like the one found in the discovery network.

Even though the node composition of these modules was not identical (due to the aforementioned differences in microarray technologies), they were functionally coincident. At the level of inquiry used here, some findings were not replicated in the validation network, in particular a module



associated to cell signaling processes in the largest connected component.

4. CONCLUSIONS

The modular structure of transcriptional networks and its relationship to biological functionality are topics of current biomedical interest. In this work, we have implemented a hierarchical module detection method to identify the highest resolution of modular structure of the transcriptional network of HER2+ breast cancer, and the functions associated with each network module.

Using this approach, we have identified biological features associated with different levels of modular structure in this network. At the highest level of modularity we observe a distribution of genes into different connected components, with more than half the genes detected belonging to a giant connected component. Furthermore, we may observe connectivity among genes of the HER2-amplicon, the most important genomic element associated with the development of HER2+ breast cancer, in an independent, specific connected component.

At a higher modular resolution, we identified communities in the largest connected component, some of which are statistically associated with sets of Gene Ontology categories related to

specific biological functions: immune system processes, cellular shape and remodeling, and cell signaling. Furthermore, some of these Gene Ontologies are associated with more than one non-overlapping module, indicating the need for joint regulation from specialized sets of genes.

Finally, submodular structures in the HER2+ network reveal finer details of the processes involved in the pathological state. For instance, specific modules responsible of the regulation of intracellular signaling, micro-RNA assembly, and viral infection were identified inside, more general modules. As such, we show that a higher modular resolutions allows for the emergence of more specific biological function regulation.

Our work showcases the importance and usefulness of analyzing the phenomenon of transcriptional regulation using a complex network approach. We present cases in which the study of the modular structures beneath the transcriptional network architecture unveil mechanisms associated with the pathological state, which may lead to insights relevant to the biomedical community.

AUTHOR CONTRIBUTIONS

SAA-C contributed with methodological tools, performed calculations and analyses, collaborated in the theoretical design and discussion of results, and collaborated in writing and revising the paper. JE-E contributed with methodological tools, collaborated in the theoretical design and discussion of results,

and collaborated in writing and revising the paper. GdA-J contributed with methodological tools, collaborated in the theoretical design and discussion of results, and collaborated in writing and revising the paper. EH-L contributed to the general design and supervision of the project, contributed with methodological tools, collaborated in the theoretical design and discussion of results, and collaborated in writing and revising the paper.

FUNDING

This work was supported by CONACYT (grant no. 285544/2016), as well as by federal funding from the National Institute of Genomic Medicine (Mexico). Additional support has been granted by the Laboratorio Nacional de Ciencias de la Complejidad (grant no. 232647/2014 CONACYT). SAA-C is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 391657 from CONACYT. EH-L is recipient of the 2016 Marcos Moshinsky Fellowship in the Physical Sciences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2018.01423/full#supplementary-material>

REFERENCES

- Adamcsek, B., Palla, G., Farkas, I. J., Derényi, I., and Vicsek, T. (2006). Cfinder: locating cliques and overlapping modules in biological networks. *Bioinformatics* 22, 1021–1023. doi: 10.1093/bioinformatics/btl039
- Alcalá-Corona, S. A., De Anda Jáuregui, G., Espinal-Enríquez, J., and Hernández-Lemus, E. H.-L. (2017). Network modularity in breast cancer molecular subtypes. *Front. Physiol.* 8:915. doi: 10.3389/fphys.2017.00915
- Alcalá-Corona, S. A., Velázquez-Caldelas, T. E., Espinal-Enríquez, J., and Hernández-Lemus, E. (2016). Community structure reveals biologically functional modules in mef2c transcriptional regulatory network. *Front. Physiol.* 7:184. doi: 10.3389/fphys.2016.00184
- Alves, N. A. (2007). Unveiling community structures in weighted networks. *Phys. Rev. E* 76:036101. doi: 10.1103/PhysRevE.76.036101
- Amini, A. A., Chen, A., Bickel, P. J., and Levina, E. (2013). Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Statist.* 41, 2097–2122. doi: 10.1214/13-AOS1138
- Arenas, A., Fernandez, A., Fortunato, S., and Gomez, S. (2008). Motif-based communities in complex networks. *J. Phys. A Math. Theor.* 41:224001. doi: 10.1088/1751-8113/41/22/224001
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Methodol.* 57, 289–300.
- Binder, S. C., Eckweiler, D., Schulz, S., Bielecka, A., Nicolai, T., Franke, R., et al. (2016). Functional modules of sigma factor regulons guarantee adaptability and evolvability. *Sci. Reports* 6:22212. doi: 10.1038/srep22212
- Bonsang-Kitziz, H., Sadacca, B., Hamy-Petit, A. S., Moarii, M., Pinheiro, A., Laurent, C., et al. (2016). Biological network-driven gene selection identifies a stromal immune module as a key determinant of triple-negative breast carcinoma prognosis. *Oncoimmunology* 5:e1061176. doi: 10.1080/2162402X.2015.1061176
- Brin, S., and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* 30, 107–117. doi: 10.1016/S0169-7552(98)00110-X
- Burstein, H. J. (2005). The distinctive nature of her2-positive breast cancers. *N. Engl. J. Med.* 353, 1652–1654. doi: 10.1056/NEJMp058197
- Cantini, L., Medico, E., Fortunato, S., and Caselle, M. (2015). Detection of gene communities in multi-networks reveals cancer drivers. *Sci. Reports* 5:17386. doi: 10.1038/srep17386
- Clauset, A., Newman, M. E., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E* 70:066111. doi: 10.1103/PhysRevE.70.066111
- Curtis, C., Shah, S. P., Chin, S.-F., Turashvili, G., Rueda, O. M., Dunning, M. J., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* 486:346. doi: 10.1038/nature10983
- Daugaard, I., and Hansen, T. B. (2017). Biogenesis and function of ago-associated rnas. *Trends Genet.* 33, 208–219. doi: 10.1016/j.tig.2017.01.003
- de Anda-Jáuregui, G., Mejía-Pedroza, R. A., Espinal-Enríquez, J., and Hernández-Lemus, E. (2015). Crosstalk events in the estrogen signaling pathway may affect tamoxifen efficacy in breast cancer molecular subtypes. *Comput. Biol. Chem.* 59, 42–54. doi: 10.1016/j.compbiolchem.2015.07.004
- de Anda-Jáuregui, G., Velázquez-Caldelas, T. E., 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
- Dey, G., and Meyer, T. (2015). Phylogenetic profiling for probing the modular architecture of the human genome. *Cell Syst.* 1, 106–115. doi: 10.1016/j.cels.2015.08.006

- Drago-Garca, D., Espinal-Enrriquez, J., and Hernandez-Lemus, E. (2017). Network analysis of emt and met micro-rna regulation in breast cancer. *Sci. Reports* 7:13534. doi: 10.1038/s41598-017-13903-1
- Espinal-Enrriquez, J., Fresno, C., Anda-Jáuregui, G., and Hernández-Lemus, E. (2017). Rna-seq based genome-wide analysis reveals loss of inter-chromosomal regulation in breast cancer. *Sci. Reports* 7:1760. doi: 10.1038/s41598-017-01314-1
- Feng, L., Tong, R., Liu, X., Zhang, K., Wang, G., Zhang, L., et al. (2016). A network-based method for identifying prognostic gene modules in lung squamous carcinoma. *Oncotarget* 7:18006-20. doi: 10.18632/oncotarget.7632
- Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., et al. (2014). *Globocan 2012 v1. 0, Cancer Incidence and Mortality Worldwide: Iarc Cancerbase no. 11. 2013*. Available online at: globocan.iarc.fr
- Fortunato, S. (2010). Community detection in graphs. *Phys. Reports* 486, 75–174. doi: 10.1016/j.physrep.2009.11.002
- Fortunato, S., and Barthelemy, M. (2007). “Resolution limit in community detection.” *Proc. Natl. Acad. Sci. U.S.A.* 104, 36–41. doi: 10.1073/pnas.0605965104
- Fortunato, S., and Hric, D. (2016). Community detection in networks: a user guide. *Phys. Reports* 659, 1–44. doi: 10.1016/j.physrep.2016.09.002
- Girvan, M., and Newman, M. E. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7821–7826. doi: 10.1073/pnas.122653799
- Kauraniemi, P., and Kallioniemi, A. (2006). Activation of multiple cancer-associated genes at the erbb2 amplicon in breast cancer. *Endocr. Relat. cancer* 13, 39–49. doi: 10.1677/erc.1.01147
- Krämer, A., Green, J., Pollard, J., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* 30, 523–530. doi: 10.1093/bioinformatics/btt703
- Lancichinetti, A., Fortunato, S., and Kertész, J. (2009). Detecting the overlapping and hierarchical community structure in complex networks. *N. J. Phys.* 11:033015. doi: 10.1088/1367-2630/11/3/033015
- Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., and Califano, A. (2006). Reverse engineering cellular networks. *Nat. Protoc.* 1:662. doi: 10.1038/nprot.2006.106
- Miecznikowski, J. C., Gaile, D. P., Chen, X., and Tritchler, D. L. (2016). Identification of consistent functional genetic modules. *Stat. Applic. Genet. Mol. Biol.* 15, 1–18. doi: 10.1515/sagmb-2015-0026
- Network, T. C. G. A. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* 490:61–70. doi: 10.1038/nature11412
- Perou, C. M., Sorlie, 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
- Polyak, K. (2011). Heterogeneity in breast cancer. *J. Clin. Invest.* 121, 3786–3788. doi: 10.1172/JCI60534
- Porter, M. A., Onnela, J.-P., and Mucha, P. J. (2009). Communities in networks. *Notic. AMS* 56, 1082–1097. Available online at: <https://www.ams.org/journals/notices/200909/rtx090901082p.pdf>
- R Core Team, (2013). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Rosvall, M., and Bergstrom, C. T. (2011). Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS ONE* 6:e18209. doi: 10.1371/journal.pone.0018209

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 © 2018 Alcalá-Corona, Espinal-Enrriquez, de Anda-Jáuregui 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) and the copyright owner(s) 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.