



# Luminal A Breast Cancer Co-expression Network: Structural and Functional Alterations

Diana García-Cortés<sup>1,2</sup>, Enrique Hernández-Lemus<sup>1,3</sup> and Jesús Espinal-Enríquez<sup>1,3\*</sup>

<sup>1</sup> Computational Genomics Division, National Institute of Genomic Medicine, Mexico City, Mexico, <sup>2</sup> Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México, Mexico City, Mexico, <sup>3</sup> Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico

## OPEN ACCESS

### Edited by:

Kimberly Glass,  
Brigham and Women's Hospital and  
Harvard Medical School,  
United States

### Reviewed by:

Rebekka Burkholz,  
Harvard University, United States  
Alexander Lachmann,  
Icahn School of Medicine at Mount  
Sinai, United States

### \*Correspondence:

Jesús Espinal-Enríquez  
jespinal@inmegen.gob.mx

### Specialty section:

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

**Received:** 14 November 2020

**Accepted:** 17 March 2021

**Published:** 20 April 2021

### Citation:

García-Cortés D, Hernández-Lemus E  
and Espinal-Enríquez J (2021) Luminal  
A Breast Cancer Co-expression  
Network: Structural and Functional  
Alterations. *Front. Genet.* 12:629475.  
doi: 10.3389/fgene.2021.629475

Luminal A is the most common breast cancer molecular subtype in women worldwide. These tumors have characteristic yet heterogeneous alterations at the genomic and transcriptomic level. Gene co-expression networks (GCNs) have contributed to better characterize the cancerous phenotype. We have previously shown an imbalance in the proportion of intra-chromosomal (*cis*-) over inter-chromosomal (*trans*-) interactions when comparing cancer and healthy tissue GCNs. In particular, for breast cancer molecular subtypes (Luminal A included), the majority of high co-expression interactions connect gene-pairs in the same chromosome, a phenomenon that we have called loss of *trans*- co-expression. Despite this phenomenon has been described, the functional implication of this specific network topology has not been studied yet. To understand the biological role that communities of co-expressed genes may have, we constructed GCNs for healthy and Luminal A phenotypes. Network modules were obtained based on their connectivity patterns and they were classified according to their chromosomal homophily (proportion of *cis*-/*trans*- interactions). A functional overrepresentation analysis was performed on communities in both networks to observe the significantly enriched processes for each community. We also investigated possible mechanisms for which the loss of *trans*- co-expression emerges in cancer GCN. To this end we evaluated transcription factor binding sites, CTCF binding sites, differential gene expression and copy number alterations (CNAs) in the cancer GCN. We found that *trans*- communities in Luminal A present more significantly enriched categories than *cis*- ones. Processes, such as angiogenesis, cell proliferation, or cell adhesion were found in *trans*- modules. The differential expression analysis showed that FOXM1, CENPA, and CIITA transcription factors, exert a major regulatory role on their communities by regulating expression of their target genes in other chromosomes. Finally, identification of CNAs, displayed a high enrichment of deletion peaks in *cis*- communities. With this approach, we demonstrate that network topology determine, to at certain extent, the function in Luminal A breast cancer network. Furthermore, several mechanisms seem to be acting together to avoid *trans*- co-expression. Since this phenomenon has been observed in other cancer tissues, a remaining question is whether the loss of long distance co-expression is a novel hallmark of cancer.

**Keywords:** loss of long range co-expression, gene co-expression networks, Luminal A breast cancer, breast cancer, transcription factor analysis, CTCF binding site analysis

## 1. BACKGROUND

Gene co-expression networks (GCN) enable the study of interactions of highly correlated genes in a transcriptional program, capturing global and local connectivity properties emerging from those interactions (Sonawane et al., 2019). These type of networks are built from gene expression profiles, a measurable output of transcription. Therefore, they outline the contribution of the regulatory elements operating at different levels of the transcription process to ensure the expression of specific sets of genes. In this sense, GCNs might provide insights about shared regulatory mechanisms and their alterations in a disease, such as cancer (Emmert-Streib et al., 2014; Yang et al., 2014; Wu et al., 2019; Liao et al., 2020). Those alterations in cancer disrupt the transcriptional process and lead to altered gene expression and the promotion of tumor progression (Garraway and Lander, 2013; Lee and Young, 2013).

There are multiple studies where GCNs are constructed and important aspects of the connectivity structure are analyzed to identify genes prognosis markers (Hsu et al., 2019), metabolic deregulation (Serrano-Carbajal et al., 2020), and differences in transcriptional profiles (van Dam et al., 2018).

In breast cancer GCNs, there is an imbalance in the proportion of intra-chromosomal (*cis*-) over inter-chromosomal (*trans*-) gene co-expression interactions, meaning that the majority of high co-expression links connect gene-pairs in the same chromosome (Espinal-Enríquez et al., 2017; de Anda-Jáuregui et al., 2019a; Dorantes-Gilardi et al., 2020). This phenomenon has been called loss of long distance co-expression. Furthermore, a highly localized co-expression pattern associated with chromosome cytobands has been observed (García-Cortés et al., 2020). These features are not present in the healthy tissue GCN. In the entire set of co-expression interactions, the loss of long distance co-expression in breast cancer (measured in base pairs) subtypes is displayed as a decay in the *cis*- co-expression values dependent on gene physical distance (de Anda-Jáuregui et al., 2019b; García-Cortés et al., 2020).

The structural characteristics evaluated in the co-expression networks are different for each breast cancer molecular subtype, displaying another instance of their emblematic heterogeneity (Alcalá-Corona et al., 2017, 2018a). The four breast cancer molecular subtypes, Luminal A, Luminal B, HER2+ and Basal-like, are classified according to their gene expression profiles and they represent different cancer manifestations, with distinct molecular traits, genomic alterations, and prognosis (Perou et al., 2000; Prat and Perou, 2011; Berger et al., 2018). Hormone status, evaluated through the expression of estrogen and progesterone receptors (ER and PR correspondingly), and the presence of human epidermal growth factor receptor 2 (HER2), play a major role for breast cancer molecular subtypes characterization and the election of therapeutic strategies (Zhang et al., 2014).

Luminal A is the most frequent breast cancer molecular subtype. Almost a half of the total cases of breast cancer correspond to this phenotype (Fan et al., 2006). These tumors

are often positive to estrogen receptor (ER) and negative to ERBB2 receptor, and they also present overexpression on the ER-regulated genes. This subtype is associated with highest median survival, best prognosis (Hu et al., 2006), and lower recurrence rates (Arvold et al., 2011; Metzger-Filho et al., 2013).

Nevertheless, clinical and molecular heterogeneity is present within Luminal A tumors, where differences in genomic alterations have been potentially associated with resistance to endocrine therapy (Ciriello et al., 2013).

Additionally, the Luminal A GCN presents the least dissimilar structure compared with the healthy GCN (García-Cortés et al., 2020). A relevant measure to analyze differences in cancer GCNs, is the size of connected components. In the case of healthy GCN, as well as in the case of Luminal A GCN, they present a giant component (a set of connected genes that contains more than the half of the total amount of nodes in the networks). The other breast cancer subtype GCNs have only small intra-chromosomal connected components. Furthermore, Luminal A GCN is the one with the highest number of inter-chromosomal (*trans*-) interactions.

The structure of a GNC is often organized into *communities* or *modules* (Alcalá-Corona et al., 2016), this is, subsets of connected genes so that the density of within-connections is higher than that of between-connections (Girvan and Newman, 2002; Porter et al., 2009; Fortunato and Hric, 2016; Alcalá-Corona et al., 2018a). In the case of GCNs, communities may correspond to a co-regulated set of genes (Wilkinson and Huberman, 2004; Zhu et al., 2008; Cantini et al., 2015). The structure of said modules may capture the phenomenology behind biological processes (Alcalá-Corona et al., 2017, 2018a,b).

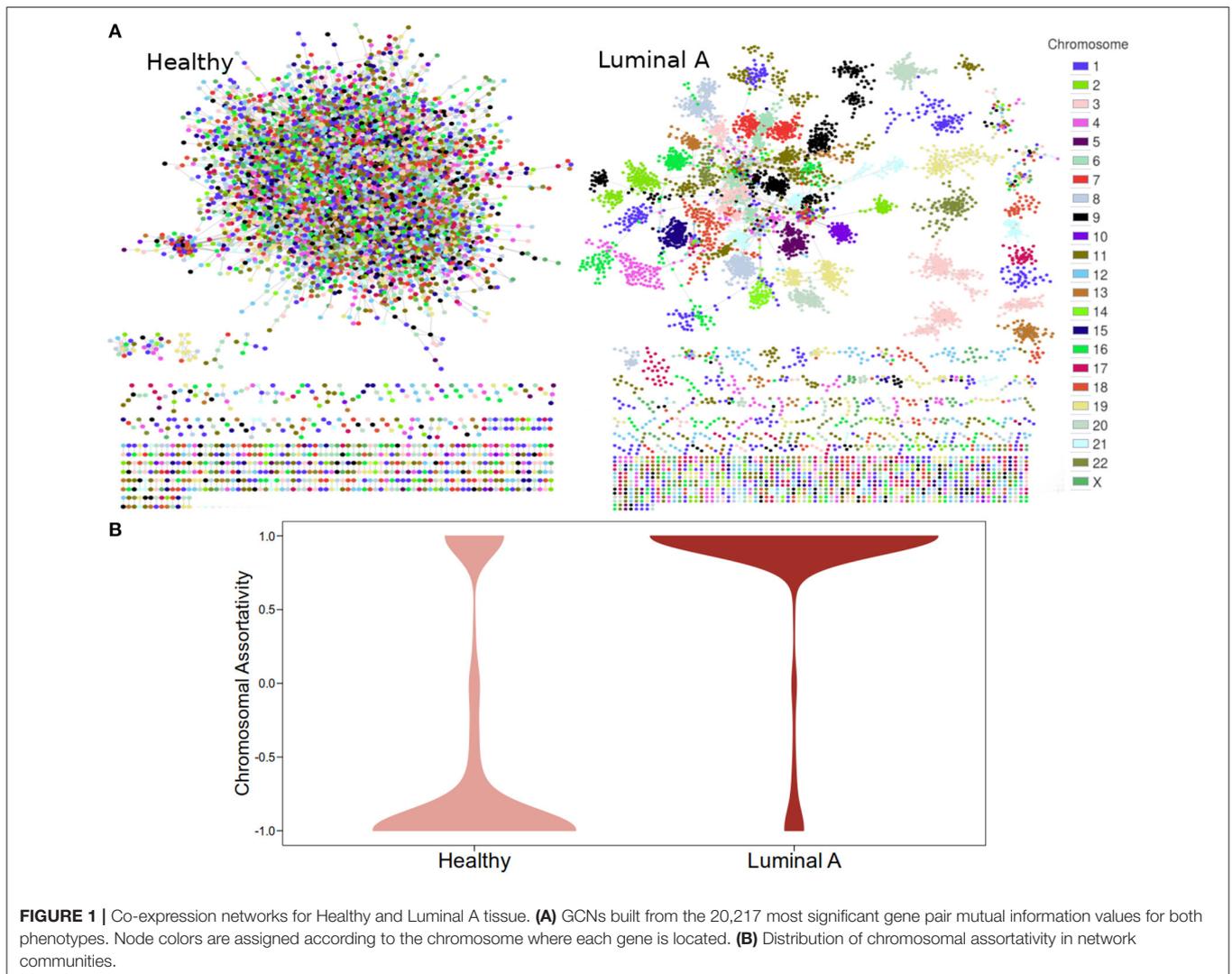
Being the subtype with the best prognosis, the most similar co-expression network, and taking into account that community structure in GCN may be implicated in the functional regulation of a cancerous phenotype, in this work we analyzed the structure of communities of the Luminal A GCN, in order to determine the relevance of the loss of long distance co-expression in the biological functions associated to that network. Additionally, we evaluated possible mechanisms for which we observe the preference for *cis*- interactions in this breast cancer subtype. We analyzed the influence of differential gene expression, transcription factor binding sites, copy number alterations, and CTCF binding sites, in order to understand the regulatory mechanisms underlying the appearance of the loss of long distance interactions in cancer GCNs.

## 2. RESULTS AND DISCUSSION

### 2.1. Community Structure Displays Loss of *trans*- Co-expression

Figure 1A displays GCNs built from the 20,217 (see Methods section) most significant mutual information interactions in the Luminal A and the Healthy co-expression profiles. Genes are colored according to the chromosome where they are located. As previously reported, the Healthy GCN has a giant component with interactions linking genes from different chromosomes. The Luminal A network also has a giant component but the

**Abbreviations:** CNA, copy number alteration; GCN, gene co-expression network; GTRD, gene transcription regulation database; LFC, Log2 fold change.



layout suggests that genes from the same chromosome are preferentially linked.

To evaluate the previous observation, communities were detected in both networks using four algorithms for weighted networks implemented in the *igraph* package: Fast Greedy, Infomap, Leading Eigenvector, and Louvain. **Supplementary Material 1** presents results for all algorithms. Jaccard indexes were calculated among communities detected by the four algorithms. More than 95% of the total number of communities detected by Fast Greedy, Leading Eigenvector, and Louvain have a Jaccard Index equal to 1, while Infomap displays more dissimilar results. Given that Louvain presents the highest modularity values, results for this algorithm are presented in the main text. **Table 1** contains the number of communities and modularity values for the four algorithms applied to the Healthy and the Luminal A network.

Chromosomal assortativity,  $ASS_{chr}$  was calculated by taking the number of intra-chromosomal links minus the number of inter-chromosomal links divided by the total number of links in

a community. **Figure 1B** displays the distribution of the  $ASS_{chr}$  in both networks in the form of violin plots. The differences in the distributions allow us to confirm the loss of *trans*- interactions in the Luminal A GCN.

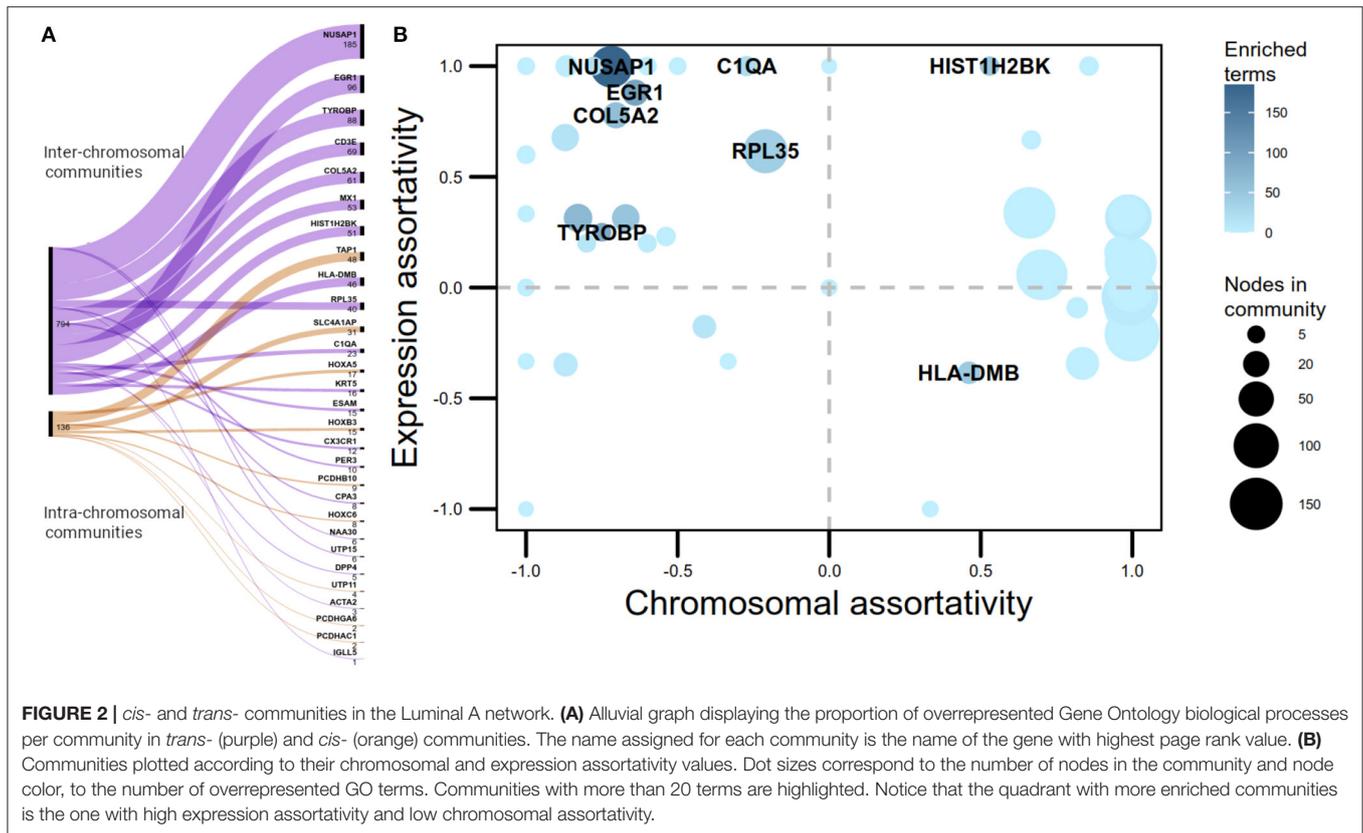
## 2.2. Specific *trans*- Communities in the Luminal A GCN Are Highly Associated With Biological Processes

To identify the functional role of the highly co-expressed groups of genes identified by network communities, an overrepresentation analysis was performed, using the biological process category in Gene Ontology (GO). Results for all algorithms are presented in **Table 1**. *-cis* communities are the ones having  $ASS_{chr}$  equals to 1.

Half of the *-trans* communities with more than five nodes extracted by the Louvain algorithm in the Luminal A GCN were associated with biological processes. However, only 12% of the *-cis* communities were enriched. Despite having a larger

**TABLE 1** | Features of *cis*- and *trans*- chromosomal communities in the Luminal A and the Healthy gene co-expression network.

Algorithm	Healthy							Luminal A								
	Modularity		Communities		Size $\geq 5$		Enriched communities		Modularity		Communities		Size $\geq 5$		Enriched communities	
	<i>cis</i> -	<i>trans</i> -	<i>cis</i> -	<i>trans</i> -	<i>cis</i> -	<i>trans</i> -	<i>cis</i> -	<i>trans</i> -	<i>cis</i> -	<i>trans</i> -	<i>cis</i> -	<i>trans</i> -	<i>cis</i> -	<i>trans</i> -		
Fast Greedy	0.703	75	325	0	50	0	14	0.934	614	87	77	40	9	20		
Infomap	0.674	83	768	1	386	1	47	0.907	826	93	194	39	16	20		
Leading Eigenvector	0.696	71	283	1	32	1	18	0.892	594	84	58	37	9	20		
Louvain	0.752	71	291	0	41	0	17	0.935	614	87	77	40	9	20		



number of intra-chromosomal *cis*- communities in the Luminal A network, the majority of communities with statistically significant biological processes associated are *trans*-. **Figure 2A** presents a visual representation in the form of an alluvial plot. There, the width of each line corresponds to the number of significantly enriched processes for a given community, named by the gene with highest page rank centrality. The difference in the amount of *cis*- and *trans*- communities with associated functions, may reflect that the set of biological processes annotated in GO do not tend to exhibit a bias toward an specific chromosome contrary to what it is observed in the Luminal A GCN communities.

There is a wide variety in the biological enriched processes in the Luminal A *trans*- communities. Processes associated with regulation of transcription, telomere maintenance, and

regulation of cell division as well as gene silencing are found. **Supplementary Table 1** contains the entire set of significantly overrepresented processes for Luminal A and healthy GCNs, as well as the shared enriched terms between both networks.

On the other hand, the enriched Luminal A *cis*- communities are mainly composed of gene families located at the same regions in the genome. In this group we have the HOXA, HOXB, and HOXC genes, which are important for embryogenesis. They have been found to be expressed in normal and neoplastic breast tissue (Cantile et al., 2003), with altered patterns of expression levels in breast cancer molecular subtypes. In particular, HOXA genes in Luminal A subtype, have shown underexpression associated with the acquisition of repressive epigenetic marks, such as hypermethylation (Novak et al., 2006; Kamalakaran et al., 2011; Hur et al., 2014).

Protocadherins (PCDHA, PCDHB, and PCDHG genes) were also identified as three distinct *cis*-communities in the Luminal A network. Protocadherin genes were previously identified as the most densely connected component (almost a clique) in a breast cancer network (Espinal-Enríquez et al., 2017). There, it was also shown that all protocadherins resulted underexpressed. The observed underexpression of this cluster coincides with a reported hypermethylation of protocadherins in breast cancer (Novak et al., 2008).

In the Healthy network 41% of the *trans*-communities were associated with biological processes, and no *cis*-communities were enriched due to the fact that *cis*-communities identified in this network have <5 genes (the threshold set for the overrepresentation analysis, see Methods). The set of terms includes mostly metabolism-associated process, cell division, and mitochondrial functions.

The Healthy and the Luminal A GCN share 24 communities of only two nodes. Additionally, there is one community named HLA-DRB1 in the Healthy GCN, and HLA-DMB in the Luminal A GCN, with a Jaccard Index of 0.916. This community is associated with activation of the immune response, and it is composed by MHC class II HLA genes located on chromosome 6 region p21.32, plus CIITA (Class II Major Histocompatibility Complex Transactivator), on Chromosome 16, and CD74, located on chromosome 5, only in the Luminal A community.

One pair of communities named CPA3 in both networks share the set of associated processes, but displays a Jaccard index of 0.705 regarding their gene sets. Processes include peptide hormone processing and regulation of systemic arterial blood pressure. Members of this community, such as TPSAB1, CMA1, CTSG, CPA3, HDC, and MS4A2, are commonly found in Mast Cells expression, part of the immune response and usually recruited to breast tumors (Aponte-López et al., 2020). The presence of these immune-system associated communities as high co-expression sets in both networks might be an instance of multiple cell types present in the sample.

### 2.3. *trans*-Communities in the Luminal A Network Present Different Patterns of Differential Expression

Once we observed that biological processes were significantly associated with *trans*-communities, a differential expression analysis was performed to assess the influence of altered gene expression in *trans*-communities and their processes. **Supplementary Figure 1** presents the differential expression representation in the GCN and **Supplementary Table 2** contains the log<sub>2</sub> fold change (LFC) values for each gene in the network.

The number of links joining genes with the same sign of LFC, minus the number of links between genes with different sign of LFC, over the total number of links, was computed per community as a measure of differential gene expression assortativity ( $ASS_{dge}$ ). **Figure 2B** plots  $ASS_{dge}$  and  $ASS_{chr}$  for *trans*-communities, as well as the number of associated GO terms. Highly enriched communities (>20 GO terms) are highlighted. The majority of these communities are placed in the first quadrant of the plot, meaning that their genes tend

to have similar differential expression but they are placed in different chromosomes. Moreover, those communities are not in the top-10 regarding size, hence functional association in *-trans* communities appears to be influenced by high  $ASS_{dge}$  and low  $ASS_{chr}$  values.

The community with the highest number of enriched GO terms is the NUSAP1 community which also contains highly overexpressed genes only (**Figure 3A**). Its enriched terms are associated with nuclear division, DNA replication, chromatid segregation, and cell cycle checkpoints, i.e., cell division processes. This community shares a Jaccard index of 0.5 regarding gene members and 0.718 regarding GO associated terms with the MKI67 community in the Healthy network.

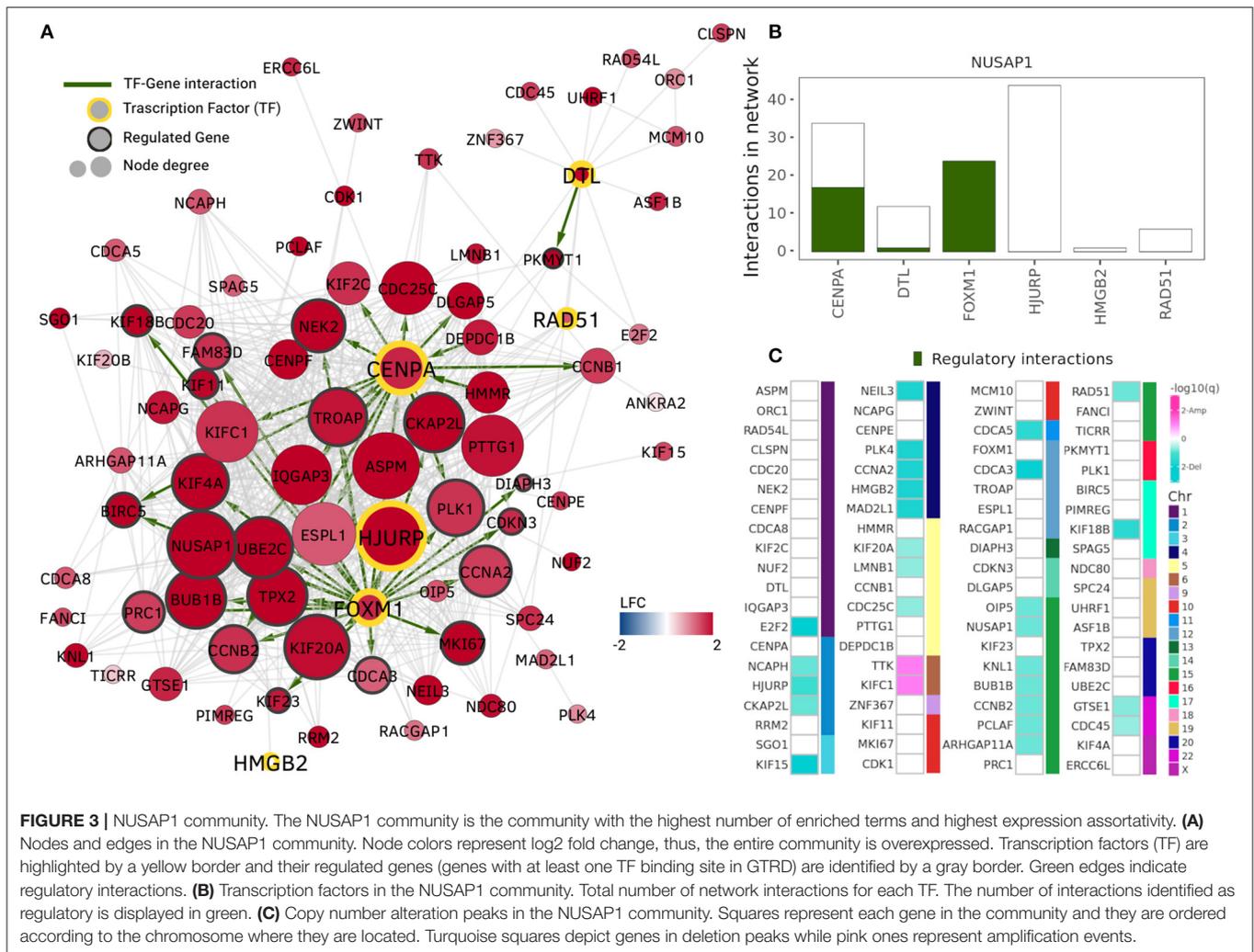
NUSAP1 has already been identified as a hub gene in a network of ER positive breast cancer tumor tissues of patients treated with tamoxifen, and derived from a similar methodology but using micro-array data (Liu et al., 2015). In that study, five hub genes with high expression levels strongly associated with poor survival were identified, and four of them: CDK1, DLGAP5, NUSAP1 and RRM2, belong to this particular community.

High expression of several genes in this community, including NUSAP1, was also observed in patients with Luminal A breast cancer and obesity (Nuncia-Cantarero et al., 2018). Nuncia-Cantarero et al. reported 39 genes related with a poor outcome group for patients with both conditions and 26 are found in this community, including FOXM1 (Forkhead box protein M1), a transcription factor that has been identified as a potential therapeutic target for breast cancer (Lu et al., 2018), highly associated with luminal tumors and ER expression (Millour et al., 2010; Carr et al., 2012).

**Table 2** shows the 39 genes reported in Nuncia-Cantarero et al. (2018). The coincident genes found in our network community are bold and their corresponding log<sub>2</sub> fold change values are displayed. Interestingly, none of the genes presented in Nuncia-Cantarero et al. (2018) are in the Luminal A GCN but those found in the NUSAP1 community.

From the highly enriched communities, RPL35 is the one with more genes. The majority of them are ribosomal proteins; therefore, among the enriched GO terms we find ribosome biogenesis, large and small ribosomal subunit assembly, as well as regulation of ubiquitin-protein transferase activity. Riboproteins in this community are mostly underexpressed (**Supplementary Figure 2**). Low levels of expression have been reported in breast cancer for RPL5 and RPL11, associated with a mechanism of apoptosis inhibition through P53 degradation (Tong et al., 2020), and induction of proliferation in MCF7 cells, a Luminal A-derived cell type (Fancello et al., 2017). It has been shown that riboproteins have high co-expression values in other gene co-expression networks (Prieto et al., 2008; Wang et al., 2020a,b). The finding of highly co-expressed cluster of riboproteins reported here, reinforces the fact that these GCNs are coherent and represent with some accuracy the actual co-expression landscape in Luminal A breast cancer.

To our knowledge, coordinated underexpression of ribosomal genes in a breast cancer subtype has not previously been described. On the contrary, an increased ribosomal content has been recently found to contribute to proliferative and



metastatic potential in breast cancer circulating tumor cells (Ebright et al., 2020). This discrepancy may be due to the fact that the overexpression of RPL transcripts, such as RPL15 observed in Ebright et al. (2020), was reported for circulating tumor cells. These tumor cells present additional alterations in their transcriptional profile, and they have acquired a highly proliferative capacity. Hence, the underexpression of ribosomal genes in the Luminal A network may be an indicative that the tumors are not as invasive as other subtypes. It is worth noticing again that Luminal A breast cancer subtype is the less aggressive, the one with the best prognostic and also the best in terms of response to therapy.

## 2.4. Effects of Transcription Factors and CNAs in *trans*- Communities

The general overexpression trend observed in the NUSAP1 community, and underexpression in the RPL35 module, suggested a contribution of altered mechanisms of transcriptional regulation promoting the formation of high co-expression clusters. To evaluate this, we analyzed the contribution of regulatory interactions from transcription

factors (TFs) and the presence of deletion and amplification peaks in the Luminal A network communities.

TFs in the ten highlighted communities from **Figure 2B** were identified using data from the Gene Transcription Regulation Database (GTRD) (Yevshin et al., 2018). Five communities included at least one gene reported as TF in GTRD. The total number of interactions for these genes in the NUSAP1 community is presented in **Figure 3B**, where the number of genes having at least one binding site in the promoter region (1,000 bp upstream, 100 bp downstream from starting point) is shown in green. It can be observed that FOXM1 transcription factor has its entire set of adjacent links marked as regulatory interactions.

As stated in the previous section, the NUSAP1 community contains interactions that have been reported in luminal associated breast cancer phenotypes. Particularly, the FOXM1 transcriptional network was identified as the largest regulon by GPU-ARACNE, the accelerated parallel implementation of ARACNE, the algorithm used here to infer the gene co-expression networks (He et al., 2017). He et al. identified 121 FOXM1 interactions with 14 experimentally validated targets.

**TABLE 2** | Previously reported genes in the NUSAP1 community.

Gene	Gene name	LFC
<b>NEK2</b>	<b>Serine/threonine-protein kinase Nek2</b>	3.564
<b>KIF4A</b>	<b>Kinesin Family Member 4</b>	3.098
<b>ASPM</b>	<b>Abnormal spindle-like microcephaly-associated protein</b>	2.567
<b>CENPF</b>	<b>Centromere protein F</b>	2.567
<b>TPX2</b>	<b>Protein TPX2</b>	2.567
<b>KIF18B</b>	<b>Kinesin Family Member 18B</b>	2.396
<b>CDC25C</b>	<b>M-phase inducer phosphatase</b>	2.316
<b>DLGAP5</b>	<b>Disks large-associated protein 5</b>	2.297
<b>NUSAP1</b>	<b>Nucleolar and spindle-associated protein</b>	2.223
<b>MKI67</b>	<b>Proliferation marker protein Ki-67</b>	2.191
<b>UBE2C</b>	<b>Ubiquitin-conjugating enzyme E2</b>	2.173
<b>HMMR</b>	<b>Hyaluronan mediated motility receptor</b>	2.162
<b>BUB1B</b>	<b>Mitotic checkpoint serine/threonine-protein kinase</b>	2.157
<b>BIRC5</b>	<b>Baculoviral IAP repeat-containing protein</b>	2.057
<b>CDK1</b>	<b>Cyclin-dependent kinase</b>	2.012
<b>KIF11</b>	<b>Kinesin Family Member 11</b>	1.963
<b>RRM2</b>	<b>Ribonucleoside-diphosphate reductase subunit M2</b>	1.961
<b>KIF20A</b>	<b>Kinesin Family Member 20</b>	1.898
<b>ISG15</b>	<b>Ubiquitin-like protein ISG15</b>	1.789
<b>GTSE1</b>	<b>G2 and S phase-expressed protein</b>	1.714
<b>FOXM1</b>	<b>Forkhead box protein M1</b>	1.699
<b>CCNB2</b>	<b>G2/mitotic-specific cyclin-B2</b>	1.621
<b>CCNB1</b>	<b>G2/mitotic-specific cyclin-B</b>	1.523
<b>PRC1</b>	<b>Protein regulator of cytokinesis</b>	1.504
<b>KIF15</b>	<b>Kinesin Family Member 15</b>	1.425
<b>ZWINT</b>	<b>ZW10 interactor</b>	1.416
<b>OIP5</b>	<b>Protein Mis18-beta</b>	1.299
BUB1	Mitotic checkpoint serine/threonine-protein kinase BUB1	
CEP55	Centrosomal protein of 55 kDa	
EZH2	Histone-lysine N-methyltransferase EZH2	
GDP-15	Growth/differentiation factor 15	
KIAA0101	PCNA-associated factor	
MELK	Maternal embryonic leucine zipper kinase	
MMP1	Matrix Metalloproteinase	
MYBL1	MYB Proto-Oncogene Like	
PBK	PDZ Binding Kinase	
RIPPLY3	Protein ripply3	
TOP2A	DNA topoisomerase 2-alpha	
TYMS	Thymidylate synthase	

39 Genes reported in Nuncia-Cantarero et al. (2018), related with poor outcome group for patients with obesity and Luminal A breast cancer. Highlighted genes are present in the NUSAP1 community. Their corresponding log2 fold change value is also displayed. Notice that all concordant genes are overexpressed.

In the NUSAP1 community, FOXM1 has 24 co-expression interactions with other genes in the module. All of these interacting genes contain a FOXM1 binding site in their promoter region according to the data gather by GTRD. From these 24 regulated genes, eight intersect with the experimentally validated targets reported in He et al. (2017).

Centromere protein A or CENPA, is another important transcription factor with overexpression in the NUSAP1

community. It regulates centromere integrity and chromosome segregation. This TF was identified in a mRNA signature correlated with lower survival ratio in Luminal A breast cancer (Xiao et al., 2018). One of its interacting proteins, HJURP, required for CENPA centromeric localization, is also a member of this community. HJURP mRNA expression level has been significantly associated with estrogen and progesterone receptor, and reported as clinically relevant for Luminal A breast cancer patients (Hu et al., 2010; Montes de Oca et al., 2015). Although HJURP is the transcription factor with more adjacent links in the NUSAP1 community, none of them was identified as a regulatory interaction; instead, HJURP was identified as regulated by FOXM1.

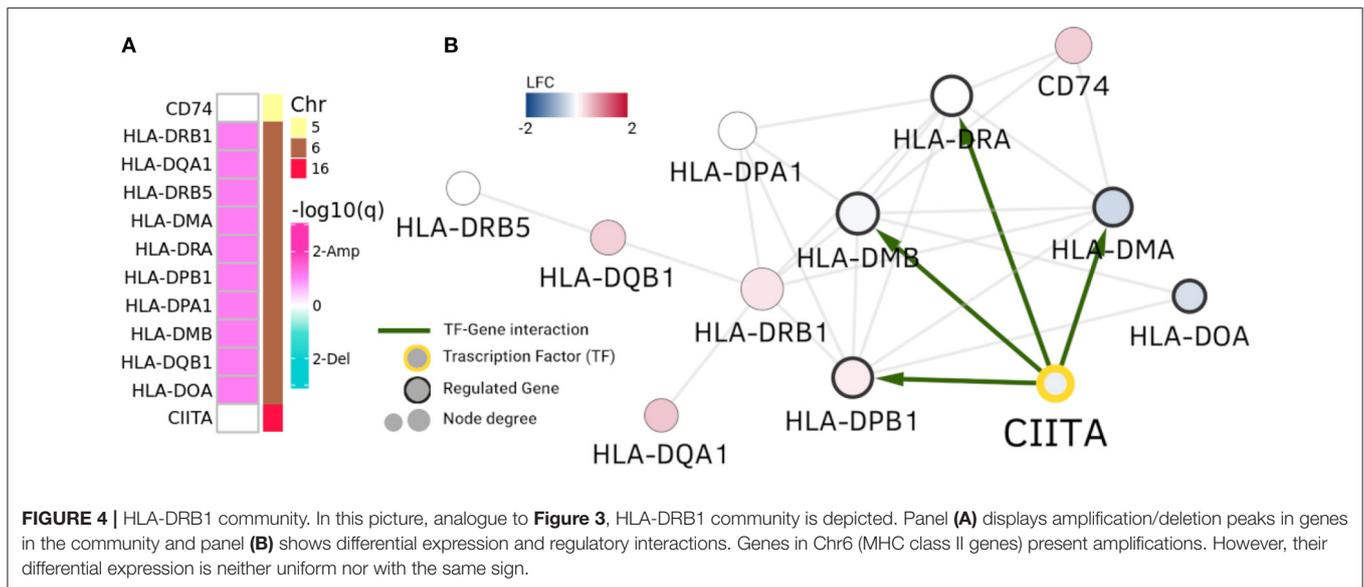
The remaining overexpressed TFs in the NUSAP1 community have also been found to play a role in the luminal breast cancer phenotype. Increased mRNA expression of RAD51, a gene in the double-strand breaks repair pathway, is associated with higher risk of tumor relapse and distant metastases in estrogen receptor positive breast cancer tumors (Barbano et al., 2011; Nieto-Jiménez et al., 2017). Overexpression of DTL and HMGB2 has also been associated with tumor progression in breast cancer (Perez-Peña et al., 2017; Fu et al., 2018), and resistance to endocrine therapies (Redmond et al., 2015). These results suggest a strong contribution of TFs, particularly from FOXM1 and CENPA, and their interactions found in the NUSAP1 community, to the process of tumorigenesis and progression in Luminal A breast cancer.

Gene copy number alteration (CNA) is a common trait of genomic instability in cancer and their presence has therapeutic relevance in breast cancer, specially for the Her2 enriched subtype (Andre et al., 2009; Inaki et al., 2014). Different levels of correlation have been identified between DNA amplification and deletion events, mRNA, and protein expression values in breast cancer, (Myhre et al., 2013), showing that it is not an homogeneous mechanism of altered expression. However, given the possible effect and importance for the breast cancer phenotype, amplification and deletion peaks may play a role in the formation of high co-expression clusters in the Luminal A network. For instance, in the case of breast cancer, correlation between CNVs and gene expression could reach until 25% (Lachmann, 2016).

Those gene expression alterations may influence importantly in the co-expression landscape. In Lachmann (2016), it was reported that CNVs may impact importantly the co-expression program, in particular for transcription factor targets.

To evaluate the role of CNVs in the Luminal A GCN, we obtained amplification and deletion peaks using the GISTIC2 algorithm (Mermel et al., 2011). **Figure 3C** presents the results for the NUSAP1 community. Turquoise squares represent genes in which a deletion has been observed, meanwhile amplifications are depicted in pink squares. Since the NUSAP1 community is *trans-*, the chromosome in which those genes are located is also depicted.

As observed, the majority of genes with copy number alterations correspond to deletions. Only two genes, TTK and KIF1C1 (Chr6) present amplifications. However, 52 out of 80 genes do not present changes in copy number. This result shows



that, at least for the NUSAP1 community, which is the one with the most differentially expressed genes, CNAs do not significantly influence neither expression nor co-expression patterns.

However, in the case of HLA-DRB1 community in **Figure 4**, we observe the opposite phenomenon: genes are not differentially expressed, but the ones that are placed in Chr6 belong to a clearly amplified region. This cluster is composed of MHC class II HLA genes. Interestingly, CIITA gene is a TF that regulates some of these human leukocyte antigen genes. As it can be observed in **Figure 4**, four of these genes have a CIITA binding site in their promoter region.

In this case CNAs and the CIITA regulation appear to exert a concomitant action with the observed copy number alterations to generate the community of MHC class II genes, independently of their differential expression. It is worth mentioning that CIITA (Class II Major Histocompatibility Complex Transactivator) is located at Chromosome 16, but clearly regulates the transcriptional and functional characteristics of HLA genes. The same representation for the RPL35 community is shown in **Supplementary Figure 2**. It is worth to stress that the HLA-DRB1 community in Luminal A GCN is almost identical to a community of the healthy GCN (Jaccard index = 0.916).

## 2.5. *cis*- Communities Are Enriched With Deletion Peaks

The presence of deletion and amplification peaks, and their effect in gene altered expression was also evaluated for *cis*- communities. **Figure 5** presents the results of an overrepresentation analysis where GISTIC2 peaks were analyzed. As it can be observed, communities are mostly enriched with deletion peaks, and their effect in the average log<sub>2</sub> fold change in *cis*- communities varies. **Supplementary Figure 3** presents the entire set of alterations in these communities.

The pattern of amplification in the q arm of chromosome 1 and deletion in chromosome 16q, previously reported in a

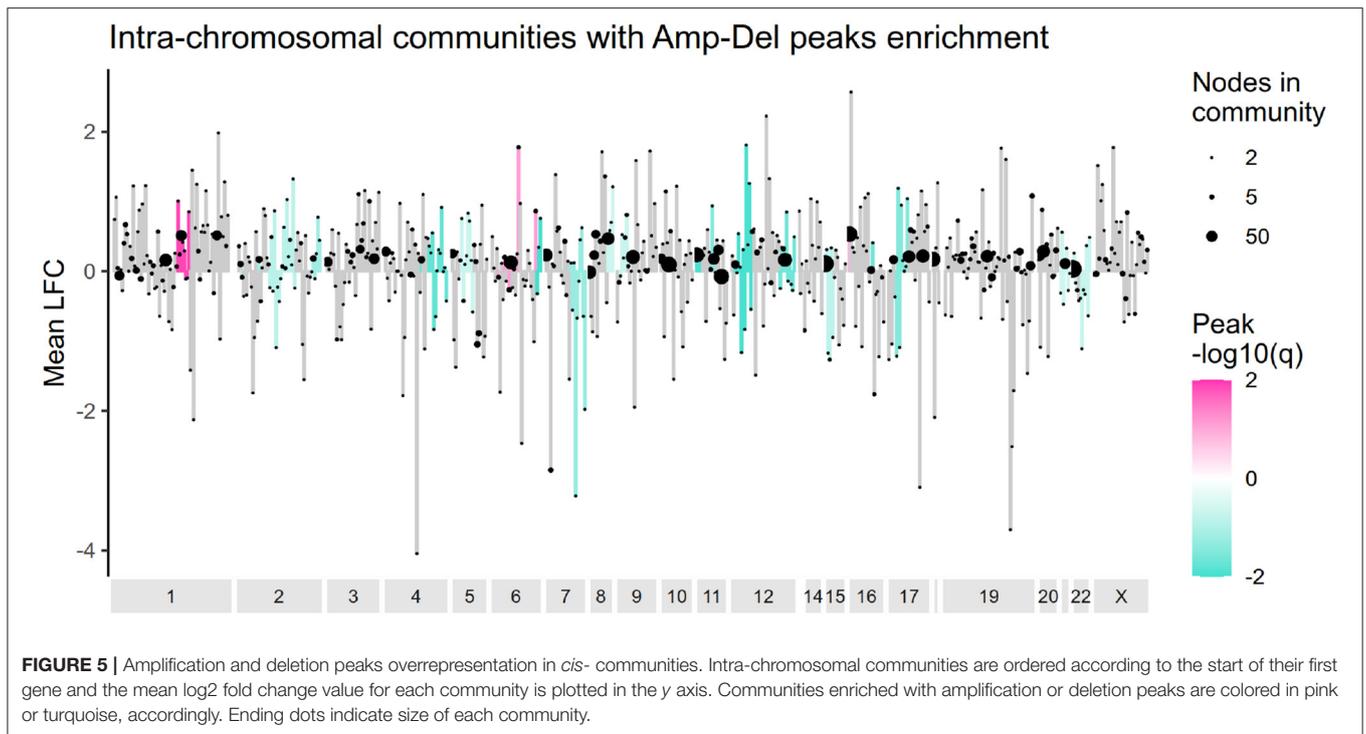
subset of Luminal A tumors (Ciriello et al., 2013) is also observed here. However, no other alteration matched that particular study. Luminal A tumors tend to have the lowest frequency of CNAs among breast cancer subtypes (Gatza et al., 2014), and as evaluated by our methodology, amplification and deletion peaks do not *a priori* determine the formation of *cis*- communities.

It is important to mention that copy number alterations are a key element affecting the gene expression of large sections of the genome (Freeman et al., 2006; Redon et al., 2006; McCarroll and Altshuler, 2007), specially in cancer (Shlien and Malkin, 2009; Lachmann, 2016; Shao et al., 2019). A large part of a chromosome being altered by a gain or loss of copy number, will trigger an equally abrupt change in several genes along that portion of the genome.

## 2.6. *cis*- Communities Are Not Bound by CTCF Binding Sites

The three-dimensional structure of DNA is another regulator of gene expression in eukaryotic cells. Regions with active transcription are characterized by open chromatin, whereas closed chromatin indicates regions of inactive or repressed transcription (Achinger-Kawecka et al., 2016; Corces and Corces, 2016). Furthermore, the regulatory effect of regions, such as enhancers and promoters, usually requires the formation of long distance chromatin loops that bring together distant genomic loci. These loops are maintained and regulated by architectural proteins, such as CTCF and cohesin, among others (Achinger-Kawecka and Clark, 2017; Pugacheva et al., 2020). Given the fact that CTCF proteins are able to modify the chromatin landscape, they may be underlying the appearance of a large amount of *cis*- communities in breast cancer.

To evaluate the role of CTCF in the appearance of *cis*- clusters of genes in the Luminal A breast cancer gene co-expression network, we calculated the number of CTCF binding sites at the boundaries of *cis*- communities. This was done using a previously



reported dataset containing Chip-seq peaks in MCF7 cells, a Luminal A breast cancer cell line (Fiorito et al., 2016).

The number of binding sites in a window of 50k base pairs before the first gene and after the last one in a community was compared to the average number of binding sites in same size windows spanning the community region (see Methods). The distribution of these binding sites is shown in **Supplementary Figure 4**. No significant difference was found in the distribution of the number of binding sites in the boundaries and the middle sections of the communities. Actually, out of the 416 *cis*- communities with at least one CTCF binding site associated, only 197 had more binding sites at the boundaries than in middle regions.

## 2.7. Loss of Long-Distance Co-expression Does Not Depend on the Correlation Measure

We decided to construct GCNs for Luminal A and healthy phenotypes using Pearson correlation, to observe whether the phenomenon of loss of long-distance co-expression was maintained using other correlation measure. The results can be observed in the form of a heatmap in **Figure 6**. There, genes are placed according to its position in the chromosome. The color of the heatmap is proportional to the correlation value. The results show that, as observed with mutual information-inferred networks, the highest correlation values occur between genes from the same chromosome.

Additionally, it can also be appreciated that the Pearson correlation values are in general higher in the healthy matrix than

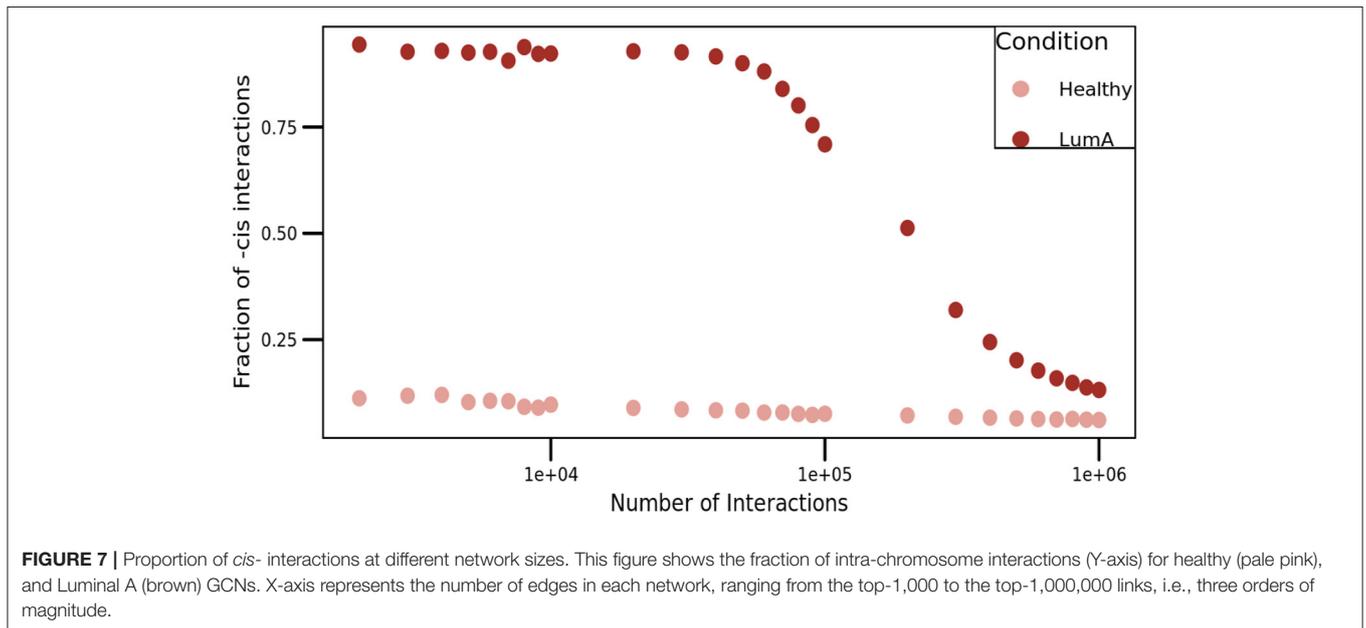
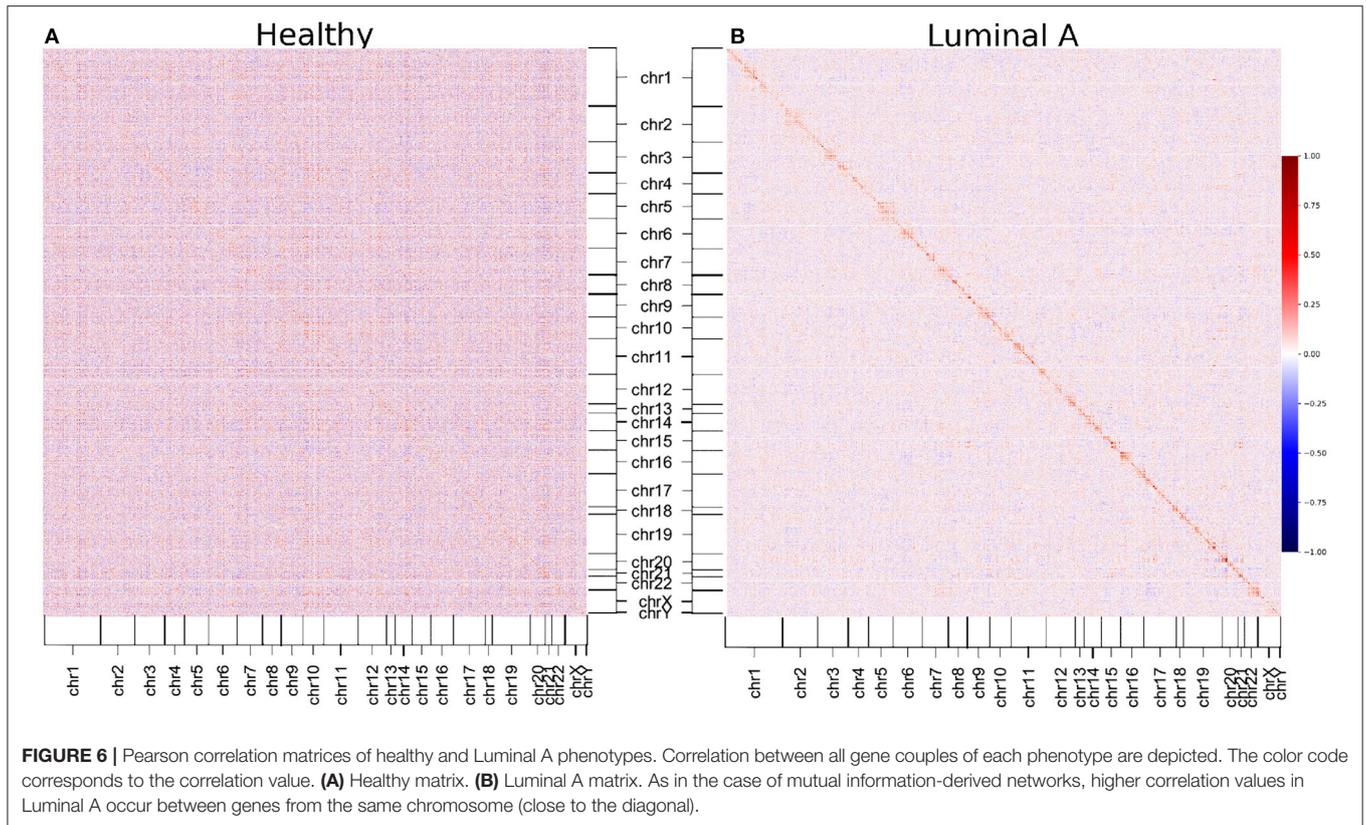
in the Luminal A breast cancer one (except for those values close to the diagonal, which represent *cis*- interactions).

## 2.8. Loss of Long-Distance Co-expression Does Not Depend on the MI Threshold Value

Setting a threshold on the weight of edges so as to discard edges with strength less than a certain value is a well-known open problem in graph theory and network science. Determination of this threshold can be made by choosing among a number of methods. For instance, if an accurate measure of the signal-to-noise ratio in the correlations of the data under consideration can be obtained, one possible way to set the threshold is by allowing all edges valued above the noise-level. In most practical applications, however, this is not feasible.

To overcome this situation, we presented a comparison of *cis/trans* proportion in both networks. For this purpose, we constructed networks with different threshold values, ranging from the top-1,000 to the top-1,000,000 higher edges (**Figure 7**). As it can be appreciated in the figure, the proportion of *cis*-interactions is always higher in Luminal A network than in the healthy GCN.

Additionally, to assess the influence of the MI threshold value in the phenomenon of loss of long-distance co-expression in Luminal A breast cancer, we observed the distribution of MI values in both networks. We constructed (a) the histograms of all interactions (20,217) in both networks, (b) the histograms for only *cis*- interactions, and (c) the histogram for *trans*- edges in both phenotypes (**Figure 8**). There, it can be observed that

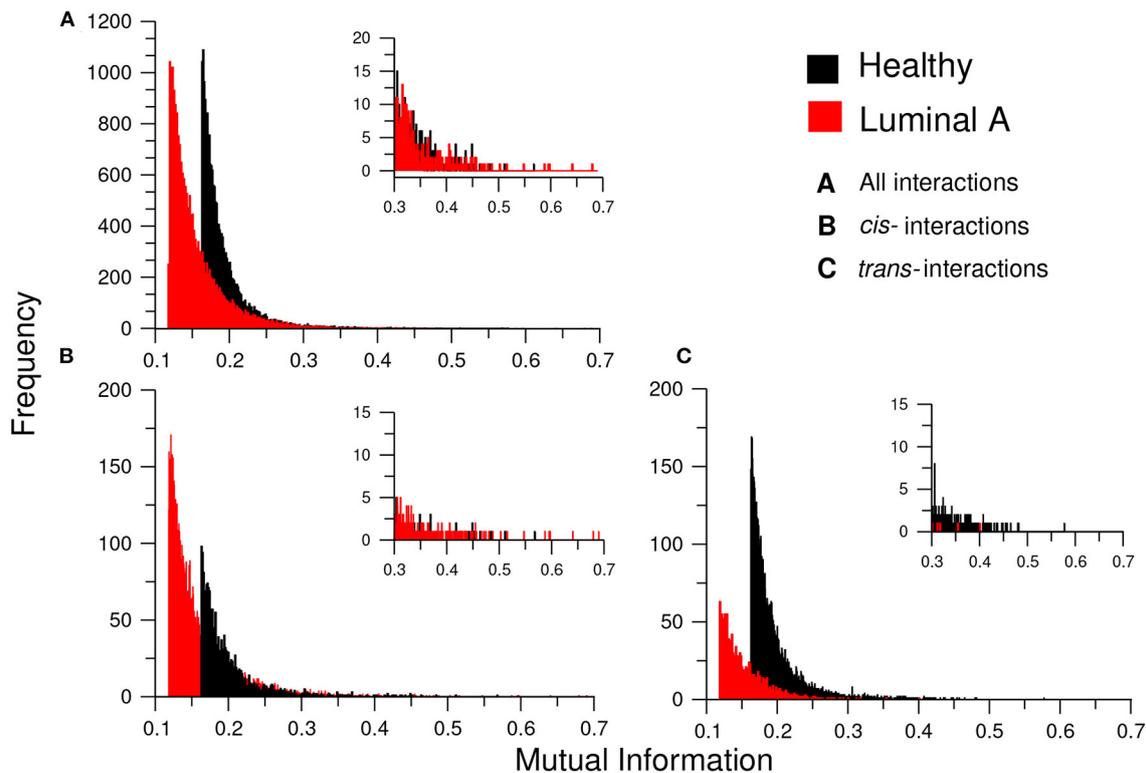


independently of the threshold, healthy interactions have higher MI values.

The above mentioned result coincides with the one presented in the matrices of **Figure 6**. Correlation values (independent on the correlation measure), are in general higher in the

healthy phenotype than in cancer, but for a subset intra-chromosome interactions.

Complementarily, in **Figure 8** we inserted a zoom of those histograms in the higher MI value region (0.3–0.7). There, it is shown that for *cis*-interactions, the Luminal A network has



**FIGURE 8** | Distribution of MI values in the GCNs. This plot shows the histograms for the MI values of the healthy (black) and Luminal A (red) GCNs. **(A)** The total of MI values. **(B)** Only *cis*- edges. **(C)** Only *trans*- interactions. Each histogram also contains an inset with a zoom of the highest interactions for each condition. Notice the absence of *trans*- interactions in the Luminal A case in the inserts of **(B,C)**; this reflects the loss of *trans*- co-expression in the cancer GCN.

more and higher interactions in the highest values; conversely, for the *trans*- interactions, the higher and more abundant links are observed in the healthy phenotype.

We have shown previously that the threshold value is not determinant to observe the loss of long-distance co-expression in other clear cell renal carcinoma (Zamora-Fuentes et al., 2020), as well as in lung cancer (Andonegui-Elguera et al., 2021). We have demonstrated for these cancer GCNs that the particular value of the threshold, affects the size and sparsity of the networks as expected. However, the proportion of inter- and intra-chromosomal links remains largely unchanged.

## 2.9. Implications of Network Topology in the Context of Luminal a Breast Cancer

We have shown that in Luminal A breast cancer, the already mentioned *loss of trans*- co-expression is not as strong as in other breast cancer subtype GCNs, but the effect is perceived. Actually, several *trans*- interactions appear in the top co-expressed pairs. Luminal A GNC topology allows us to:

- identify functional communities (mostly *trans*-)
- differentiate enriched functions between healthy and cancer GCNs
- observe mechanisms that may influence the appearance of this loss of long distance co-expression

- observe specific differential expression patterns depending on the community

The identification of significant biological processes, associated with particular sets of highly co-expressed genes is one of the most relevant improvements of using network topology to analyze the functional implications of RNA-Seq-based genome-wide multi-sample sets for a given phenotype. The use of network communities improves the specificity of the enrichment analysis over using the whole genome or using differentially expressed genes.

The number of enriched processes in *cis*- communities is significantly lower than the ones associated with *trans*- communities, given the total number of communities for each type. However, the functions that are significant for *cis*- communities, are also relevant for cell maintenance. For instance, HOXA community, whose genes are relevant for organism development. These genes are found together in chromosome 7p15.2, and they are all underexpressed. Analogously, the protocadherin cluster is found to be related to cell adhesion, which is one of the non-shared processes between Luminal A GCN and the healthy GCN (**Supplementary Material 1**).

From the alluvial diagram of **Figure 2** it can be observed that out of the 11 enriched *cis*- communities, 6 correspond to HOX and protocadherin clusters. This could be an indicative of the

importance of the conjugated action that these set of genes may have for the phenotype. Additionally, these clusters appear with the same differential expression trend.

### 3. CONCLUDING REMARKS

Based on the previous analysis, we may conclude that for the establishment of the regulatory program observed in the Luminal A subtype gene co-expression network, compared with the healthy GCN, several DNA modifications and regulatory elements must participate. DNA modifications (copy number alterations, transcription factor regulation, CTCF binding sites) should exert, to a certain extent, influence over the gene co-expression interactions. Additionally, differential gene expression is a relevant element to take into account, specially for *trans*-communities. We can establish that, for the manifestation of the *loss of trans-co-expression* in cancer it is not only necessary to observe separately differential gene expression, transcription factor regulation, CNAs, or CTCF binding sites, but to take them all into account.

Other regulatory elements should also participate in modifying the co-expression patterns between a healthy and a cancer co-expression network: micro-RNA regulation (Drago-García et al., 2017; de Anda-Jáuregui et al., 2018), topologically associated domains and their boundaries (Rafique et al., 2015; Achinger-Kawecka et al., 2020; Khoury et al., 2020), long non-coding RNAs (Hung et al., 2011; Zhang et al., 2019), the methylation profiles (Paz et al., 2003; Hernández-Lemus et al., 2019), among others, might delineate these imbalance between *cis*- and *trans*-genetic relationships.

More investigation regarding the aforementioned elements is also important in order to have an integral picture of the regulatory landscape in the cancer genome, and provide hypotheses that could explain the phenomenon of loss of long distance genetic interactions in cancer.

It is likely plausible that the *loss of trans-co-expression* observed in breast cancer (and breast cancer molecular subtypes) responds to a physical/mechanical principle in which the transcriptional machinery is somehow altered. Recently, we have observed the loss of long distance co-expression in clear cell renal carcinoma (Zamora-Fuentes et al., 2020), and in lung adenocarcinoma, as well as in squamous cell lung cancer (Andonegui-Elguera et al., 2021).

The ubiquity of this disruption of the *normal* transcriptional landscape led us to hypothesize that the physical principle behind this global alteration is the same in all of these cancer tissues. The consistency and relevance of this loss could be considered as a possible emergent hallmark of cancer. Further investigation toward this particular issue must be achieved beforehand, however, further investigation is required.

## 4. METHODS

### 4.1. Databases

Gene expression values for Luminal A and Healthy samples were retrieved from our previous publication (García-Cortés et al., 2020), with RNA-seq data obtained from The Cancer Genome

Atlas (TCGA) breast invasive carcinoma dataset (Tomczak et al., 2015), downloaded from the Genomic Data Commons (GDC) Data Portal. The GDC Data portal case identifiers for Luminal A were used to download “Masked Copy Number Segment Files” for the GISTIC2 pipeline. The Chip-seq data was downloaded from the Gene Expression Omnibus dataset GSE85106 (Fiorito et al., 2016), and only the control sample for CTCF was used. The Homo sapiens genes promoter dataset from the Gene Transcription Regulation Database (GTRD) (Yevshin et al., 2018) was used to identify transcription factors and their regulatory interactions.

### 4.2. Data Processing

As detailed in García-Cortés et al. (2020), 113 samples for Healthy tissue and 1,102 cancer samples were acquired and pre-processed to  $\log_2$  normalized gene expression values. After applying the PAM50 algorithm using the Permutation-Based Confidence for Molecular Classification (Fresno et al., 2017) as implemented in the `pbcmc` R package (Fresno et al., 2016), and multidimensional noise reduction using ARSyN R implementation (Nueda et al., 2012), 217 samples for Luminal A breast cancer were identified.

The “Masked Copy Number Segment Files” were downloaded from GDC and integrated into one segmentation file to run `gistic2` (Mermel et al., 2011). The parameters suggested in the Copy Number Variation Analysis Pipeline from GDC and the GDC reference sequence, and markers file were used. The identified amplification and deletion regions in the lesions output file with 0.99 confidence were re-mapped to keep genes spanned entirely by peaks.

### 4.3. Network Construction

The ARACNE (Margolin et al., 2006) algorithm was used to calculate mutual information (MI) to quantify statistical dependence between pairs of genes. The method associates a significance value (*p*-value) to each MI value based on permutation analysis, as a function of the sample size. Only the highest interactions in terms of their statistical significance ( $P \leq 1e^{-8}$ ) were kept for further analysis. The total number of interactions in the Luminal A and the Healthy network were reduced to 20,127, the number of significant interactions in the Healthy network.

### 4.4. Community Detection and Assortativity Calculation

Four community detection algorithms were evaluated: Fast Greedy (Clauset et al., 2004), Infomap (Rosvall and Bergstrom, 2008), Leading Eigenvector (Newman, 2006), and Louvain (Blondel et al., 2008; Rahimnejad et al., 2019). MI values were used as link weights. Their implementation in the `igraph` (Csardi and Nepusz, 2006) R package was used. Algorithm results were compared using the Jaccard index, a coefficient that measures similarity between two finite sets, defined as the size of their intersection divided by the size of their union. Genes in a community constitute a set and all communities identified by one algorithm were compared against communities identified by another one. The same approach was used to

**TABLE 3** | CTCF binding sites location classification.

	Promoter	Gene body	Intergenic region
Dataset	868	8,047	11,438
In Luminal A network	177	1,343	887

compare the set of GO terms associated per community in the overrepresentation analysis.

$$J(C_1, C_2) = \frac{(C_1 \cap C_2)}{(C_1 \cup C_2)} \quad (1)$$

To calculate chromosomal assortativity, the chromosome location for each gene was used. For each community, the number of links joining genes in the same chromosome (*-cis* links) minus the number of links joining genes in different chromosomes (*-trans* links), was divided by the total number of links in the community. Expression assortativity was calculated in the same manner, using the log<sub>2</sub> fold change sign to classify genes into overexpressed or underexpressed as the assortativity attribute.

$$ASS_{chr} = \frac{|\{\{x, y\} \mid x, y \in C_i \text{ and } x.chr = y.chr\}| - |\{\{x, y\} \mid x, y \in C_i \text{ and } x.chr \neq y.chr\}|}{|\{\{x, y\} \mid x, y \in C_i\}|}$$

$C_i$  = community  $i$  in network.

#### 4.5. Overrepresentation Analysis

The `enrichGO` function from the `clusterProfiler` (Yu et al., 2012) R package was used to identify over-represented or enriched terms in the Biological Process category in Gene Ontology (GO). Enrichment was performed for communities with five or more genes and GO terms with a minimum size of ten were retained. Genes in the original expression matrix defined the universe set. Terms with adjusted  $p$ -value below 0.005 using the Benjamini and Hochberg method for multiple testing were kept. The overrepresentation analysis for amplification and deletion peaks was conducted using the generic function `enricher` from the same package. The same universe set was used and no size threshold for communities or peaks was defined. An adjusted  $p$ -value of 0.05 was set as cutoff.

#### 4.6. Differential Expression Analysis

Differential expression analysis was performed as described in (Espinal-Enriquez et al., 2017). The `limma` package (Ritchie et al., 2015) in R was used to determine overexpressed or underexpressed genes, by adjusting a gene based linear model. An absolute difference of log<sub>2</sub> fold change  $\geq 0.5$  and a  $p$ -value  $< 0.05$  was set as threshold.

#### 4.7. Transcription Factors Identification

The entire set of gene promoters in the smallest region available,  $[-100, +10]$  base pairs from starting site was downloaded from the Gene Transcription Regulation Database (GTRD) (Yevshin et al., 2018). For the selected communities, gene members that matched transcription factors (TF) in GTRD were extracted and

their neighboring genes were compared to the set of annotated genes that had at least one binding site from that TF in the ChIP-seq data.

#### 4.8. CTCFs

We took the CTCFs in genes and promoters in the *cis*- Luminal A network communities that were not in other genes or promoters. For the Inter-regional CTCFs, we took the ones that were in a region  $< 50k$  bps from the extreme of the promoter and the extreme of the gene.

Once filtered, the binding sites were classified according to their location. CTCFs in gene bodies, promoters (+1,000, -500 bps) and intergenic region were identified. **Table 3** displays the classified binding sites for the complete dataset, as well as the binding sites present in genes comprising the Luminal A *trans*-communities. For the intergenic region, only CTCF binding sites in a window of 50k base pairs upstream the first gene and downstream the last one in *cis*- communities were kept.

#### DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: Genomic Data Commons Data Portal <https://portal.gdc.cancer.gov/>, Gene Expression Omnibus dataset GSE85106 Gene Transcription Regulation Database, <http://gtrd.biouml.org/>, Relevant data used, and the scripts to generate the results and figures can be found in the following repository: <https://github.com/diannaeluma>.

#### AUTHOR CONTRIBUTIONS

DG-C performed the computational analyses, developed and implemented the programming code, performed the pre-processing and low-level data analysis, made the figures, drafted the manuscript. EH-L contributed to the theoretical analysis, co-supervised the project, contributed to the writing of the manuscript. JE-E conceived and designed the project, co-supervised the project, discussed the results, drafted the manuscript. All authors read and approved the final version of the manuscript.

#### FUNDING

This work was supported by CONACYT (558985 student grant to DG-C, 285544/2016, and 2115/2018 to JE-E), 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 (232647/2014 CONACYT). JE-E was recipient of the 2018 Miguel Alemán Fellowship in Health Sciences. EH-L was a recipient of the 2016 Marcos Moshinsky Fellowship in the Physical Sciences.

#### ACKNOWLEDGMENTS

DG-C was a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM). This work is part of her Ph.D. Thesis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.629475/full#supplementary-material>

**Supplementary Material 1** | Results of community detection algorithms evaluation.

**Supplementary Table 1** | List of Gene Ontology biological processes overrepresented in the Healthy and the Luminal A gene co-expression networks (GCN), as well as the shared enriched terms between both networks.

**Supplementary Table 2** | List of genes in the Luminal A GCN and the Healthy GCN with their chromosomal location, associated log<sub>2</sub> fold change (LFC) value, and corresponding community for each algorithm.

**Supplementary Figure 1** | Differential expression in the Luminal A GCN. The NUSAP1 community is highlighted.

**Supplementary Figure 2** | RPL35 community. Left panel presents amplification and deletion peaks identified by GISTIC2, through pink and turquoise squares. Genes are ordered according to their corresponding chromosome. Right panel displays differential expression and regulatory interactions in genes in the community.

**Supplementary Figure 3** | Amplification and deletion peaks in *cis*-communities. Entire set of copy number alterations identified in intra-chromosomal communities. Genes are displayed according to their starting site.

**Supplementary Figure 4** | CTCF binding sites distribution over *cis*-communities. Binding sites at a distance of no more than 50,000 base pairs from a gene in the community are displayed.

## REFERENCES

- Achinger-Kawecka, J., and Clark, S. J. (2017). Disruption of the 3D cancer genome blueprint. *Epigenomics* 9, 47–55. doi: 10.2217/epi-2016-0111
- Achinger-Kawecka, J., Taberlay, P. C., and Clark, S. J. (2016). Alterations in three-dimensional organization of the cancer genome and epigenome. *Cold Spring Harbor Symp. Quant. Biol.* 81, 41–51. doi: 10.1101/sqb.2016.81.031013
- Achinger-Kawecka, J., Valdes-Mora, F., Luu, P.-L., Giles, K. A., Caldon, C. E., Qu, W., et al. (2020). Epigenetic reprogramming at estrogen-receptor binding sites alters 3D chromatin landscape in endocrine-resistant breast cancer. *Nat. Commun.* 11, 1–17. doi: 10.1038/s41467-019-14098-x
- Alcalá-Corona, S. A., de Anda-Jáuregui, G., Espinal-Enríquez, J., and Hernández-Lemus, E. (2017). Network modularity in breast cancer molecular subtypes. *Front. Physiol.* 8:915. doi: 10.3389/fphys.2017.00915
- Alcalá-Corona, S. A., de Anda-Jáuregui, G., Espinal-Enríquez, J., Tovar, H., and Hernández-Lemus, E. (2018a). “Network modularity and hierarchical structure in breast cancer molecular subtypes,” in *International Conference on Complex Systems* (Cham: Springer), 352–358. doi: 10.1007/978-3-319-96661-8\_36
- Alcalá-Corona, S. A., Espinal-Enríquez, J., De Anda Jáuregui, G., and Hernandez-Lemus, E. (2018b). The hierarchical modular structure of HER2+ breast cancer network. *Front. Physiol.* 9:1423. doi: 10.3389/fphys.2018.01423
- 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
- Andonegui-Elguera, S. D., Zamora-Fuentes, J., Espinal-Enríquez, J., and Hernandez-Lemus, E. (2021). Loss of long-distance co-expression in lung cancer. *Front. Genet.* 12:192. doi: 10.3389/fgene.2021.625741
- Andre, F., Job, B., Dessen, P., Tordai, A., Michiels, S., Liedtke, C., et al. (2009). Molecular characterization of breast cancer with high-resolution oligonucleotide comparative genomic hybridization array. *Clin. Cancer Res.* 15, 441–51. doi: 10.1158/1078-0432.CCR-08-1791
- Aponte-López, A., Enciso, J., Muñoz-Cruz, S., and Fuentes-Pananá, E. M. (2020). An *in vitro* model of mast cell recruitment and activation by breast cancer cells supports anti-tumoral responses. *Int. J. Mol. Sci.* 21:5293. doi: 10.3390/ijms21155293
- Arvold, N. D., Taghian, A. G., Niemierko, A., Abi Raad, R. F., Sreedhara, M., Nguyen, P. L., et al. (2011). Age, breast cancer subtype approximation, and local recurrence after breast-conserving therapy. *J. Clin. Oncol.* 29:3885. doi: 10.1200/JCO.2011.36.1105
- Barbano, R., Copetti, M., Perrone, G., Paziienza, V., Muscarella, L. A., Balsamo, T., et al. (2011). High RAD51 mRNA expression characterize estrogen receptor-positive/progesteron receptor-negative breast cancer and is associated with patient's outcome. *Int. J. Cancer* 129, 536–545. doi: 10.1002/ijc.25736
- Berger, A. C., Korkut, A., Kanchi, R. S., Hegde, A. M., Lenoir, W., Liu, W., et al. (2018). A comprehensive pan-cancer molecular study of gynecologic and breast cancers. *Cancer Cell* 33, 690–705.e9. doi: 10.1016/j.ccell.2018.03.014
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008:P10008. doi: 10.1088/1742-5468/2008/10/P10008
- Cantile, M., Pettinato, G., Procino, A., Feliciello, I., Cindolo, L., and Cillo, C. (2003). *In vivo* expression of the whole HOX gene network in human breast cancer. *Eur. J. Cancer* 39, 257–264. doi: 10.1016/S0959-8049(02)00599-3
- Cantini, L., Medico, E., Fortunato, S., and Caselle, M. (2015). Detection of gene communities in multi-networks reveals cancer drivers. *Sci. Rep.* 5:17386. doi: 10.1038/srep17386
- Carr, J. R., Kiefer, M. M., Park, H. J., Li, J., Wang, Z., Fontanarosa, J., et al. (2012). FoxM1 regulates mammary luminal cell fate. *Cell Rep.* 1, 715–29. doi: 10.1016/j.celrep.2012.05.005
- Ciriello, G., Sinha, R., Hoadley, K. A., Jacobsen, A. S., Reva, B., Perou, C. M., et al. (2013). The molecular diversity of luminal A breast tumors. *Breast Cancer Res. Treat.* 141, 409–420. doi: 10.1007/s10549-013-2699-3
- 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
- Corces, M. R., and Corces, V. G. (2016). The three-dimensional cancer genome. *Curr. Opin. Genet. Dev.* 36, 1–7. doi: 10.1016/j.gde.2016.01.002
- Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Interf. Complex Syst.* 5, 1–9.
- de Anda-Jáuregui, G., Espinal-Enríquez, J., Drago-García, D., and Hernández-Lemus, E. (2018). Nonredundant, highly connected micrornas control functionality in breast cancer networks. *Int. J. Genomics* 2018:9585383. doi: 10.1155/2018/9585383
- de Anda-Jáuregui, G., Espinal-Enríquez, J., and Hernández-Lemus, E. (2019a). Spatial organization of the gene regulatory program: an information theoretical approach to breast cancer transcriptomics. *Entropy* 21:195. doi: 10.3390/e21020195
- de Anda-Jáuregui, G., Fresno, C., García-Cortés, D., Espinal-Enríquez, J., and Hernández-Lemus, E. (2019b). Intrachromosomal regulation decay in breast cancer. *Appl. Math. Nonlin. Sci.* 4, 217–224. doi: 10.2478/AMNS.2019.1.00020
- Dorantes-Gilardi, R., García-Cortés, D., Hernández-Lemus, E., and Espinal-Enríquez, J. (2020). Multilayer approach reveals organizational principles disrupted in breast cancer co-expression networks. *Appl. Netw. Sci.* 5, 1–23. doi: 10.1007/s41109-020-00291-1
- Drago-García, D., Espinal-Enríquez, J., and Hernández-Lemus, E. (2017). Network analysis of *emt* and *met* micro-RNA regulation in breast cancer. *Sci. Rep.* 7:13534. doi: 10.1038/s41598-017-13903-1
- Ebright, R. Y., Lee, S., Wittner, B. S., Niederhoffer, K. L., Nicholson, B. T., Bardia, A., et al. (2020). Deregulation of ribosomal protein expression and translation promotes breast cancer metastasis. *Science* 367, 1468–1473. doi: 10.1126/science.aay0939
- Emmert-Streib, F., de Matos Simoes, R., Mullan, P., Haibe-Kains, B., and Dehmer, M. (2014). The gene regulatory network for breast cancer: integrated regulatory landscape of cancer hallmarks. *Front. Genet.* 5:15. doi: 10.3389/fgene.2014.00015
- Espinal-Enríquez, 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. Rep.* 7:1760. doi: 10.1038/s41598-017-01314-1

- Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S., Nobel, A. B., et al. (2006). Concordance among gene-expression-based predictors for breast cancer. *N. Engl. J. Med.* 355, 560–569. doi: 10.1056/NEJMoa052933
- Fancello, L., Kampen, K. R., Hofman, I. J., Verbeek, J., and De Keersmaecker, K. (2017). The ribosomal protein gene RPL5 is a haploinsufficient tumor suppressor in multiple cancer types. *Oncotarget*. 8, 14462–14478. doi: 10.18632/oncotarget.14895
- Fiorito, E., Sharma, Y., Gilfillan, S., Wang, S., Singh, S. K., Satheesh, S. V., et al. (2016). CTCF modulates estrogen receptor function through specific chromatin and nuclear matrix interactions. *Nucleic Acids Res.* 44, 10588–10602. doi: 10.1093/nar/gkw785
- Fortunato, S., and Hric, D. (2016). Community detection in networks: a user guide. *Phys. Rep.* 659, 1–44. doi: 10.1016/j.physrep.2016.09.002
- Freeman, J. L., Perry, G. H., Feuk, L., Redon, R., McCarroll, S. A., Altshuler, D. M., et al. (2006). Copy number variation: new insights in genome diversity. *Genome Res.* 16, 949–961. doi: 10.1101/gr.3677206
- Fresno, C., González, G. A., Llera, A. S., and Fernández, E. A. (2016). *PBCMC: Permutation-Based Confidence for Molecular Classification*. R package version 1.2.
- Fresno, C., González, G. A., Merino, G. A., Flesia, A. G., Podhajcer, O. L., Llera, A. S., et al. (2017). A novel non-parametric method for uncertainty evaluation of correlation-based molecular signatures: its application on PAM50 algorithm. *Bioinformatics* 33, 693–700. doi: 10.1093/bioinformatics/btw704
- Fu, D., Li, J., Wei, J., Zhang, Z., Luo, Y., Tan, H., et al. (2018). HMGB2 is associated with malignancy and regulates Warburg effect by targeting LDHB and FBP1 in breast cancer. *Cell Commun. Signal.* 16, 1–10. doi: 10.1186/s12964-018-0219-0
- García-Cortés, D., de Anda-Jáuregui, G., Fresno, C., Hernández-Lemus, E., and Espinal-Enríquez, J. (2020). Gene co-expression is distance-dependent in breast cancer. *Front. Oncol.* 10:1232. doi: 10.3389/fonc.2020.01232
- Garraway, L. A., and Lander, E. S. (2013). Lessons from the cancer genome. *Cell* 153, 17–37. doi: 10.1016/j.cell.2013.03.002
- Gatza, M. L., Silva, G. O., Parker, J. S., Fan, C., and Perou, C. M. (2014). An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nat. Genet.* 46, 1051–1059. doi: 10.1038/ng.3073
- 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
- He, J., Zhou, Z., Reed, M., and Califano, A. (2017). Accelerated parallel algorithm for gene network reverse engineering. *BMC Syst. Biol.* 11:83. doi: 10.1186/s12918-017-0458-5
- Hernández-Lemus, E., Reyes-Gopar, H., Espinal-Enríquez, J., and Ochoa, S. (2019). The many faces of gene regulation in cancer: a computational oncogenomics outlook. *Genes* 10:865. doi: 10.3390/genes10110865
- Hsu, H. M., Chu, C. M., Chang, Y. J., Yu, J. C., Chen, C. T., Jian, C. E., et al. (2019). Six novel immunoglobulin genes as biomarkers for better prognosis in triple-negative breast cancer by gene co-expression network analysis. *Sci. Rep.* 9:4484. doi: 10.1038/s41598-019-40826-w
- Hu, Z., Fan, C., Oh, D. S., Marron, J., He, X., Qaqish, B. F., et al. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* 7:96. doi: 10.1186/1471-2164-7-96
- Hu, Z., Huang, G., Sadanandam, A., Gu, S., Lenburg, M. E., Pai, M., et al. (2010). The expression level of HJURP has an independent prognostic impact and predicts the sensitivity to radiotherapy in breast cancer. *Breast Cancer Res.* 12, 1–15. doi: 10.1186/bcr2487
- Hung, T., Wang, Y., Lin, M. F., Koegel, A. K., Kotake, Y., Grant, G. D., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* 43, 621–629. doi: 10.1038/ng.848
- Hur, H., Lee, J. Y., Yun, H. J., Park, B. W., and Kim, M. H. (2014). Analysis of HOX gene expression patterns in human breast cancer. *Mol. Biotechnol.* 56, 64–71. doi: 10.1007/s12033-013-9682-4
- Inaki, K., Menghi, F., Woo, X. Y., Wagner, J. P., Jacques, P. É., Lee, Y. F., et al. (2014). Systems consequences of amplicon formation in human breast cancer. *Genome Res.* 24, 1559–1571. doi: 10.1101/gr.164871.113
- Kamalakaran, S., Varadan, V., Giercksky Russnes, H. E., Levy, D., Kendall, J., Janevski, A., et al. (2011). DNA methylation patterns in luminal breast cancers differ from non-luminal subtypes and can identify relapse risk independent of other clinical variables. *Mol. Oncol.* 5, 77–92. doi: 10.1016/j.molonc.2010.11.002
- Khoury, A., Achinger-Kawecka, J., Bert, S. A., Smith, G. C., French, H. J., Luu, P.-L., et al. (2020). Constitutively bound ctf sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat. Commun.* 11, 1–13. doi: 10.1038/s41467-019-13753-7
- Lachmann, A. (2016). *Confounding effects in gene expression and their impact on downstream analysis* (Ph.D. thesis), Columbia University, New York, NY, United States.
- Lee, T. I., and Young, R. A. (2013). Transcriptional regulation and its misregulation in disease. *Cell* 152, 1237–1251. doi: 10.1016/j.cell.2013.02.014
- Liao, Y., Wang, Y., Cheng, M., Huang, C., and Fan, X. (2020). Weighted gene coexpression network analysis of features that control cancer stem cells reveals prognostic biomarkers in lung adenocarcinoma. *Front. Genet.* 11:311. doi: 10.3389/fgene.2020.00311
- Liu, R., Guo, C. X., and Zhou, H. H. (2015). Network-based approach to identify prognostic biomarkers for estrogen receptor-positive breast cancer treatment with tamoxifen. *Cancer Biol. Ther.* 16, 317–324. doi: 10.1080/15384047.2014.1002360
- Lu, X.-F., Zeng, D., Liang, W.-Q., Chen, C.-F., Sun, S.-M., and Lin, H.-Y. (2018). FoxM1 is a promising candidate target in the treatment of breast cancer. *Oncotarget* 9, 842–852. doi: 10.18632/oncotarget.23182
- Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., and Califano, A. (2006). Reverse engineering cellular networks. *Nat. Protoc.* 1, 662–671. doi: 10.1038/nprot.2006.106
- McCarroll, S. A., and Altshuler, D. M. (2007). Copy-number variation and association studies of human disease. *Nat. Genet.* 39, S37–S42. doi: 10.1038/ng2080
- Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhi, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12, 1–14. doi: 10.1186/gb-2011-12-4-r41
- Metzger-Filho, O., Sun, Z., Viale, G., Price, K. N., Crivellari, D., Snyder, R. D., et al. (2013). Patterns of recurrence and outcome according to breast cancer subtypes in lymph node-negative disease: results from international breast cancer study group trials viii and ix. *J. Clin. Oncol.* 31:3083. doi: 10.1200/JCO.2012.46.1574
- Millour, J., Constantinidou, D., Stavropoulou, A. V., Wilson, M. S., Myatt, S. S., Kwok, J. M., et al. (2010). FOXM1 is a transcriptional target of ER $\alpha$  and has a critical role in breast cancer endocrine sensitivity and resistance. *Oncogene* 29, 2983–2995. doi: 10.1038/onc.2010.47
- Montes de Oca, R., Gurard-Levin, Z. A., Berger, F., Rehman, H., Martel, E., Corpet, A., et al. (2015). The histone chaperone HJURP is a new independent prognostic marker for luminal A breast carcinoma. *Mol. Oncol.* 9, 657–674. doi: 10.1016/j.molonc.2014.11.002
- Myhre, S., Lingjærde, O. C., Hennessy, B. T., Aure, M. R., Carey, M. S., Alsner, J., et al. (2013). Influence of DNA copy number and mRNA levels on the expression of breast cancer related proteins. *Mol. Oncol.* 7, 704–718. doi: 10.1016/j.molonc.2013.02.018
- Newman, M. E. (2006). Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* 74:036104. doi: 10.1103/PhysRevE.74.036104
- Nieto-Jiménez, C., Alcaraz-Sanabria, A., Páez, R., Pérez-Peña, J., Corrales-Sánchez, V., Pandiella, A., et al. (2017). DNA-damage related genes and clinical outcome in hormone receptor positive breast cancer. *Oncotarget* 8, 62834–62841. doi: 10.18632/oncotarget.10886
- Novak, P., Jensen, T., Oshiro, M. M., Watts, G. S., Kim, C. J., and Futscher, B. W. (2008). Agglomerative epigenetic aberrations are a common event in human breast cancer. *Cancer Res.* 68, 8616–8625. doi: 10.1158/0008-5472.CAN-08-1419
- Novak, P., Jensen, T., Oshiro, M. M., Wozniak, R. J., Nouzova, M., Watts, G. S., et al. (2006). Epigenetic inactivation of the HOXA gene cluster in breast cancer. *Cancer Res.* 66, 10664–10670. doi: 10.1158/0008-5472.CAN-06-2761
- Nueda, M. J., Ferrer, A., and Conesa, A. (2012). ARSYn: a method for the identification and removal of systematic noise in multifactorial time course microarray experiments. *Biostatistics* 13, 553–566. doi: 10.1093/biostatistics/kxr042
- Nuncia-Cantarero, M., Martínez-Canales, S., Andrés-Pretel, F., Santpere, G., Ocaña, A., and Galan-Moya, E. M. (2018). Functional transcriptomic annotation and protein-protein interaction network analysis identify NEK2,

- BIRC5, and TOP2A as potential targets in obese patients with luminal A breast cancer. *Breast Cancer Res. Treat.* 168, 613–623. doi: 10.1007/s10549-017-4652-3
- Paz, M. F., Fraga, M. F., Avila, S., Guo, M., Pollan, M., Herman, J. G., et al. (2003). A systematic profile of DNA methylation in human cancer cell lines. *Cancer Res.* 63, 1114–1121.
- Perez-Peña, J., Corrales-Sánchez, V., Amir, E., Pandiella, A., and Ocana, A. (2017). Ubiquitin-conjugating enzyme E2T (UBE2T) and denticleless protein homolog (DTL) are linked to poor outcome in breast and lung cancers. *Sci. Rep.* 7:17530. doi: 10.1038/s41598-017-17836-7
- Perou, C. M., Sorile, T., Eisen, M. B., Van De Rijn, M., Jeffrey, S. S., Ress, C. A., et al. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747–752. doi: 10.1038/35021093
- Porter, M. A., Onnela, J.-P., and Mucha, P. J. (2009). Communities in networks. *Not. AMS* 56, 1082–1097.
- Prat, A., and Perou, C. M. (2011). Deconstructing the molecular portraits of breast cancer. *Mol. Oncol.* 5, 5–23. doi: 10.1016/j.molonc.2010.11.003
- Prieto, C., Riusueño, A., Fontanillo, C., and De Las Rivas, J. (2008). Human gene coexpression landscape: confident network derived from tissue transcriptomic profiles. *PLoS ONE* 3:e3911. doi: 10.1371/journal.pone.0003911
- Pugacheva, E. M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A. L., et al. (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. *Proc. Natl. Acad. Sci. U.S.A.* 117, 2020–2031. doi: 10.1073/pnas.1911708117
- Rafique, S., Thomas, J. S., Sproul, D., and Bickmore, W. A. (2015). Estrogen-induced chromatin decondensation and nuclear re-organization linked to regional epigenetic regulation in breast cancer. *Genome Biol.* 16:145. doi: 10.1186/s13059-015-0719-9
- Rahiminejad, S., Maurya, M. R., and Subramaniam, S. (2019). Topological and functional comparison of community detection algorithms in biological networks. *BMC Bioinformatics* 20:212. doi: 10.1186/s12859-019-2746-0
- Redmond, A. M., Byrne, C., Bane, F. T., Brown, G. D., Tibbitts, P., O'Brien, K., et al. (2015). Genomic interaction between ER and HMGB2 identifies DDX18 as a novel driver of endocrine resistance in breast cancer cells. *Oncogene* 34, 3871–3880. doi: 10.1038/onc.2014.323
- Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., et al. (2006). Global variation in copy number in the human genome. *Nature* 444, 444–454. doi: 10.1038/nature05329
- 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. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Natl. Acad. Sci. U.S.A.* 105, 1118–1123. doi: 10.1073/pnas.0706851105
- Serrano-Carbajal, E. A., Espinal-Enríquez, J., and Hernández-Lemus, E. (2020). Targeting metabolic deregulation landscapes in breast cancer subtypes. *Front. Oncol.* 10:97. doi: 10.3389/fonc.2020.00097
- Shao, X., Lv, N., Liao, J., Long, J., Xue, R., Ai, N., et al. (2019). Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Med. Genet.* 20:175. doi: 10.1186/s12881-019-0909-5
- Shlien, A., and Malkin, D. (2009). Copy number variations and cancer. *Genome Med.* 1, 1–9. doi: 10.1186/gm62
- Sonawane, A. R., Weiss, S. T., Glass, K., and Sharma, A. (2019). Network medicine in the age of biomedical big data. *Front. Genet.* 10:294. doi: 10.3389/fgene.2019.00294
- Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol.* 19, A68–A77. doi: 10.5114/wo.2014.47136
- Tong, D. D., Zhang, J., Wang, X. F., Li, Q., Liu, L. Y., Yang, J., et al. (2020). MeCP2 facilitates breast cancer growth via promoting ubiquitination-mediated P53 degradation by inhibiting RPL5/RPL11 transcription. *Oncogenesis* 9:56. doi: 10.1038/s41389-020-0239-7
- van Dam, S., Vösa, U., van der Graaf, A., Franke, L., and de Magalhães, J. P. (2018). Gene co-expression analysis for functional classification and gene-disease predictions. *Brief. Bioinformatics* 19, 575–592. doi: 10.1093/bib/bbw139
- Wang, J., Yi, Y., Chen, Y., Xiong, Y., and Zhang, W. (2020a). Potential mechanism of rrm2 for promoting cervical cancer based on weighted gene co-expression network analysis. *Int. J. Med. Sci.* 17:2362. doi: 10.7150/ijms.47356
- Wang, J. C., Ramaswami, G., and Geschwind, D. H. (2020b). Gene co-expression network analysis in human spinal cord highlights mechanisms underlying amyotrophic lateral sclerosis susceptibility. *bioRxiv*. doi: 10.1101/2020.08.16.253377
- Wilkinson, D. M., and Huberman, B. A. (2004). A method for finding communities of related genes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 5241–5248. doi: 10.1073/pnas.0307740100
- Wu, Y., Luo, S., Yin, X., He, D., Liu, J., Yue, Z., et al. (2019). Co-expression of key gene modules and pathways of human breast cancer cell lines. *Biosci. Rep.* 39:BSR20181925. doi: 10.1042/BSR20181925
- Xiao, B., Chen, L., Ke, Y., Hang, J., Cao, L., Zhang, R., et al. (2018). Identification of methylation sites and signature genes with prognostic value for luminal breast cancer. *BMC Cancer* 18:405. doi: 10.1186/s12885-018-4314-9
- Yang, Y., Han, L., Yuan, Y., Li, J., Hei, N., and Liang, H. (2014). Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat. Commun.* 5:3231. doi: 10.1038/ncomms4231
- Yevshin, I., Sharipov, R., Kolmykov, S., Kondrakhin, Y., and Kolpakov, F. (2018). GTRD: a database on gene transcription regulation—2019 update. *Nucleic Acids Res.* 47, D100–D105. doi: 10.1093/nar/gky1128
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterprofiler: an R package for comparing biological themes among gene clusters. *OmicS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Zamora-Fuentes, J. M., Hernandez-Lemus, E., and Espinal-Enriquez, J. (2020). Gene expression and co-expression networks are strongly altered through stages in clear cell renal carcinoma. *Front. Genet.* 11:1232. doi: 10.3389/fgene.2020.578679
- Zhang, M. H., Man, H. T., Zhao, X. D., Dong, N., and Ma, S. L. (2014). Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials. *Biomed. Rep.* 2, 41–52. doi: 10.3892/br.2013.187
- Zhang, T., Hu, H., Yan, G., Wu, T., Liu, S., Chen, W., et al. (2019). Long non-coding rna and breast cancer. *Technol. Cancer Res. Treat.* 18:1533033819843889. doi: 10.1177/1533033819843889
- Zhu, J., Zhang, B., Smith, E. N., Drees, B., Brem, R. B., Kruglyak, L., et al. (2008). Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat. Genet.* 40, 854–861. doi: 10.1038/ng.167

**Conflict of Interest:** 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 © 2021 García-Cortés, Hernández-Lemus and Espinal-Enríquez. 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.