Functional and transcriptional connectivity of communities in breast cancer co-expression networks

Guillermo de Anda-Jáuregui[1], Sergio Antonio Alcalá-Corona[2], Jesús Espinal-Enríquez[1,3] and Enrique Hernández-Lemus[1,3]*

*Correspondence:
ehernandez@inmegen.gob.mx
[1]Computational Genomics,
National Institute of Genomic
Medicine, Periférico Sur 4809,
Ciudad de México, México
[3]Centro de Ciencias de la
Complejidad, Universidad Nacional
Autónoma de México, Ciudad
Universitaria, Ciudad de México,
México
Full list of author information is
available at the end of the article

Abstract

Transcriptional co-expression networks represent the concerted gene regulation programs by means of statistical inference of co-expression patterns. The rich phenomenology of transcriptional processes behind complex phenotypes such as cancer, is often captured (at least partially) in the connectivity structure of transcriptional co-expression networks. By analyzing the community structure of these networks, we may develop a deeper understanding of that phenomenology. We identified the modular structure of a transcriptional co-expression network obtained from breast cancer gene expression as well as a non-cancer adjacent breast tissue network as a control. We then analyzed the biological functions associated to the resulting communities by means of enrichment analysis. We also generated two projected networks for both, tumor and control networks: The first one is a projection to a network in which nodes are communities and edges represent topologically adjacent communities, indicating co-expression patterns between them. For the second projection, a bipartite network was generated containing a layer of modules and a layer of biological processes, with links between modules and the functions in which they are enriched; from this bipartite network, a projection to the community layer was obtained. From the analysis of the communities and projections, we were able to discern distinctive patterns of regulation between tumors and controls. Even though the connectivity structure of transcriptional co-expression networks is quite different, the topology of the projected networks is somehow similar, indicating functional compartmentalization, in both tumor and control conditions. However, the biological functions represented in the corresponding modules resulted notably different, with the tumor network comprising functional modules enriched for well-known hallmarks of cancer.

Keywords: Breast cancer networks, Modularity, Bipartite networks, Functional enrichment

Background

Co-expression networks are graph-theoretical constructs that represent global-level regulatory interactions and expression patterns of genes. These are well-defined mathematical structures amenable for systematic analysis of its global and local properties, as well as its dynamics and functionality. The case of said networks related to complex phenotypes such as cancer has been an area of interest in recent times (De Craene

and Berx 2013; Dang et al. 2006; de Anda-Jáuregui et al. 2016). Modular structure (Girvan and Newman 2002; Newman 2006; Palla et al. 2005; Palla et al. 2007) is a quite relevant feature of co-expression networks, since it may provide some clues as to what are the actual biological mechanisms in complex phenotypes (Alcalá-Corona et al. 2016). In the case of breast cancer deregulation, functional biological organization has been shown to be related to network modularity (Alcalá-Corona et al. 2017; Alcalá-Corona et al. 2018). Such *community structure* of gene organization is characteristic of the different breast cancer molecular subtypes (Alcalá-Corona et al. 2017), so that particularities of the molecular phenotypes are well represented in the modular partition of the network (Alcalá-Corona et al. 2018).

Transcriptional co-expression networks can be probabilistically inferred from high-throughput gene expression data (Basso et al. 2005; Margolin et al. 2006; Hernández-Lemus and Siqueiros-García 2013; Hernández-Lemus and Rangel-Escareño 2011; Delgado and Gómez-Vela 2018; Kuzmanovski et al. 2018; Wong et al. 2018; Manem et al. 2018; Liu et al. 2018), and provide a representation of the expression landscape of a given phenotype. These type of regulatory networks consist of nodes representing genes and links representing co-expression (i.e. strong statistical dependency) between said genes. Given a Co-Expression Network $\mathcal{G}$, composed of gene nodes and links representing co-expression between genes, it is possible to detect non-overlapping *co-expression modules* (communities) due to its topology.

The gene set of each module M_i may be tested for association to known gene-sets of biological interest, such as biological functions, using enrichment analysis. These associations may be represented as a bipartite graph $\mathcal{B}$, with a set of module nodes M and a set of biological functions F, with links between modules and the functions in which they are enriched. With this in mind, it is possible to project $\mathcal{G}$ and $\mathcal{B}$ into two new graphs $\mathcal{G}P$ and $\mathcal{B}P$ (see methods) where nodes correspond to modules detected in the original graph $\mathcal{G}$.

These two projections recover two distinct types of relationships between groups of genes: on the one hand, whether different groups of genes have a level of *co-expression* that may be driven by biological factors, such as co-regulation; and on the other hand, whether different groups of genes are involved in the *control* of biological functions that are necessary for a given biological context (for instance, a phenotype). An interesting case is that of modules that are co-regulated and connected through shared biological functions.

In this work, we analyze two coexpression networks derived from *basal breast cancer* (tumors) and *healthy breast tissue* (controls), and explore the two modular projections described. We identify the differences in modular structure between the two phenotypes, and how these different modular structures differ in terms of the two types of intermodular relationships that we have described.

Methods

Network inference

Co-expression networks were reconstructed from gene expression data. Basal breast cancer gene expression data, along with adjacent normal expression data, were obtained from the Cancer Genome Atlas (Network and et al. 2012). Data acquisition, and pre-processing is described in (Espinal-Enriquez et al. 2017). Briefly, we used 142 Basal-like subtype breast cancer samples, along with 101 solid-tissue adjacent normal samples.

15,642 annotated genes were included in each sample, after removal of low-counts transcripts ($<$ 5 per sample). This set of un-paired data were pre-processed, normalized and bias-reduced, to have a comparable set of expression data between cancer and control samples.

Mutual Information (MI) was computed using an implementation of the ARACNE algorithm for all gene pairs (de Anda-Jáuregui et al. 2019). A suitable MI threshold was selected based on the following criteria:

- At least 80% of nodes in the genome (out of 15,642) must be present in the network by being connected to at least one other gene
- The network must have a giant connected component (i.e., the largest connected component with more than half of the nodes)
- The highest (most restrictive) MI threshold must be selected

We evaluated different MI threshold values related to quantiles of the MI distribution. Generated networks were imported as `igraph` for [R] objects. igraph version 0.71 and R version 3.5.1 were used.

Mutual information is the maximum entropy/maximum likelihood estimate of statistical dependence between two random variables (Chow and Liu 1968). It is indeed a symmetrized version of the Kullback-Leibler divergence between the joint probability distribution for two variables and the product of their marginals (Kullback and Leibler 1951) (i.e. the joint probability distribution under independence conditions). Being a maximum entropy estimate it needs the least number of assumptions on the probability distributions. Indeed the only needed assumption is that these distributions have compact support. Other correlation measures assume identically distributed variables, linearity or rank ordering among them, etc. Such assumptions are often not compliant with the nature of gene expression data such as nonlinearity, 'delays' (i.e. correlation shifts), and so on. For these reasons, mutual information has been thoroughly used for the inference of (large) gene co-expression networks. Another advantage of the use of mutual information measures to deconvolute gene regulatory networks from massive gene expression data is the fact that, in most cases (whenever Hammersley-Clifford conditions apply), the resulting graphs meet the requirements to belong to the family of Markov random fields, something that under some scenarios may be quite useful (Dobruschin 1968).

The major drawback for the use of the mutual information approach is the fact that one needs a way to reconstruct the probability distributions from empirical data. Even under the relatively 'soft' conditions imposed by Glivenko-Cantelli convergence, this means that one still have to have a somehow large number of samples (more than approx. 100 for the case of gene expression data) for the empirical distribution to be useful in order to minimize the number of false positives. These conditions are fulfilled here.

Module detection and enrichment

Modules where detected using the `Infomap` (Rosvall and Bergstrom 2007; 2008; Rosvall et al. 2009) implementation for `igraph`, using 1000 iterations to achieve convergence. We have chosen the Infomap algorithm, since it has proven to be highly efficient compared to other methods. Based on benchmarks, Infomap was the best- ranked method in runtime, accuracy and performance (Lancichinetti et al. 2009), as it was assessed in terms of the LFR benchmark (Lancichinetti et al. 2008).

The field of enrichment analysis includes a wide variety of techniques (García-Campos et al. 2015). In this work we used an Over Representation Analysis, in which a hypergeometric (or Fischer exact) test is used to identify a statistically significant association between each module's gene set, and the sets of genes involved in biological functions as described by the Gene Ontology (GO) database (Ashburner et al. 2000).
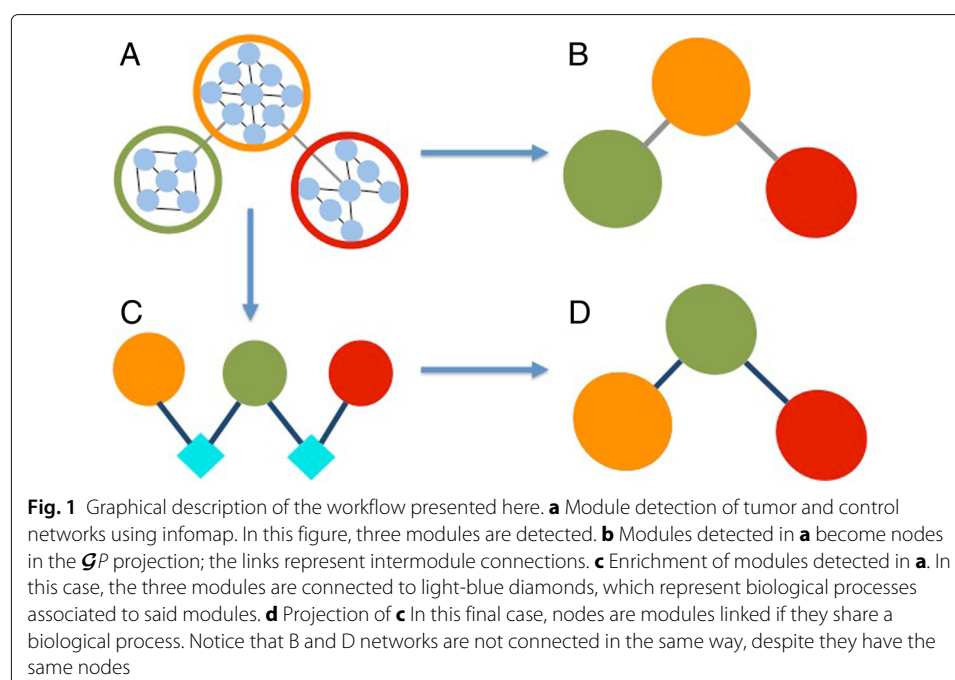
Each module gene set was tested for enrichment of Gene Ontology (Ashburner et al. 2000) terms via hypergeometric testing using the HTSanalyzer (Wang et al. 2011) package for R. GO terms were considered enriched if they had an adjusted Benjamini-Hochberg (Benjamini and Hochberg 1995) *p*-value smaller than 0.05. Enrichment relationships found were represented as a bipartite network, with a layer of modules and a layer of GO terms.

Figure 1 presents a pictorial abstraction of this process. Panel **a** represents module detection of $\mathcal{G}$ using Infomap. In panel B modules detected in panel A become nodes in the $\mathcal{G}P$ projection; links represent intermodule connections. Enrichment of modules (i.e. the $\mathcal{B}$ network) detected in panel A is presented in panel C. The three modules are connected to turquoise diamonds, which represent biological processes associated to said modules. Panel D shows a projection $\mathcal{B}P$ of $\mathcal{B}$ in which nodes are modules linked if they share a biological process.

GP and BP projections

Topological and functional neighborhoods define two projections $\mathcal{G}P$ and $\mathcal{B}P$ as previously mentioned. The first projection, $\mathcal{G}P$, is a graph where nodes are modules M, and links exist between modules M_i and M_j if there are links in $\mathcal{G}$ between genes in M_i and genes in M_j: we say these modules are *topologically adjacent* in the original network.

The second projection, $\mathcal{B}P$, is a graph where nodes are modules M and links exist between M_i and M_j if there is overlap in the neighborhoods of M_i and M_j in B: we say that these modules are *functionally adjacent*.

Fig. 1 Graphical description of the workflow presented here. **a** Module detection of tumor and control networks using infomap. In this figure, three modules are detected. **b** Modules detected in **a** become nodes in the $\mathcal{G}P$ projection; the links represent intermodule connections. **c** Enrichment of modules detected in **a**. In this case, the three modules are connected to light-blue diamonds, which represent biological processes associated to said modules. **d** Projection of **c** In this final case, nodes are modules linked if they share a biological process. Notice that B and D networks are not connected in the same way, despite they have the same nodes
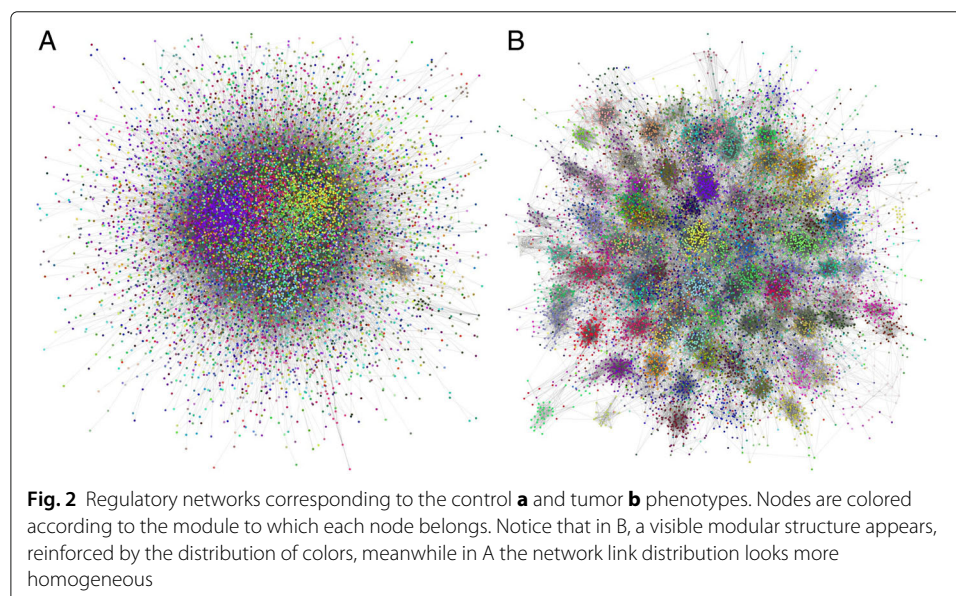
Results

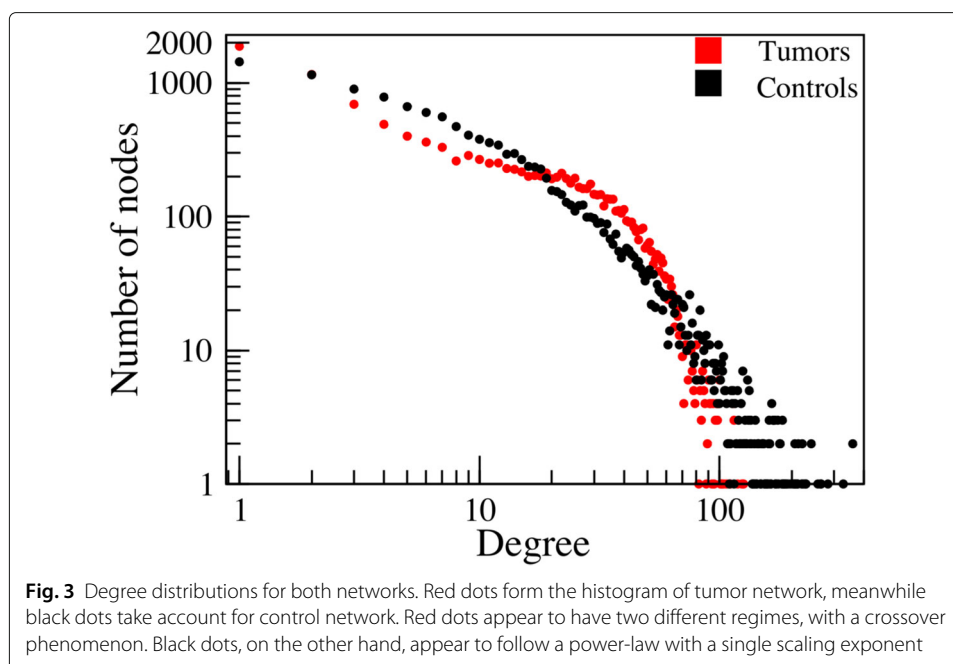### Co-expression networks for breast cancer and adjacent normal

Networks were generated from the tumors and control datasets. After scanning different threshold values for mutual information (see Additional File 1) the highest threshold for MI that covered our criteria was found at the 0.999 quantile. These networks are described in Additional File 2. Figure 2 illustrates how different the tumor and control networks are; nodes are colored by the module to which they belong. It can be observed in the tumor network, modules with nodes of the same color, whereas in the control network, modules are not observable and colors are less separated. This is further supported by the different degree distributions (Fig. 3).

Said differences can be identified even by a quick glance at the node-degree distributions (Fig. 3; χ^2 testing for differences in discrete distribution was performed, with the following results: χ^2-statistic $= 1074170, p - value = 4.99e - 10$), as well as by the observation of the force-directed network visualization. The control network is characterized by a mono-scaled regime (the degree distribution appears to follow a power-law with a single scaling exponent over the whole range of degree values) whereas the tumor network presents an evident crossover phenomenon leading to multi-scaling, i.e. the degree distribution does not follow a power-law with a single scaling exponent, but rather it seems to have several different scaling regimes, with regions containing inflection points in so-called *crossover* regions.

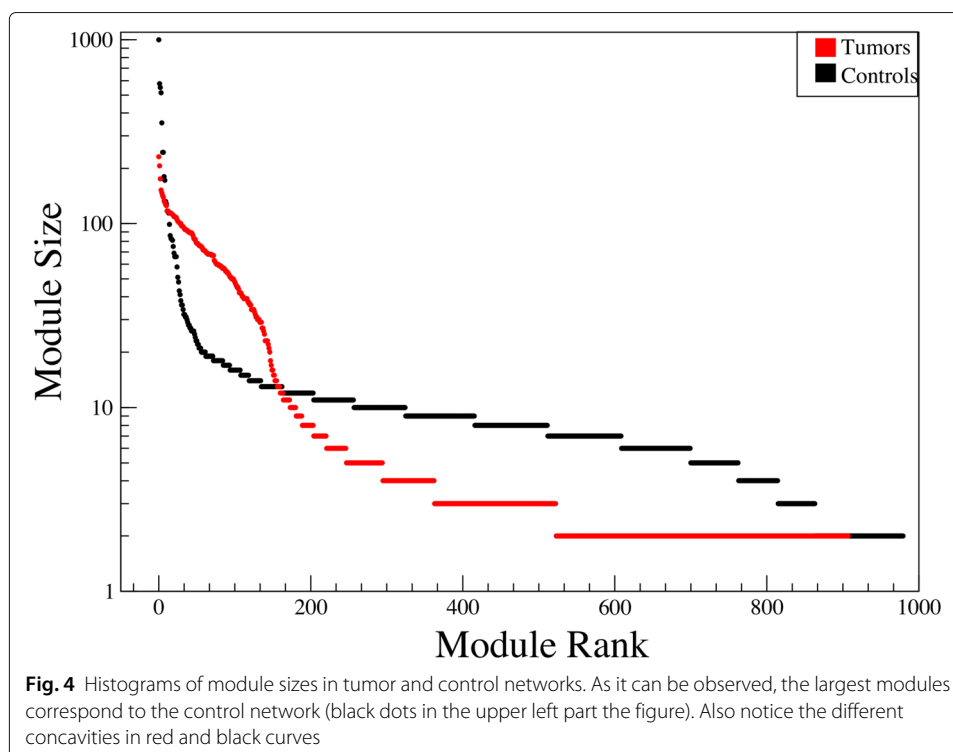### Modular structure of breast cancer and healthy breast networks

Distinct modular structures were found in each network, in agreement with previous results from our group (Alcalá-Corona et al. 2018). The partition for the tumor network has a smaller description length L (Rosvall and Bergstrom 2008) value ($L = 8.268641$) than the control network ($L = 11.80941$). In the control network, we identified 981 modules, whereas for the tumor network we found 910 modules. As it can also be observed in Figs. 2 and 3, Fig. 4 shows histograms of the different module sizes, showing the
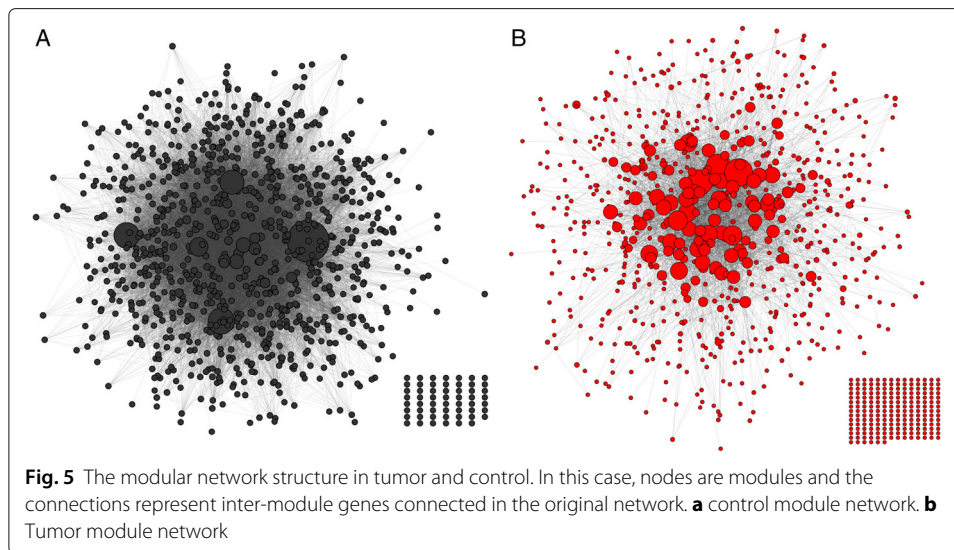

Fig. 2 Regulatory networks corresponding to the control **a** and tumor **b** phenotypes. Nodes are colored according to the module to which each node belongs. Notice that in B, a visible modular structure appears, reinforced by the distribution of colors, meanwhile in A the network link distribution looks more homogeneous

Fig. 3 Degree distributions for both networks. Red dots form the histogram of tumor network, meanwhile black dots take account for control network. Red dots appear to have two different regimes, with a crossover phenomenon. Black dots, on the other hand, appear to follow a power-law with a single scaling exponent

largest modules in control network (χ^2 testing for differences in discrete distribution was performed, with the following results: χ^2-statistic $= 40324.45$, $p-value = 4.99e-10$).

For each transcriptional co-expression network, we projected the modules identified in it to a $\mathcal{GP}$ network were adjacent modules in the original network are found. These $\mathcal{GP}_t$ (for tumors) and $\mathcal{GP}_c$ (for controls) are depicted in Fig. 5, and described in Additional file 3 (modular projection parameters). There are three main differences between these
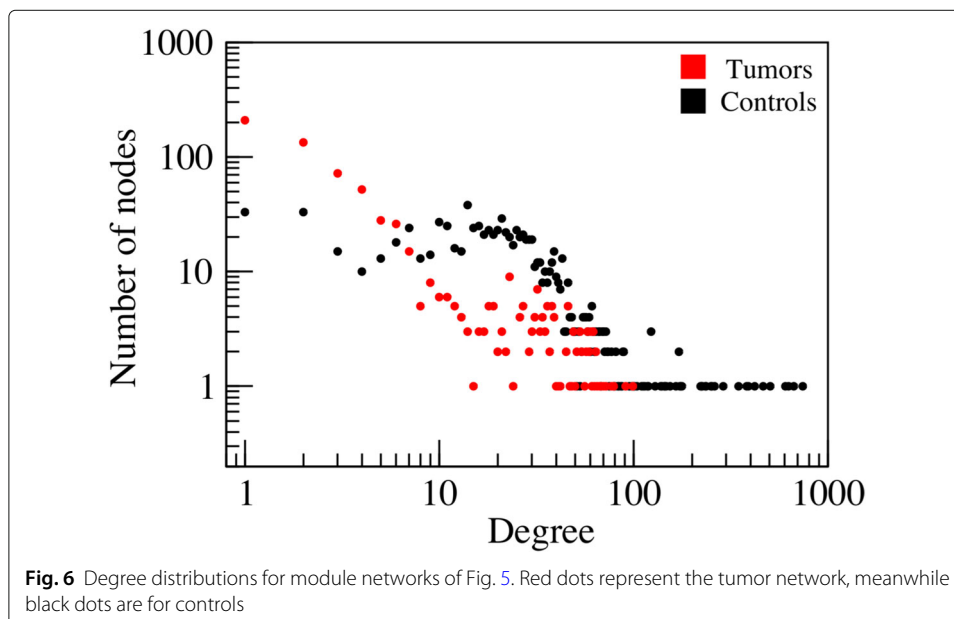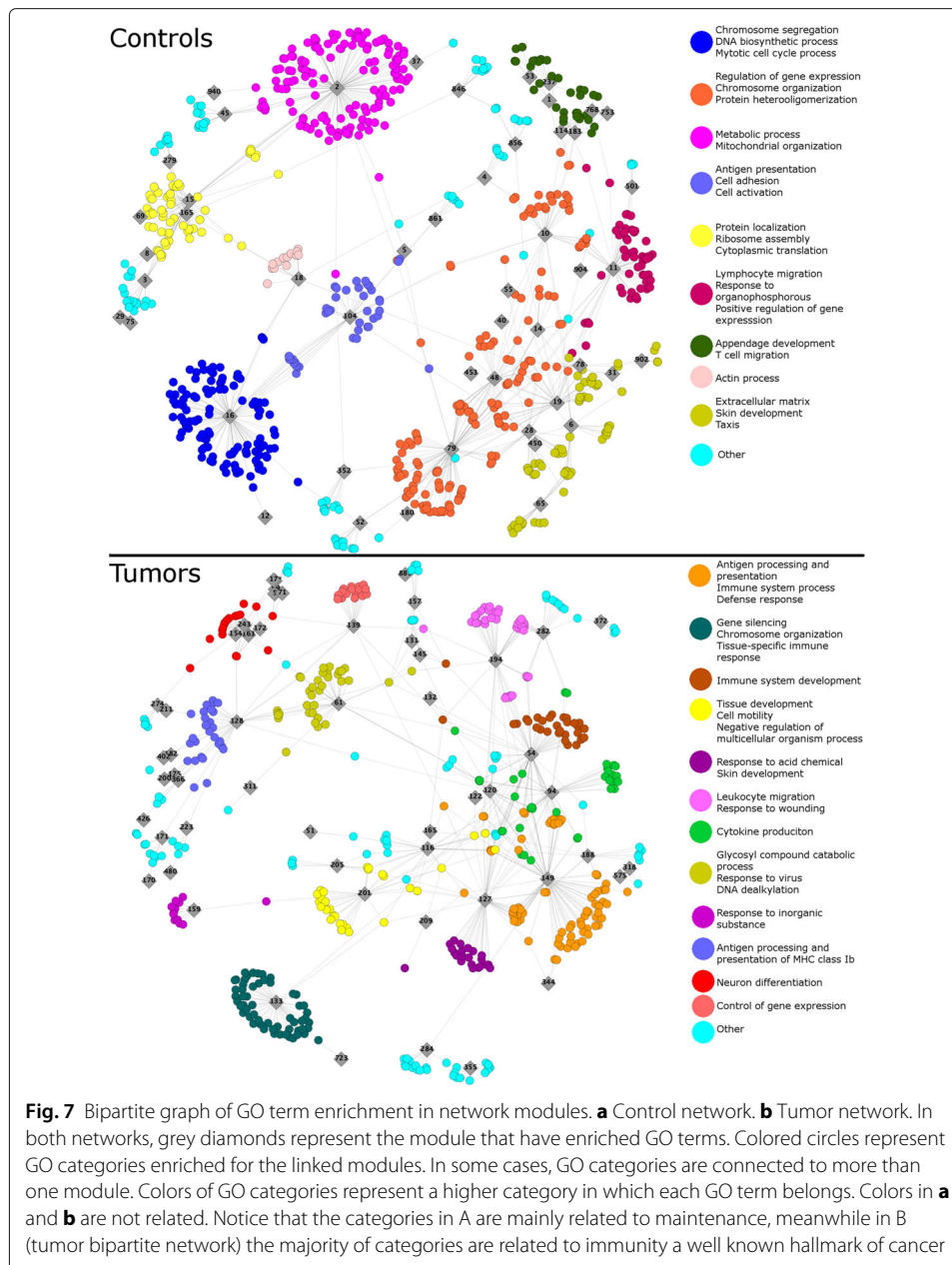

Fig. 4 Histograms of module sizes in tumor and control networks. As it can be observed, the largest modules correspond to the control network (black dots in the upper left part the figure). Also notice the different concavities in red and black curves

Fig. 5 The modular network structure in tumor and control. In this case, nodes are modules and the connections represent inter-module genes connected in the original network. **a** control module network. **b** Tumor module network

networks that may be observed: i) a characteristic degree distribution for each projection (Fig. 6 χ^2 testing for differences in discrete distribution was performed, with the following results: χ^2-statistic $= 49532.28$, $p - value = 4.99e - 10$), ii) the higher edge density in $\mathcal{GP}_c$, which is also related to iii) the higher link/node ratio in $\mathcal{GP}_c$.

Functional Enrichment

We identified a set of biological functions described as GO terms associated to modules detected in the tumor and control networks. We represented these functional associations as bipartite graphs $\mathcal{B}_t$ (for tumors) and $\mathcal{B}_c$ (for controls) that are represented in Fig. 7a and b, with parameters described in Additional file 4.

We identified 665 GO terms associated to $\mathcal{M}_t$ and 827 GO terms associated to $\mathcal{M}_c$. It is important to notice that not all modules were enriched in biological processes; in fact,
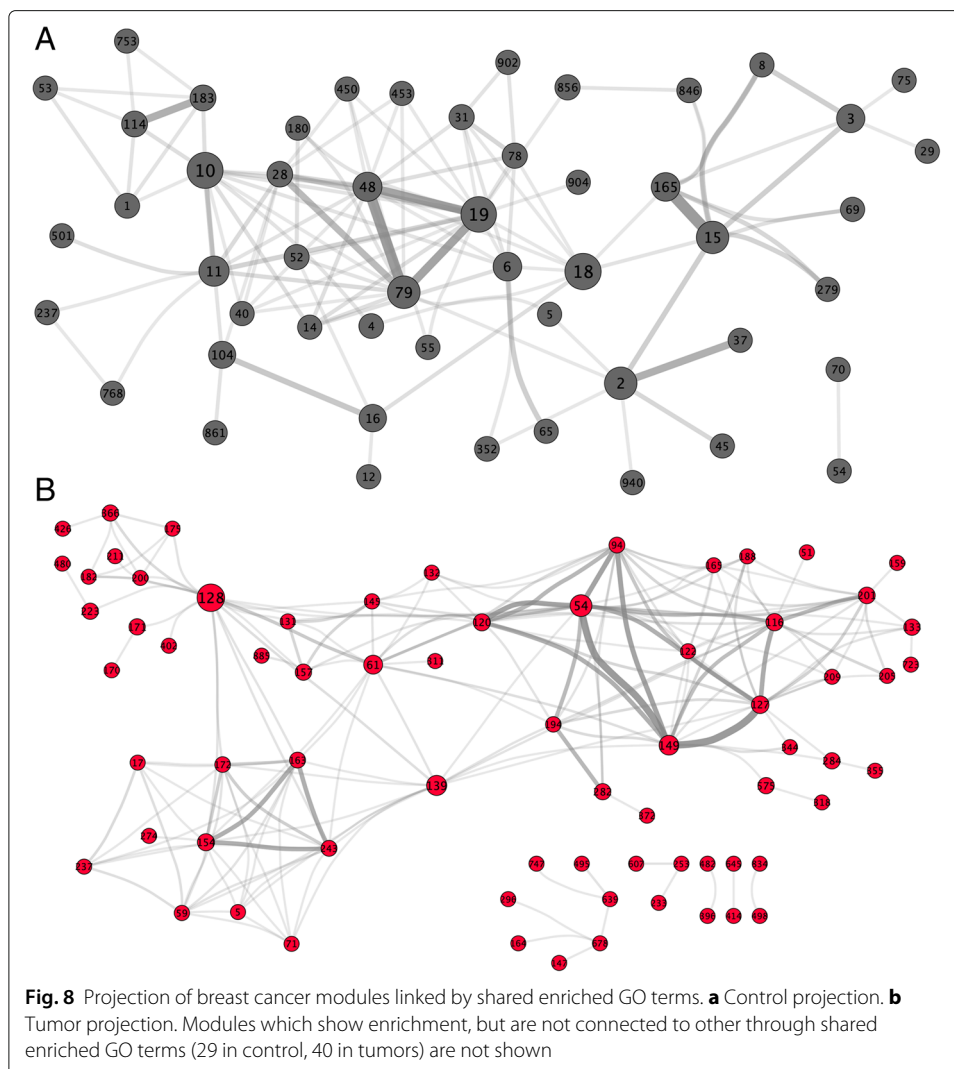

Fig. 6 Degree distributions for module networks of Fig. 5. Red dots represent the tumor network, meanwhile black dots are for controls

Fig. 7 Bipartite graph of GO term enrichment in network modules. **a** Control network. **b** Tumor network. In both networks, grey diamonds represent the module that have enriched GO terms. Colored circles represent GO categories enriched for the linked modules. In some cases, GO categories are connected to more than one module. Colors of GO categories represent a higher category in which each GO term belongs. Colors in **a** and **b** are not related. Notice that the categories in A are mainly related to maintenance, meanwhile in B (tumor bipartite network) the majority of categories are related to immunity a well known hallmark of cancer

only 110 enriched modules are found in $\mathcal{B}_t$ and 82 enriched modules were found in $\mathcal{B}_c$. Furthermore, the set of enriched GO terms $\mathcal{B}_c$ and $\mathcal{B}_t$ are different (with a Jaccard index of 0.34).

The projections of modules based on functional adjacency $\mathcal{B}P_c$ and $\mathcal{B}P_t$ are shown in Fig. 8a and b. In the figure, modules are connected if they share at least one enriched process. Node size represents the module degree. Edge width is proportional to the number of shared enriched processes between modules. In both cases there are some modules that share several enriched processes. In $\mathcal{B}P_t$ (8b), there are clusters of modules sharing GO terms, whereas in 8A the compartmentalization is less evident.

Additional file 5 shows some of the relevant parameters for these projections. It may be observed that these projections are very sparse in terms of edges: only 51 of $\mathcal{M}_c$ are

Fig. 8 Projection of breast cancer modules linked by shared enriched GO terms. **a** Control projection. **b** Tumor projection. Modules which show enrichment, but are not connected to other through shared enriched GO terms (29 in control, 40 in tumors) are not shown

connected to other modules, whereas for $\mathcal{M}_t$ the number is 70. Importantly, there are modules in tumor and control networks (40 and 29, respectively) that are associated to GO terms not shared with any other module.

Discussion

Most central modules in the GP projection are the largest ones

The most central modules in both the $\mathcal{GP}_c$ and $\mathcal{GP}_t$ projections are also the largest ones. In $\mathcal{GP}_t$, this central module has 231 genes and 5437 intra-modular links. It is connected to 99 other modules. The most central module in $\mathcal{GP}_c$, has 1000 genes and 17,583 intra-modular links. It is connected to 742 other modules.

Interestingly such highly central modules are not particularly notable in terms of their functional associations. In controls, the largest module is enriched in 6 processes of nucleic acid regulation; it is linked through processes (i.e. in $\mathcal{BP}_c$) to 6 other modules. For tumors, the largest module shows no statistically significant enrichment, and therefore is not linked to any other module in the $\mathcal{BP}_t$ projection.

Functional compartmentalization in health and disease

The bipartite graphs $\mathcal{B}$ are topologically similar between tumor and control; however, the enriched functions in each network are different. In both cases, the structures show star-like motives (Fig. 7a and b), which indicate mostly unique processes associated to a given gene module. We interpret this as evidence of compartmentalization of regulation, where each module is controlling the activity of independent sets of biological processes.

   We observe important differences in terms of the biological processes associated to the most connected (i.e., most enriched) modules. The two most connected modules (with 146 and 121 neighbors, respectively) in $\mathcal{B}_c$ are associated to metabolism and cell cycle processes, as illustrated Fig. 7a as well as in Additional files 6 and 7; meanwhile, immunity-related processes are associated for the two most connected modules (with 95 and 81 neighbors, respectively) in $\mathcal{B}_t$, which we illustrate in Additional files 8 and 9.

   As it may be observed, associated processes in $\mathcal{B}_c$ are for maintenance, meanwhile the processes associated to the $\mathcal{B}_t$ are well-known *hallmarks of cancer* (Hanahan and Weinberg 2011). The identification of *hallmark* processes in breast cancer co-expression networks derived from high-throughput data is consistent with recent reports by our own group.

Most connected modules through functional adjacency are similar in health and disease

The way modules are connected through processes is similar between health and disease, even though the modules and functions are different. The most enriched modules are not, however, the ones that are more connected to other modules in terms of functional adjacency. These are, as seen in the $\mathcal{BP}$ projections for both controls or tumors, of comparable sizes: 86 and 74 genes, with 356 and 361 intra-modular links respectively. In controls, this module is enriched in 20 processes. Through these processes, it is connected to 18 modules. It is also connected through co-expression links, as seen in the $\mathcal{BP}_c$ projection, to 123 other modules. Meanwhile, the comparable module in tumors is enriched in 81 different processes, but through these is linked to only 20 other modules. Through co-expression links, it is connected to 123 other modules. Interestingly, again there is little overlap in the processes associated to these modules, sharing only one function, *Membrane protein complex*, a general homeostatic event.

Connections between modules through functional and topological adjacency are seldom found

By comparing the set of links in the $\mathcal{GP}$ and $\mathcal{BP}$ projections, we may observe that there are very few links between modules appearing in both projection. In the case of tumors, $\mathcal{GP}_t$ and $\mathcal{BP}_t$ have 37 shared links (Additional file 9), whereas in controls, $\mathcal{GP}_t$ and $\mathcal{BP}_t$ have 51 shared links. As such, we may observe that both in health and disease, the connectivity patterns among gene modules in terms of co-expression and functionality are quite different.

Conclusions

Networks of gene regulation are known to exhibit a modular behavior. The co-expression of gene modules is a form in which cellular processes are regulated. In this work, we demonstrate that modules in transcriptional co-expression networks have different ways to interact, either through co-expression or through jointly regulating functional processes. There are instances in which modules are connected both transcriptionally and

functionally, but these are rare. transcriptional co-expression networks of cancer have a more modular structure than those found in health. Modules found in the health network have higher degrees, whereas modules in the breast cancer network are less likely to have transcriptional relationships to other modules.

We observe that the set of biological functions associated to gene modules are vastly different in breast cancer and health, with gene modules of cancer associated to functions that drive disease, whereas gene modules in health are linked to functions associated to the maintenance of homeostasis. However, we may observe that the connectivity patterns formed by associations of gene modules and biological functions are similar in both health and disease, which indicates that compartmentalization of functional regulation through gene expression remains, even though the processes that are being regulated change.

The behaviors in terms of transcriptional and functional connectivity that gene modules in transcriptional co-expression networks exhibit, may allow for the identification of important modules in terms of either transcriptional, or functional, importance associated to biological conditions of importance, such as cancer.

Available code

All the code used for the present work is available in our repository: https://github.com/guillermodeandajauregui/ModuleEnrichmentAndProjection.

Additional files

Additional file 1: Supplementary Data. Includes two TXT files with tables of the analysis parameters for both tumors and controls at different MI thresholds. The files include the threshold quantile, number of nodes, edges, components, and largest component sizes for each quantile threshold tested. (TXT 478 b)

Additional file 2: Supplementary Data. Excel file with Topological parameters for co-expression networks (TXT 478 b)

Additional file 3: Supplementary Data. Excel file with Topological parameters for topologically adjacent module networks (XLSX 34 kb)

Additional file 4: Supplementary Data. Excel file with Topological parameters for bipartite networks (XLSX 465 kb)

Additional file 5: Supplementary Data. Excel file with Topological parameters for functionally adjacent networks (PNG 630 kb)

Additional file 6: Bipartite control network for cell cycle-related processes. (PNG 413 kb)

Additional file 7: Bipartite control network for metabolism-related processes. (PNG 488 kb)

Additional file 8: Bipartite tumor network for immunity and signaling-related processes. (XLSX 11 kb)

Additional file 9: Bipartite tumor network for immune system-related processes. (PDF 208 kb)

Author details
[1] Computational Genomics, National Institute of Genomic Medicine, Periférico Sur 4809, Ciudad de México, México. [2] Department of Ecology & Evolution, University of Chicago, 1103 East 57th St., 60637 Chicago, IL, USA. [3] Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Ciudad Universitaria, Ciudad de México, México.

References
Alcalá-Corona SA, de Anda-Jáuregui G, Espinal-Enríquez J, Hernández-Lemus E (2017) Network modularity in breast cancer molecular subtypes. Front Physiol 8:915
Alcalá-Corona SA, de Anda-Jáuregui G, Espinal-Enriquez J, Tovar H, Hernández-Lemus E (2018) Network modularity and hierarchical structure in breast cancer molecular subtypes. In: Int Conf Compl Syst. Springer Nature, Cham. pp 352–358
Alcalá-Corona SA, Espinal-Enríquez J, De Anda Jáuregui G, Hernández-Lemus E (2018) The hierarchical modular structure of her2+ breast cancer network. Front Physiol 9:1423
Alcalá-Corona SA, Velázquez-Caldelas TE, Espinal-Enríquez J, Hernández-Lemus E (2016) Community structure reveals biologically functional modules in mef2c transcriptional regulatory network. Front Physiol 7:184
Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. (2000) Gene ontology: tool for the unification of biology. Nat Genet 25(1):25
Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A (2005) Reverse engineering of regulatory networks in human b cells. Nat Genet 37(4):382
Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B (Methodological):289–300
Chow C, Liu C (1968) Approximating discrete probability distributions with dependence trees. IEEE Trans Inf Theory 14(3):462–467
Dang CV, O'Donnell KA, Zeller KI, Nguyen T, Osthus RC, Li F (2006) The c-myc target gene network. Semin Cancer Biol 16:253–264. Elsevier
de Anda-Jáuregui G, Espinal-Enriquez J, Hernández-Lemus E (2019) Spatial Organization of the Gene Regulatory Program: An Information Theoretical Approach to Breast Cancer Transcriptomics. Entropy, 21 195(2):1–11
de Anda-Jáuregui G, Velázquez-Caldelas TE, Espinal-Enríquez J, Hernández-Lemus E (2016) Transcriptional network architecture of breast cancer molecular subtypes. Front Physiol 7:568
De Craene B, Berx G (2013) Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer 13(2):97
Delgado FM, Gómez-Vela F (2018) Computational methods for gene regulatory networks reconstruction and analysis: A review. Artificial intelligence in medicine 95:133–145
Dobruschin P (1968) The description of a random field by means of conditional probabilities and conditions of its regularity. Theory Probab Appl 13(2):197–224
Espinal-Enriquez J, Fresno C, Anda-Jáuregui G, Hernandez-Lemus E (2017) Rna-seq based genome-wide analysis reveals loss of inter-chromosomal regulation in breast cancer. Sci Rep 7(1):1760
García-Campos MA, Espinal-Enríquez J, Hernández-Lemus E (2015) Pathway analysis: state of the art. Front Physiol 6:383
Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821–7826
Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144(5):646–674
Hernández-Lemus E, Rangel-Escareño C (2011) The role of information theory in gene regulatory network inference. In: Deloumeaux P, Gorzalka JD (eds). Information Theory:New Research. Mathematics Research Development series, Nova Publishing, New York
Hernández-Lemus E, Siqueiros-García JM (2013) Information theoretical methods for complex network structure reconstruction. Compl Adap Syst Model 1(1):8
Kullback S, Leibler RA (1951) On information and sufficiency. Ann Math Stat 22(1):79–86
Kuzmanovski V, Todorovski L, Džeroski S (2018) Extensive evaluation of the generalized relevance network approach to inferring gene regulatory networks. GigaScience 7(11):118
Lancichinetti A, Fortunato S, Kertész J (2009) Detecting the overlapping and hierarchical community structure in complex networks. New J Phys 11(3):033015
Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. Phys Rev E 78(4):046110
Liu W, Li L, Long X, You W, Zhong Y, Wang M, Tao H, Lin S, He H (2018) Construction and analysis of gene co-expression networks in escherichia coli. Cells 7(3):19
Manem V, Adam GA, Gruosso T, Gigoux M, Bertos N, Park M, Haibe-Kains B (2018) CrosstalkNet: A visualization tool for differential co-expression networks and communities. Cancer Res 78(8):2140–2143
Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A (2006) Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7:7
Network CGA, et al. (2012) Comprehensive molecular portraits of human breast tumours. Nature 490(7418):61
Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci 103(23):8577–8582
Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. Nature 435(7043):814
Palla G, Farkas IJ, Pollner P, Derényi I, Vicsek T (2007) Directed network modules. New J Phys 9(6):186
Rosvall M, Axelsson D, Bergstrom CT (2009) The map equation. Eur Phys J Spec Top 178(1):13–23

Rosvall M, Bergstrom CT (2007) An information-theoretic framework for resolving community structure in complex networks. Proc Nat Acad Sci 104(18):7327–7331

Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Nat Acad Sci 105(4):1118–1123

Wang X, Terfve C, Rose JC, Markowetz F (2011) Htsanalyzer: an r/bioconductor package for integrated network analysis of high-throughput screens. Bioinformatics 27(6):879–880

Wong DC, Ariani P, Castellarin S, Polverari A, Vandelle E (2018) Co-expression network analysis and cis-regulatory element enrichment determine putative functions and regulatory mechanisms of grapevine atl e3 ubiquitin ligases. Sci Rep 8(1):3151