Loss of long-range co-expression is a common trait in cancer

Diana García-Cortés[1,3], Enrique Hernández-Lemus[1,2], and Jesús Espinal Enríquez[1,2,*]

[1]National Institute of Genomic Medicine
[2]Centro de Ciencias de la Complejidad. Universidad Nacional Autónoma de México
[3]Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México

Cancer cells display common traits and enabling-characteristics previously described as the Hallmarks of Cancer. These occur alongside alterations in the regulatory mechanisms controlling gene transcription. Gene co-expression networks (GCNs) identify correlated sets of genes that might share these mechanisms. We have previously reported loss of long-range co-expression for breast, lung, and kidney cancer GCNs. Here, we extend the study to fifteen tissues. Unlike in healthy phenotypes, the highest cancer gene-pair interactions are intra-chromosomal and their strength decays with base-pair distance. Communities in tumor GCNs are strongly associated with cancer related processes along with a wide presence of adaptive immune response. Riboproteins are highly co-expressed in cancer and normal GCNs, suggesting their relevance for cell viability. The loss of long-range co-expression is not observed in other chronic diseases such as Type-2 Diabetes or Alzheimer's disease. These results suggest that the loss of long-range co-expression is a common trait in cancer.

gene co-expression networks | cancer networks | loss of long-range co-expression | common traits in cancer

Correspondence: ***jespinal@inmegen.gob.mx***

Introduction

A number of distinctive shared characteristics acquired by cancer cells during progression towards tumorigenesis have been compiled over the course of two decades by Hanahan and Weinberg, giving birth to the paradigmatic Hallmarks of Cancer (1). This evolving conceptual framework has allowed us to encompass the multiple alterations that take place within this complex disease (2–4). The fundamental principles, common to most neoplastic malignancies, include: sustaining proliferative signaling, evading growth suppressors, avoiding immune destruction, enabling replicative immortality, activating invasion and metastasis, inducing or accessing vasculature, resisting cell death, and deregulating cellular metabolism, with two more recently added: unlocking phenotypic plasticity, and senescent cells (1, 5, 6).

These authors have also now identified four mechanisms that aid in the establishment of said hallmarks, called enabling characteristics: tumor-promoting inflammation, genome instability and mutation, non-mutational epigenetic reprogramming, and polymorphic microbiomes (5, 6). The disruptions detailed by these enabling characteristics achieve their realization into an oncogenic phenotype manifesting said functional hallmarks, mainly through the process of gene transcription.

Transcriptional regulation, the series of processes controlling gene transcription, requires an intricate synergy between multiple regulatory mechanisms, including cis-regulatory elements, specialized proteins such as transcription factors, and epigenetic markers operating in the context of a specific chromatin configuration (7, 8). The outcome of this regulatory landscape and its altered states in cancer can be studied by examining gene expression. Gene expression is one of the measurable outcomes of the transcription process and as such, it serves as an input to other functional and statistical studies like gene co-expression analysis (9).

The analysis of gene co-expression profiles facilitates the identification of highly correlated sets of genes (10) and it might provide insights into altered regulatory mechanisms that disrupt the expression of gene sets in a coordinated manner (11). Gene co-expression profiles are often analyzed by calculating the correlation between pairs of genes in a gene expression matrix of multiple samples from a particular phenotype. The most significant pairs of co-expression values can be extracted to construct a gene co-expression network (GCN). GCNs provide a topological characterization of the transcriptome, allowing us to compare not only global and local connectivity properties from different phenotypes, but also their functional implications (12–14).

By analyzing gene co-expression profiles and building GCNs using RNA-Seq data in cancer tissue and its normal counterpart, we have previously described a loss in the proportion of inter-chromosomal gene co-expression interactions in breast cancer and breast cancer molecular subtypes (15–18), lung cancer (19), and clear cell renal carcinoma (20). In these tissues, the majority of significant co-expression links connect gene pairs in the same chromosome and these display a highly localized co-expression pattern. Conversely, their normal tissue-derived networks, present an elevated number of inter-chromosomal interactions.

Local high co-expression is a characteristic that has been reported for normal tissues, both in bulk RNA-Seq (21) and single cell analysis (22), reinforcing the conclusion that gene order in eukaryotes is not random (23, 24). However, this phenomenon has been observed in gene pairs within a local vicinity of less than 1Mb (25, 26). Our previous results indicate that in cancer, this high co-expression expands to longer distances and its prevalence throughout different tissues suggests it may be a common trait in the disease. Moreover, in terms of the regulatory mechanisms contributing to the emergence of high co-expression regions, we did not find a decisive association with a single mechanism in the Luminal A

breast cancer subtype (27). Indeed, in normal tissues, a similar pattern is displayed, with multiple factors contributing to local co-expression (21). These results, in both cancer and normal phenotypes, demand the need to extend the analysis of gene co-expression and the mechanisms that jointly drive the transcription of multiple sets of genes.

In this work, we simultaneously evaluated the loss of long-range co-expression in fifteen tissues (8,772 samples) by integrating data from distinct data sources: TCGA (28) and UCSC Xena (29) (which includes samples from the GTEx and TCGA collaborations (30)). This in order to determine whether (i) the loss of inter-chromosomal co-expression is a common trait in cancer, (ii) there is a dependency between high co-expression and physical distance between genes located in the same chromosome, and (iii) whether these phenomena are exclusively found in cancer. To do this, we analyzed gene co-expression profiles from said samples, localizing shared regions of high-density co-expression. We also built GCNs and compared their topological features through network community extraction and identified biologically associated communities. This analysis allowed us to confirm that loss of long-range co-expression is indeed a common trait in cancer.

Results

In cancer, highly co-expressed gene pairs are intra-chromosomal. Mutual Information (MI) was calculated between pairs of genes, as a measure of co-expression, for each tissue and sample type independently. The total number of gene pairs studied ranges from 46.442 million in skin to 102.95 million in lung, due to the tissue-wise pre-processing filters applied.

Gene pair datasets have an average percentage of 5.33% of intra-chromosomal interactions. When different thresholds of top MI values are considered, this percentage remains stable in normal tissues. However, higher fractions of intra-chromosomal interactions constitute the top MI values in cancer tissues. Figure 1 presents a heatmap displaying the fraction of intra-chromosomal interactions for different thresholds of top MI values in cancer and normal tissues. For bladder samples, the same values are also presented as a line-plot. Supplementary Figure 1 contains line-plots for the remaining tissues.

Kolmogorov-Smirnov tests indicate that the distribution of intra-chromosomal fractions in cancer and normal tissues are significantly different, with p-values ranging from 1e-16 to 0.005 (Supplementary Material 2) when the top 1 million interactions are considered. This observation indicates that in all analyzed cancer types, the strongest co-expressed gene pairs belong to the same chromosome, whereas in normal tissues there is no inclination towards high co-expression intra-chromosomal interactions.

Distance dependent MI decay is a common feature in cancer. After showing that the strongest co-expression interactions in cancer occur between genes in the same chromosomes, we observed whether correlation values were

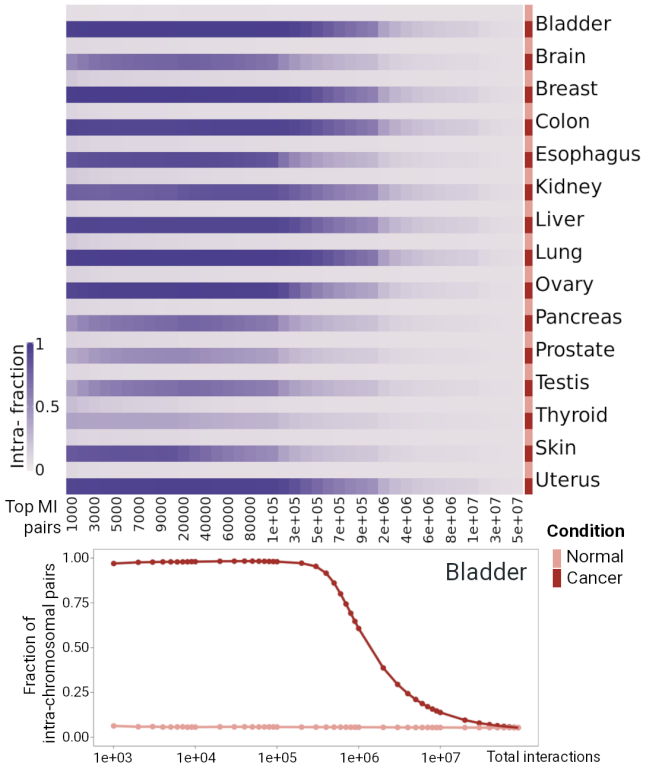

Fig. 1. Heatmap of fractions of intra-chromosomal interactions at different top mutual information (MI) thresholds for cancer and normal tissues for all fifteen datasets analyzed. In normal co-expression profiles this fraction remains stable across thresholds while, for cancer, the top MI interactions are mostly intra-chromosomal. For bladder, the same fractions for both normal and cancer phenotypes are shown in a line plot.

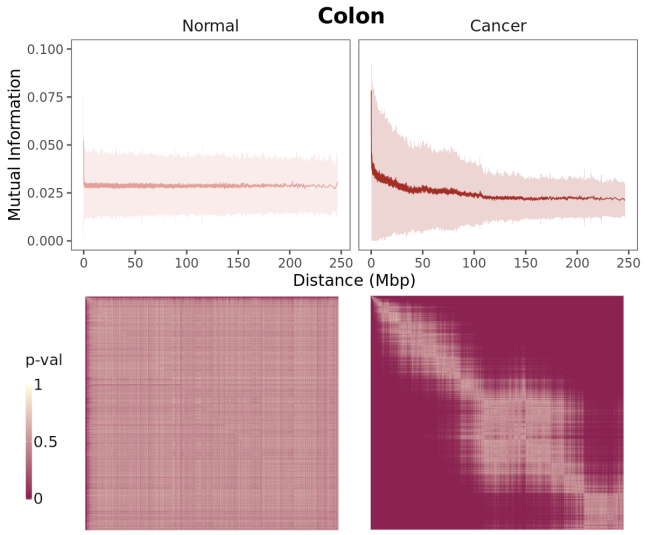

Fig. 2. Average Mutual Information and standard deviation for bins of one thousand intra-chromosomal interactions plotted against average gene-gene base pair distance in colon tissue. Lower panel displays p-values for Wilcoxon tests comparing distributions of MI values between bins. Normal tissue shows no difference in bins associated with distance while in cancer, distant bins are significantly different, but neighboring bins present similar values.

dependent on the chromosomal distance between intra-chromosome genes. We evaluated intra-chromosomal co-expression by sorting gene pair interactions in terms of base pairs distance. Figure 2 presents the average MI values for 1,000-interaction bins, with their corresponding standard deviations plotted versus distance in colon tissue. Supplementary Figure 3 does this for all fifteen tissues in the study.

In cancer, gene pairs with shorter distances show higher MI values and as distance increases there is a decay in co-expression that reaches a plateau. Conversely, in normal intra-chromosomal co-expression profiles, only genes in a very close proximity present the same behavior. Wilcoxon tests were performed to compare the distribution of MI values among bins. P-values, displayed for colon cancer in the lower panel of Figure 2 and in Supplementary Figure 4, show that, in cancer, neighboring bins are similar while distant bins are significantly different. On the contrary, normal-derived distribution of bins does not present differences associated with base pair distance. Therefore, not only the strongest co-expressed genes belong to the same chromosome, but they are also placed in close proximity.

High-density intra-cytoband regions are shared between cancer tissues. In order to identify regions with highly co-expressed intra-chromosomal interactions, we selected the cytobands that reached at least half of their total possible interactions at the top 100,000 MI values. Figure 3 presents high-density cytobands shared among at least five tissues.

In normal samples, there is only one region with those characteristics: region q11.221 in chromosome Y. This region contains five genes including DDX3Y, an RNA helicase; NLGN4Y, member of the family of neuroligins; and a peptidase, USP9Y, among others. In cancer, there are up to five high-density cytobands within a single chromosome. Thyroid is the only tissue where no hotspot cytobands were identified at this level of resolution. Chromosome 8 contains three contiguous cytobands with highly co-expressed interactions (p11.23, p11.22 and p11.21) in breast, liver, and lung cancer. The region containing these cytobands has been associated with an amplification in several cancers, with proposed driver genes including ZNF703, FGFR1 and PLPP (31), as well as a breakout region associated with ADAM32 (32). The most shared hotspot region among cancer tissues is Chromosome 1q42.11, present in thirteen (out of fifteen considered) tissues. Additionally, cytoband q21.33 in chromosome 13 is a hotspot in ten different tissues. To our knowledge, these two regions have not been previously associated with cancer.

It is worth noting that, when compared to a null model that randomly places intra-chromosomal interactions according to the real number of links found inside each chromosome, all regions displayed in Figure 3, achieved a p-value $< 1e^{-16}$ (Supplementary Material 5). This result indicates that high co-expression interactions placed in close proximity are aggregated as shared neighborhoods in multiple cancer types, and cytobands serve as structural units.

Networks. The top 100,000 MI interactions from normal and cancer tissues were used to build co-expression networks. Table 1 presents their main features, while Figure 4 displays a visual representation of networks for selected tissues. Supplementary Figure 6 displays networks for all tissues. As stated before, at this cutoff the majority of interactions in cancer join gene pairs within the same chromosomes, while normal networks resemble what we had previously observed in breast and lung cancer: a giant component with mostly inter-chromosomal links (15, 17, 19).

By considering only the fraction of intra-chromosomal links in cancer networks (Table 1), two groups can be identified. Networks from bladder, breast, esophagus, kidney, liver, lung, ovary, and uterus, with more than 83% of intra-chromosomal interactions and networks from brain, pancreas, prostate, skin, testis, and thyroid, with less than 66%. The structure of these two groups is visually different when a force directed layout is applied for network visualization (Supplementary Figure 6). In normal co-expression networks there is no such division. The mean fraction of intra-chromosomal interactions in cancer co-expression networks is 0.7582 ± 0.2327, while networks from normal tissues have an average fraction of 0.0716 ± 0.0085.

Links are mostly unique in normal co-expression networks, but not in cancer. To compare co-expression networks further, a hierarchical clustering was performed, taking into account the tissues where each link appears. Sets with more than 1000 shared links are shown in Figure 5A.

Unsupervised hierarchical clustering divides cancer tissues into three major clusters. There, ovary and esophagus are placed along the networks with lower fraction of intra-chromosomal links. Additionally, there are two clusters of high intra-chromosomal fraction networks: one includes breast, bladder, lung, and uterus, while the other one is constituted by networks from colon, kidney and liver cancer. The biggest set of shared interactions occurs between normal networks in esophagus and skin.

However, even though normal networks look very similar in Figure 4 and have similar fractions of intra-chromosomal links, they share less interactions among them than cancer networks, as shown in the heatmap (Figure 5A). The number of unique interactions per network in the horizontal bar plots are shown in Figure 5B. The fifteen normal networks have an average number of unique interactions of 83,642 ($\pm 10,362$). Vertical bar plots in Figure 5B display the twenty sets with the highest number of shared links. In accordance with their composition, normal co-expression networks share mostly inter-chromosomal interactions, while cancer networks share same-chromosome links. Lung, breast, and bladder cancer are the three cancer co-expression networks with the lowest number of unique interactions (5,621, 7,894 and 8,607 respectively), while the thyroid cancer network presents the largest number of unique interactions with 67,077.

There is an aspect associated with the source of the data that is highlighted in the vertical bar plots: the top 14th sets of shared interactions in the normal phenotype belong exclusively to tissues from the UCSC Xena dataset. This situa-

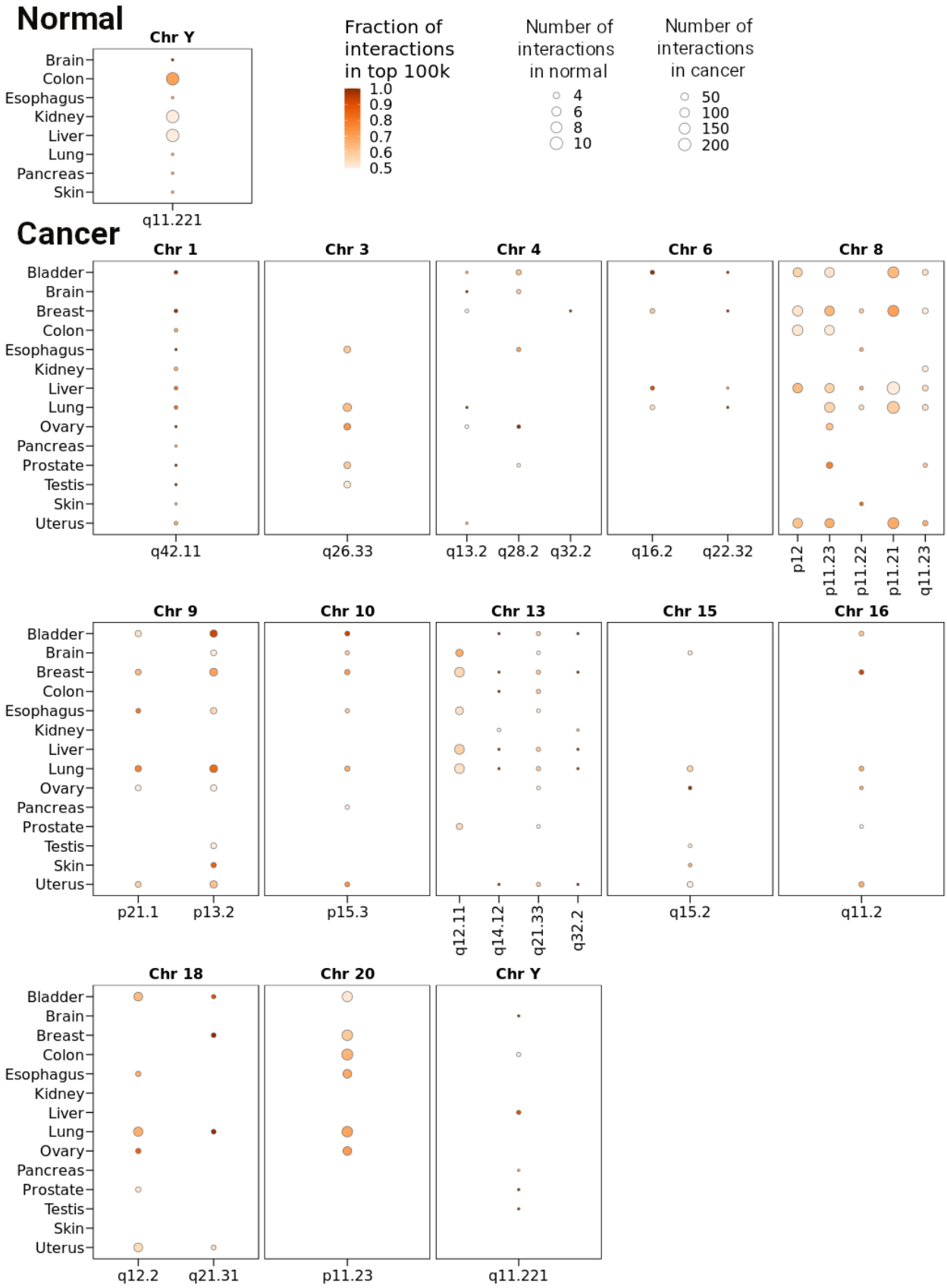

Fig. 3. High-density intra-cytobands shared in tissues. Cytobands with more than 50% of their total possible interactions present in the top-100,000 MI values and shared among, at least, five tissues.

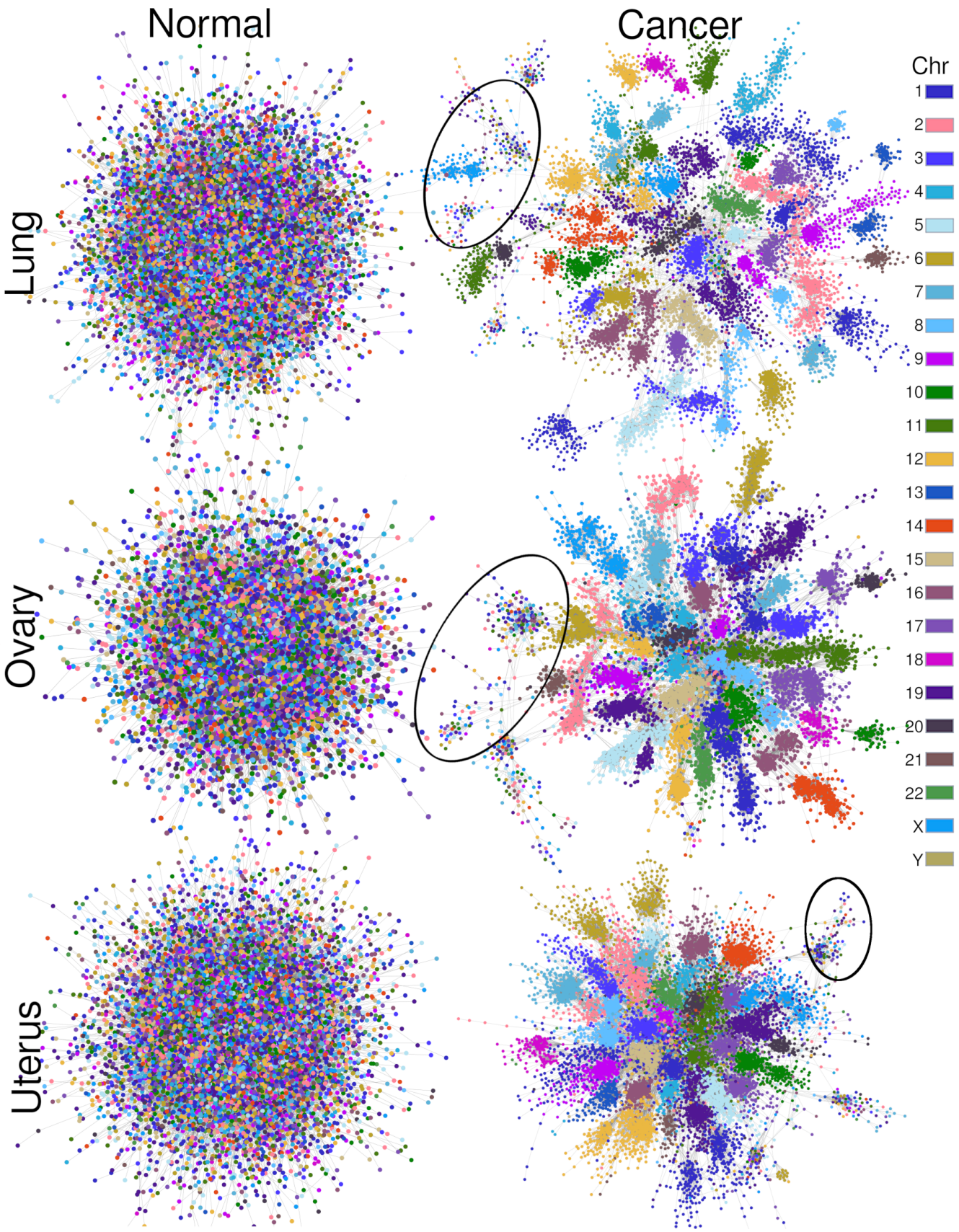

Fig. 4. Co-expression networks from the top one hundred thousand interactions in selected tissues displayed with a force directed layout. In cancer networks, modules concentrating intra-chromosomal interactions appear, while this pattern is not visible in normal networks. Circles identify communities associated with the Gene Ontology process of *adaptive immune response*.

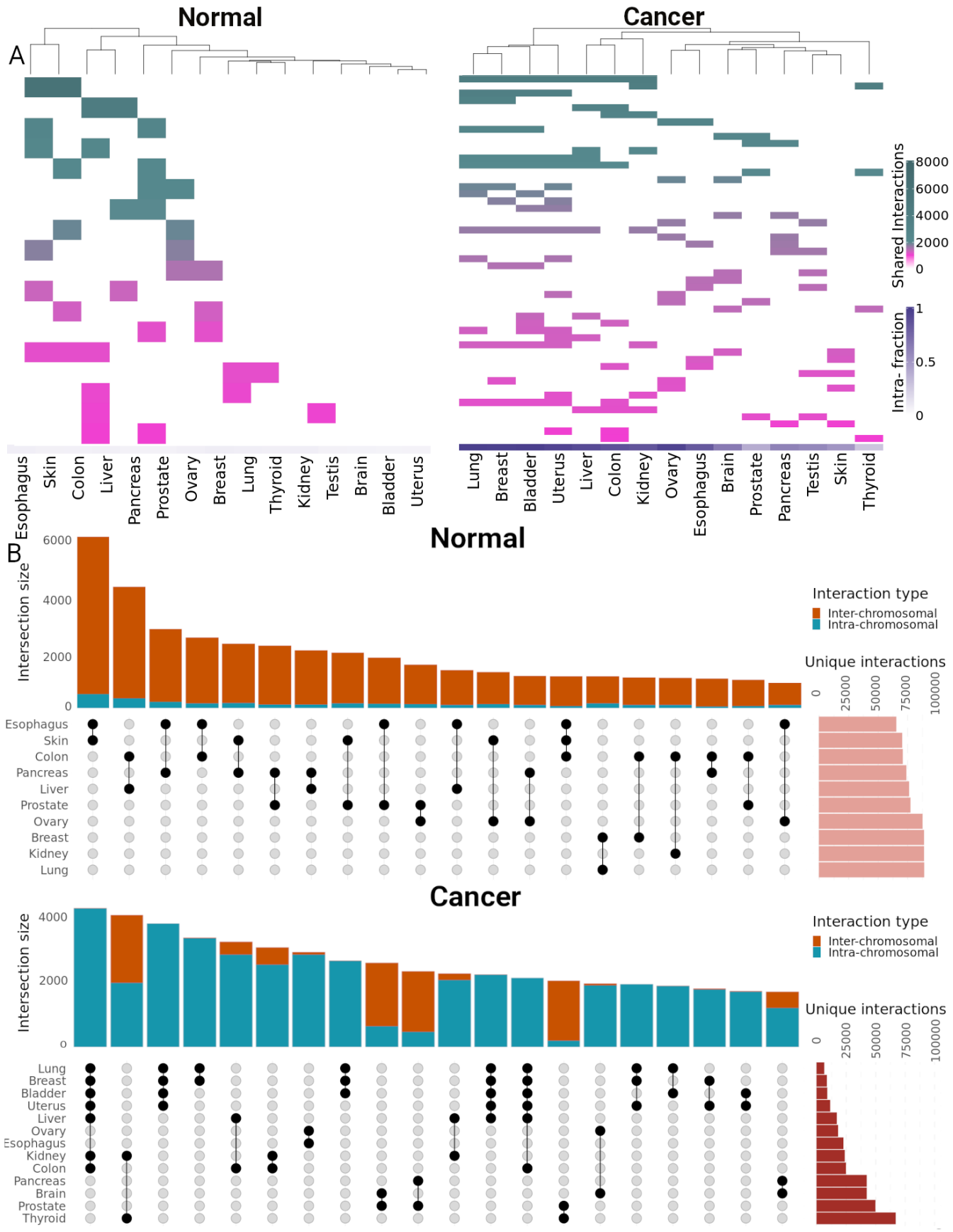

Fig. 5. A Hierarchical clustering of interactions shared among tissues in normal and cancer co-expression networks showing sets with more than one thousand interactions. **B** Bars and upset blots displaying the top twenty sets of shared interactions. Normal co-expression networks have a higher number of unique interactions while cancer networks share mostly intra-chromosomal interactions.

Table 1. Main features of normal and cancer co-expression networks for all fifteen tissues in the study. Networks are built by taking the top 100,000 MI interactions, although they differ in the number of nodes (Genes). The Intra- fraction refers to the fraction of links joining genes in the same chromosome. Communities were computed using the Louvain algorithm (see Methods). Up and Down refers to genes upregulated and downregulated ($pval_{adj} < 0.05$) in the differential gene expression analysis.

Tissue	Condition	Genes	Intra- fraction	Communities	Up	Down
Bladder	Normal	13221	0.05709	34		
	Cancer	9403	0.98007	321	4618	1995
Brain	Normal	8897	0.06249	68		
	Cancer	8738	0.65396	129	4735	3436
Breast	Normal	11902	0.08626	93		
	Cancer	10155	0.97363	423	5571	3435
Colon	Normal	10645	0.07128	176		
	Cancer	8958	0.87639	373	5111	3264
Esophagus	Normal	8325	0.08036	85		
	Cancer	9117	0.83717	139	4580	3856
Kidney	Normal	10955	0.08052	89		
	Cancer	9612	0.85921	394	5168	3486
Liver	Normal	11952	0.07664	144		
	Cancer	9622	0.91591	357	4984	3551
Lung	Normal	12469	0.07914	80		
	Cancer	10116	0.97844	424	6950	2321
Ovary	Normal	9071	0.07302	44		
	Cancer	7729	0.95778	193	3626	3496
Pancreas	Normal	8057	0.06685	97		
	Cancer	9659	0.61381	51	4604	4469
Prostate	Normal	9614	0.06704	26		
	Cancer	8645	0.39908	108	4406	3629
Skin	Normal	7370	0.07258	97		
	Cancer	9841	0.51385	29	5164	4098
Testis	Normal	6825	0.06484	93		
	Cancer	10174	0.58501	73	5605	4128
Thyroid	Normal	14095	0.07634	50		
	Cancer	10259	0.27775	372	4509	4189
Uterus	Normal	11935	0.05935	91		
	Cancer	9956	0.95163	330	4778	3312

tion does not occur in cancer, where the biggest set contains tissues from both sources.

Riboprotein genes constitute a cluster shared by cancer and normal co-expression networks. To identify co-expression patterns that appear in both normal and cancer co-expression networks from multiple tissues, we integrated a common network with the most frequent interactions, defined as those present in > 10 tissues. These networks are displayed in Figure 6. The cancer network has 12,802 links, while there are only 355 edges in the normal network. The intersection of normal and cancer networks corresponds to 92% of genes in the common normal network and 59% of its links. According to the previously discussed nature of cancer networks, the large majority of clusters in the cancer common network are composed of genes from the same chromosome; however, small subsets of nodes are formed by inter-chromosomal interactions.

The largest connected component in the intersection of normal and cancer networks consists of thirty-nine genes, located at the center of the giant component of the cancer network (circled in black in Figure 6). Thirty-seven genes in this cluster encode for molecules in the ribosomal protein family, with members of both the small and large subunits. The remaining two are EEF1B2, involved in the transfer of aminoacylated tRNAs to the ribosome, and NACA, which prevents mistranslocation of aberrant nascent proteins to the endoplasmic reticulum. The presence of this cluster indicates that ribosomal genes retain a pattern of high co-expression among them in all analyzed tissues and in both conditions, cancer and normal.

The community structure of co-expression networks is associated with biological processes. The force directed layout applied to the co-expression networks for network visualization (Figure 4 and Supplementary Figure 6) suggests a modular structure influenced by chromosomal gene location, at least in the cancer co-expression networks with higher fraction values of intra-chromosomal links. Therefore, we performed a community structure analysis and evaluated the results from different algorithms (see Methods). Table 1 indicates the number of communities identified in each network with the Louvain algorithm. For each community, two types of nominal assortativity were calcu-

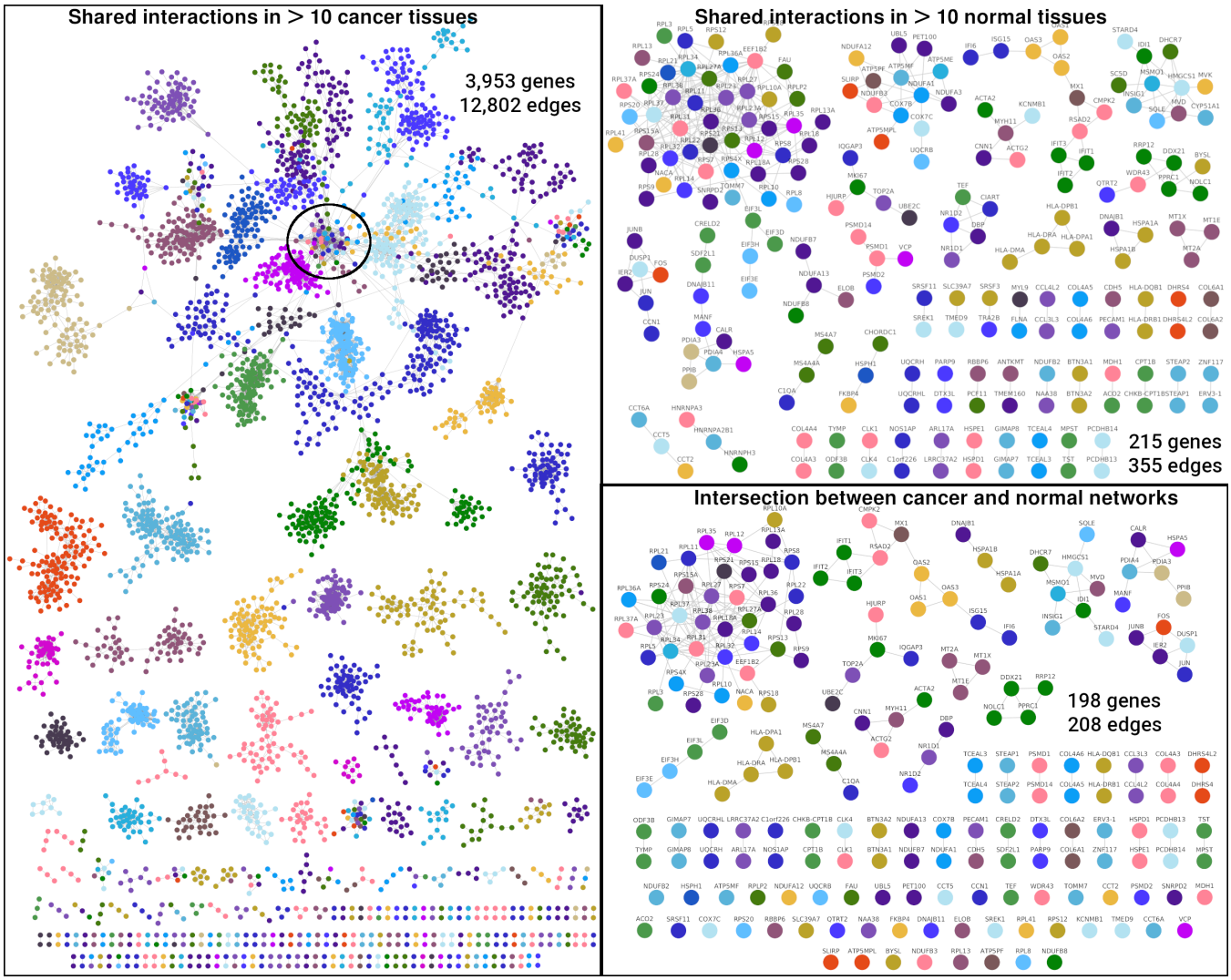

Fig. 6. Cancer and normal common networks integrated with interactions repeated in more than ten tissues as well as the intersection between these two networks, visualized using a force directed layout. Genes encoding ribosomal proteins, joined by inter-chromosomal links, constitute the largest component in the intersection and are enclosed in a black circle at the center of the largest component in the cancer network.

lated: chromosomal assortativity and differential expression assortativity (only in cancer networks). These assortativity metrics allowed us to quantify the community bias towards a single chromosome or a single differential expression fold change direction.

The distributions of chromosomal assortativities for cancer and normal co-expression networks are presented as violin plots in Figure 7. They show that network communities in cancer have a tendency towards 1, meaning that genes in a community belong mostly to one single chromosome. In normal co-expression networks, there is an opposite trend, with chromosomal assortativity values gathering heavily towards -1 in tissues such as bladder, breast, liver, ovary, prostate, thyroid and uterus, and with values partitioned between -1 and 1 in other tissues, thus indicating either link bias towards different chromosomes or mixed trends respectively.

Supplementary Figure 7 contains violin plots for expression assortativity in cancer co-expression networks. Here, there is a higher variation than in those related to chromosomal assortativity. Multiple tissues have communities with an expression assortativity close to 1, meaning that communities have either over-expressed or under-expressed genes. In pancreas and skin, most communities have values around 0, while only esophagus has a similar number of communities with differential expression assortativity of -1, 0 and 1.

We performed Gene Ontology (GO) over-representation analyses in order to identify the biological processes associated with each community in the co-expression networks. Given that the number of communities for each network is large, we decided to be quite stringent in the statistical significance for the enrichment analyses ($p_{adj} < 1e^{-10}$) mentioned above.

The fraction of statistically significant enriched communities is different for each tissue and condition. For all fifteen normal co-expression networks, the total of unique enriched terms is 300, while cancer co-expression networks have 360 different GO terms associated, with 146 terms at the intersection (62% of normal GO terms and 44% of terms in cancer). For normal phenotypes, colon, esophagus, and skin are the top three enriched tissues with 121, 113, and 97 processes,

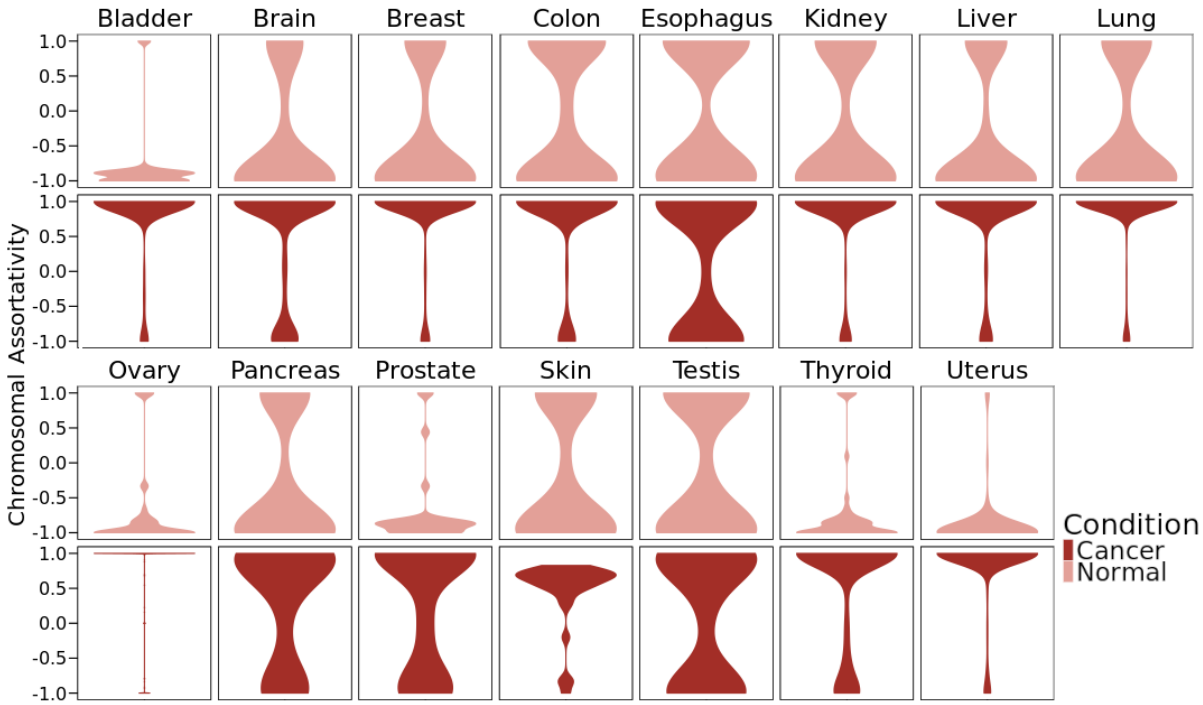

Fig. 7. Chromosomal assortativity in network communities of all fifteen tissues in the study. Normal tissues display a tendency towards -1, meaning that communities are composed of links joining genes in different chromosomes while, in cancer tissues, communities have mostly intra-chromosomal interactions, with chromosomal assortativity values close to 1.

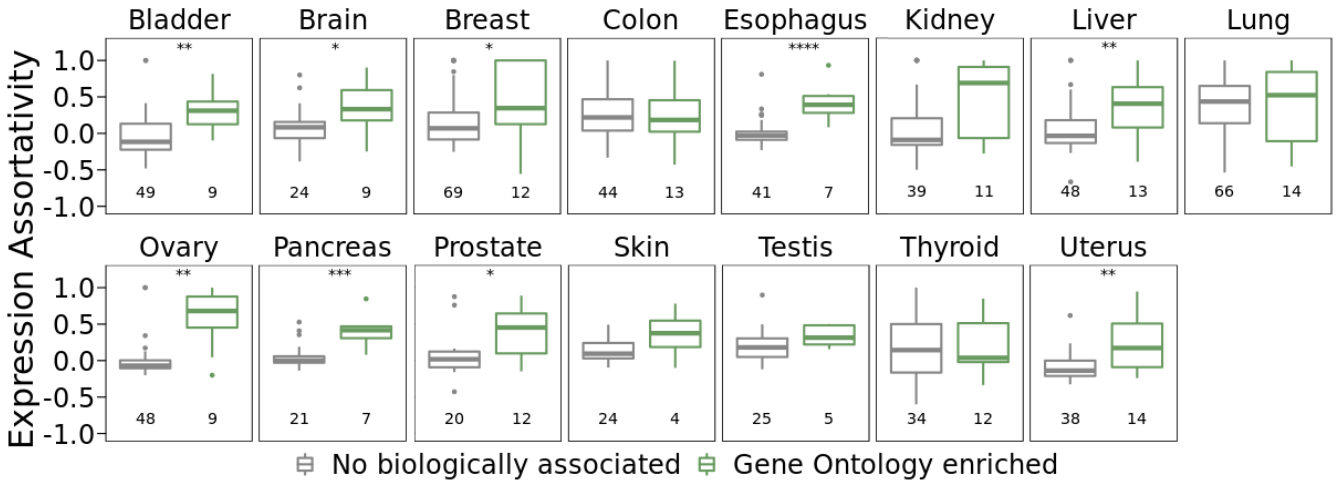

Fig. 8. Differential expression assortativity for both GO enriched and not enriched communities in the cancer co-expression networks. Biologically associated communities tend to have higher values of expression assortativity meaning that genes in communities share an under- or over-expression trend

while bladder and uterus have no statistically significant GO processes associated. In cancer, colon, liver, and thyroid are the tissues with the highest number of enriched terms, with 135, 128, and 127 respectively.

We first observed expression assortativity differences between cancer GO enriched communities and non-biologically associated communities, as shown in Figure 8. In almost all tissues, average expression assortativity in biologically associated communities is higher than assortativity in not enriched communities, suggesting that genes in biologically associated communities are likely to have similar differential expression trends.

For several cancer tissues, chromosomal assortativity, presented for both conditions in Supplementary figure 8, differs between GO enriched and not enriched communities. Not enriched communities display a chromosomal assortativity close to 1, especially in networks with a higher fraction of intra-chromosomal interactions, while enriched communities display a broader range of chromosomal assortativity values, with a preference for negative values. In the normal phenotypes, there is no difference between enriched and not enriched communities in terms of chromosomal assortativity because they all share values close to -1, given that the majority of links in these networks are inter-chromosomal.

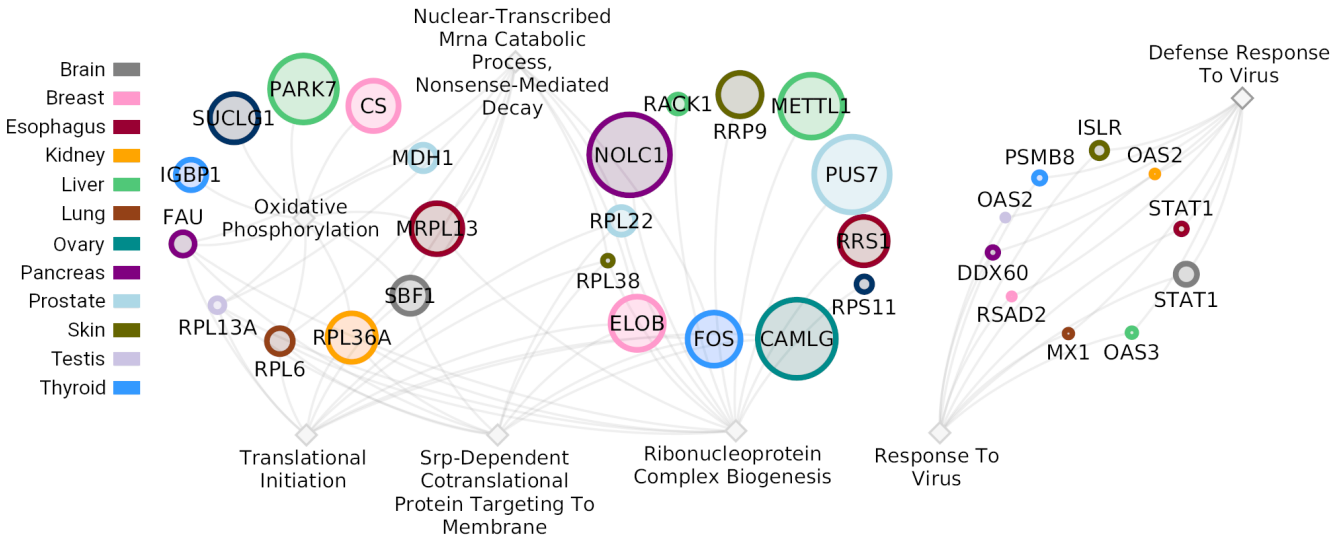

Fig. 9. Bipartite network of the most common Gene Ontology biological processes with a degree greater than ten, joined to their communities in the normal phenotype. Diamonds represent enriched processes and circles represent communities colored according to their tissue.

Shared GO processes in co-expression networks' communities. By associating communities from the fifteen networks with their enriched biological processes, we created a bipartite network for normal and cancer phenotypes and selected the processes with a high degree ($D > 10$), this to obtain the most common GO processes on each phenotype.

In the normal bipartite network (displayed in Figure 9), diamonds represent enriched processes and circles represent communities, colored according to their corresponding tissues. There, two connected components are observed, the biggest one is associated with translation and establishment of protein localization with GO elements such as *ribonucleoprotein complex biogenesis* and *translational initiation*. Several communities in this component have ribosomal protein genes as their highest page rank element (see Methods), like RPL6 in lung or RRP9 in skin.

In cancer (Figure 10), the bipartite subnetwork of common GO processes contains five connected components. One of them resembles the ribosome-associated component in the normal network, with GO processes such as *translational initiation* and *establishment of protein localization to endoplasmic reticulum*, suggesting again a prevalent pattern of high co-expression among ribosomal genes. There are eleven communities in this component, with a mean fraction of intra-chromosomal interactions of 0.4285 ± 0.282.

The HOX genes, with clusters A, B, C, and D identified in mammals, play a role in embryonic development; their expression is bound to spatiotemporal regulation, related to the position of each gene within its cluster (33). In the cancer common GO processes, we found twelve communities with members of the HOXA and the HOXB gene families in bladder, breast, colon, kidney, liver, lung, and uterus cancer (Supplementary Figure 9), associated with processes such as *embryonic skeletal system morphogenesis*. These communities are entirely intra-chromosomal, with genes placed in chromosomes 7 (HOXA) and 17 (HOXB).

HOXA communities present an underexpression trend in breast, colon and uterus cancer. In breast and colon cancer this pattern has been related to aberrant methylation and epigenetic silencing (34, 35), while the underexpression of HOXA10 has been reported in leiomyomas, the most common benign uterine tumor (36). In lung cancer the cluster has a mixed trend, although the overexpression of HOXA10 was found to promote growth and metastasis of lung adenocarcinoma (37) and it was also correlated to the clinical stage of lung squamous cell carcinoma (38).

The overexpression presented by HOXB genes has been reported for multiple cancer types. The increased expression of HOXB9 in uterine cancer was associated with histological grade and lymph node metastasis (39). In colon cancer, overexpression of HOXB genes, in particular HOXB8, has been associated with the activity of a super-enhancer occupied by MYC (40). For liver cancer, a high expression of HOXB7 results in proliferation, migration and cancer progression (41). Finally, the aberrant expression of HOXB genes in breast cancer has been associated with cancer heterogeneity and the process of tumor development (42).

Another module is highly associated with histone-related activities, with GO processes such as *nucleosome assembly*, *protein-dna complex assembly*, and *chromatin organization involved in negative regulation of transcription*. It is composed of thirteen communities with a mean fraction of intra-chromosomal links of 0.6442 ± 0.2228, tending to link genes in chromosome 6. Genes with the highest page-rank belong to the H2A and H2B histone families; however, communities include members from the H3C, H4C, H1 and H4 families. Supplementary Figure 10 presents communities with less than a hundred nodes in this module and their associated differential gene expression. Genes in these communities have an overexpression trend in bladder, brain, breast, esophagus, kidney, liver, lung, ovary, and uterus cancer. Aberrant expression of histone variants has been reported in multiple

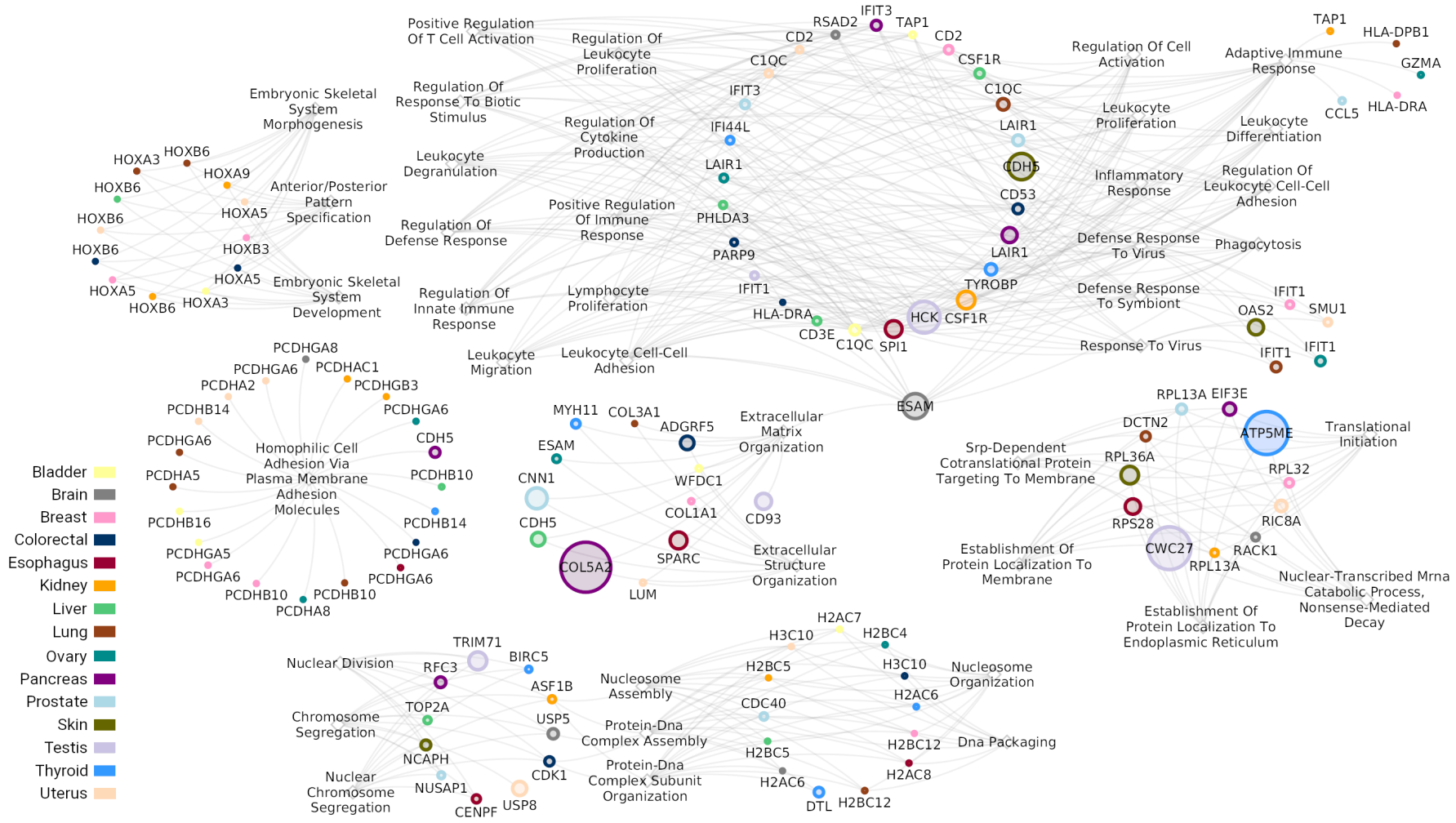

Fig. 10. Bipartite network of the most common Gene Ontology biological processes, with a degree greater than ten, joined to their communities in the cancer phenotype. Diamonds represent enriched processes and circles represent communities colored according to their tissue.

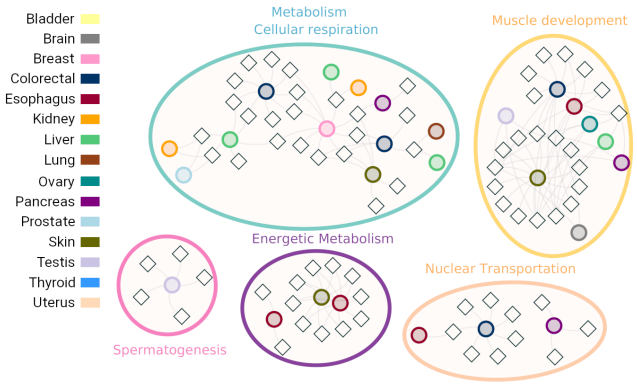

Fig. 11. Bipartite network of Gene Ontology processes and their associated communities found exclusively in the normal phenotype.

cancer types (43). Their upregulated expression in breast cancer has been associated with acquired resistance to treatment and poor overall survival (44, 45).

Genes in the PCDHG and PCDHB protocadherins clusters constitute another component, associated with the GO term *homophilic cell adhesion via plasma membrane adhesion molecules*. This cluster includes communities from twelve cancer types such as bladder, breast, kidney, ovary, and thyroid, among others. Only three out of the twenty communities in this component are not entirely intra-chromosomal. Genes located in chromosome 5 present a pattern of underexpression, specially in bladder, breast, and uterus, as previously reported by our group in Luminal A breast cancer (27). These communities are displayed in Supplementary Figure 11. Low expression of these genes has been associated with epigenetic alterations through a mechanism of long-range epigenetic silencing by hypermethylation in different types of cancer, including breast, colon, liver, and lung, promoting processes such as migration and proliferation (46–48).

Normal- and cancer-only associated processes. The bipartite networks also allow us to obtain the processes associated with either normal co-expression networks or cancer co-expression networks but that do not appear in networks from the other condition. Sets of similar normal-only processes are shown in Figure 11 and cancer-only ones in Figure 12. Again, diamonds represent enriched processes in networks communities ($p_{adj} < 1e^{-10}$), and colored circles represent communities in their corresponding tissues.

In normal networks communities, we have identified five sets. The biggest one: metabolism and cellular respiration, which includes communities from eight different tissues with processes such as *mitochondrial translation*, *regulation of lipid metabolic process*, and *cholesterol homeostasis*. The next one, muscle development, also includes eight tissues, as well as processes like *cardiac cell development* and *myofibril assembly*. The remaining sets are: *spermatogenesis*, *energetic metabolism* and *nuclear transportation*. All these processes are necessary for cell maintenance.

The largest set of cancer-only processes is closely related to immune response. It contains communities from thirteen different tissues and is associated with processes that include *antigen processing and presentation*, *interleukin-1 production*, *lymphocyte differentiation*, *response to chemokine*, among many others (fifty-eight in total). There are two other sets associated with activation of the immune response: *T-cell activation* and *response to viruses*. Moreover, the network includes other processes known to be altered in cancer: *cell migration*, *cell cycle* and *chromatin organization*.

As for the *stress response to metal ions processes*, it has been reported that metallothioneins are altered in the methylation profiles in breast (49), kidney (50), lung (51, 52) , and ovarian

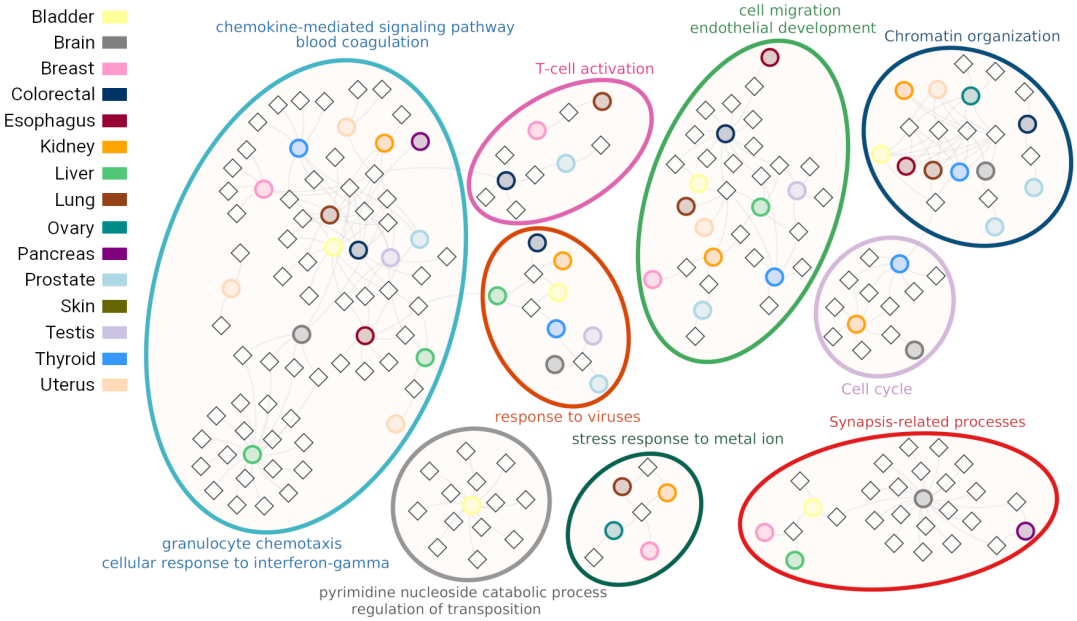

Fig. 12. Bipartite network of Gene Ontology processes and their associated communities found exclusively in the cancer phenotype. Similar GO terms were grouped. Diamonds represent enriched processes and circles represent communities colored according to their tissue.

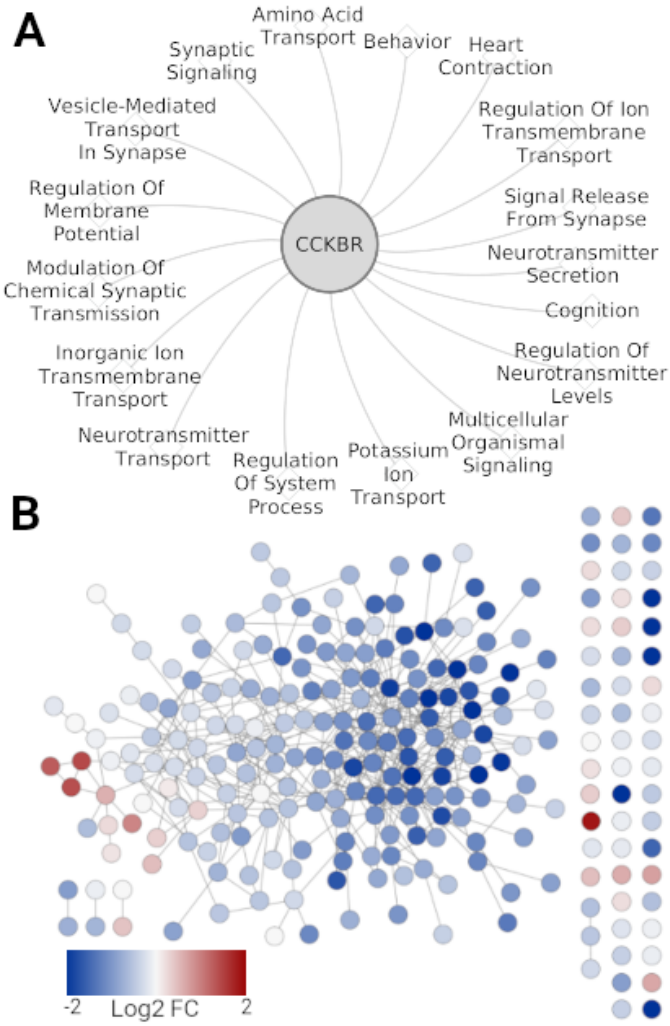

Fig. 13. A. Gene ontology terms uniquely associated to CCKBR community in the brain cancer co-expression network. **B**. Genes in the CCKBR community participating in the GO enrichments colored according to their $log_2 FC$ values.

cancer (53), they are also highly correlated to poor outcome and malignancy. Gene networks of these four tissues have communities associated with response to metal ions.

Single tissue enriched processes. Finally, GO terms uniquely associated with a single tissue in both normal and cancer networks were extracted. These unique terms describe the individual features of each co-expression network and some of them are related to tissue-specific known biological processes.

For instance, in the brain cancer co-expression network, the CCKBR community is related to processes such as *neurotransmitter secretion*, *synaptic signaling*, and *regulation of membrane potential*, among others. The uniquely associated terms in the brain cancer network and their associated communities are displayed in Figure 13A. The 257 genes that participate in the enrichments from the CCKBR community (Figure 13B) have an underexpression trend, with an average log_2 fold change of -0.703.

In the testis normal co-expression network, the community labeled Boll or BOULE, a gene required for sperm development (54), is related to multiple sperm associated processes, such as: *cilium movement*, *axoneme assembly*, *spermatogenesis*, and *single fertilization*. This community has a negative chromosomal assortativity (-0.887), meaning that the majority of its links are inter-chromosomal. In this community, 133 genes participate in the biological processes and, out of those 133 genes, ninety seven are also present in the cancer co-expression network (in multiple communities). These genes are mostly underexpressed with an average log_2 fold change value of -2.614.

In the skin normal co-expression network, enriched terms in the MLANA community include *melanin biosynthetic process*, *melanin metabolic process*, and *pigment biosynthetic process*. Out of the nineteen genes in this community, seven participate in those GO terms, and are also found in the cancer co-expression network with an over-expression tendency (average log_2 fold change of 1.3). The gene with the highest page rank value, MLANA or Melan-A, encodes a protein involved in melanosome biogenesis and it is useful for melanoma diagnosis (55). However, it is not annotated in any GO process and it is also present in the cancer co-expression network with a 1.515 $log_2 FC$.

These unique GO terms for testis and skin, as well as genes in the previously mentioned communities, are displayed in Supplementary Figure 12.

Adaptive immune response is a common and isolated process in cancer. The Gene Ontology process *adaptive immune response* is the most repeated process in the bipartite community-process cancer network. It is enriched in twenty-one communities from all cancer types, except for skin cancer. Some of these communities include genes in chromosome 6, encoding almost all members from the major histocompatibility complex II, such as HLA-DP, HLA-DM, HLA-DOA, HLA-DQ, and HLA-DR, plus CIITA (chromosome 16) and CD74 (chromosome 5). CIITA regulates the transcription of the HLA genes, acting as a transcriptional coactivator (56), while CD74 mediates the assembly and subcellular trafficking of the MHCII complex (57).

Additionally, cancer networks reveal another detail regarding adaptive immune response-related communities: those modules appear in the periphery of networks, and their associated genes present a low degree. Inter-chromosomal external clusters, enclosed by black ellipses in Figure 4, in cancer networks of lung, ovary and uterus tissues correspond to the communities associated with *adaptive immune response*. To quantify this observation, we calculated k-core decomposition, also named coreness, for each node in each network (58). This analysis extracts the maximal subgraph with all nodes having at least degree k. The communities associated with *Adaptive immune response* display lower mean coreness than the rest of the communities in cancer, as it can be seen in Figure 14.

The loss of inter-chromosomal co-expression is not present in other diseases. To evaluate whether the loss of inter-chromosomal co-expression is a phenomenon exclusively present in cancer, we analyzed two RNA-Seq datasets down-

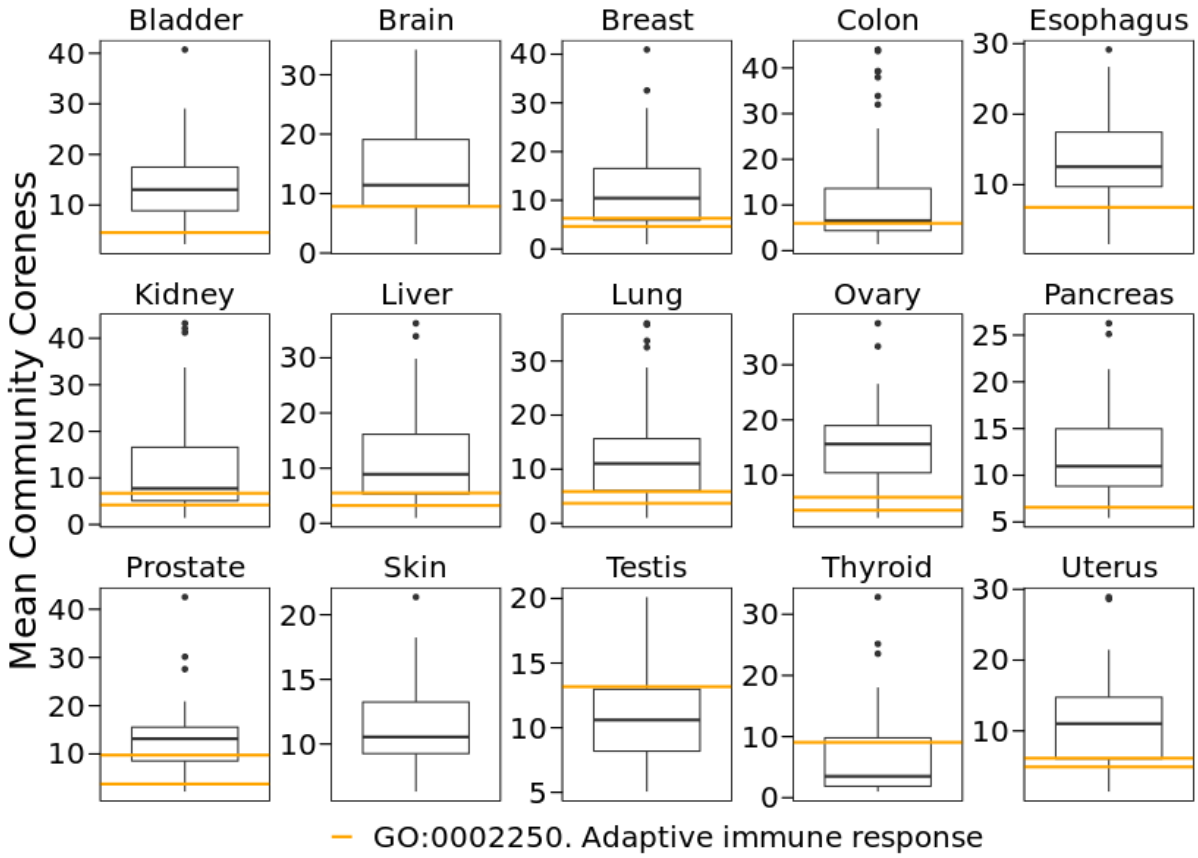

Fig. 14. Distribution of mean coreness values per community in cancer co-expression networks with highlighted values for communities associated with the Gene Ontology process: *Adaptive immune response*. These communities, placed in the periphery of co-expression networks, are also highlighted in Figure 4.

loaded from the Gene Expression Omnibus database: one for Alzheimer's disease (AD) (59), composed of brain tissue samples, and one for whole blood Type-2 Diabetes (T2D) (60) samples. Mutual information between gene pairs was calculated using the same algorithm used for cancer datasets, and the top 100,000 interactions were extracted to build co-expression networks.

In these diseases, the fraction of intra-chromosomal interactions in co-expression networks for the normal and disease phenotypes do not differ. For AD, the normal co-expression network has an intra-chromosomal fraction of 0.05997, while for the disease phenotype it is 0.05579. In the T2D network, the intra-chromosomal fraction for the normal phenotype is 0.07327 while the disease network has a same-chromosome link fraction of 0.07336.

Furthermore, the topology of the AD networks, both normal and disease, present a very similar structure to the topology of the non-cancer tissues in terms of the size of the giant component, as well as the smaller connected components separated from it. However, the structure of the T2D networks (normal and disease) differs from that of the AD networks, as illustrated in Figure 15. These observations demonstrate that the loss of inter-chromosomal co-expression is not present in these chronic degenerative diseases.

Discussion

In this work we have characterized two cancer associated phenomena: the loss of inter-chromosomal co-expression and the distance-dependent decay in intra-chromosome gene co-expression. By setting multiple Mutual Information (MI) thresholds as a measure of statistical dependence, we have determined that the strongest correlation values occur between genes located within the same chromosome and, by associating intra-chromosomal interactions with the physical distance between genes (measured in base pairs), we have concluded that, in cancer, high correlation values are placed in close neighborhoods. Conversely, for normal tissues (i.e. healthy adjacent tissue used as control in TCGA and samples tagged as Normal Tissue in GTEx-Xena), top-correlations are not biased towards intra- or inter-chromosome interactions nor to any characteristic physical distance between gene pairs, except for very close distances, as shown in the left side of the distance plots for normal tissue profiles (Figure 2 and Supplementary Figure 3).

The difference between cancer and normal profiles regarding strength of gene co-expression and intra-chromosomal distance can be appreciated in Figure 2 and Supplementary Figure 3. All cancer tissues present a significant decay in the mean value of MI with respect to base pair distance between same-chromosome genes. On the other hand, MI values in normal networks reach a plateau from the very beginning of

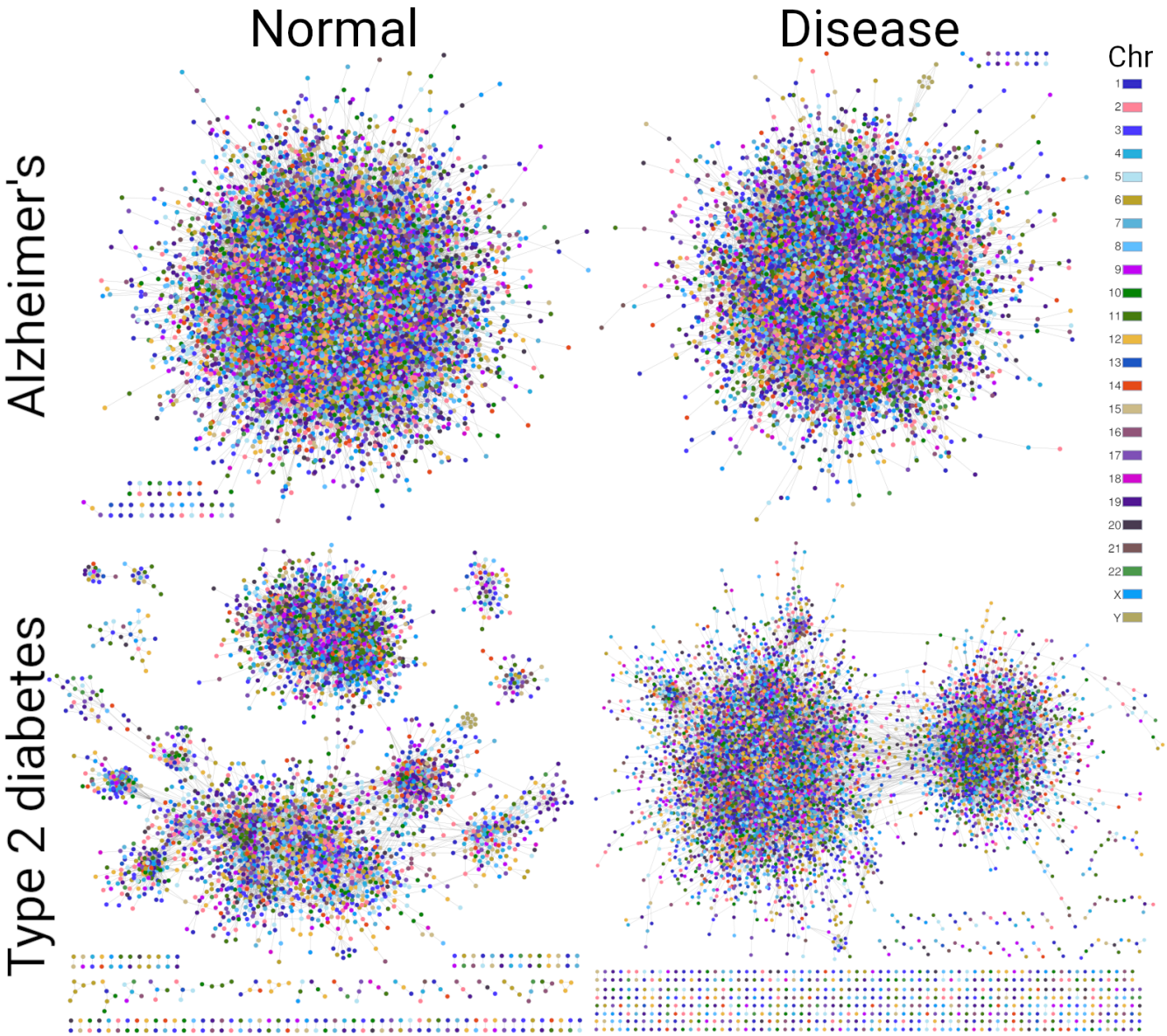

Fig. 15. Co-expression networks from Alzheimer's disease and type-2 diabetes composed by the top 100,000 MI interactions.

the plots. In both cases, standard deviation profiles reinforce the findings. How can we interpret this decay? Which mechanisms are underlying this behavior in cancer?

High local gene co-expression has been previously reported for normal human tissues between genes that are located near each other (within 1 Mb), regardless of strand or transcription orientation (22, 23, 25, 26), showing that gene transcription is influenced by its physical location. Moreover, there is evidence that the transcription process contributes to the formation of gene clusters and influences chromatin organization at a fine scale (61, 62). In the cancer profiles shown here, highly co-expressed gene pairs span a much greater distance, some of them reaching their decay plateau at around 50 Mb.

The assessment of the contribution of different regulatory elements such as enhancers, transcription factor binding sites, CTCF binding sites, and Hi-C contacts, to the local gene co-expression in normal tissues done by (21) indicated that no feature can uniquely explain the co-expression patterns observed, but they all contribute in a complementary fashion. We have shown, for Luminal A and B breast cancer subtypes, that intra-chromosomal communities are not bound by CTCF binding sites and that copy number alterations do not influence co-expressed network communities with a similar differential expression trend (27) nor do they affect intra-cytoband co-expression patterns (63). Our group has also analyzed epigenetic mechanisms in the context of cancer gene co-expression, in particular the role of miRNA regulation, showing that specific miRNA families alter co-expression in genes related to the epithelial-mesenchymal transition (64, 65), they also promote subtype-specific pathways in breast cancer (66), and are associated with progression stage in clear cell renal carcinoma (67). However, these examples come from inter-chromosomal interactions. Hence, further research is needed to clarify the regulatory mechanisms that enable local co-expression in normal tissues and the alterations that broaden co-expression regions in cancer.

It is worth recalling that, in our mutual information calculations, each gene expression vector (in normal or cancer tissue) is compared to another vector independently; there is no influence regarding chromosomal position or phenotype. Therefore, the fact that neighboring intra-chromosome genes show higher statistical dependency than those more distant, could indicate that, in cancer, there is an ubiquitous phenomenon allowing or enhancing close expression patterns across long regions of the chromosomes. We hypothesize that decay behavior in cancer could be related to functional or mechanical features of the transcriptional machinery that allow or impede gene transcription.

Although all fifteen cancer co-expression profiles present a loss of long-range co-expression, Supplementary Figures 1 and 3 display differences in the extent to which these happen among tissues. For instance, both cancer and normal co-expression profiles in thyroid and prostate present less remarkable differences, while tissues such as bladder, breast, lung, and ovary show both greater distances between fractions of intra-chromosomal interactions and a more pronounced decay. These differences could be associated with particular features of cancer manifestation in each tissue.

Focusing on thyroid cancer, loss of distant co-expression is not as evident as in the rest of the cancer profiles: the slope of its MI vs distance decay (Supplementary Figure 3) is less pronounced and its network has the highest number of unique interactions. Moreover, although this cancer network has protocadherins and histone communities, it does not present other common features such as the presence of HOX genes or high-density intra-cytoband regions. According to GLOBOCAN, thyroid cancer is placed among the cancer types with lowest mortality rates: 0.5 per 100,000 in women, and 0.3 per 100,000 in men (68). The most common thyroid subtype is papillary, followed by follicular and, in a smaller proportion, the most aggressive subtypes: medullary and anaplastic (69). The first three subtypes have a 5-years survival rate of almost 100% (70, 71). We have shown that, in papillary and follicular thyroid cancer, relevant pathways are significantly less deregulated than in the anaplastic subtype (72). Therefore, the particular behavior in the structure of the thyroid cancer network might reflect the less aggressive nature of the disease. With this in mind, further investigation involving subtype-specific cancer networks, could be appealing.

The high-density intra-cytoband regions we found by identifying clusters of strong (top 100,000) intra-chromosomal interactions, indeed appear in all cancer tissues, except thyroid. Moreover, there are multiple cytobands, such as 1q42.11 and 13q21.33, with co-expression hotspots shared among more than ten tissues and present only in cancer profiles. As cytobands *per se*, these regions have not been associated with cancer, although some of their genes have. For instance, NVL in 1q42.11 has been identified as a prognostic outlier for prostate cancer (73) while, in the same region, the CNIH4 gene has been found to promote metastasis in colon cancer (74). The q21.22 region in chromosome 13 includes BORA, an activator of the protein kinase Aurora A (AURKA), which has been found to have a role in bladder cancer (75), ovarian cancer (76), as well as breast, lung, and gastric adenocarcinomas (77).

Normal and cancer co-expression networks built from the top 100,000 MI interactions are even visually different when a force directed layout is applied, especially for the networks with a higher fraction of intra-chromosomal links: bladder, breast, esophagus, kidney, liver, lung, ovary, and uterus (Table 1 and Supplementary Figure 6). Evidence regarding the appearance of co-expression clusters from gene expression data and their functional association has been reported elsewhere (78–80), as well as a loss of connectivity in cancer co-expression networks derived from microarray data (81). However, the loss of long-range gene interactions and the gain of close distance links displayed by the top 100,000 MI interaction networks had not been previously described as a common trait in cancer.

When intersections among networks are extracted, there is a decisive difference between cancer and normal phenotypes: normal co-expression networks share less interactions than cancer networks. We have already reported this for lung cancer, where the networks for squamous cell lung carcinoma and lung adenocarcinoma shared more interactions than their normal counterparts (19). The co-expression program is a manifestation of the state of a phenotype and the fact that this program is highly similar among cancer tissues from different origins and locations reflects functional similarities. We argue that the large set of shared intra-chromosomal interactions is associated with the dedifferentiation processes that cells suffer during tumor development (82–84).

Despite the small number of connections shared among normal co-expression networks, there is an important cluster that constitutes the largest component in the common normal tissue network (Figure 6): riboprotein encoding genes. This module also appears in the common cancer network, exhibiting that ribosomes preserve a high co-expression pattern both in normal and cancer tissues. The ribosome is an essential cellular component for all organisms. It is widely known that its encoding gene sequences are highly conserved among species (85). This fact may explain its prevalence in almost all co-expression networks; additionally, it exhibits the capability of co-expression networks to capture important characteristics associated with a phenotype. Its presence reflects the significance of the translation process. The appearance of this cluster in co-expression networks from other clinical conditions and even other organisms requires further investigation.

Regarding network structure, cancer tissues present an observable modular behavior (Supplementary Figure 6) and, indeed, their modularity values are higher than normal networks (Table 3), especially in high intra-fraction networks. This observation could indicate that, in normal phenotypes, genes interact in a less compartmentalized manner with transcriptional events, and thus their co-expression values, being mostly driven by cellular functions and environmental signals. In contrast, segregated modules in cancer co-expression networks might illustrate a partial loss of information spread or a gain in specific co-expression regions.

Chromosomal assortativity (Figure 7) serves as a metric to summarize the following phenomenon: in normal networks communities there is a strong tendency towards inter-chromosomal interactions, while cancer communities are mostly conformed by intra-chromosomal links. Furthermore, there are no networks with chromosomal assortativity values tending towards zero, meaning that communities are built either from intra- or inter- chromosomal links. The significance of both chromosomal and differential expression assortativity can be appreciated when the differences between biologically associated and non-biologically associated communities are compared: functionally enriched communities in cancer display higher expression assortativity values. This illustrates how structural features in co-expression networks are related to functional alterations, with genes in biological processes being expressed in a coordinated fashion while also collectively under- or over-expressed.

Modular composition, distance-associated co-expression decay, as well as high chromosomal and expression assortativity in enriched network communities suggest a manifestation of highly orchestrated co-expression events involving genes in close vicinities in cancer networks. The loss of long-distance communication at the transcriptomic level, and the appearance of these co-expression hotspots might derive in the loss of relevant cellular functions.

The latter can be appreciated in the bipartite networks of Gene Ontology processes shared in more than ten tissues, either normal or cancer, and in the bipartite networks of normal-only and cancer-only processes (Figures 9, 10, 11, 12). GO terms in normal networks are associated with cellular functions such as translation, metabolism and development.

On the other hand, although we have found an important presence of immune response associated processes in inter-chromosome cancer communities, we also found same-chromosome communities comprising gene families such as HOXA and HOXB genes, protocadherins, histone variants, and metallothioneins. These modules present a similar differential expression trend. It is worth noting that some of these gene families also display a high co-expression in normal tissues, calculated through a different method (21).

The structure of some cancer co-expression networks reveals a tissue-specific association. For instance, the network community observed in Figure 13 presents a set of neural system-related processes with predominantly underexpressed genes. This example implies an association among differential gene expression, gene co-expression and functional features of a concrete network community, providing a phenotype-specific manifestation.

Communities associated with *adaptive immune response* appear in all cancer tissues and present a low coreness value (genes are placed at the external layers of the networks), suggesting minor communication with the rest of the network. Additionally, those communities have immune-cell specific genes. This may imply that the function associated with those communities is acquired through the infiltration of immune cells into all cancer tissues. Further investigation regarding the nature and origin of those co-expression interactions is essential. However, this result reflects the importance that gene co-expression analysis has in terms of functional discoveries in cancer research.

The affirmation that loss of long-range co-expression is a cancer-specific phenomenon is reinforced with the results presented in Figure 15. Two chronic-degenerative diseases such as Alzheimer's disease and Type-2 Diabetes present strong similarities in their co-expression landscapes, not only after a visual inspection, but also in their chromosomal assortativity values. The presence of loss of long-range co-expression in other diseases certainly must be evaluated with additional disease-specific datasets.

Despite their peculiarities, the results presented here are consistent throughout fifteen tissues in cancer, contrasted against their normal counterparts. Moreover, there are no significant differences associated with data source: both TCGA and UCSC Xena datasets have tissues with high and low levels of loss of distant co-expression; and the number of samples apparently does not influence network structure, something we had already assessed in breast (15) and lung cancer (19). However, there are some considerations that need to be addressed in future analyses. For instance, the hierarchical clustering shown in Figure 5 does not group according to tissue similarity or cancer origin. Further evaluation of the nature of the displayed similarities is needed. In that same figure, the biggest intersection sets which include esophagus, skin, and pancreas interactions, belong to the UCSC Xena dataset, with data from GTEx; we should evaluate whether the preprocessing steps added some bias.

Our research on this topic will benefit from additional datasets. Single cell RNA-Seq in cancer tissues will allow us to identify the cell types that present loss of long-range co-expression. With different cancer multi-omics datasets, we will be able to assess the influence of altered regulatory mechanisms of gene transcription in the loss of distant gene co-expression.

Concluding remarks

The loss of distant gene co-expression in cancer consists of two main manifestations: a loss of inter-chromosome gene co-expression and an intra-chromosomal co-expression decay dependent on physical distance between genes. We observed this phenomena in fifteen cancer tissues, but not in normal phenotypes. We have shown this phenomenon does not depend on the co-expression cutoff or the number of samples, and it is also independent of the data source. The structure observed in cancer networks is associated with common functional traits in the disease. Riboprotein genes constitute a particular set of inter-chromosome interactions shared by more than twenty tissues, both cancer and normal, thus indicating the high relevance of co-regulation of theses genes to preserve cell viability. Biologically associated network communities in cancer present high differential expression assortativity, i.e., genes of the same community have the same differential expression trend. External communities to the largest component in the majority of cancer networks are as-

sociated with adaptive immune response and this could mean that these communities are not a structural part of the network, suggesting an immune cell infiltration. Finally, the loss of distant co-expression is not present in other chronic degenerative diseases, such as Type-2 Diabetes or Alzheimer's disease. We believe these phenomena might be a manifestation of a hallmark of cancer not previously reported. Further investigation, mainly consisting in targeted experimental studies, is necessary to unveil the mechanisms behind this co-expression program alteration. However, we are confident that the disappearance of long-range correlation in cancer phenotypes is a matter that could reflect a deeper anomaly that has not been previously explored.

Methods and Materials

Databases. RNA-Seq data from The Cancer Genome Atlas (28) was downloaded in *HTSeq-Counts* format from the Genomic Data Commons (GDC) Data Portal for bladder, breast, kidney, lung, thyroid and uterus tissues. Samples tagged as Primary Tumor constitute the cancer data set, while the normal data set is integrated by samples with sample type: Solid Tissue Normal. Gene expression *RSEM expected counts* were downloaded from the UCSC Xena browser (29), a database that integrates TCGA, TARGET and GTEx expression data through the implementation of an RNA-seq pipeline, using Toil, an open source workflow software (30). For brain, colon, esophagus, liver, ovary, pancreas, prostate, skin, and testis, normal counts were obtained from GTEx-Xena, tagged as Normal Tissue, while cancer data for these tissues was identified as Primary Tumor from TCGA-Xena. RSEM expected counts were transformed from $log_2(expectedcount + 1)$ to integer $expectedcount$ values.

Data processing. The pipeline to process RNA-Seq data and to perform differential expression analysis was adapted from (15). A Snakemake (86) workflow was implemented to create a reproducible analysis: `https://github.com/ddiannae/tcga-xena-pipeline`. A second pipeline was created for the analysis of co-expression profiles and co-expression networks, `https://github.com/ddiannae/distance-analysis`.

Pre-processing. Steps in pre-processing were performed per tissue and sample type. TCGA samples were annotated using their annotation reference file, GENCODE v22, `https://gdc.cancer.gov/about-data/gdc-data-processing/gdc-reference-files`, while UCSC Xena uses GENCODE v23. Only protein coding genes, conventional chromosomes and genes that appear in the GENCODE v37 (April, 2021) annotation, were kept. After the raw count matrices were assembled and integrated with their corresponding annotations, datasets from both sources share the remaining pipeline steps.
The `NOISeq` R library (87) was used for quality control. Genes with low expression values were removed by identifying genes with a count mean expression < 10 and genes with a zero-count expression value in more than 50% of the samples. Gene counts expression boxplots revealed that some samples from the XENA database had extremely low values; hence, samples were removed from further analysis if their mean expression value fell two standard deviations over or under the total mean.

Data normalization. Bias detection plots were obtained to assess mean gene length and %GC and RNA content bias, and a combination of different within/between normalization strategies were tested to remove bias presence using the `EDASeq` package (88). The best alternative for each tissue was selected after visual inspection of the `NOISeq` plots. Principal Component Analysis plots showed overlap between experimental groups, therefore `ARSyNSeq`, also from the `NOISeq` library, was used for multidimensional noise reduction and batch effect removal with default parameters.

Mutual Information (MI) calculation. The ARACNE algorithm (89) was used to calculate mutual information (MI) in order to quantify statistical dependence between pairs of genes. The complete set of gene pair MI values was obtained per tissue and sample type. A `Singularity` container was created to run ARACNE inside the Snakemake workflow. The source code can be found here: `https://github.com/ddiannae/ARACNE-multicore`

Differential expression analysis. Differentially expressed genes in cancer samples compared to normal ones were identified using the `limma` (90) R package. P-values were obtained with a hypothesis tests based on empirical Bayes and were adjusted to control multiple comparisons using the False Discovery Rate (FDR); $p_{adj} < 0.05$ was set as threshold.

Co-expression profile analysis.

Identification of intra-chromosomal fraction. The total set of gene pairs for each tissue and phenotype was sorted in decreasing value of MI and the fraction of intra-chromosomal interactions was calculated at different thresholds of top MI pairs (1E3, 2E3, ... 1E8). Distribution of intra-chromosomal fractions in cancer and normal phenotypes was compared by taking the values from each threshold up to the last one, using Kolmogorov-Smirnov tests.

Distance versus MI analysis. The set of same chromosome gene pairs from the MI dataset or intra-chromosomal interactions were grouped into 1000 interactions bins, according to distance in base pairs between genes, and mean MI value, as well as mean distance; then the MI standard deviation for each bin was obtained. The 1000 MI values for each bin per phenotype were compared against other bins using Wilcoxon tests.
In the cytoband analysis, the total number of possible intra-cytoband pairs was calculated by considering the genes in the expression matrix for each tissue. MI values were sorted and only the top 100,000 were used to identify shared cytobands with more than 50% of their total possible interactions. The null model to evaluate the significance of these results takes the total number of interactions in a chromosome and assigns

Table 2. Number of samples and their data sources for each tissue in the study. The number of genes (Genes) in the expression matrix changes per tissue due to processing filters.

Tissue	Source	Condition	Samples	Genes
Bladder	TCGA	Normal	19	13390
		Cancer	398	
Brain	UCSC Xena	Normal	88	10608
		Cancer	508	
Breast	TCGA	Normal	111	14326
		Cancer	1047	
Colon	UCSC Xena	Normal	163	15571
		Cancer	287	
Esophagus	UCSC Xena	Normal	269	10438
		Cancer	178	
Kidney	TCGA	Normal	123	14314
		Cancer	839	
Liver	UCSC Xena	Normal	107	14756
		Cancer	358	
Lung	TCGA	Normal	101	14750
		Cancer	995	
Ovary	UCSC Xena	Normal	88	10236
		Cancer	413	
Pancreas	UCSC Xena	Normal	165	10386
		Cancer	177	
Prostate	UCSC Xena	Normal	93	10001
		Cancer	473	
Skin	UCSC Xena	Normal	231	9908
		Cancer	88	
Testis	UCSC Xena	Normal	162	11342
		Cancer	138	
Thyroid	TCGA	Normal	56	14637
		Cancer	471	
Uterus	TCGA	Normal	35	13335
		Cancer	591	

those interactions to random pairs of genes in that chromosome. The real number of intra-cytoband interactions per cytoband was compared to the distribution of 10,000 random assignations.

Network Analysis.

Community Detection. Co-expression networks were built using the top 100,000 Mutual Information interactions per tissue and samples type. Four community detection algorithms were tested through their implementations in the `igraph` R package (91): Louvain, Fast Greedy, Infomap, and Leading Eigenvalue. MI values were used as link weights. Only communities from the Louvain algorithm were used for further analysis given that it consistently obtained the highest modularity values (Table 3). Moreover, this algorithm has proved to be successful for biological networks and their association to functional features (92, 93)

Nominal assortativity. To calculate chromosomal assortativity, we used the chromosome location for each gene. For each community, we divided the number of links joining genes in the same chromosome (intra-chromosomal interactions) minus the number of links joining genes in different chromosomes (inter-chromosomal interactions), by the total number of links in the community. In a similar manner, differential expression assortativity, defined only for cancer co-expression networks, takes the number of links between genes with same log_2FC sign and subtracts the links joining genes with a different sign.

GO Overrepresentation Analysis. The `enrichGO` function from the `clusterProfiler` R package (94) was used to identify over-represented or enriched terms in the Biological Process category in Gene Ontology (GO) (95). Enrichment was performed for communities with five or more genes; GO terms with a minimum size of ten were retained. Genes in the original expression matrix for each tissue defined the universe set. We kept terms with $p_{adj} < 1E^{-10}$ using the Benjamini and Hochberg method for multiple testing.

Coreness Decomposition. The k-core or coreness decomposition is used as a proxy to analyze the core-periphery position of a community in the co-expression networks. In a

Table 3. Modularity values obtained by the four community detection algorithms that were tested. The Louvain partition was used given that it achieves the greatest scores.

Tissue	Condition	Louvain	Fast Greedy	Infomap	Leading Eigenvalue
Bladder	Normal	0.554	0.471	0.466	0.369
	Cancer	0.967	0.964	0.935	0.020
Brain	Normal	0.505	0.452	0.445	0.449
	Cancer	0.796	0.758	0.752	0.193
Breast	Normal	0.515	0.459	0.448	0.334
	Cancer	0.965	0.960	0.930	0.109
Colon	Normal	0.717	0.639	0.661	0.005
	Cancer	0.923	0.896	0.903	0.023
Esophagus	Normal	0.624	0.584	0.579	0.002
	Cancer	0.896	0.872	0.868	0.002
Kidney	Normal	0.556	0.492	0.497	0.477
	Cancer	0.894	0.845	0.868	0.222
Liver	Normal	0.596	0.530	0.532	0.003
	Cancer	0.945	0.934	0.922	0.437
Lung	Normal	0.545	0.500	0.477	0.386
	Cancer	0.971	0.968	0.933	0.122
Ovary	Normal	0.448	0.421	0.401	0.406
	Cancer	0.959	0.952	0.917	0.005
Pancreas	Normal	0.506	0.458	0.466	0.003
	Cancer	0.746	0.711	0.706	0.306
Prostate	Normal	0.436	0.382	0.395	0.372
	Cancer	0.697	0.656	0.655	0.203
Skin	Normal	0.578	0.497	0.549	0.002
	Cancer	0.620	0.594	0.580	0.000
Testis	Normal	0.630	0.518	0.580	0.511
	Cancer	0.747	0.713	0.704	0.001
Thyroid	Normal	0.477	0.445	0.410	0.001
	Cancer	0.689	0.632	0.645	0.014
Uterus	Normal	0.542	0.433	0.465	0.456
	Cancer	0.946	0.936	0.915	0.174

network, k-core refers to a subset of its nodes in which all of them have at least degree k (58). Nodes with low k-core value have low degree, while nodes with high k-core are highly connected to each other. To obtain this decomposition, the `coreness` function in `igraph` was used over the entire network and the mean values per community were calculated.

Supplementary Material

1. Fraction of intra-chromosomal interactions at different thresholds for the fifteen tissues in the study.

2. Kolmogorov-Smirnov p-values for the comparison between fraction of intra-chromosomal interactions in normal and cancer tissues at different MI thresholds.

3. Average MI values of bins of one thousand intra-chromosomal interactions plotted against base pair distance.

4. P-values of Wilcoxon tests comparing the distribution of MI values for bins of intra-chromosomal interactions.

5. Statistical significance of the number of intra-cytoband interactions, compared to a null model.

6. Co-expression networks from the top 100 thousand MI interactions for normal and cancer tissues.

7. Differential expression assortativity for communities in cancer co-expression networks.

8. Boxplots for chromosomal assortativity for GO enriched and not enriched network communities for all tissues in normal and cancer phenotypes.

9. HOXA-HOXB family communities in cancer with genes colored according to their differential expression.

10. H2A, H2B family communities in cancer with genes colored according to their differential expression.

11. PCDHG and PCDHB family communities in cancer with genes colored according to their differential expression.

12. Unique enrichments in skin and testis normal co-expression networks with genes associated to those enrichments in cancer with their corresponding differential expression.

Bibliography

1. Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *cell*, 100(1):57–70, 2000.
2. Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer. an organizing principle for cancer medicine. *Primer of the molecular biology of cancer. 2nd ed. Philadelphia: Wolters Kluwer*, pages 28–57, 2015.
3. Nadine Darwiche. Epigenetic mechanisms and the hallmarks of cancer: An intimate affair. *American journal of cancer research*, 10(7):1954, 2020.
4. Levi A Garraway and Eric S Lander. Lessons from the cancer genome. *Cell*, 153(1):17–37, 2013.
5. Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011.
6. Douglas Hanahan. Hallmarks of cancer: new dimensions. *Cancer discovery*, 12(1):31–46, 2022.
7. Patrick Cramer. Organization and regulation of gene transcription. *Nature*, 573(7772):45–54, 2019.
8. Bas van Steensel and Eileen EM Furlong. The role of transcription in shaping the spatial organization of the genome. *Nature reviews Molecular cell biology*, 20(6):327–337, 2019.
9. Enrique Hernández-Lemus, Helena Reyes-Gopar, Jesús Espinal-Enríquez, and Soledad Ochoa. The many faces of gene regulation in cancer: a computational oncogenomics outlook. *Genes*, 10(11):865, 2019.
10. Sipko Van Dam, Urmo Vosa, Adriaan van der Graaf, Lude Franke, and Joao Pedro de Magalhaes. Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in bioinformatics*, 19(4):575–592, 2018.
11. Paola Paci, Giulia Fiscon, Federica Conte, Rui-Sheng Wang, Lorenzo Farina, and Joseph Loscalzo. Gene co-expression in the interactome: moving from correlation toward causation via an integrated approach to disease module discovery. *NPJ systems biology and applications*, 7(1):1–11, 2021.
12. Abhijeet R Sonawane, Scott T Weiss, Kimberly Glass, and Amitabh Sharma. Network medicine in the age of biomedical big data. *Frontiers in Genetics*, 10:294, 2019.
13. Yang Yang, Leng Han, Yuan Yuan, Jun Li, Nainan Hei, and Han Liang. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nature communications*, 5(1):1–9, 2014.
14. Guillermo de Anda-Jáuregui, Sergio Antonio Alcalá-Corona, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Functional and transcriptional connectivity of communities in breast cancer co-expression networks. *Applied Network Science*, 4(1):1–13, 2019.
15. Jesús Espinal-Enríquez, Cristóbal Fresno, Guillermo Anda-Jáuregui, and Enrique Hernández-Lemus. RNA-Seq based genome-wide analysis reveals loss of inter-chromosomal regulation in breast cancer. *Scientific Reports*, 7(1):1–19, 2017. ISSN 20452322.
16. Guillermo de Anda-Jáuregui, Cristobal Fresno, Diana García-Cortés, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Intrachromosomal regulation decay in breast cancer. *Applied Mathematics and Nonlinear Sciences*, 4(1):217–224, 2019.
17. Diana García-Cortés, Guillermo de Anda-Jáuregui, Cristóbal Fresno, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Gene Co-expression Is Distance-Dependent in Breast Cancer. *Frontiers in Oncology*, 10(July):1–13, 2020.
18. Alfredo González-Espinoza, Jose Zamora, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Gene co-expression in breast cancer: a matter of distance. *Frontiers in oncology*, page 4743, 2021.
19. Sergio Daniel Andonegui-Elguera, José María Zamora-Fuentes, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Loss of long distance co-expression in lung cancer. *Frontiers in genetics*, page 192, 2021.
20. Jose María Zamora-Fuentes, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Gene expression and co-expression networks are strongly altered through stages in clear cell renal carcinoma. *Frontiers in genetics*, 11:578679, 2020.
21. Diogo M Ribeiro, Simone Rubinacci, Anna Ramisch, Robin J Hofmeister, Emmanouil T Dermitzakis, and Olivier Delaneau. The molecular basis, genetic control and pleiotropic effects of local gene co-expression. *Nature communications*, 12(1):1–13, 2021.
22. Diogo M Ribeiro, Chaymae Ziyani, and Olivier Delaneau. Shared regulation and functional relevance of local gene co-expression revealed by single cell analysis. *Communications Biology*, 5(1):1–11, 2022.
23. Laurence D Hurst, Csaba Pál, and Martin J Lercher. The evolutionary dynamics of eukaryotic gene order. *Nature Reviews Genetics*, 5(4):299–310, 2004.
24. Barak A Cohen, Robi D Mitra, Jason D Hughes, and George M Church. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature genetics*, 26(2):183–186, 2000.
25. Avazeh T Ghanbarian and Laurence D Hurst. Neighboring genes show correlated evolution in gene expression. *Molecular biology and evolution*, 32(7):1748–1766, 2015.
26. Miki Ebisuya, Takuya Yamamoto, May Nakajima, and Eisuke Nishida. Ripples from neighbouring transcription. *Nature cell biology*, 10(9):1106–1113, 2008.
27. Diana García-Cortés, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Luminal A Breast Cancer Co-expression Network: Structural and Functional Alterations. *Frontiers in genetics*, 12(July):629475, 2021. ISSN 1664-8021.
28. John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.
29. Mary J Goldman, Brian Craft, Mim Hastie, Kristupas Repečka, Fran McDade, Akhil Kamath, Ayan Banerjee, Yunhai Luo, Dave Rogers, Angela N Brooks, et al. Visualizing and interpreting cancer genomics data via the xena platform. *Nature biotechnology*, 38(6):675–678, 2020.
30. John Vivian, Arjun Arkal Rao, Frank Austin Nothaft, Christopher Ketchum, Joel Armstrong, Adam Novak, Jacob Pfeil, Jake Narkizian, Alden D Deran, Audrey Musselman-Brown, et al. Toil enables reproducible, open source, big biomedical data analyses. *Nature biotechnology*, 35(4):314–316, 2017.
31. Ioannis A Voutsadakis. Amplification of 8p11. 23 in cancers and the role of amplicon genes. *Life Sciences*, 264:118729, 2021.
32. Stefanie Marczok, Birgit Bortz, Chong Wang, and Heike Pospisil. Comprehensive analysis of genome rearrangements in eight human malignant tumor tissues. *PloS one*, 11(7): e0158995, 2016.
33. Jordi Garcia-Fernàndez. The genesis and evolution of homeobox gene clusters. *Nature Reviews Genetics*, 6(12):881–892, 2005.
34. Petr Novak, Taylor Jensen, Marc M Oshiro, Ryan J Wozniak, Marcella Nouzova, George S Watts, Walter T Klimecki, Christina Kim, and Bernard W Futscher. Epigenetic inactivation of the hoxa gene cluster in breast cancer. *Cancer research*, 66(22):10664–10670, 2006.
35. Muhiddin Ishak, Rashidah Baharudin, Loh Teng-Hern Tan, Learn-Han Lee, and Nurul-Syakima Ab Mutalib. Landscape of hoxa genes methylation in colorectal cancer. *Progress In Microbes & Molecular Biology*, 3(1), 2020.
36. Hongling Du and Hugh S Taylor. The role of hox genes in female reproductive tract development, adult function, and fertility. *Cold Spring Harbor perspectives in medicine*, 6(1): a023002, 2016.
37. Tong Lu, Tong Qiu, Bin Han, Yuanyong Wang, Xiao Sun, Yi Qin, Ao Liu, Nan Ge, and Wenjie Jiao. Circular rna circcsnk1g3 induces hoxa10 signaling and promotes the growth and metastasis of lung adenocarcinoma cells through hsa-mir-143-3p sponging. *Cellular Oncology*, 44(2):297–310, 2021.
38. Yi-Nan Guo, Bin Luo, Wen-Jie Chen, Xin Chen, Zhi-Gang Peng, Kang-Lai Wei, and Gang Chen. Comprehensive clinical implications of homeobox a10 in 3,199 cases of non-small cell lung cancer tissue samples combining qrt-pcr, rna sequencing and microarray data. *American Journal of Translational Research*, 11(1):45, 2019.
39. Junhu Wan, Hongyang Liu, Quanling Feng, Jun Liu, and Liang Ming. Hoxb9 promotes endometrial cancer progression by targeting e2f3. *Cell death & disease*, 9(5):1–18, 2018.
40. Ying Ying, Yejun Wang, Xiaoyan Huang, Yanmei Sun, Junbao Zhang, Meiqi Li, Junhui Zeng, Maolin Wang, Wenjun Xiao, Lan Zhong, et al. Oncogenic hoxb8 is driven by myc-regulated super-enhancer and potentiates colorectal cancer invasiveness via bach1. *Oncogene*, 39 (5):1004–1017, 2020.
41. Kwei-Yan Liu, Li-Ting Wang, Shih-Hsien Hsu, and Shen-Nien Wang. Homeobox genes and hepatocellular carcinoma. *Cancers*, 11(5):621, 2019.
42. Simone Aparecida de Bessa Garcia, Mafalda Araújo, Tiago Pereira, João Mouta, and Renata Freitas. Hox genes function in breast cancer development. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1873(2):188358, 2020.
43. Fátima Liliana Monteiro, Tiago Baptista, Francisco Amado, Rui Vitorino, Carmen Jerónimo, and Luisa A Helguero. Expression and functionality of histone h2a variants in cancer. *Oncotarget*, 5(11):3428, 2014.
44. Shweta R Nayak, Emily Harrington, David Boone, Ryan Hartmaier, Jian Chen, Thushangi N Pathiraja, Kristine L Cooper, Jeffrey L Fine, Joseph Sanfilippo, Nancy E Davidson, et al. A role for histone h2b variants in endocrine-resistant breast cancer. *Hormones and Cancer*, 6 (5):214–224, 2015.
45. Wenting Xie, Jiajia Zhang, Peng Zhong, Shanshan Qin, Han Zhang, Xin Fan, Yuzhen Yin, Ruipeng Liang, Yali Han, Yina Liao, et al. Expression and potential prognostic value of histone family gene signature in breast cancer. *Experimental and therapeutic medicine*, 18 (6):4893–4903, 2019.
46. Xinbing Sui, Da Wang, Shumin Geng, Gongli Zhou, Chao He, and Xiaotong Hu. Methylated promoters of genes encoding protocadherins as a new cancer biomarker family. *Molecular biology reports*, 39(2):1105–1111, 2012.
47. Ming Shan, Yonghui Su, Wenli Kang, Ruixin Gao, Xiaobo Li, and Guoqiang Zhang. Aberrant expression and functions of protocadherins in human malignant tumors. *Tumor Biology*, 37 (10):12969–12981, 2016.
48. Ana Florencia Vega-Benedetti, Eleonora Loi, Loredana Moi, Sylvain Blois, Antonio Fadda, Manila Antonelli, Antonella Arcella, Manuela Badiali, Felice Giangaspero, Isabella Morra, et al. Clustered protocadherins methylation alterations in cancer. *Clinical epigenetics*, 11 (1):1–20, 2019.
49. Mary Ann Sens, Seema Somji, Scott H Garrett, C Larry Beall, and Donald A Sens. Metallothionein isoform 3 overexpression is associated with breast cancers having a poor prognosis. *The American journal of pathology*, 159(1):21–26, 2001.
50. Imad Alkamal, Odiljon Ikromov, Angelika Tölle, Tom Florian Fuller, Ahmed Magheli, Kurt Miller, Hans Krause, and Carsten Kempkensteffen. An epigenetic screen unmasks metallothioneins as putative contributors to renal cell carcinogenesis. *Urologia Internationalis*, 94 (1):99–110, 2015.
51. Bozena Werynska, Bartosz Pula, Beata Muszczynska-Bernhard, Aleksandra Piotrowska, Aleksandra Jethon, Marzena Podhorska-Okolow, Piotr Dziegiel, and Renata Jankowska. Correlation between expression of metallothionein and expression of ki-67 and mcm-2 proliferation markers in non-small cell lung cancer. *Anticancer research*, 31(9):2833–2839, 2011.
52. Gui-You Liang, Sheng-Xun Lu, Gang Xu, Xing-Da Liu, Jian Li, and Deng-Shen Zhang. Expression of metallothionein and nrf2 pathway genes in lung cancer and cancer-surrounding tissues. *World Journal of Surgical Oncology*, 11(1):1–5, 2013.
53. JG Hengstler, H Pilch, M Schmidt, H Dahlenburg, J Sagemüller, I Schiffer, F Oesch, PG Knapstein, B Kaina, and B Tanner. Metallothionein expression in ovarian cancer in relation to histopathological parameters and molecular markers of prognosis. *International journal of cancer*, 95(2):121–127, 2001.
54. Chirag Shah, Michael JW VanGompel, Villian Naeem, Yanmei Chen, Terrance Lee, Nicholas Angeloni, Yin Wang, and Eugene Yujun Xu. Widespread presence of human boule homologs among animals and conservation of their ancient reproductive function. *PLoS genetics*, 6(7):e1001022, 2010.

55. Jinyan Du, Arlo J Miller, Hans R Widlund, Martin A Horstmann, Sridhar Ramaswamy, and David E Fisher. Mlana/mart1 and silv/pmel17/gp100 are transcriptionally regulated by mitf in melanocytes and melanoma. *The American journal of pathology*, 163(1):333–343, 2003.
56. Dinah Singer and Ballachanda Devaiah. CIITA and Its Dual Roles in MHC Gene Transcription. *Frontiers in Immunology*, 4:476, 2013. ISSN 1664-3224.
57. Bernd Schröder. The multifaceted roles of the invariant chain CD74 — More than just a chaperone. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1863(6, Part A):1269–1281, 2016. ISSN 0167-4889.
58. Stephen B Seidman. Network structure and minimum degree. *Social networks*, 5(3):269–287, 1983.
59. Samuel Morabito, Emily Miyoshi, Neethu Michael, Saba Shahin, Alessandra Cadete Martini, Elizabeth Head, Justine Silva, Kelsey Leavy, Mari Perez-Rosendahl, and Vivek Swarup. Single-nucleus chromatin accessibility and transcriptomic characterization of alzheimer's disease. *Nature genetics*, 53(8):1143–1155, 2021.
60. Hung-Hsin Chen, Lauren E Petty, Kari E North, Joseph B McCormick, Susan P Fisher-Hoch, Eric R Gamazon, and Jennifer E Below. Novel diabetes gene discovery through comprehensive characterization and integrative analysis of longitudinal gene expression changes. *Human Molecular Genetics*, 2022.
61. Bas van Steensel and Eileen EM Furlong. The role of transcription in shaping the spatial organization of the genome. *Nature reviews Molecular cell biology*, 20(6):327–337, 2019.
62. María E Soler-Oliva, José A Guerrero-Martínez, Valentina Bachetti, and José C Reyes. Analysis of the relationship between coexpression domains and chromatin 3d organization. *PLoS computational biology*, 13(9):e1005708, 2017.
63. Candelario Hernández-Gómez, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. The role of copy number variants in gene co-expression patterns for luminal b breast tumors. *Frontiers in genetics*, 13:806607–806607, 2022.
64. Diana Drago-García, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Network analysis of emt and met micro-rna regulation in breast cancer. *Scientific reports*, 7(1):1–17, 2017.
65. Guillermo de Anda-Jáuregui, Jesús Espinal-Enríquez, Diana Drago-García, and Enrique Hernández-Lemus. Nonredundant, highly connected micrornas control functionality in breast cancer networks. *International journal of genomics*, 2018, 2018.
66. Guillermo de Anda-Jáuregui, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Highly connected, non-redundant microrna functional control in breast cancer molecular subtypes. *Interface Focus*, 11(4):20200073, 2021.
67. Jose María Zamora-Fuentes, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Oncogenic role of mir-217 during clear cell renal carcinoma progression. *Frontiers in oncology*, 12:934711–934711, 2022.
68. Hyuna Sung, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, Ahmedin Jemal, and Freddie Bray. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 71(3):209–249, 2021.
69. Adalberto Miranda-Filho, Joannie Lortet-Tieulent, Freddie Bray, Bochen Cao, Silvia Franceschi, Salvatore Vaccarella, and Luigino Dal Maso. Thyroid cancer incidence trends by histology in 25 countries: a population-based study. *The lancet Diabetes & endocrinology*, 9(4):225–234, 2021.
70. American Cancer Society. Thyroid cancer survival rates, by type and stage, 2016.
71. Luigino Dal Maso, A Tavilla, F Pacini, D Serraino, BAC Van Dijk, MD Chirlaque, R Capocaccia, Nerea Larranaga, M Colonna, Dominic Agius, et al. Survival of 86,690 patients with thyroid cancer: a population-based study in 29 european countries from eurocare-5. *European Journal of Cancer*, 77:140–152, 2017.
72. Jesús Espinal-Enríquez, Said Muñoz-Montero, Ivan Imaz-Rosshandler, Aldo Huerta-Verde, Carmen Mejía, and Enrique Hernández-Lemus. Genome-wide expression analysis suggests a crucial role of dysregulation of matrix metalloproteinases pathway in undifferentiated thyroid carcinoma. *BMC genomics*, 16(1):1–23, 2015.
73. Shuang G Zhao, Joseph R Evans, Vishal Kothari, Grace Sun, Ashley Larm, Victor Mondine, Edward M Schaeffer, Ashley E Ross, Eric A Klein, Robert B Den, et al. The landscape of prognostic outlier genes in high-risk prostate cancerhigh-risk prostate cancer prognostic outlier gene landscape. *Clinical cancer research*, 22(7):1777–1786, 2016.
74. Sonakshi Mishra, Carolina Bernal, Marianna Silvano, Santosh Anand, and Ariel Ruiz i Altaba. The protein secretion modulator tmed9 drives cnih4/tgfα/gli signaling opposing tmed3-wnt-tcf to promote colon cancer metastases. *Oncogene*, 38(29):5817–5837, 2019.
75. Songtao Cheng, Tianchen Peng, Xiaolu Zhu, Fenfang Zhou, Gang Wang, Lingao Ju, Yu Xiao, Xuefeng Liu, and Xinghuan Wang. Bora regulates cell proliferation and migration in bladder cancer. *Cancer cell international*, 20(1):1–10, 2020.
76. Alfonso Parrilla, Marta Barber, Blanca Majem, Josep Castellví, Juan Morote, José Luis Sánchez, Asunción Pérez-Benavente, Miguel F Segura, Antonio Gil-Moreno, and Anna Santamaria. Aurora borealis (bora), which promotes plk1 activation by aurora a, has an oncogenic role in ovarian cancer. *Cancers*, 12(4):886, 2020.
77. Qiong-Xia Zhang, Rui Gao, Jin Xiang, Zhong-Yu Yuan, Yuan-Min Qian, Min Yan, Zi-Feng Wang, Quentin Liu, Hai-Dong Zhao, and Chang-Hong Liu. Cell cycle protein bora serves as a novel poor prognostic factor in multiple adenocarcinomas. *Oncotarget*, 8(27):43838, 2017.
78. Wencheng Yin, Luis Mendoza, Jimena Monzon-Sandoval, Araxi O Urrutia, and Humberto Gutierrez. Emergence of co-expression in gene regulatory networks. *PloS one*, 16(4): e0247671, 2021.
79. Pawel Michalak. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics*, 91(3):243–248, 2008.
80. Sergio Antonio Alcalá-Corona, Guillermo de Anda-Jáuregui, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Network modularity in breast cancer molecular subtypes. *Frontiers in physiology*, 8:915, 2017.
81. Roberto Anglani, Teresa M Creanza, Vania C Liuzzi, Ada Piepoli, Anna Panza, Angelo Andriulli, and Nicola Ancona. Loss of connectivity in cancer co-expression networks. *PloS one*, 9(1):e87075, 2014.
82. Dinorah Friedmann-Morvinski and Inder M Verma. Dedifferentiation and reprogramming: origins of cancer stem cells. *EMBO reports*, 15(3):244–253, 2014.
83. Arnatchai Maiuthed, Narumol Bhummaphan, Sudjit Luanpitpong, Apiwat Mutirangura, Chatchawit Aporntewan, Arthitaya Meeprasert, Thanyada Rungrotmongkol, Yon Rojanasakul, and Pithi Chanvorachote. Nitric oxide promotes cancer cell dedifferentiation by disrupting an oct4: caveolin-1 complex: A new regulatory mechanism for cancer stem cell formation. *Journal of Biological Chemistry*, 293(35):13534–13552, 2018.
84. Jinyang Li and Ben Z Stanger. How tumor cell dedifferentiation drives immune evasion and resistance to immunotherapy. *Cancer research*, 80(19):4037–4041, 2020.
85. Maki Yoshihama, Tamayo Uechi, Shuichi Asakawa, Kazuhiko Kawasaki, Seishi Kato, Sayomi Higa, Noriko Maeda, Shinsei Minoshima, Tatsuo Tanaka, Nobuyoshi Shimizu, et al. The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome research*, 12(3):379–390, 2002.
86. Felix Mölder, Kim Philipp Jablonski, Brice Letcher, Michael B Hall, Christopher H Tomkins-Tinch, Vanessa Sochat, Jan Forster, Soohyun Lee, Sven O Twardziok, Alexander Kanitz, et al. Sustainable data analysis with snakemake. *F1000Research*, 10, 2021.
87. Sonia Tarazona, Pedro Furió-Tarí, David Turrà, Antonio Di Pietro, María José Nueda, Alberto Ferrer, and Ana Conesa. Data quality aware analysis of differential expression in rna-seq with noiseq r/bioc package. *Nucleic acids research*, 43(21):e140–e140, 2015.
88. Davide Risso, Katja Schwartz, Gavin Sherlock, and Sandrine Dudoit. Gc-content normalization for rna-seq data. *BMC bioinformatics*, 12(1):1–17, 2011.
89. Adam A Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In *BMC bioinformatics*, volume 7, pages 1–15. BioMed Central, 2006.
90. Matthew E Ritchie, Belinda Phipson, DI Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47–e47, 2015.
91. Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal*, Complex Systems:1695, 2006.
92. Sara Rahiminejad, Mano R. Maurya, and Shankar Subramaniam. Topological and functional comparison of community detection algorithms in biological networks. *BMC Bioinformatics*, 20(1):1–25, 2019. ISSN 14712105.
93. Sergio Antonio Alcala-Corona, Santiago Sandoval-Motta, Jesus Espinal-Enriquez, and Enrique Hernandez-Lemus. Modularity in biological networks. *Frontiers in Genetics*, 12, 2021.
94. Guangchuang Yu, Li-Gen Wang, Yanyan Han, and Qing-Yu He. clusterprofiler: an r package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*, 16(5):284–287, 2012.
95. Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000.