



# Deciphering the Tissue-Specific Regulatory Role of Intronless Genes Across Cancers

Katia Aviña-Padilla<sup>1</sup> , José Antonio Ramírez-Rafael<sup>1</sup> , Octavio Zambada-Moreno<sup>1</sup> , Gabriel Emilio Herrera-Oropeza<sup>2</sup> , Guillermo Romero<sup>3</sup> , Ishaan Gupta<sup>4</sup> , and Maribel Hernández-Rosales<sup>1</sup>

<sup>1</sup> CINVESTAV-Irapuato, Libramiento Norte Carretera Irapuato León Kilómetro 9.6, 36821 Irapuato, Guanajuato, Mexico  
[maribel.hr@cinvestav.mx](mailto:maribel.hr@cinvestav.mx)

<sup>2</sup> King's College London, Strand, London WC2R 2LS, UK

<sup>3</sup> Data-Pop Alliance, 99 Madison Avenue, New York, NY 11211, USA

<sup>4</sup> Indian Institute of Technology - Delhi, Main Road, IIT Campus, Hauz Khas, New Delhi 110016, India

**Abstract.** Intronless genes (IGs) or single-exon genes lacking introns are found across Eukaryotes. IGs are not regulated by the splicing machinery and may be subject to lower post-transcriptional gene expression variability. Therefore, IGs might be potential candidates for biomarkers with better predictability and easier regulation as targets for therapy. Cancer is a complex disease that relies on progressive uncontrolled cell division linked with multiple dysfunctional biological processes. Tumor heterogeneity remains the most challenging feature in cancer diagnosis and treatment. Given the clinical relevance of IGs, we aim to identify their unique expression profiles and interactome, that may act as functional signatures across eight different cancers. We identified 940 protein-coding IGs in the human genome, of which about 35% were differentially expressed across the analyzed cancer datasets. Specifically, 78% of differentially expressed IGs were undergoing transcriptional reprogramming with elevated expression in tumor cells. Remarkably, in all the studied tumors, a highly conserved induction of a group of deacetylase-histones located in a region of chromosome 6 enriched in nucleosome and chromatin condensation processes. This study highlights that differentially expressed human intronless genes across cancer types are prevalent in epigenetic regulatory roles participating in specific protein-protein interaction (PPI) networks for ESCA, GBM, and LUAD tumors. We determine that IGs play a key role in the tumor phenotype at transcriptional and post-transcriptional levels, with important mechanisms such as interactomics rewiring.

**Keywords:** Single-exon genes · Comparative genomics · Cancer genomics · Epigenetics · Tumor heterogeneity · Cancer evolutionary genomics

---

Supported by CONACYT.

## 1 Introduction

Most eukaryotic gene structures contain exons interrupted by non-coding introns that are removed by RNA splicing to generate the mature mRNA. Although the most prevalent class of genes in the human genome contain Multiple Exon Genes (MEGs), about 5% of the genes are Intronless (IGs) or Single-Exon (SEGs) that lack introns. Due to the absence of introns and associated post-transcriptional splicing, IGs may be subject to lower post-transcriptional gene expression variability suggesting a potential role as clinical biomarkers and drug targets that deserve careful consideration in diseases such as cancer [13, 24, 27]. Several previous studies have identified the role of IGs in cancer [2, 5, 13, 17, 36]. For example the *RPRM* gene increased cell proliferation and tumor suppression activity in gastric cancer [2]; *CLDN8* gene is associated with colorectal carcinoma and renal cell tumors [5] while *ARLTS1* is upregulated in melanoma; *PURA* & *TAL2* are upregulated in leukemia [13] and protein kinase *CK2α* gene is up-regulated in all human cancers [17].

A remarkable instance of IGs acting in a clinical role is *SOX11*, a member of *SOXC* (SRY-related HMG-box) gene family of transcription factors involved in embryonic development and tissue remodeling by participating in cell fate determination [4]. *SOX11* has been associated with tumorigenesis, with aberrant nuclear protein expression in Mantle Cell Lymphoma (MCL) patients [8, 30, 35, 37]. This TF is not expressed in normal lymphoid cells or other mature B cell lymphomas (except Burkitt lymphoma), but it is highly expressed in conventional MCL, including the cyclin D1-MCL subtype [36]. Hence, *SOX11* represents a widely used marker in the differential diagnosis of MCL and other types of small B-cell neoplasias in the clinical hemato-oncology practice [18, 31].

However, to date, a comparative analysis integrating transcriptomic and interactomics profiles of IGs across diverse types of cancer is missing. Hence, this work aims to identify and characterize their expression, functional role and interactomics profiles across different selected cancers.

## 2 Results

In this study we use RNA-sequencing data from 3880 tumor samples belonging to 8 cancer types from the cancer genome atlas or TCGA (Appendix Table 1). We selected the four most prevalent cancers, namely Breast Invasive Carcinoma (BRCA), Colon Adenocarcinoma (COAD), Lung Adenocarcinoma (LUAD) and Prostate Adenocarcinoma (PRAD), along with the four most aggressive cancers with high intrinsic heterogeneity, namely Bladder Urothelial Carcinoma (BLCA), Esophageal Carcinoma (ESCA), Glioblastoma Multiforme (GBM) and Kidney Renal Clear Cell Carcinoma (KIRC). We found that 338 out of the 940 genes identified as IG-encoded proteins are undergoing differential regulation in tumors compared to the normal tissue. GBM had the most number of differentially expressed (DE)-IGs at 168, followed by KIRC at 130, and both PRAD and COAD at 116. While in BRCA 104, LUAD 87, BLCA 97 and ESCA 86 were determined.

## 2.1 Functional Assignment and Gene Expression of IGs in Normal Tissue

First, we performed a functional assignment analysis to characterize the biological role and the expression profiles of IG-encoded proteins in normal tissue. Our results determined that 940 proteins in the human genome are classified as IGs. Secondly, when performing a functional enrichment analysis, it stands out their relevance in chromosome and genetic material organization as well as their response to sensory stimulus in biological processes (Fig. 1).

Furthermore, we examined their constitutive expression on healthy tissue samples corresponding to those of the analyzed tumors. This analysis based on the GTex database, is depicted in Fig. 2. We observe that the identified tissue samples have more similar transcriptional profiles when belonging to the same tissue. There is a clear separation of clusters belonging to the brain, with very different profiles in comparison to the rest of tissues. Kidney cortex, lung, prostate and breast, esophagus mucosa, colon transverse clusters are very well distinguished from the rest; while all the gastrointestinal tissues are clustered in the same expression area. In contrast, bladder samples do not seem to have a tissue-specific expression profile. Hence, IGs have tissue-specific expression in most of the tissues corresponding to the cancer diseases/tumors studied here.

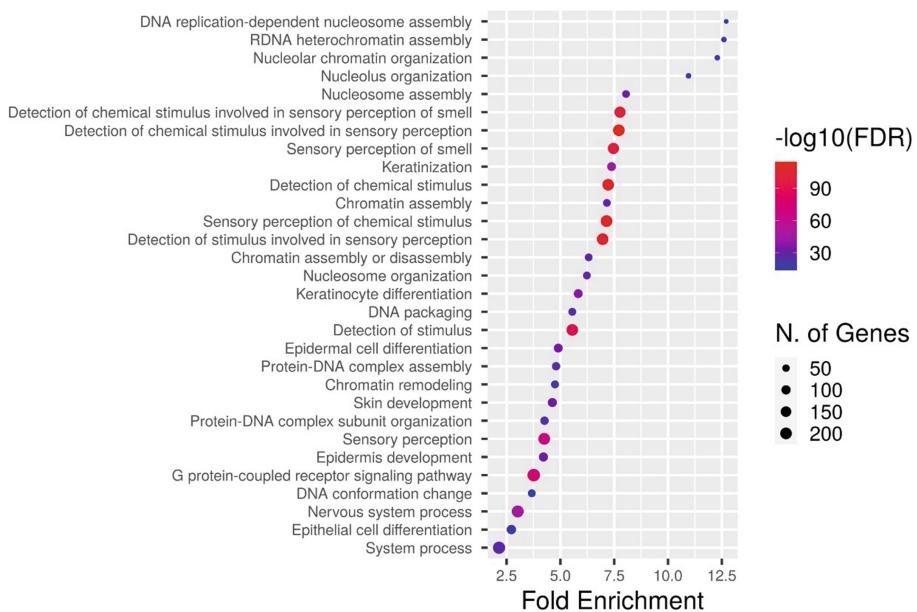
## 2.2 IGs Tend to Have a More Induced Gene Expression Pattern When Compared to MEGs

In order to characterize the distinct biological behavior of IGs, we studied the overall gene expression patterns of IGs compared to those of MEGs in all the tumors. Notably, a greater percentage of upregulated genes are found among DE-IGs than DE-MEGs in all the different types of cancers analyzed.

For instance, when comparing both populations of genes, we identified statistical significance for an over-representation of the up-regulated IGs over the MEGs group among DEGs in PRAD ( $p\text{-value} = 2.8108 \times 10^{-7}$ ), ESCA ( $p\text{-value} = 7.0858 \times 10^{-5}$ ), LUAD ( $p\text{-value} = 0.0015$ ) and BRCA ( $p\text{-value} = 0.0276$ ) cancers. Moreover, we aimed to compare if the upregulation levels are higher in IGs than in MEGs transcripts. Our analysis revealed that in BRCA, BLCA, ESCA and GBM IGs tend to express in higher levels than MEGs, as shown in Appendix Fig. 8.

## 2.3 Upregulated IGs Across Cancer Types Encode for Highly Conserved HDAC Deacetylate Histones Involved in Negative Gene Regulation

To dig insight into the role of the prevalent upregulation mechanism identified for the DE-IGs, we aimed to characterize the groups of induced IGs across the analyzed cancer types. 338 out of 940 (35%) IG encoded proteins in the human genome are deregulated in at least one of the eight analyzed tumor types, where 106 (30%) are found to be up-regulated in one cancer type and down-regulated

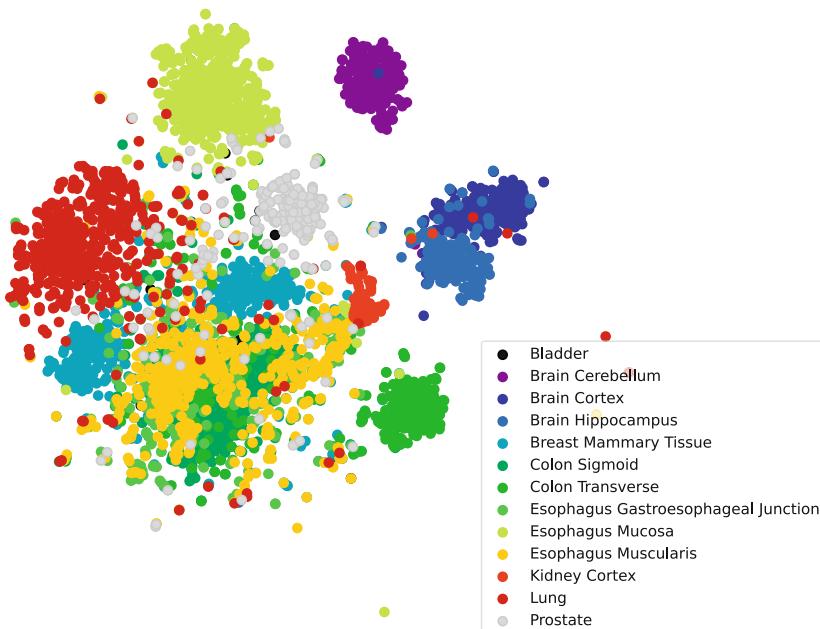


**Fig. 1. Functional enrichment of IG-encoded proteins in the human genome.** Genes are depicted by a dot, the size of the dot represents the number of genes. FDR is calculated based on a nominal *p*-value from the hypergeometric test. Fold Enrichment is defined as the percentage of *introneless genes* belonging to a pathway, divided by the corresponding percentage in the background. FDR reports how likely the enrichment is by chance, higher values are colored on a scale of red to blue. In the x-axis, Fold Enrichment indicates how drastically genes of a certain pathway are overrepresented.

in another type. In a higher number of them, 222 (65%), upregulation is conserved in two or more cancers, which suggests they are undergoing transcriptional reprogramming with higher rates of upregulated levels of their mRNAs as an outcome. Moreover, this upregulation mechanism is highly shared among the different cancers (Fig. 3), for instance, most of the genes upregulated in BLCA and ESCA are shared with the other tumors, while tumors with a significant but less shared upregulation are GBM, PRAD and KIRC, which could be expected given their remarkable heterogeneity (Fig. 3).

To delve insights into the conservation of the activation of this negative gene expression mechanism in the cancer genomes, we identified the most conserved upregulated histones in all the diseases. We identified a group of 13 histone-related genes whose expression is highly conserved among the cancers (Appendix Fig. 10 and Appendix Fig. 11). Moreover, this group of upregulated genes is statistically enriched in a proximal region (26.1968–27.8927 position in Mbps) in chromosome 6 in the human genome (Appendix Fig. 12).

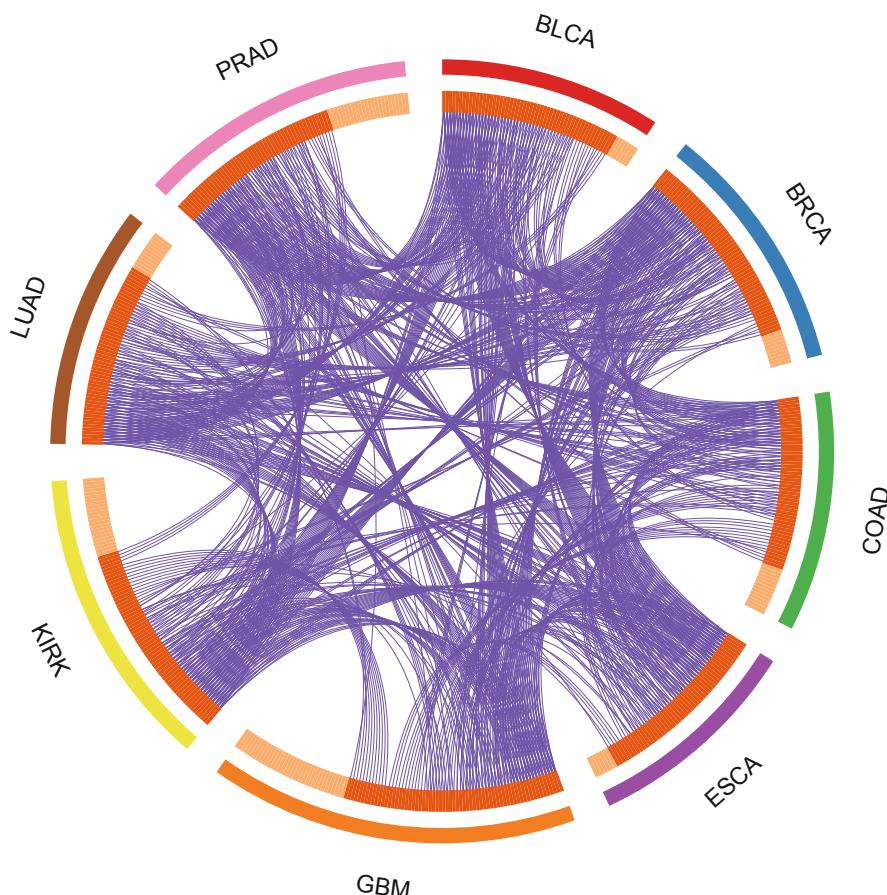
*HASPIN* was found to be the only DE-IG shared among all the tumor genomes, highlighting its relevance in tumorigenesis. This protein kinase is



**Fig. 2.** IG expression profile across healthy tissues. t-SNE algorithm summarises the expression pattern of different samples in a given tissue. Each sample is represented as a dot; color indicates the tissue. Dots distance represents transcriptional profile similarity: the closer they are in the plane, the more similar is their transcriptional profile.

known to be required for histone H3 phosphorylation, necessary for chromosomal passenger complex accumulation at centromeres of mitotic cells [10,34]. In addition to its chromosomal association, it is associated with centrosomes and spindles during mitosis. The overexpression of this gene is related to a delayed progression through early mitosis [10].

Functional enrichment analysis was carried out for the deregulated IGs in all the studied diseases (Appendix Table 2). GBM is the disease with the most DE-IGs and with the most diverse functional roles for the induced genes. The GBM-specific enriched terms are pathways of neurodegeneration, cell-cell adhesion, and gland development. The cancer-specific enriched terms chaperone-mediated protein folding and regulation of neuron apoptotic processes were found for esophagus and colon cancers, respectively. Further, the Reactome pathways *R-HSA-321481* deacetylases histones (HDACs) and *R-HSA-3214858* RMTs methylate histone arginines were also enriched in IGs while the GO:0006335 term *DNA replication-dependent chromatin assembly* was also enriched suggesting an essential role in cancer biology.



**Fig. 3. Upregulated IGs across the analyzed cancer genomes.** The Circos plot depicts how upregulated genes for each cancer overlap. On the outside shell, each arc represents a cancer genome: BLCA in red, BRCA in blue, COAD in green, ESCA in purple, GBM in orange, KIRC in yellow, LUAD in brown, and PRAD in pink. On the inside shell, the dark orange color represents the proportion of upregulated genes shared with other cancers, while the light orange color represents the proportion of genes that are uniquely upregulated in a cancer. Purple arcs represent shared upregulated genes, the greater the number of arcs and the longer the dark orange area imply a greater overlap among the upregulated IGs across cancers. (Color figure online)

## 2.4 IG Downregulation Is Conserved in Breast and Colon Cancers and Is Involved in Signaling and Cell-Specific Functions

When comparing the downregulated IGs and their enriched functional terms among the diseases, it was pointed out that the breast and colon cancers share all the repressed IGs with their functional pattern. Downregulated genes in those cancers are involved in the regulation of cellular localization, sensory organ devel-

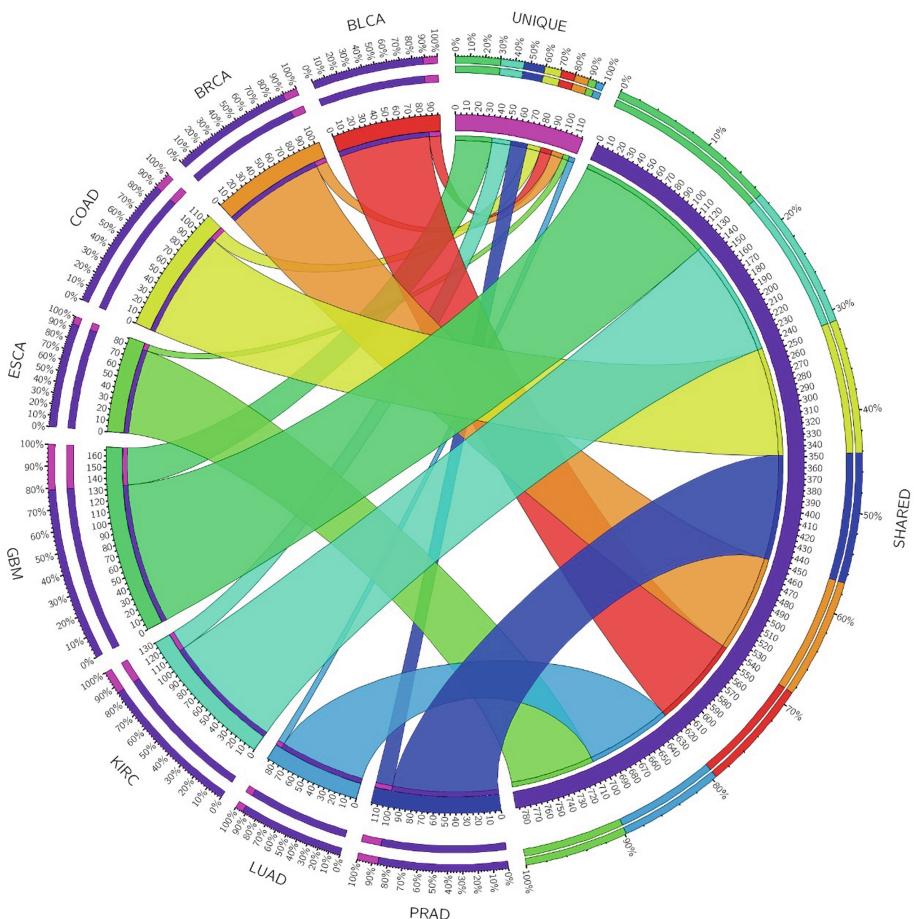
opment, regulation of membrane potential, adipogenesis, epithelial to mesenchymal transition in colorectal cancer, growth, actin filament base process, positive regulation of cell death, class A/1 (rhodopsin-like receptors) and the regulation of secretion by cell biological processes (Appendix Table 2). Notably, the most shared downregulated pathway among cancer genomes is the regulation of anatomical structural size, shared among BRCA, PRAD, GBM, LUAD and BLCA tumors. Additionally, unique roles are found in lung and prostate cancers. They have downregulated genes involved in the vitamin D receptor pathway and the wounding response, respectively.

## 2.5 Cancer-Specific Differentially Expressed IGs

To study the specificity in the regulation of human IGs, we quantified the number of cancer-specific DE-IGs (Fig. 4). We also built a bipartite network to identify cancer-specific and shared deregulated IGs, where upper nodes represent DE-IGs and bottom nodes represent cancer types. We link an IG to a cancer if it was found to be deregulated in that cancer (Appendix Fig. 9). Among the 338 DE-IGs, we identified that 35% were specific for a cancer type (Appendix Table 3). GBM followed by KIRC and PRAD with 35, 18, and 16 DE-IGs, respectively, displayed the highest specificity for IG expression.

**GBM.** Among the specific DE-IGs found in GBM, an evident functional enriched first group includes mainly cellular proliferation and cytoskeleton-related genes. An enrichment of microtubule-based movement, transport along the microtubule, cytoskeleton-dependent intracellular transport and microtubule-based transport is identified ( $FDR = 0.0132$ ). Moreover, the second group of specific deregulated genes is enriched in functions related to tumor suppression, a negative regulatory mechanism of cell growth control, which usually inhibits tumor development. We identified GBM-specific deregulated genes with crucial involvement in the *p53* pathway, such as *KLLN*, *INSM2* and *SFN*.

**KIRC.** The cancer-specific deregulated IGs in KIRC include the tissue-specific *KAAG1* gene (kidney-associated antigen 1) along with the transcription factors *POU5*, *TAL* and *bHLH*. Moreover, transmembrane proteins such as the proto-cadherin beta 10 (*PCDH10*), related to the *Wnt* signaling, were also upregulated. *PCDH10* has been classified as a tumor suppressor in multiple cancers. It induces cell cycle retardation and increases apoptosis by regulating the *p53/p21/Rb* axis and *Bcl-2* expression [19].



**Fig. 4. Shared and unique differentially expressed IGs across cancers.** Circo plot depicts the number of shared DE-IGs across the 8 types of tumors under study. In the left hemisphere, the total number of DE-IGs in each cancer is represented by an arc: BLCA (red), BRCA (orange), COAD (yellow), ESCA (light green), GBM (green), KIRC (turquoise), LUAD (light blue), and PRAD (dark blue). The unique (fuchsia) and shared (purple) groups are depicted in the right hemisphere. The inner arcs represent the distribution of the DE-IGs in each cancer, while the outside arcs represent the intersection of both hemispheres on a scale of 0–100%. Data was obtained from the TCGA database. (Color figure online)

**PRAD.** IGs specifically downregulated in PRAD include *RAB6D* (member RAS oncogene) and *MARS2* (methionyl-TRNA Synthetase 2) related to amino acid metabolism, while *PBOV1*, a 135 amino acid protein with a transmembrane domain was highly upregulated ( $\log FC = 3.1699$ ). This gene was mapped to 6q23-q24, a region associated with loss of androgen dependence in prostate

tumors [3]. Other upregulated IGs include transcription factors such as *FOXB2* (Forkhead Box B2) and *ATOH1* a *bHLH*, which is associated with goblet cell carcinoid and Merkel cell carcinoma diseases [33]. Moreover, *FRAT2*, an essential positive regulator of the *Wnt* signaling and *LENG9*, a member of the conserved cluster of receptors of the leukocyte-receptor complex (*LRC*) were also upregulated [26].

**BRCA.** Few cancer-specific differentially expressed IGs that code for TFs were identified for BRCA, while specific gene expression is more abundant for transmembrane proteins in this cancer, as is the case of a group of glycoproteins and transmembrane receptors involved in signal transduction. For instance, upregulation of *PIGM* (phosphatidylinositol glycan anchor biosynthesis class M), *GP5* (glycoprotein V platelet), *CALML5* (calmodulin-like 5), coiled-coil domain-containing proteins *CCDC96* and *CCDC87* were identified specifically for this cancer.

**LUAD.** Regarding the specific differential expression found for lung tumors, the following TFs are significantly upregulated: *TAF1L*, a transcription factor that is related to chronic inflammatory diseases by regulating apoptotic pathways including regulation of *TP53* activity, and *FOXE1*, a forkhead TF previously related to thyroid cancer [11]. Interestingly, specific transmembrane signaling receptors are upregulated. *MAP10* is a microtubule-associated protein and *THBD*, a thrombomodulin endothelial-specific type I membrane receptor that binds thrombin. Among *THBD* related pathways are collagen- and fibrin clot formation as well as calcium ion binding and cell surface interactions at the vascular wall.

**BLCA.** In BLCA, only *ZXDB* (Zinc Finger X-Linked Duplicated B) TF is identified as specifically deregulated. Transcripts for signal transduction proteins are specifically upregulated for *OR52E8* (Signaling by GPCR), *LRRC10B* (Leucine-Rich Repeat Containing 10B) and *MAS1L*, a proto-oncogene which is a peptide ligand-binding receptor linked to sensor neurons for pain stimuli detection.

**ESCA and COAD.** No specific TFs are specifically differentially expressed for colon and esophageal cancer. Gastrointestinal adenocarcinomas of the tubular gastrointestinal tract, including esophagus, stomach, colon, and rectum, share a spectrum of genomic features, including TF-guided genetic regulation [28]. In agreement with this, we found that transmembrane proteins related to GPCR signaling pathways are specifically undergoing differential expression for both cancers; for instance, in colon, *GPR25* (G protein-coupled receptor 25), *CCDC85B* (coiled-coil domain containing 85B), *CCDC184* (coiled-coil domain containing 184), *SPRR1A* (small proline-rich protein 1A), *PROB1* (proline-rich essential protein 1). Meanwhile, for esophagus tumors, *OR51B4* related to signaling by GPCRs and *LDHAL6B* involved in glucose metabolism and respiratory electron transport were identified.

Remarkably, when comparing IGs deregulation across the eight cancer types, we identified that ESCA and LUAD shared 93% of the DE-IGs, while BRCA, BLCA, COAD, PRAD and KIRC shared more than 86% at least with another cancer. Finally, and presumably given its “*multiforme*” nature, GBM is the cancer type identified with a significant but lesser percentage of shared DE-IGs (79%). Notwithstanding, our results show that GBM represents the tumor where the IGs have more differential expression, specificity and concerted functional roles. This could be explained due to their gene expression tissue-specific relevance in the brain.

## 2.6 Proteins Encoded by Cancer-Specific Deregulated IGs Interact with Distinct Groups of Proteins in PPI Networks

Gene expression is a phenomenon where coupled biochemical interactions take place to transcribe mRNA for protein production. There is a high regulation in the balance of this event. Hence differences in protein composition, production or abundance are a consequence of disruptions in cell phenotypes. The differential expression pattern of the mRNA is key in determining a cell state at the molecular and physiological level [1]. It has been shown that genes involved in “*similar diseases*” share protein-protein interactions (PPI) and higher expression profiling similarity [12].

To determine if the DE-IGs play a crucial role in inhibiting or exacerbating biological reactions at a physical level, we obtained the Protein-Protein Interactions (PPIs) with highest confidence (according to STRING database) where DE-IGs for each cancer are involved. Our results show that a considerable proportion of proteins encoded by DE-IGs in each cancer interacts with specific groups of proteins, due to the low percentage of PPIs that are shared between two or more cancers (Fig. 6). In this analysis, we will consider that an interaction is shared between cancers if a DE-IG in one cancer is also deregulated in another cancer, and the interaction of the IG-encoded protein with another protein is reported by STRING, otherwise, we classify that interaction as cancer-specific.

For instance, 100% of the interactions found in LUAD are cancer-specific, followed by the high specificity of interactions found in ESCA (99.3865%) with only a unique shared PPI of centromere complex with BLCA (*HIST1H2BJ*, *CENPA*). GBM has a similar pattern with 95.8333% of unique interactions and only 19 shared ones. In BLCA, 59.16% of the analyzed PPI are specific to this cancer type. PRAD is a condition that possesses 53.1339% cancer-specific interactions and COAD has 42.1455%. For BRCA, 30.62% of its interactions are only found in this cancer, the lowest fraction of unique interactions. KIRC has the second-lowest with only 31.55% of unique interactions.

If two cancers share PPIs related to the cancer-specific DE-IGs, there could be an underlying affected process characteristic for such diseases. To determine those specific processes, protein complexes involved in shared interactions were examined, finding that most of the identified proteins belong to families of core

histones. Overall, the analyzed PPIs indicate a very distinct pattern of interactomics for each cancer. The tumors that share the greatest proportion of PPIs are breast and colon, exhibiting 145 common interactions which represents 44.61% of all interactions identified for such diseases. The second higher similarity found is for prostate and kidney tumors, sharing 227 interactions (28.64% of all). The rest of the tumors share at most 21.32% interactions (Appendix Fig. 13 and Appendix Table 4).

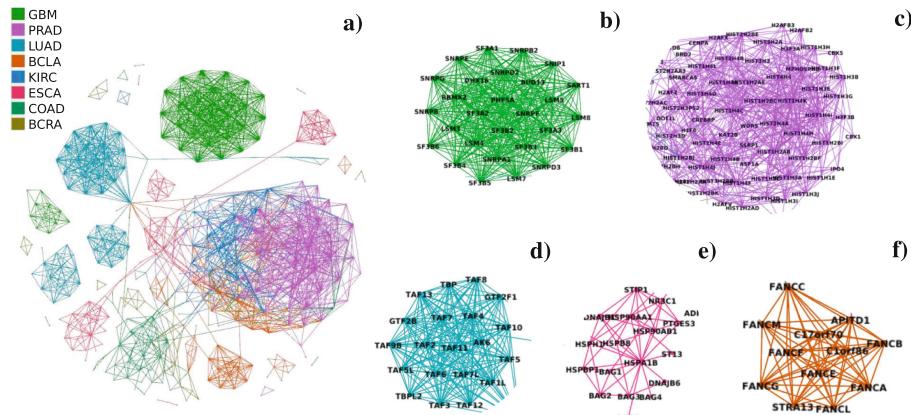
These cancer-specific interactions were analyzed deeper to delve into their functional role in each cancer (Appendix Fig. 14). As it could be expected due to their high specificity at the PPI level, this network approach shows specific and defined clusters for glioblastoma (Fig. 5b) and lung (Fig. 5d). In less-defined clusters, esophageal and bladder cancer can be observed in panels e) and f). In contrast, in the case of prostate, bladder and kidney tumors, the interactions are linked by common DE-IGs in a cluster. On the other hand, we identified that the communities in the network are defined primarily by the type of proteins. Functional enrichment shows specific biological processes intrinsic to the physical interactions implied for the genes in each cancer (Appendix Fig. 14). For instance, proteins involved in BLCA-specific PPIs are conducting mainly chromatin organization and DNA repair reactions, while in colon tumors, proteins play a concerted role in the regulation of transcriptional processes. In keeping with this, in esophageal tissue, the interactors are involved in the regulation of different classes of non-coding RNAs. Meanwhile, in glioblastoma and lung tumors, proteins are specifically interacting for splicing activity and transcription initiation processes, respectively. Kidney and prostate tumors show specific protein-protein interactions for DNA-replication and protein-protein complex formation, and DNA organization and packaging. In contrast, proteins with specific interactions in breast tumors, have an important activity linked to cellular growth and apoptotic processes.

Intrinsic of the PPI networks is the topological information capable of characterizing cancer proteins [16]. Surprisingly, our comparison of IGs with their physical interactors at a protein-protein level shows a tendency to lower betweenness centrality, lower degree and a few interactions with other IGs. Therefore, even though IGs are not hubs in the network of interactions where they participate, changes in their regulatory role can cause a cascade of disruptions in interactions that might lead to malignancy.

Altogether, these results indicate opposite patterns to those reported for cancer genes, which suggest that the behavior tendency of IG proteins is to interact with *oncogenic genes* [16].

For a closer approach, the nearest fifty interactors of the DE-IG encoded proteins in each cancer were analyzed. In Fig. 6 we can observe that cancer-specific DE-IG encoded proteins interact with specific groups of proteins in distinct cancers, suggesting that DE-IGs affect particular post-transcriptional processes.

The most important protein complexes for each type of cancer were determined for cancer-specific deregulated IGs and their interactors (Appendix Table 5). In total, we detected 27 protein complexes where cancer-specific DE-IGs and



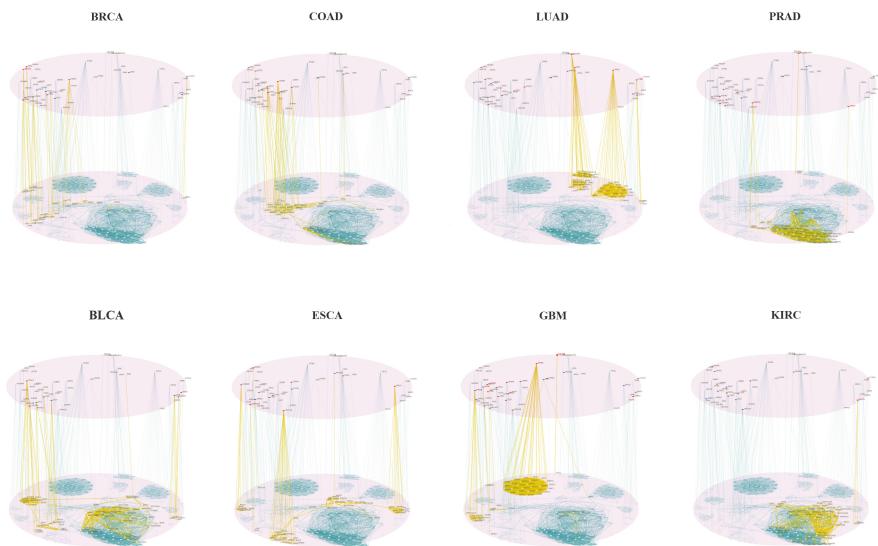
**Fig. 5. Protein-protein interaction network of DE-IGs and their 50 closest interactors across cancers.** a) Unique and shared PPI across cancer are shown, each cancer has a specific color for its interactions. Close-ups of the biggest and more isolated clusters in the network are presented: b) GBM, c) PRAD, d) LUAD, e) ESCA, and f) BLCA. PPI data was retrieved from the String database.

their interactors are found. Colon cancer is the disease containing more DE-IGs in the same complex: 9 out of 13 corresponding to the disassembly of the beta-catenin Wnt signaling pathway. Bladder DE-IGs are interacting mainly with the proteins of the CD40 signaling pathway and are involved in multi-monicellular organism processes. Likewise colon, breast cancer-specific IGs are related to Wnt signaling pathway, but also to follicle-stimulating hormone signaling.

Glioblastoma-specific proteins have a major participation in splicing and vesicle transport. Kidney's are more related to regulation of feeding behavior. Esophageal's IGs are involved in protein folding, and tRNA and ncRNA processing. Lung cancer proteins participate in transcription preinitiation complex assembly, as well as in the BMP signaling pathway. Prostate cancer DE-IGs are found mainly related to chromatid sister cohesion and translation and posttranslational processes. The protein complex related to Wnt signaling pathway is associated with a large group of deregulated genes among Bladder, Breast, Colon, Kidney and Glioblastoma, being a total of 17 IGs with disrupted expression.

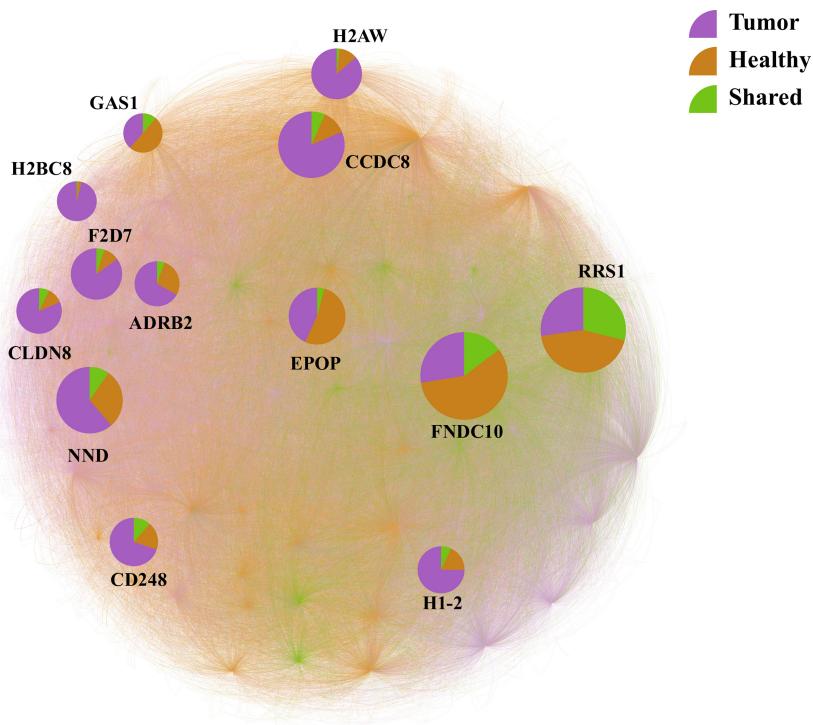
## 2.7 DE-IGs Participate in the Genetic “rewiring” of Cancer Cells

Cancer cells undergo significant genetic “rewiring” as they acquire metastatic traits and adapt to survival in multiple environments with varying nutrient availability, oxygen concentrations and extracellular signals. Therefore, to effectively treat metastatic cancer, it is important to understand the strategies that cancer cells adopt during the metastatic process. Finally, we focus on studying the “rewiring” between healthy and tumor samples using BRCA as a model (Fig. 7). We used breast cancer since it is the only dataset that fits the criteria for a mutual



**Fig. 6. Multilayer networks of IG-encoded proteins and their cancer-specific interactors.** Protein-protein interactions were found for deregulated IGs in each cancer. The cancer-specific deregulated IGs that encode for proteins interact with exclusive groups of proteins in each cancer. Upper layer shows deregulated IGs across cancer, highlighting the cancer-specific upregulated (in red) and downregulated (in blue). Links to the bottom layer connect IG-encoded proteins to proteins found in PPIs by STRING. Yellow links highlight interactions that can be affected by deregulation of IGs in a specific type of cancer. (Color figure online)

information analysis (at least 100 samples for each condition are required). A network of DE-IGs and their co-expressed genes was built and analyzed for each condition. Our results show that the tumor co-expression network is composed of a total of 62,462 interactions among 15,347 genes, while the healthy network has 13,037 genes with 45,941 interactions. There are only 9,615 interactions shared between the two network topologies. All the differences between these two networks are potential rewiring-caused co-expression interactions that may be part of the mechanism that BRCA cells follow to achieve the characteristics of their phenotype. For instance, as seen in Fig. 7, IGs like *RRS1*, *FNDC10*, *EPOP* and *NND* are highly affected by this “rewiring” behavior in cancer. According to differential expression and network analysis, we observe that histones play a key role in rewiring the co-expression mechanism from healthy to cancer tissue. Examples of this are *H2AW* and *H2BC8*, which are core histones that have only few healthy tissue-specific interactions that are almost lost during cancer development, which creates many new interactions with other genes.



**Fig. 7. Co-expression network of healthy and tumoral breast tissue.** In this network, the co-expression of DE-IGs found in BRCA in healthy tissue (depicted by brown links) and breast tumor tissue (purple links) is shown. Shared co-expression relations are depicted with green links. Pie charts for IG nodes of a higher degree are shown. Each of them indicates the percentage of each type of interaction: in purple tumor-specific interaction, in brown healthy-specific interaction, in green conserved interaction. We can see the effect of rewiring by observing that a high percentage of interactions that are present in healthy tissue is not observed in tumor tissue. Moreover, many more new interactions emerge in the tumor tissue. (Color figure online)

### 3 Discussion

Here, we present a comprehensive analysis of differentially expressed Intronless genes (IGs) across eight cancers using a standardized bioinformatics pipeline. We uncover signature changes in gene expression patterns of IGs such as enhanced expression changes when compared to multi-exon genes, unique and shared transcriptional signatures among different cancer types, and association with specific protein-protein interactions that might have potential effects in post-transcriptional processes.

Our results show that IGs tend to have tissue-specific transcriptional profiles in normal tissues. In the case of cancers, the upregulated IGs were related to negative regulation of gene silencing and negative regulation of cell differentiation. Furthermore, we found a core set of chromatin-modifying genes such as *HASPIN* and Histone deacetylases (HDACs), located on chromosome 6, that were upregulated across all the analyzed cancers. We can observe in Fig. 4 that the expression of these genes in cancer differentiate from that of normal tissue in most of the cases, indicating deregulation in the diseased phenotype. On the other hand, the distinct expression among healthy tissues is evident, this may suggest that these histone-related genes present tissue-specific expression levels.

Although the pattern of DE-IGs was found to be not specifically shared across the different tumor types, highlighting their specificity modulating cancer signaling and proliferative pathways. For example, beta-catenin *Wnt* signaling was regulated in kidney and prostate tumors while GPCR signaling critical to inflammation [6,22,23] was regulated in gastrointestinal esophageal and colon adenocarcinomas, and *p53* pathways.

This biological behavior could be due to the tumor's cellular origin and heterogeneity [15], and tissue-specific expression [4,13]. Due to their remarkable heterogeneity, glioblastoma, prostate, and kidney tumors were found with a less representative shared behavior in the upregulation of IGs. Notably, glioma tumors are the malignant microenvironment where IGs showed the most characteristic gene expression regulation profile with 35 unique DE-IGs, more than 50% greater than other cancers. Notwithstanding, our results show that GBM represents the tumor where the IGs have more differential expression, specificity and concerted functional roles. This could be explained due to the "multiforme" nature of GBM that is driven by brain-specific gene expression as is further evidenced by least shared DE-IGs (79%) with other cancers [15].

In conclusion, IGs drive massive transcriptional rewiring, as observed in co-expression analysis, that may drive tumorigenesis and may represent novel therapeutic targets for gene therapy and in-silico drug design such as the popular GPCRs or emerging HDAC inhibitors [9]. Although we focus on 8 cancers based on their heterogeneity and aggressivity, we expect that using our analysis paradigm, IGs may also display characteristic expression profiles in other cancers.

## 4 Materials and Methods

### 4.1 Data Extraction and Curation for IG, and MEG Datasets

Data was extracted using Python scripts ([https://github.com/GEmilioHO/intronless\\_genes](https://github.com/GEmilioHO/intronless_genes)), and the Application Programming Interface (API). Homo sapiens genome was assembled at a chromosome level and was accessed at the Ensembl REST API platform (<http://rest.ensembl.org/>) accessed using Python

with the ensembl\_rest package). The pipeline process was as follows: protein-coding genes with CDS identifiers for transcripts for all chromosomes were retrieved and classified into two datasets named “single-exon genes” (SEGs), and “multiple exon genes” (MEGs) depending on exon and transcript count. The SEG dataset was submitted to the Intron DB comparison (<http://www.nextgenbioinformatics.org/IntronDB>) to separate those with UTR introns and referred to as uiSEGs. The output of the pipeline was a third dataset containing only intronless genes (IGs). After data extraction, a manual curation step in the IG and uiSEG datasets was followed to discard incomplete annotated protein sequences and mitochondrial encoded proteins. The final IG dataset contained 940 protein-coding genes with only one exon and one transcript.

## 4.2 Gene Expression Profiles in Healthy Tissue Tissue

The 424 collected samples of IGs expression (see Sect. 4.4) were scattered in a plane (Fig. 2) by using t-SNE algorithm, which takes samples in a high dimensional space (defined as the expression level of the IGs) and returns a two-dimensional representation where two samples are close if their expression patterns are similar. This process was performed using the Scikit-learn python package [29].

## 4.3 Bipartite Network and Quantification of Shared and Unique DE-IGs

The bipartite network was constructed with upper nodes representing IGs and bottom nodes representing cancer types. We place links connecting each IG to the cancer types where such a gene is differentially expressed (DE). Nodes corresponding to genes in the bipartite network were sorted by degree aimed to visually identify intronless genes cancer-specific deregulated, and those whose expression is disrupted in most of the diseases.

Quantification of shared and cancer-specific deregulated IGs is computed by identifying the number of deregulated IGs in every pair of cancers, and dividing such quantity by the total of the disrupted genes in any of the compared diseases, this metric is known as the Jaccard similarity coefficient. The results are reported as a heatmap.

## 4.4 Data Source and Differential Expression Analysis Across Cancer

Currently, the (The Cancer Genome Atlas) TCGA possesses data for the study of 37 different tumor types. We selected 8 types of cancer primary malignant tumors and their respective adjacent tissues for the present study. Gene expression data from patients for BRCA, BLCA, COAD, ESCA, GBM, KIRC, LUAD

and PRAD cancers was downloaded from the NIH website (<https://portal.gdc.cancer.gov/>) using the TCGAbiolinks R package [7] with the following restriction criteria: samples types primary tumor and solid tissue normal (control); results of RNAseq experimental strategy; and workflow type HTSeq-counts format. Differential expression analysis was carried out using the TCGAutilis R package [25], indicating which of the obtained samples correspond to tumors and which to control, and establishing a filtering threshold of FDR = 0.05 and logFC = 1 (absolute value) to consider a gene significantly differentially expressed. For comparative analysis, normal healthy tissue data was obtained from the dataset gtex.v8 of the GTEx Portal in June 2020 by using the GTEx API. We found gene expression data for 424 IGs in 13 different tissues: bladder (21 samples), brain cerebellum (241), brain cortex (255), brain hippocampus (197), breast mammary tissue (459), colon sigmoid (373), colon transverse (406), esophagus gastroesophageal junction (375), esophagus mucosa (555), esophagus muscularis (515), kidney cortex (85), lung (578), and prostate (245).

#### 4.5 Upregulation Significant Differences of IGs and MEGs Among Cancers

We obtained a dataset of all significantly upregulated IGs and MEGs for each type of cancer studied, in these datasets we compared IGs and MEGs gene expression employing Python scripts using a hypergeometric test. In order to find significant differences in upregulation between IGs and MEGs, the Wilcoxon test was carried out for each type of tumor. Additionally, to confirm that the datasets fulfilled all assumptions a Levene's test was performed for each individual cancer data set.

#### 4.6 Functional Enrichment Analysis of Differentially Expressed IGs

The functional enrichment was conducted using the over-representation analysis of the functional assignment (ORA). Genes with differential expression up to one log2-fold change values were considered as up-regulated with a *p*- and *q-value* set at 0.05 and 0.10, respectively. First, the functional enrichment of the 338 differentially expressed human IG proteins (up-regulated and downregulated separately) was performed using all human IGs proteins as a background “universe” (selecting input as species: *Homo sapiens*, universe 940 human IGs). The comparative functional enrichment analyses were performed using Metascape (<https://metascape.org/>) for the biological process category, including KEGG and Reactome pathways. To delve insight into the role of specific IGs in the affected biological processes ORA was assessed to determine category barplots. The Circos software [20] was employed for data analysis and visualization.

#### 4.7 DE-IGs PPI Network Construction and Protein Complex Identification

Network analysis and visualization were performed using python scripts (See repository <https://gitlab.com/jarr.tecn/de-igs-cancers/-/tree/master>) and the Gephi software. STRINGDB platform [32] was used to download physical and functional interactions data. The highest confidence scores (0.9) were filtered for this study, keeping the most probable interactions. Then, interactions were requested for the set of unique DE-IG of each cancer type, downloading relations between those genes, and also their interactors in the first shell (up to 50 interactors). All this data is assembled into a single network. Network metrics such as degree distribution, closeness, and betweenness centralities were computed using Python networkx library [14]. For protein complex detection, we assumed that highly clustered nodes in the PPI network correspond to protein complexes. Therefore, 27 protein complexes were predicted using the Louvain method to find communities implemented in Gephi.

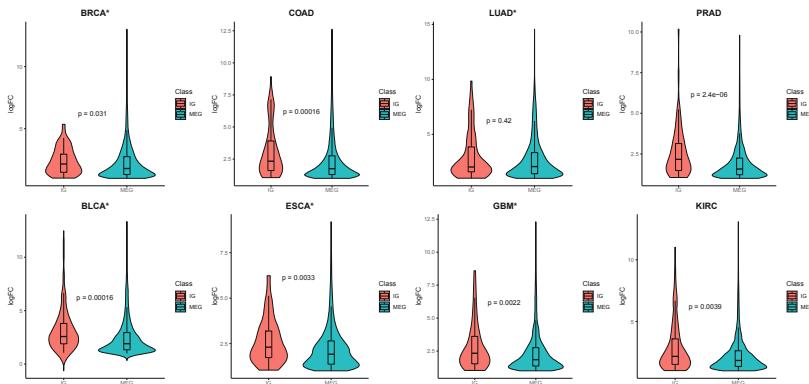
#### 4.8 BRCA Network Deconvolution

Co-expression networks were inferred using ARACNe-AP [21], a mutual information-based tool for gene regulatory network deconvolution. In this analysis, we used separated submatrices for healthy and tumor tissue and the list of every DE-IG in the BRCA dataset. For this study, a *p-value* of  $1 \times 10^{-8}$  was set up and 100 bootstraps were carried out. Then, they were consolidated into a single network, getting an inferred co-expression network for each condition.

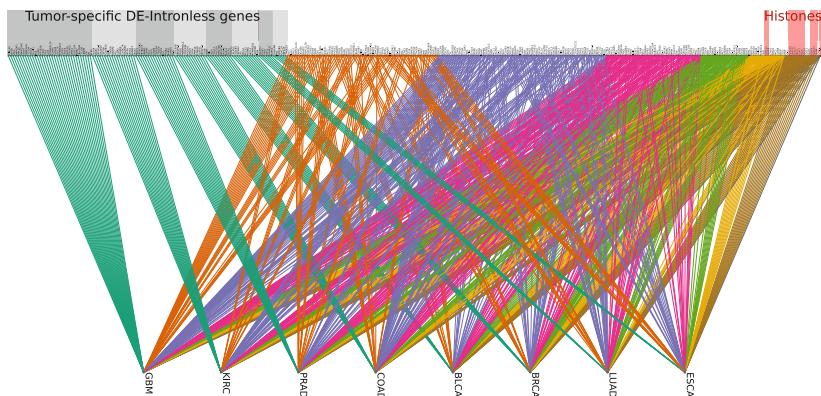
**Acknowledgments.** We would like to thank Roddy Jorquera and Carolina González for fruitful discussions. Special thanks to Fernando Flores for helping with designing and constructing the co-expression network. We are also thankful to Carlos González for visualization and technical support.

This research was funded by Conacyt Ciencia Basica Project 254206. K.A.P (CVU:227919), J.A.R.R (CVU: 1147711), and O.Z.M (CVU: 1147042) received financial support from CONACyT. K.A.P is a current holder of a fellowship from the Fulbright Comexus García-Robles foundation.

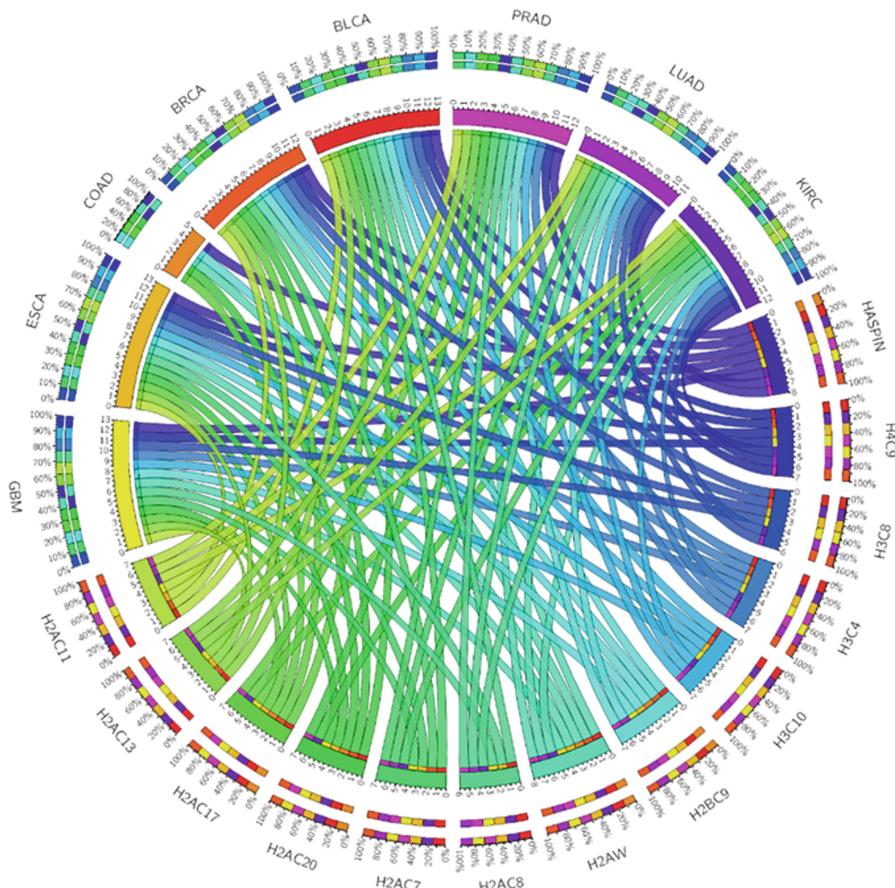
## Appendix



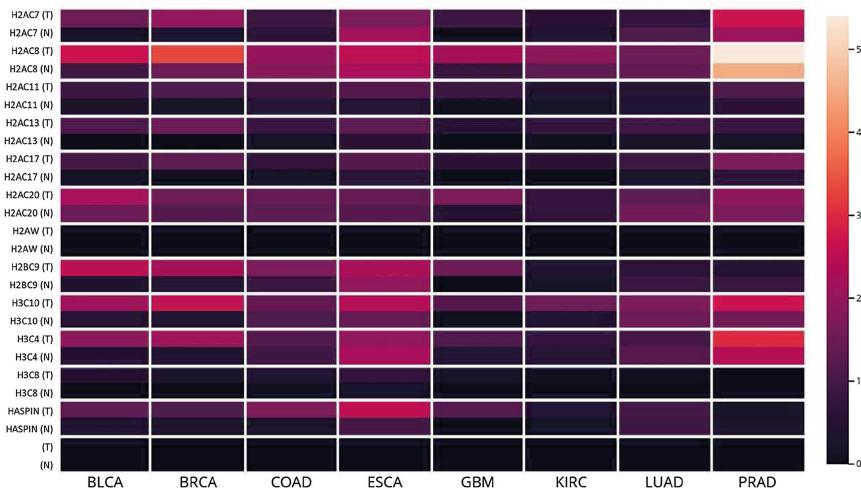
**Fig. 8. Upregulated gene expression levels among cancers.** A violin plot is shown for each cancer type, separated by gene type (IGs or MEGs); expression levels are shown on the y-axis using log<sub>2</sub>fold change values. The *p*-value for the Wilcoxon test is shown in order to indicate the statistical difference between both populations. Cancer types with asterisk represent those datasets that meet all the assumptions of the Wilcoxon test for this comparison. Levene's test was performed to measure shape similarity between IG and MEG groups. Only up-regulated genes log fold change equal or greater than 1 are selected.



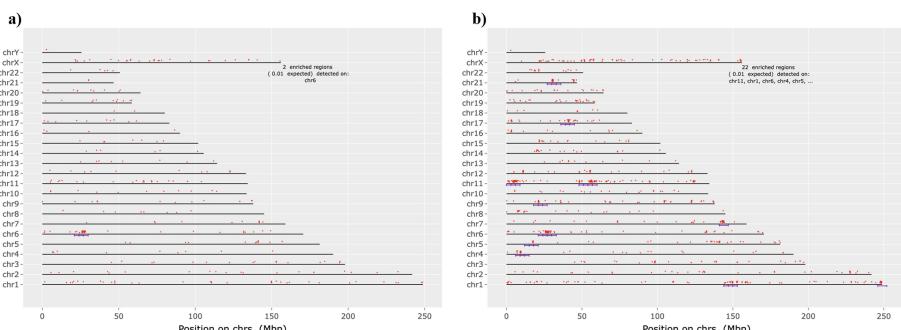
**Fig. 9. IG-cancer type bipartite network.** Intronless genes are shown in the upper group of nodes, while cancers are located in the bottom group. A link joining an IG and a cancer type indicates that the expression level of the IG is found disrupted in the linked cancer type. Upper nodes are color-sorted by their cancer specificity: genes deregulated in a single cancer are shown at the left in blue, followed by genes deregulated in two cancers, three cancers, and so on. At the right of the upper nodes, there is a single gene (*HASPIN*) whose expression is disrupted in all tissues. In general, all histones are deregulated in almost all cancers. (Color figure online)



**Fig. 10. A high conserved group of up-regulated histones across tumors.** Circo plot depicts the conservation of the deregulation of a group of histones across cancers. In the upper hemisphere, the total number of histone-related DE-IGs in each cancer is represented by an arc: GBM (yellow), ESCA (mustard), COAD (orange), BRCA (dark orange), BLCA (red), PRAD (pink), LUAD (violet), and KIRC (purple). Histone encoded IGs are depicted in the lower hemisphere. The inner arcs represent the distribution of the histones in each cancer (0–13 genes), while the outside arcs represent the tumors that have that particular gene upregulated (on a scale of 0–100%). Data was obtained from the TCGA database. (Color figure online)



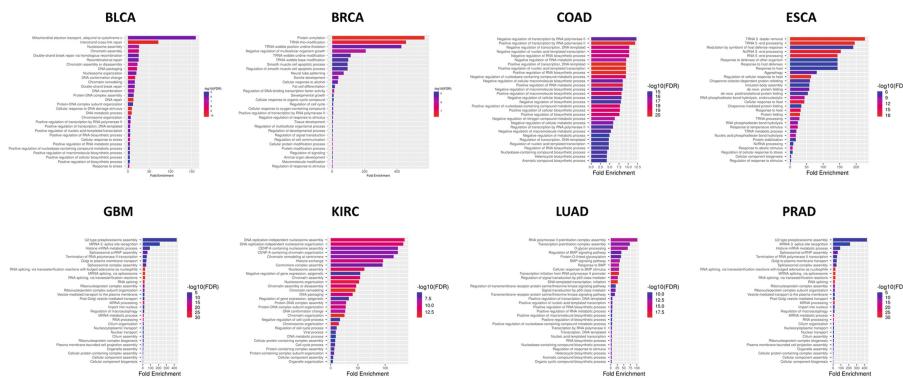
**Fig. 11. Expression of histones among tumors and normal tissues.** Heatmap depicting a comparison of the expression of the 13 histone-related genes identified with upregulated conserved expression among the studied cancers. The expression of each gene is shown for each type of tumor (T), as well as its expression in normal tissue from GTEx data (N). The color key from darker to light indicates low to high transcript per million (TPM) values.



**Fig. 12. Comparison of genome location of highly conserved DE-IGs.** a) Genome location of IGs in the human genome. b) Genome location of highly conserved DE-IGs. DE-IGs are represented by red dots. The purple lines indicate regions where these genes are statistically enriched, compared to the density of genes in the background. The hypergeometric test is used to determine if the presence of the genes is significant. Essentially, the genes in each region define a gene set/pathway. The chromosomes may be only partly shown as the last gene's location to draw the line is used. Data was obtained from the ENSEMBL database.



**Fig. 13. Shared PPI heatmap.** Color intensity indicates a higher percentage of shared PPI related to DE IGs of the compared diseases. LUAD does not share any DE IG interaction with other cancer types, and ESCA only one with BLCA. GMB shares almost no interaction with any other cancer (a maximum of 1.64% with BLCA). COAD and BRCA are the two cancers where more DE IGs are shared, therefore having more interactions in common.



**Fig. 14. PPI networks ontologies.** FDR is calculated based on nominal  $p$ -value from the hypergeometric test. Fold Enrichment is defined as the percentage of genes in each cancer belonging to a pathway, divided by the corresponding percentage in the background. FDR indicates how likely the enrichment is by chance. In x-axis the Fold Enrichment indicates how drastically genes of a certain pathway are overrepresented. FDR values are depicted red (high) to blue (low). (Color figure online)

**Table 1.** Available expression data for each type of cancer

Cancer type	Primary solid tissue samples	Solid tissue normal samples
BLCA	414	19
BRCA	1102	113
COAD	478	41
ESCA	161	11
GBM	156	5
KIRC	538	72
LUAD	533	59
PRAD	498	52

**Table 2.** Gene Ontology (GO) enrichment of deregulated IGs for each cancer

Upregulated IGs		Log10P							
GO	Description	BLCA	BRCA	COAD	ESCA	GBM	KIRC	LUAD	PRAD
GO:0061077	Chaperone-mediated protein folding	0.00	0.00	0.00	-3.54	0.00	0.00	0.00	0.00
GO:0043523	Regulation of neuron apoptotic process	0.00	0.00	-3.41	0.00	0.00	0.00	0.00	0.00
GO:0000280	Nuclear division	0.00	0.00	0.00	0.00	0.00	0.00	-3.40	0.00
GO:0000904	Cell morphogenesis involved in differentiation	0.00	0.00	0.00	0.00	0.00	0.00	-3.68	0.00
GO:0030900	Forebrain development	0.00	0.00	0.00	0.00	0.00	0.00	-2.19	0.00
GO:0002009	Morphogenesis of an epithelium	0.00	0.00	0.00	0.00	0.00	-2.73	-3.17	0.00
R-HSA-420499	Class C/3 (Metabotropic glutamate/pheromone receptors)	0.00	0.00	0.00	0.00	0.00	-3.50	0.00	0.00
GO:0008285	Negative regulation of cell population proliferation	0.00	0.00	0.00	-2.68	-4.14	0.00	-2.46	0.00
GO:0007423	Sensory organ development	-2.60	0.00	0.00	0.00	-5.21	0.00	-2.38	0.00
hsa05022	Pathways of neurodegeneration - multiple diseases	0.00	0.00	0.00	0.00	-3.00	0.00	0.00	0.00
GO:0045669	Positive regulation of osteoblast differentiation	0.00	0.00	0.00	0.00	-3.79	0.00	0.00	0.00
GO:0098609	Cell-cell adhesion	0.00	0.00	0.00	0.00	-2.57	0.00	0.00	0.00
GO:0048732	Gland development	0.00	0.00	0.00	0.00	-4.33	0.00	0.00	0.00
GO:1903706	Regulation of hemopoiesis	0.00	-2.54	0.00	-2.08	-3.87	-2.13	0.00	0.00
GO:0045596	Negative regulation of cell differentiation	0.00	-2.02	-2.12	-2.33	-2.34	-2.23	0.00	0.00
GO:0016570	Histone modification	-2.34	-2.14	0.00	-2.37	-3.26	0.00	-2.19	-3.94
GO:0007517	Muscle organ development	-3.88	0.00	0.00	-2.74	-2.82	-2.20	-2.56	0.00
R-HSA-3214815	HDACs deacetylate histones	-10.77	-12.90	0.00	-9.84	-5.83	-9.04	-3.25	-15.15
R-HSA-3214858	RMTs methylate histone arginines	-5.62	-7.73	0.00	-7.63	-4.60	-8.42	-2.90	-8.98
GO:0006335	DNA replication-dependent chromatin assembly	-2.26	-4.69	0.00	-3.15	-2.65	-5.06	0.00	-4.32
Downregulated IGs		Log10P							
GO	Description	BLCA	BRCA	COAD	ESCA	GBM	KIRC	LUAD	PRAD
GO:0060341	Regulation of cellular localization	0.00	-3.03	-3.03	0.00	0.00	0.00	0.00	0.00
GO:0007423	Sensory organ development	0.00	-2.85	-2.85	0.00	0.00	0.00	0.00	0.00
GO:0042391	Regulation of membrane potential	0.00	-4.93	-4.93	0.00	0.00	0.00	0.00	0.00
WP236	Adipogenesis	0.00	-4.11	-4.11	0.00	0.00	0.00	0.00	0.00
WP4239	Epithelial to mesenchymal transition in colorectal cancer	0.00	-3.87	-3.87	0.00	0.00	-3.26	0.00	0.00
GO:0040007	Growth	0.00	-5.90	-5.90	0.00	0.00	-2.70	0.00	-3.01
GO:0030029	Actin filament-based process	-2.04	-2.08	-2.08	0.00	0.00	0.00	0.00	-4.23

(continued)

**Table 2.** (*continued*)

Upregulated IGs									
GO	Description	Log10P							
		BLCA	BRCA	COAD	ESCA	GBM	KIRC	LUAD	PRAD
GO:0010942	Positive regulation of cell death	-2.04	-2.08	-2.08	0.00	0.00	0.00	0.00	-3.01
GO:0090066	Regulation of anatomical structure size	-2.70	-8.04	-8.04	0.00	-3.49	0.00	-2.44	-2.57
R-HSA-373076	Class A/1 (Rhodopsin-like receptors)	0.00	-2.36	-2.36	0.00	-3.90	0.00	-4.62	-2.16
GO:0042752	Regulation of circadian rhythm	0.00	0.00	0.00	0.00	-2.77	0.00	0.00	0.00
GO:0009725	Response to hormone	0.00	0.00	0.00	0.00	-3.11	-2.70	0.00	0.00
GO:0043269	Regulation of ion transport	-2.70	0.00	0.00	0.00	-3.49	0.00	0.00	0.00
GO:0007193	Adenylate cyclase-inhibiting G protein-coupled receptor signaling pathway	-2.66	0.00	0.00	0.00	-3.01	0.00	-3.41	0.00
GO:1903530	Regulation of secretion by cell	-2.58	-2.62	-2.62	0.00	-3.30	0.00	0.00	0.00
GO:0008285	Negative regulation of cell population proliferation	0.00	0.00	0.00	0.00	0.00	-2.16	0.00	0.00
GO:2000027	Regulation of animal organ morphogenesis	0.00	0.00	0.00	0.00	0.00	-2.84	0.00	0.00
GO:0030162	Regulation of proteolysis	0.00	0.00	0.00	0.00	0.00	-3.98	0.00	0.00
GO:0009611	Response to wounding	0.00	0.00	0.00	0.00	0.00	0.00	0.00	nan
WP2877	Vitamin D receptor pathway	0.00	0.00	0.00	0.00	0.00	0.00	-3.04	0.00

**Table 3.** Cancer-specific deregulated genes

Cancer	Cancer-specific deregulated genes	Number of genes
BLCA	ZXDB, TRIL, ZXDA, FTHL17, OR52E8, LRRC10B, KRTAP2-3, ERVV-1, MAS1L, OXER1, HDGFL1	11
BRCA	CALML5, CCDC87, CAPZA3, GP5, ASCL4, MOCS3, PKDREJ, H4C12, MTRNR2L8, CCDC96, PIGM	11
COAD	CCDC85B, KRTAP3-3, PROB1, HNRNPA1L2, CCDC184, MAGEH1, GPR25, P2RY4, SPRR1A, KCNA10, DUSP21, KRTAP1-1, RTL8B	13
ESCA	LDHAL6B, IFNE, OR51B4, H2AC14, RNF225, CTAGE15	6
GBM	PPIAL4A, RAP2B, SOX3, ANKRD63, SPPL2C, ANKRD34C, ATOH7, CSNK1A1L, TTC30A, H3C14, OR14I1, IRS4, SF3B5, PPP1R2B, FFAR1, F8A3, KLLN, MTRNR2L10, PGAM4, KLHL34, USP27X, MTRNR2L6, SSTR4, EXOC8, H2AC19, ATP5MGL, GPR52, BLOC1S4, INSM2, C6orf226, TTC30B, CENPB, SFN, SRY, MC4R	35
KIRC	OR2A4, NPAP1, KRTAP5-8, TAS2R13, TSPYL6, PCDHB10, DDI1, CTAGE9, POU5F2, TAL2, ELOA2, TAS2R3, KLHDC7A, TAS2R30, RXFP4, KAAG1, MAGEF1, EBLN2	18
LUAD	GALNT4, FOXE1, THBD, TAF1L, PPP4R3C, MAP10	6
PRAD	PBOV1, KRTAP13-2, KBTBD13, RAB6D, KDM4E, H3C1, FRAT2, LENG9, CCDC54, CTAGE8, ATOH1, IER5, FOXB2, MARS2, SUMO4, H1-6	16

**Table 4. Shared PPIs.** The first column contains pairs of cancers, followed by the number of PPI related to unique DE IG of the compared diseases, finally, in the third column is shown the percentage of shared interactions. Note that two cancer may share more genes than other couples of genes but have a smaller percentage, this is due to a difference in the total number of interactions associated with the compared tumors, for example, PRAD and KIRC share 227 PPIs, representing a percentage of 28.64, while COAD and BRCA have less common interactions (145) but a bigger percentage: 44.61.

Cancers	Percentage of total interactions in compared cancers	Interactions shared
COAD, BRCA	44.62	145
KIRC, PRAD	28.65	277
KIRC, BRCA	21.32	132
KIRC, COAD	20.75	138
BLCA, KIRC	14.17	118
BRCA, PRAD	11.64	95
COAD, PRAD	10.94	95
BLCA, PRAD	9.35	95
BLCA, BRCA	7.48	43
BLCA, COAD	6.86	43
BLCA, GBM	1.65	14
PRAD, GBM	1.31	15
KIRC, GBM	0.91	9
BRCA, GBM	0.30	2
COAD, GBM	0.28	2
BLCA, ESCA	0.18	1

**Table 5.** Identified PPI complexes

Protein complex	Enriched functions	Cancers	FDR
CD40 signaling pathway	Regulation of protein monoubiquitination; Regulation of CD40 signaling pathway	BLCA	1.80E-02
Multi-monicellular organism process	Response to corticosterone; Multi-monicellular organism process; Myeloid leukocyte differentiation; Positive regulation of transcription by RNA polymerase II	BLCA	1.40E-05
Electron transport chain	Mitochondrial electron transport, ubiquinol to cytochrome c; Pons development; Pyramidal neuron development	BLCA	6.70E-03
Chromatid sister cohesion	Positive regulation of maintenance of sister chromatid cohesion; Positive regulation of maintenance of mitotic sister chromatid cohesion; Regulation of maintenance of sister chromatid cohesion	BLCA, PRAD	5.10E-05
Dendrite morphogenesis	Positive regulation of dendrite morphogenesis; Regulation of dendrite morphogenesis; Regulation of mitotic nuclear division	BRCA	5.70E-04
tRNA modifications	Protein urmylation; tRNA thio-modification; tRNA wobble position uridine thiolation	BRCA	1.50E-10
Protein localization to nucleolus	Ribosomal large subunit export from nucleus; Protein localization to nucleolus	BRCA	5.50E-03

(continued)

**Table 5.** (*continued*)

Protein complex	Enriched functions	Cancers	FDR
Wnt signaling pathways	Canonical Wnt signaling pathway; Wnt signaling pathway; Cell-cell signaling by wnt	KIRC, COAD, BRCA, BLCA, GBM	3.90E-19
Follicle-stimulating hormone signaling	Desensitization of G protein-coupled receptor signaling pathway by arrestin; Norepinephrine-epinephrine-mediated vasodilation involved in regulation of systemic arterial blood pressure; Follicle-stimulating hormone signaling pathway; Negative regulation of multicellular organism growth	BRCA	6.50E-05
GPI anchor	Preassembly of GPI anchor in ER membrane; GPI anchor biosynthetic process; GPI anchor metabolic process	BRCA	3.40E-08
Hormone-mediated apoptosis	Hormone-mediated apoptotic signaling pathway; Somatostatin receptor signaling pathway; Somatostatin signaling pathway; Positive regulation of T cell anergy	COAD	2.20E-02
Light stimulus	Wnt signaling pathway, calcium modulating pathway; Detection of light stimulus; Phototransduction	ESCA	1.10E-03
Vasodilatation	Glutamate catabolic process to 2-oxoglutarate; Response to transition metal nanoparticle; Negative regulation of collagen biosynthetic process; Regulation of angiotensin levels in blood; Angiotensin maturation	ESCA	2.10E-02
Protein folding	Regulation of cellular response to heat; Chaperone cofactor-dependent protein refolding; Inclusion body assembly; de novo protein folding	ESCA	4.10E-08
Heart morphogenesis	Coronary vein morphogenesis; Negative regulation of cell proliferation involved in heart valve morphogenesis; Mitral valve formation; Cardiac right atrium morphogenesis; Cell proliferation involved in heart valve morphogenesis	ESCA	9.90E-03
tRNA and ncRNA processing	tRNA 5'-leader removal; tRNA 5'-end processing; ncRNA 5'-end processing; Endonucleolytic cleavage involved in tRNA processing	ESCA	6.90E-03
Splicing	U2-type pre-spliceosome assembly; mRNA 3'-splice site recognition; Histone mRNA metabolic process; Spliceosomal snRNP assembly	GBM	9.40E-07
Habituation	Habituation; Nonassociative learning; Regulation of glycogen (starch) synthase activity	GBM	1.20E-02
Vesicle transport	Golgi to plasma membrane transport; Vesicle docking involved in exocytosis; Vesicle docking; Exocytic process	GBM	8.50E-07
Intracellular transport	Intracellular anterograde transport; Negative regulation of myotube differentiation; Intracellular transport	GBM	1.90E-02
Regulation of feeding behavior	Positive regulation of feeding behavior; Regulation of feeding behavior; Positive regulation of behavior; Negative regulation of myotube differentiation	KIRC	5.00E-04
Transcription preinitiation complex assembly	RNA polymerase II preinitiation complex assembly; Transcription preinitiation complex assembly	BLCA, LUAD, GBM	2.90E-15
BMP signaling pathway	Pharyngeal system development; Endocardial cushion development; Positive regulation of bone mineralization; Negative regulation of BMP signaling pathway	LUAD	8.80E-12
Innate immune response	Innate immune response activating cell surface receptor signaling pathway; Innate immune response-activating signal transduction; Activation of innate immune response	LUAD	8.90E-13
Blood coagulation system	Positive regulation of blood coagulation; Positive regulation of hemostasis; Positive regulation of coagulation	LUAD	6.30E+00
Translation and posttranslational processes	Translation; Peptide biosynthetic process; Amide biosynthetic process	PRAD	1.30E-03
Beta-catenin destruction complex	Negative regulation of type B pancreatic cell development; Superior temporal gyrus development; Beta-catenin destruction complex assembly; Regulation of type B pancreatic cell development; Beta-catenin destruction complex disassembly	PRAD	4.10E-02

## References

1. Ademuwagun, I., et al.: Contributors. In: Forero, D.A., Patrinos, G.P. (eds.) *Genome Plasticity in Health and Disease*, pp. xi–xiv. Translational and Applied Genomics, Academic Press (2020). <https://doi.org/10.1016/B978-0-12-817819-5.01002-0>. <https://www.sciencedirect.com/science/article/pii/B9780128178195010020>

2. Amigo, J.D., et al.: The reproto gene family: a novel gene lineage in gastric cancer with tumor suppressive properties. *Int. J. Mol. Sci.* **19**(7), 1862 (2018). <https://doi.org/10.3390/ijms19071862>. <https://www.mdpi.com/1422-0067/19/7/1862>
3. An, G., et al.: Cloning and characterization of UROC28, a novel gene overexpressed in prostate, breast, and bladder cancers1. *Cancer Res.* **60**(24), 7014–7020 (2000)
4. Aviña-Padilla, K., et al.: Evolutionary perspective and expression analysis of intronless genes highlight the conservation of their regulatory role. *Front. Genet.* **12**, 1101 (2021). <https://doi.org/10.3389/fgene.2021.654256>. <https://www.frontiersin.org/article/10.3389/fgene.2021.654256>
5. Bujko, M., Kober, P., Mikula, M., Ligaj, M., Ostrowski, J., Siedlecki, J.: Expression changes of cell-cell adhesion-related genes in colorectal tumors. *Oncol. Lett.* **9**, 2463–2470 (2015). <https://doi.org/10.3892/ol.2015.3107>. <https://doi.org/10.3109/10428194.2010.514968>. pMID 20919851
6. Chen, K., Bao, Z., hua Gong, W., Tang, P.C., Yoshimura, T., Wang, J.M.: Regulation of inflammation by members of the formyl-peptide receptor family. *J. Autoimmun.* **85**, 64–77 (2017). <https://doi.org/10.1016/j.jaut.2017.06.012>. <https://pubmed.ncbi.nlm.nih.gov/28689639/>
7. Colaprico, A., et al.: TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**(8), e71–e71 (2015). <https://doi.org/10.1093/nar/gkv1507>
8. Croci, G.A., et al.: Reproducibility of histologic prognostic parameters for mantle cell lymphoma: cytology, Ki67, p53 and SOX11. *Virchows Arch.* **477**(2), 259–267 (2020). <https://doi.org/10.1007/s00428-020-02750-7>
9. Dai, E., Zhu, Z., Wahed, S., Qu, Z., Storkus, W., Guo, Z.S.: Epigenetic modulation of antitumor immunity for improved cancer immunotherapy. *Mol. Cancer* **20**, 141–152 (2021). <https://doi.org/10.1186/s12943-021-01464-x>. <https://link.springer.com/article/10.1007/s10555-020-09944-0>
10. Dai, J., Sultan, S., Taylor, S.S., Higgins, J.M.: The kinase haspin is required for mitotic histone H3 Thr 3 phosphorylation and normal metaphase chromosome alignment. *Genes Dev.* **19**, 472–488 (2005). <https://doi.org/10.1101/gad.1267105>
11. Denny, J.C., et al.: Variants near foxe1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenotype-wide studies. *Am. J. Hum. Genet.* **89**(4), 529–542 (2011). <https://doi.org/10.1016/j.ajhg.2011.09.008>. <https://www.sciencedirect.com/science/article/pii/S0002929711003983>
12. Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M., Barabási, A.L.: The human disease network. *Proc. Natl. Acad. Sci.* **104**(21), 8685–8690 (2007). <https://doi.org/10.1073/pnas.0701361104>. <https://www.pnas.org/doi/abs/10.1073/pnas.0701361104>
13. Grzybowska, E.A.: Human intronless genes: functional groups, associated diseases, evolution, and mrna processing in absence of splicing. *Biochem. Biophys. Res. Commun.* **424**(1), 1–6 (2012). <https://doi.org/10.1016/j.bbrc.2012.06.092>. <https://www.sciencedirect.com/science/article/pii/S0006291X12011874>
14. Hagberg, A., Swart, P., Chult, D.: Exploring network structure, dynamics, and function using networkx, pp. 11–15 (2008)
15. Herrera-Oropeza, G.E., Angulo-Rojo, C., Gástelum-López, S.A., Varela-Echavarría, A., Hernández-Rosales, M., Aviña-Padilla, K.: Glioblastoma multiforme: a multi-omics analysis of driver genes and tumour heterogeneity. *Interface Focus* **11**(4), 20200072 (2021). <https://doi.org/10.1098/rsfs.2020.0072>. <https://royalsocietypublishing.org/doi/abs/10.1098/rsfs.2020.0072>

16. Huang, S.H., Lo, Y.S., Luo, Y.C., Tseng, Y.Y., Yang, J.M.: A homologous mapping method for three-dimensional reconstruction of protein networks reveals disease-associated mutations. *BMC Syst. Biol.* **12**(S2), 8685–8690 (2018)
17. Hung, M.S., et al.: Functional polymorphism of the ck2 $\alpha$  intronless gene plays oncogenic roles in lung cancer. *PLoS ONE* **5**(7), 1–10 (2010). <https://doi.org/10.1371/journal.pone.0011418>
18. Inamdar, A.A., et al.: Mantle cell lymphoma in the era of precision medicine-diagnosis, biomarkers and therapeutic agents. *Oncotarget* **7**(30), 48692–48731 (2016). <https://doi.org/10.18632/oncotarget.8961>. <https://www.oncotarget.com/article/8961/>
19. Jao, T.M., et al.: PCDH10 exerts tumor-suppressor functions through modulation of EGFR/AKT axis in colorectal cancer. *Cancer Lett.* **499**, 290–300 (2021). <https://doi.org/10.1016/j.canlet.2020.11.017>. <https://www.sciencedirect.com/science/article/pii/S0304383520306108>
20. Krzywinski, M., et al.: Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009). <https://doi.org/10.1101/gr.092759.109>. <https://genome.cshlp.org/content/19/9/1639>
21. Lachmann, A., Giorgi, F.M., Lopez, G., Califano, A.: ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**(14), 2233–2235 (2016). <https://doi.org/10.1093/bioinformatics/btw216>
22. Le, Y., Murphy, P.M., Wang, J.M.: Formyl-peptide receptors revisited. *Trends Immunol.* **23**(11), 541–548 (2002). [https://doi.org/10.1016/S1471-4906\(02\)02316-5](https://doi.org/10.1016/S1471-4906(02)02316-5). <https://www.sciencedirect.com/science/article/pii/S1471490602023165>
23. Liang, W., et al.: The contribution of chemoattractant GPCRs, formylpeptide receptors, to inflammation and cancer. *Front. Endocrinol.* **11**, 17 (2020). <https://doi.org/10.3389/fendo.2020.00017>. <https://www.readcube.com/articles/10.3389/fendo.2020.00017>
24. Liu, X.Y., et al.: Methylation of SOX1 and vim promoters in serum as potential biomarkers for hepatocellular carcinoma. *Neoplasma* **64**, 745–753 (2017). [https://doi.org/10.4149/neo\\_2017\\_513](https://doi.org/10.4149/neo_2017_513)
25. Marcel, R., Lucas, S., Sean, D., Levi, W.: TCGAutils: TCGA utility functions for data management (2021). <https://bioconductor.org/packages/release/bioc/html/TCGAutils.html>. R package version 1.14.4
26. Martin, A.M., Kulski, J.K., Witt, C., Pontarotti, P., Christiansen, F.T.: Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends Immunol.* **23**(2), 81–88 (2002). [https://doi.org/10.1016/S1471-4906\(01\)02155-X](https://doi.org/10.1016/S1471-4906(01)02155-X). <https://www.sciencedirect.com/science/article/pii/S147149060102155X>
27. Ohki, R., et al.: Reprimo, a new candidate mediator of the P53-mediated cell cycle arrest at the G2 phase\*. *J. Biol. Chem.* **275**(30), 22627–22630 (2000). <https://doi.org/10.1074/jbc.C000235200>. <https://www.sciencedirect.com/science/article/pii/S0021925819661077>
28. Pan, J., et al.: Lineage-specific epigenomic and genomic activation of oncogene HNF4A promotes gastrointestinal adenocarcinomas. *Cancer Res.* **80**(13), 2722–2736 (2020). <https://doi.org/10.1158/0008-5472.CAN-20-0390>
29. Pedregosa, F., et al.: Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2012). <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>

30. Sander, B., et al.: Mantle cell lymphoma—a spectrum from indolent to aggressive disease. *Virchows Arch.* **468**(3), 245–257 (2015). <https://doi.org/10.1007/s00428-015-1840-6>
31. Sander, B.: Mantle cell lymphoma: recent insights into pathogenesis, clinical variability, and new diagnostic markers. *Semin. Diagn. Pathol.* **28**(3), 245–255 (2011). <https://doi.org/10.1053/j.semdp.2011.02.010>. <https://www.sciencedirect.com/science/article/pii/S0740257011000153>. Seminars on Lymphomas, Part II
32. Szklarczyk, D., et al.: The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**(D1), D605–D612 (2020). <https://doi.org/10.1093/nar/gkaa1074>
33. Verhaegen, M.E., et al.: Merkel cell polyomavirus small T antigen initiates merkel cell carcinoma-like tumor development in mice. *Cancer Res.* **77**(12), 3151–3157 (2017). <https://doi.org/10.1158/0008-5472.CAN-17-0035>
34. Wang, F., et al.: Histone H3 Thr-3 phosphorylation by haspin positions aurora b at centromeres in mitosis. *Science* **330**(6001), 231–235 (2010). <https://doi.org/10.1126/science.1189435>. <https://www.science.org/doi/abs/10.1126/science.1189435>
35. Wang, Y.Y., et al.: Expression of SOX11 mRNA in mantle cell lymphoma and its clinical significance. *Zhonghua xue ye xue za zhi = Zhonghua xueyexue zazhi* **33**(7), 556–560 (2012). <http://europepmc.org/abstract/MED/22967417>
36. Xu, W., Li, J.Y.: SOX11 expression in mantle cell lymphoma. *Leuk. Lymphoma* **51**(11), 1962–1967 (2010). <https://doi.org/10.3109/10428194.2010.514968>. pMID 20919851
37. Yao, Z., et al.: The role of tumor suppressor gene SOX11 in prostate cancer. *Tumour Biol.* **8**(36), 6133–6138 (2015). <https://doi.org/10.1007/s13277-015-3296-3>. <http://europepmc.org/abstract/MED/22967417>