

Deciphering the tissue-specific regulatory role of intronless genes across cancers

Katia Aviña-Padilla¹, José Antonio Ramírez-Rafael¹, Gabriel Emilio Herrera-Oropeza², Guillermo Romero³, Octavio Zambada-Moreno¹, Ishaan Gupta⁴, and Maribel Hernández-Rosales¹

¹CINVESTAV-Irapuato; ²King's College London; ³Data-Pop Alliance NY; ⁴Indian Institute of Technology, Delhi.

maribel.hr@cinvestav.mx

Intronless genes (IGs) or single-exon genes lacking an intron are found across most Eukaryotes. Notably, IGs display a higher transcriptional fidelity as they are not regulated through alternative splicing, suggesting better predictability biomarkers 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 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.

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 entail a higher transcriptional fidelity suggesting a potential role as clinical biomarkers and drug targets that deserve careful consideration in diseases such as cancer [1-3]. Several previous studies have identified the role of IGs in cancer [2, 4-7]. For example the *RPRM* gene increased cell proliferation and tumor suppression activity in gastric cancer [7]; *CLDN8* gene is associated with colorectal carcinoma and renal cell tumors [6]; while *ARLTS1* is upregulated in melanoma; *PURA* & *TAL2* are upregulated in leukemia [2] and protein kinase *CK2α* gene is up-regulated in all human cancers [4];

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 [8]. *SOX11* has been associated with tumorigenesis, with aberrant nuclear protein expression in Mantle Cell Lymphoma (MCL) patients [9-12]. 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 [5]. 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 [13-14].

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 (**Supplementary Table 1**). We selected the four most prevalent cancers, namely Breast Invasive Carcinoma (BRCA), Colon Adenocarcinoma (COAD), Lung Adenocarcinoma (LUAD), 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 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 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 the multi-exonic genes (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 in means of expression ranges are higher in IGs than in MEGs transcripts. Our analysis revealed that in almost all the cancers under study, except in lung adenocarcinoma, IGs tend to express in higher ranges than MEGs, as shown in **Supplementary Figure 1**. Further, we found a high proportion of IGs encodes for histone proteins involved in chromatin structure and other proteins essential to the regulation of development, growth, and proliferation (**Supplemental Figure 2**).

2.2 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 differentially expressed across the eight analyzed tumor types. In a high number of them, 222 (~78%), 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 (**Figure 1**). 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 (**Figure 1a**).

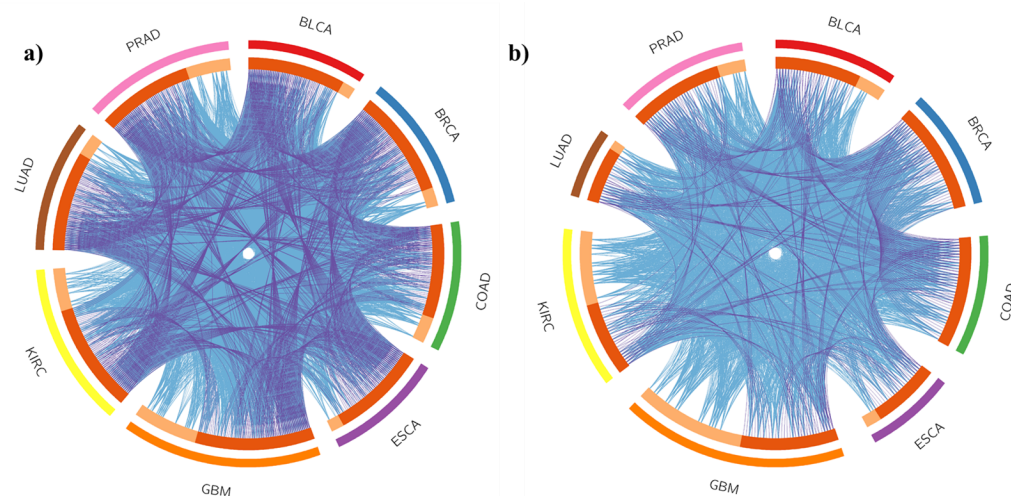


Figure 1. Deregulated IGs across the analyzed cancer genomes. The Circos plot depicts how deregulated genes for each cancer overlap. On the outside shell, each arc represents a cancer genome: BLCA in red, BRCA in blue, COAD in green, ESCA in purple, GBM in orange, KIRC in yellow, LUAD in brown, and PRAD in pink. On the inside shell, the dark orange color represents the proportion of DE genes shared with other cancers, while the light orange color represents the proportion of genes that are uniquely deregulated in a cancer. Purple arcs represent shared DE genes, while blue arcs represent shared functional enrichment terms among cancers. The greater the number of purple arcs and the longer the dark orange area imply a greater overlap among the deregulated IGs across cancers. (a) shows the upregulated genes, while (b) the downregulated.

To delve insights into the conservation of the activation of this negative gene expression mechanism in the cancer genomes, we identify 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 (**Supplementary Figure 3**). Moreover, this group of upregulated genes is statistically enriched in a proximal region (26.1968-27.8927 position in Mb) in chromosome 6 in the human genome (**Supplementary Figure 4**).

Haspin was found to be the only DE-IG shared among all the tumor genomes, highlighting its relevance in tumorigenesis. This gene is known to be required for histone H3 phosphorylation, necessary for chromosomal passenger complex accumulation at centromeres of mitotic cells [15-16]. 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 [15].

Functional enrichment analysis was carried out for the deregulated IGs in all the studied diseases (**Supplementary 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 GO:0006335 term: DNA replication-dependent chromatin assembly was also enriched suggesting an essential role in cancer biology.

2.3 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 development, 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, (**Supplementary 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 wounding response respectively.

2.4 Cancer-specific differentially expressed IGs

To study the specificity in the regulation of human IGs, we quantified the number of cancer-specific DE-IGs (**Figure 2**). 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 (**Supplementary Figure 2**). Among the 338 DE-IGs, we identified that ~35% were specific for a cancer type (**Supplementary Table 3**). GBM followed by KIRC and PRAD with 35, 18, and 16 DE-IGs respectively displayed the highest specificity for IG expression.

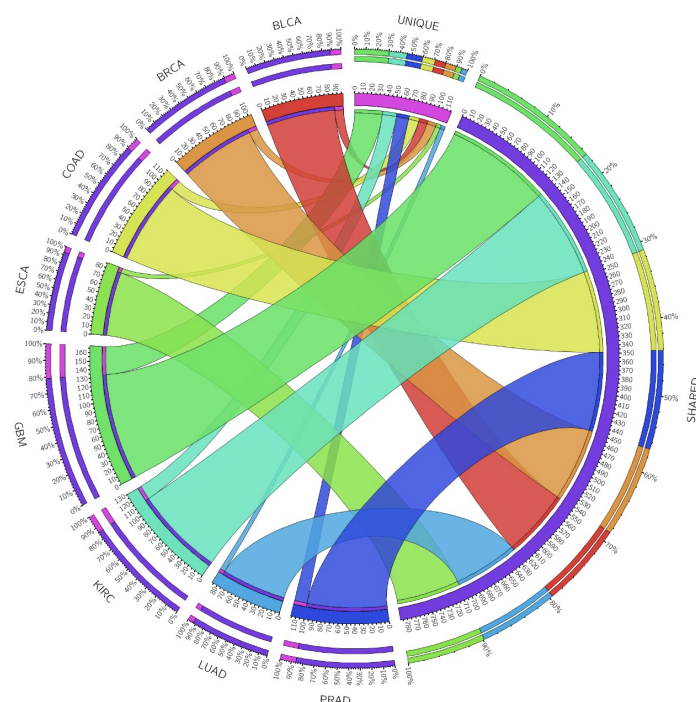


Figure 2. Shared and unique differentially expressed IGs across cancers. Circa plot depicts the number of shared DE-IGs across the 8 types of tumors under study. In the left hemisphere, the total number of DE-IGs in each cancer is represented by an arc: BLCA (red), BRCA (orange), COAD (yellow), ESCA (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.

2.4.1 GBM

Among the specific DE-IGs found in GBM, an evident 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 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*.

2.4.2 KIRC

The cancer-specific deregulated IGs in KIRC include the tissue-specific *KAAG1* gene (kidney-associated antigen 1) along with transcription factors *POU5*, *TAL*, and *bHLH*. Moreover, transmembrane proteins such as the protocadherin 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 [17].

2.4.3 PRAD

IGs specifically downregulated in PRAD include *RAB6D* (member RAS oncogene) and *MARS2* (methionyl-TRNA Synthetase 2) related to amino acid metabolism, while *PBOVI*, a 135 amino acid protein with a transmembrane domain was highly upregulated (logFC=3.1699). This gene is mapped to 6q23-q24, a region associated with loss of androgen dependence in prostate tumors [18]. 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 (Verhaegen et al 2017). 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 [19].

2.4.4 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, such 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*, are identified specifically for this cancer.

2.4.5 LUAD

Regarding the specific differential expression found for lung tumors, the following TFs are significantly upregulated: *TAFIL*, a transcription factor that is related to chronic inflammatory diseases by regulating apoptotic pathways including regulation of *TP53* activity, and *FOXEl*, a forkhead TF previously related to thyroid cancer [20]. Interestingly, specific transmembrane signaling receptors are upregulated. *MAP10* is a microtubule-associated protein and *THBD* thrombomodulin endothelial-specific type I membrane receptor that binds thrombin. Among *THBD* related pathways are collagen formation and formation of fibrin clot, as well as calcium ion binding and cell surface interactions at the vascular wall.

2.4.7 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), *LRRClOB* (Leucine-Rich Repeat Containing 10B), and *MASIL*, a proto-oncogene which is a peptide ligand-binding receptor linked to sensor neurons for pain stimuli detection.

2.4.6 ESCA & 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 [21]. In keeping 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 IG's 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.

2.5 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 a key in determining a cell state at the molecular and physiological level [22]. It has been shown that genes involved in “*similar diseases*” depict shared protein-protein interactions (PPI), and higher expression profiling similarity [23].

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 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 DE-IGs that are shared between two or more cancers (**Supplementary Figure 5**). In this work, 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, having 145 common interactions representing 44.61 % of the total 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 (**Supplementary Figure 5, Supplementary Table 4**).

These cancer-specific interactions were analyzed deeper to delve into their functional role in each cancer (**Figure 3**). As it could be expected due to their high specificity at the PPI level, this network approach shows specific clusters well defined for glioblastoma (**Fig 3b**) and lung (**Fig 3d**). 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 (Supplementary Figure 6). 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 [24]. 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.

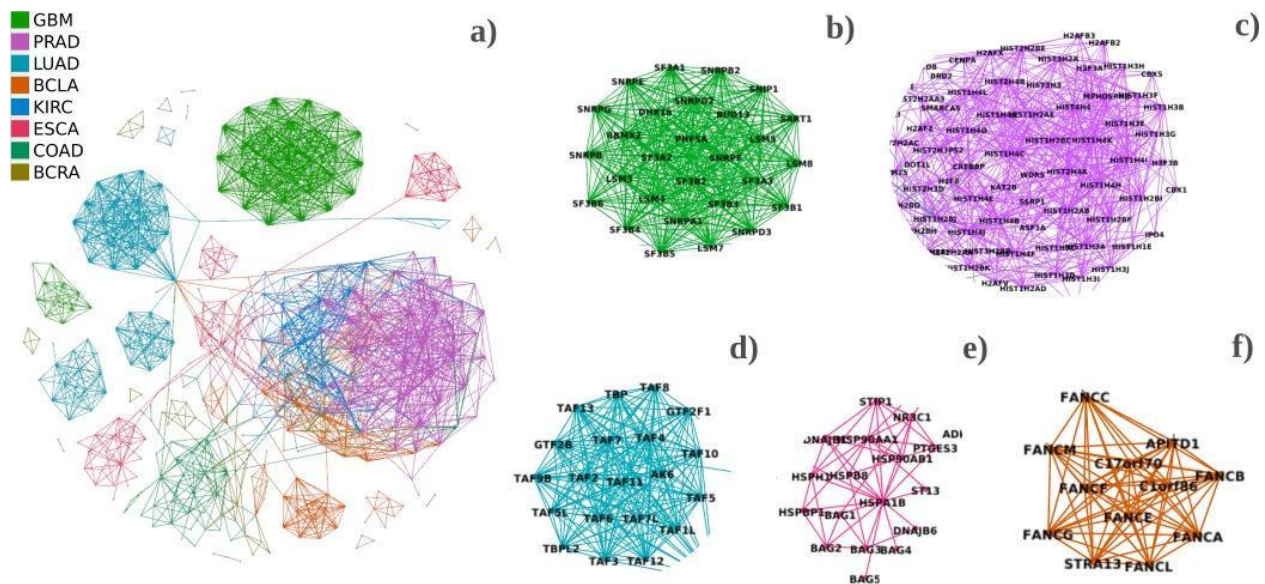


Figure 3. Protein-protein interaction network of DE-IGs and their 50 closest interactors across cancers. a) Unique and shared PPI across cancer are shown, each cancer has a specific color for its interactions. Close-ups of the biggest and more isolated clusters in the network are presented: b) GBM, c) PRAD, d) LUAD, e) ESCA, and f) BLCA. Data was obtained from the String database.

Altogether, these results are the opposite of the patterns reported for cancer genes, which suggests that the behavior tendency of IG proteins is to interact with *oncogenic* genes [24].

For a closer approach, the nearest fifty interactors of the DE-IG encoded proteins in each cancer were analyzed. In **Figure 4** 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 (**Supplementary Table 5**). In total, we detected 27 protein complexes where cancer-specific DE-IGs and 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-multicellular 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 postranslation 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.

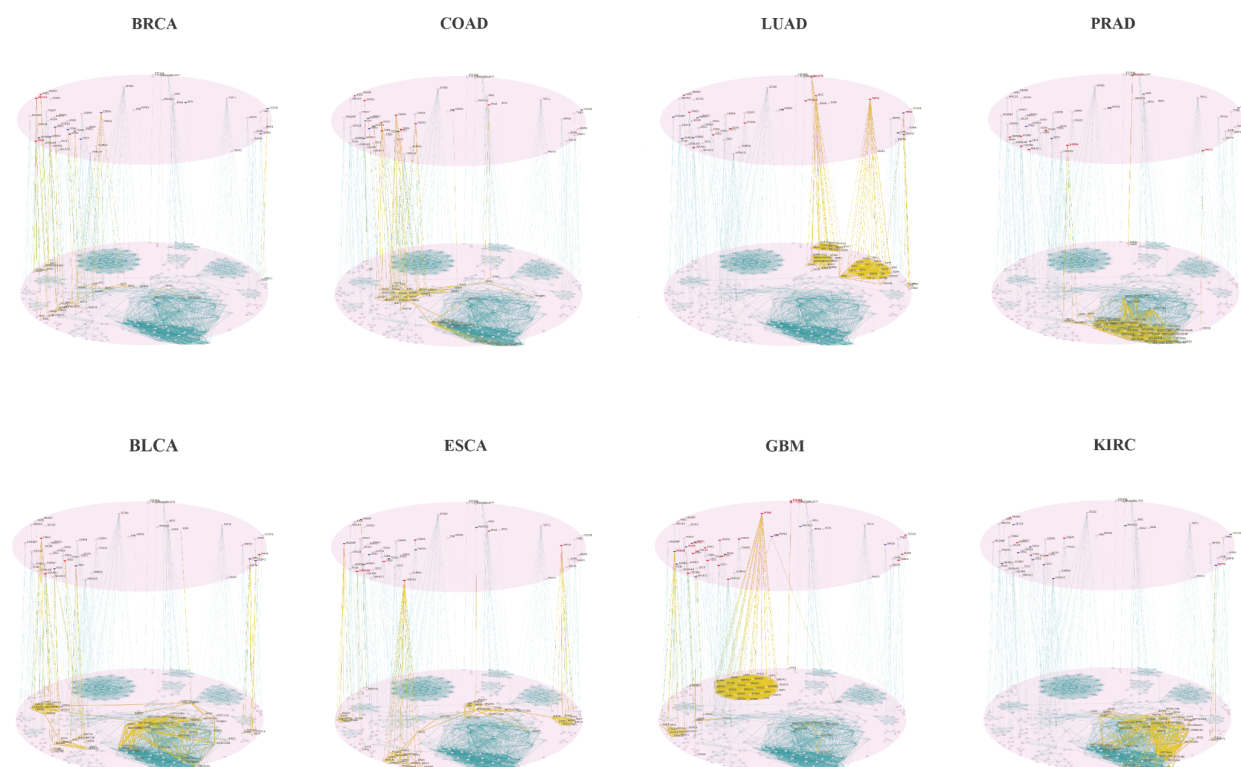


Figure 4. Multilayer networks of IG-encoded proteins and their cancer-specific interactors. Protein-protein interactions were found for deregulated IGs in each cancer. The cancer-specific deregulated IGs that encode for proteins interact with exclusive groups of proteins in each cancer. Upper layer shows deregulated IGs across cancer, highlighting the cancer-specific upregulated IGs 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.

2.6 DE-IGs participate in the genetic “rewiring” of cancer cells

Cancer cells undergo significant genetic “rewiring” as they acquire metastatic traits and adapt to survive 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 adopted during the metastatic process. Finally, we focus on studying the “rewiring” between healthy and tumor samples using BRCA as a model (**Figure 5**). We used breast cancer since it is the only dataset that fits the criteria for a mutual 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 among 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 **Figure 5**, IGs like *RRS1*, *FNDCl0*, *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.

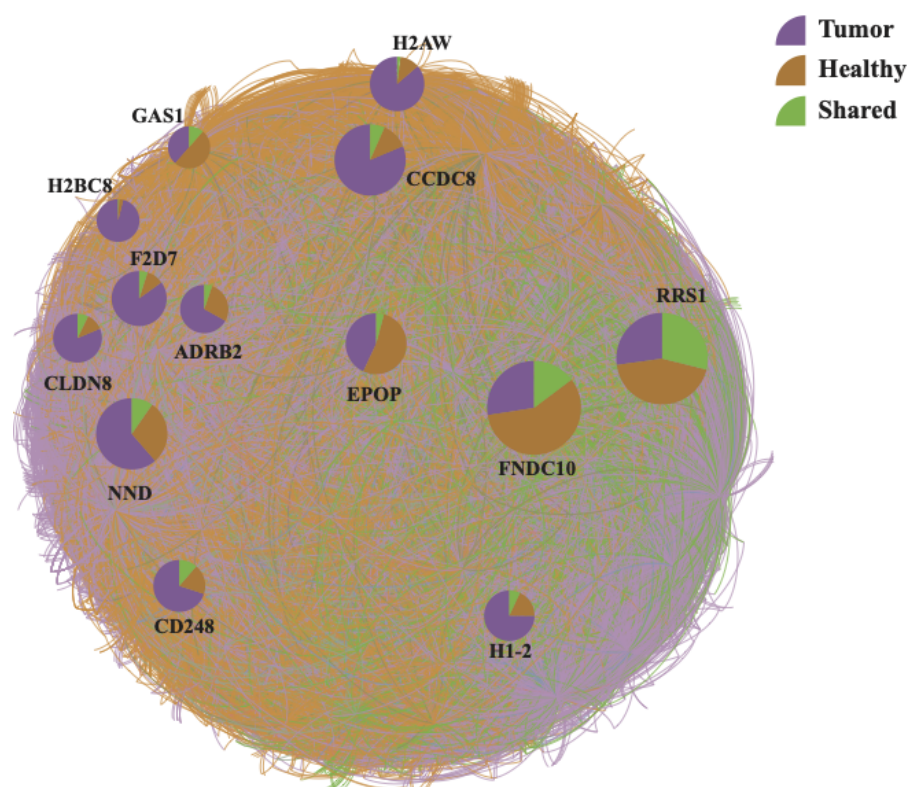


Figure 5. Co-expression network of healthy and tumoral breast tissue. In this network, the co-expression of DE-IGs found in BRCA in healthy tissue (depicted by brown links) and breast tumor tissue (purple links) is shown. Shared co-expression relations are depicted with green links. Pie charts for IG 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.

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 between different cancer types, and association with specific protein-protein interactions that might have potential effects in post-transcriptional processes.

Strikingly, our results show that the upregulated IGs across the eight different cancer types were related to negative regulation of gene expression (gene silencing) and negative regulation of cell differentiation. Further, 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. Although the pattern of differentially expressed (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 [25-27] was regulated in gastrointestinal esophageal and colon adenocarcinomas, and p53 pathways.

This biological behavior could be due to tumor's cellular origin and heterogeneity [28], and tissue-specific expression [2, 8]. 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.

In conclusion, IGs drive massive transcriptional rewiring, as observed in co-expression analysis, that may drive tumorigenesis and may act as novel therapeutic targets for gene therapy and in-silico drug design such as popular GPCRs or emerging HDAC inhibitors [29].

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), 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 Bipartite network & 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. A heatmap for shared 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.

4.3 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 found the primary tumor and adjacent normal tissue transcriptomic data for 18 types. Among those, we selected 8 types 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 [30] 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 [31], 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.

4.4 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 [32] was employed for data analysis and visualization.

4.5 DE-IGs PPI network construction & protein complex identification

Network analysis and visualization were performed using python scripts (See repository) and the Gephi software. StringDB platform [33] 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 [34]. 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.

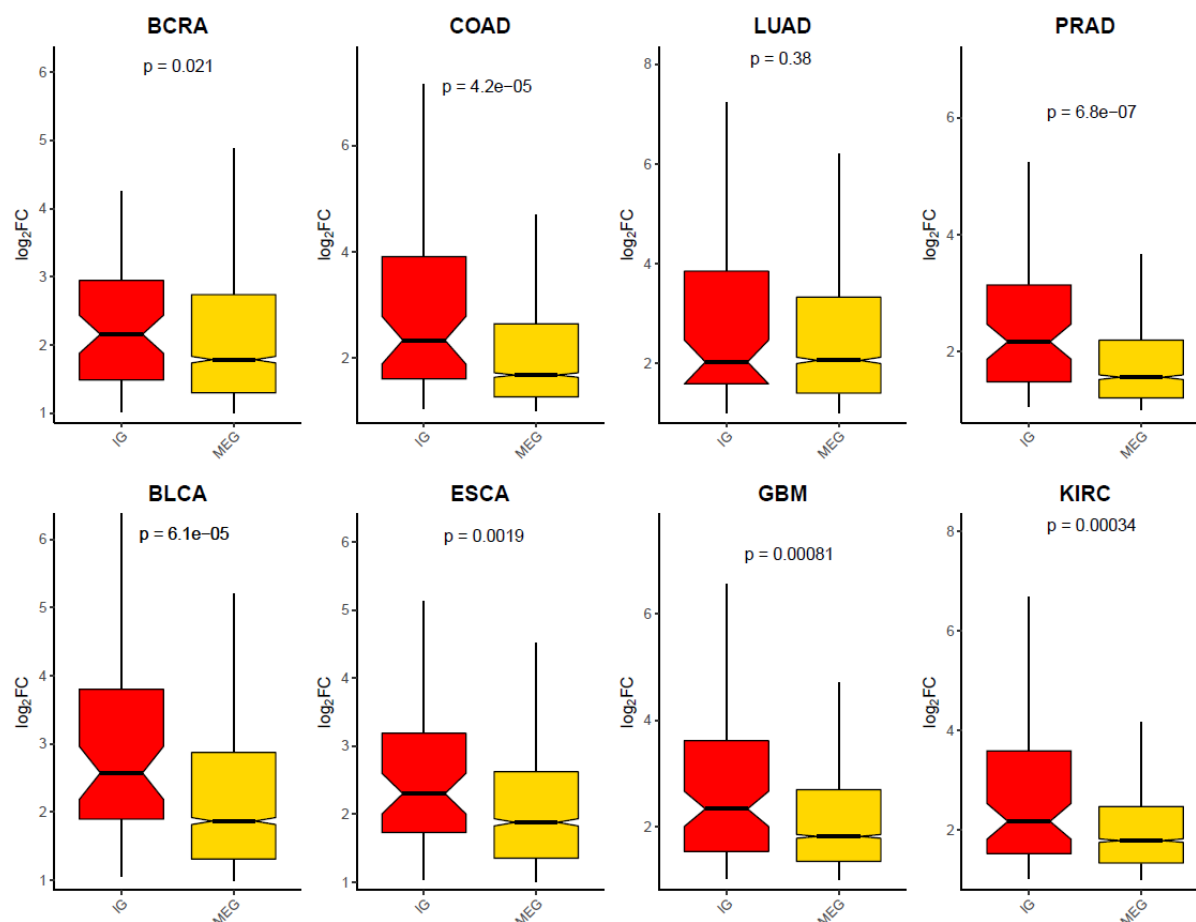
4.6 BRCA network deconvolution

Co-expression networks were inferred using ARACNe-AP [45], 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 1×10^{-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.

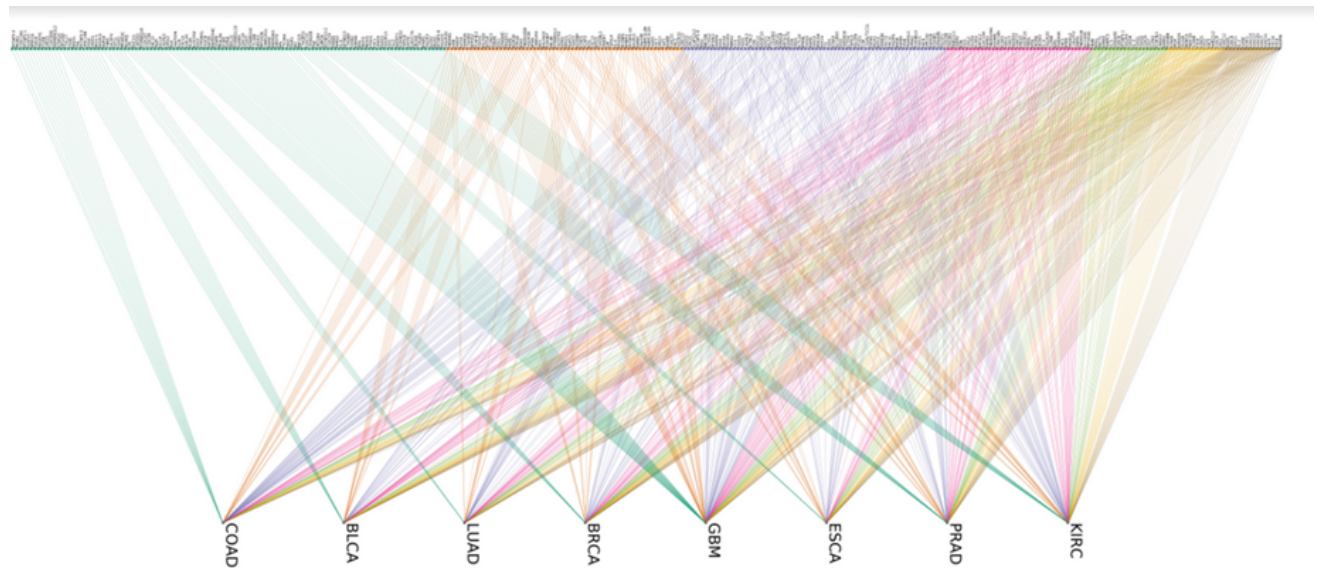
Acknowledgments: We would like to thank Roddy Jorquera and Carolina González for fruitful discussions. Special thanks to Fernando Flores for helping with designing and constructing the co-expression network. We are also thankful to Carlos González for visualization and technical support.

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.

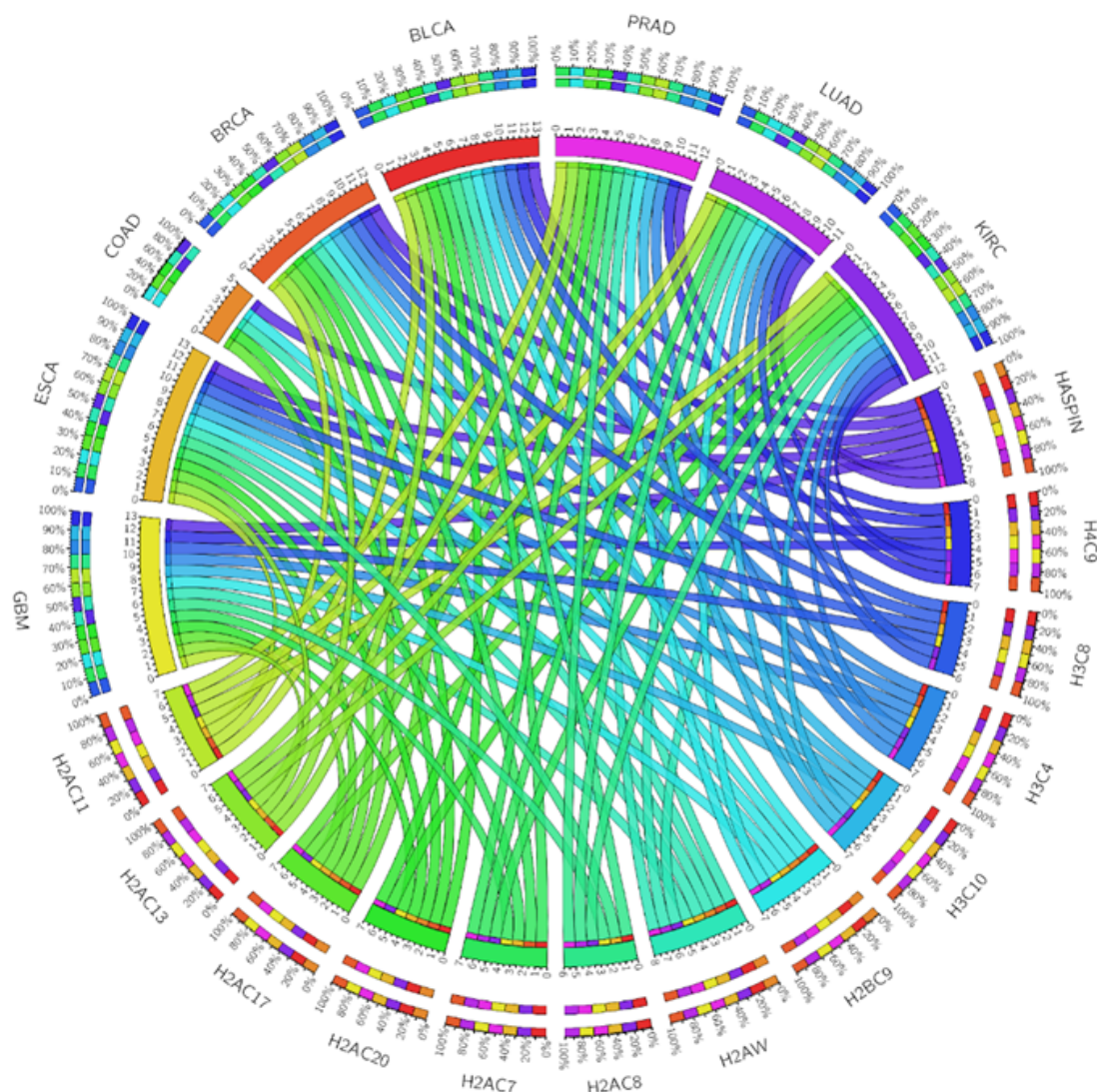
Appendix



Supplementary Figure 1. Upregulated gene expression ranges among cancers. A boxplot is shown for each cancer type, separated by gene type (IGs or MEGs); expression ranges are shown on the y-axis using log₂fold change values. The boxes present a notch corresponding to the confidence interval for each population. In the upper part of the graph, the *p*-value for the Wilcoxon test is shown in order to indicate the statistical difference between both populations.

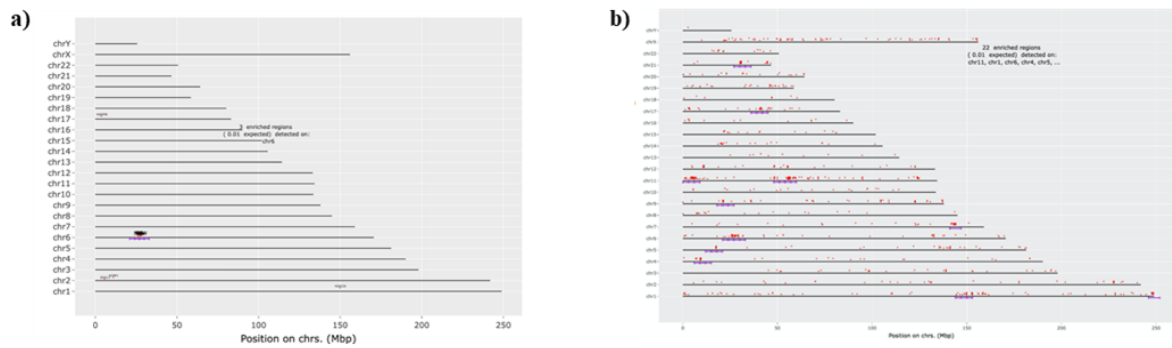


Supplementary Figure 2. IG-cancer type bipartite network. Intronless genes are shown in the upper group of nodes, while cancers are located in the bottom group. A link joining an IG and a cancer type indicates that the expression level of the IG is found disrupted in the linked cancer type. Upper nodes 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, tree cancers, and so on. At the right of the upper nodes, there is a single gene (HASPIN) whose expression is disrupted in all tissues.

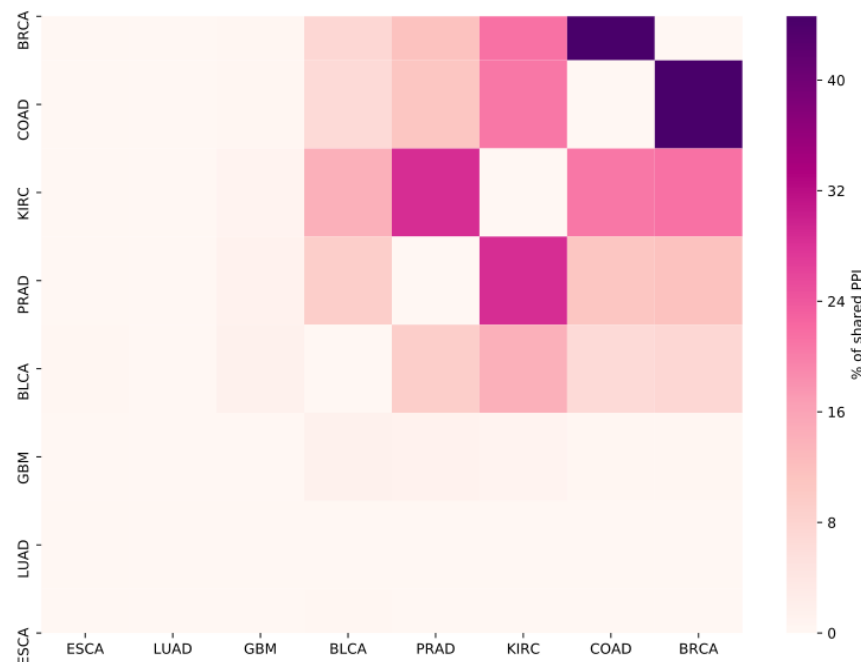


Supplementary Figure 3. A high conserved group of up-regulated histones across tumors.

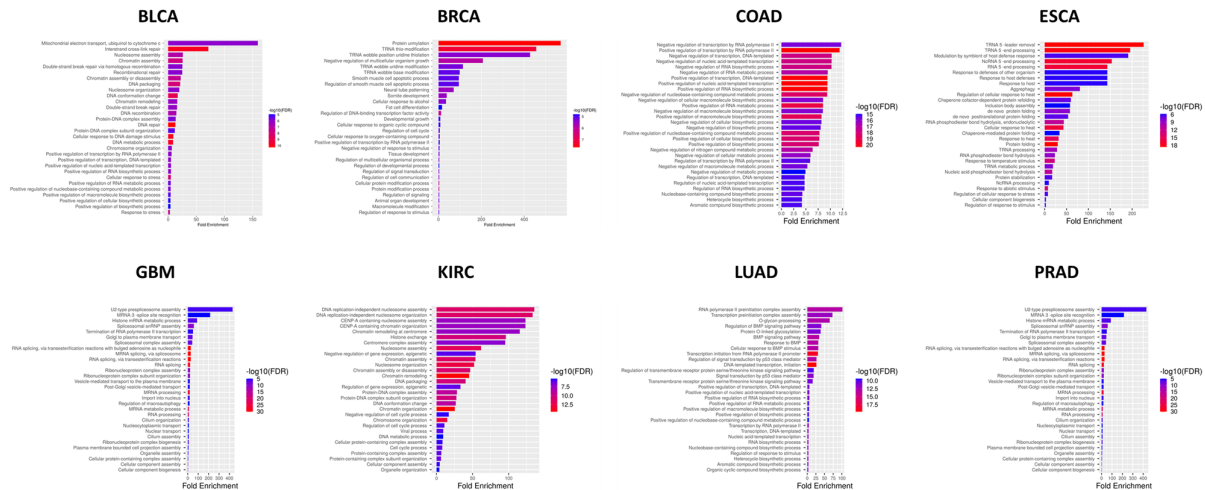
Circos plot depicts the conservation of the deregulation of a group of histones across cancers. In the upper hemisphere, the total number of histone-related DE-IGs in each cancer is represented by an arc: GBM (yellow), ESCA (mustard), COAD (orange), BRCA (dark orange), BLCA (red), PRAD (pink), LUAD (violet), 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



Supplementary Figure 4. 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 IGs encoded proteins 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.



Supplementary Figure 5. 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.



Supplementary figure 6. 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).

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

Cancer Type	Primary Solid Tumor 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

Upregulated lgs										
GO	Description	BLCA	BRCA	COAD	ESCA	GBM	KIRC	LUAD	PRAD	
GO:0061077	chaperone-mediated protein folding	0	0	0	-3.53586873	0	0	0	0	0
GO:0043523	regulation of neuron apoptotic process	0	0	-3.40758971	0	0	0	0	0	0
GO:0000280	nuclear division	0	0	0	0	0	0	-3.40379439	0	0
GO:0000904	cell morphogenesis involved in differentiation	0	0	0	0	0	0	-3.68114644	0	0
GO:0030900	forebrain development	0	0	0	0	0	0	-2.18767642	0	0
GO:0002009	morphogenesis of an epithelium	0	0	0	0	0	-2.72650097	-3.16890112	0	0
R-HSA-420499	Class C/3 (Metabotropic glutamate/pheromone receptors)	0	0	0	0	0	-3.50484037	0	0	0
GO:0008285	negative regulation of cell population proliferation	0	0	0	-2.68098765	-4.14284021	0	-2.45972703	0	0
GO:0007423	sensory organ development	-2.59703879	0	0	0	-5.21338499	0	-2.38111052	0	0
hsa05022	Pathways of neurodegeneration - multiple diseases	0	0	0	0	-3.00071112	0	0	0	0
GO:0045669	positive regulation of osteoblast differentiation	0	0	0	0	-3.79340483	0	0	0	0
GO:0098609	cell-cell adhesion	0	0	0	0	-2.57250443	0	0	0	0
GO:0048732	gland development	0	0	0	0	-4.33244685	0	0	0	0
GO:1903706	regulation of hemopoiesis	0	-2.54469388	0	-2.08227155	-3.87423413	-2.12814219	0	0	0
GO:0045596	negative regulation of cell differentiation	0	-2.02051142	-2.11733152	-2.32977561	-2.34478984	-2.23435568	0	0	0
GO:0016570	histone modification	-2.33988818	-2.14045077	0	-2.36694073	-3.2598089	0	-2.18767642	-3.94085955	0
GO:0007517	muscle organ development	-3.88450456	0	0	-2.74045414	-2.81631242	-2.20409831	-2.55543381	0	0
R-HSA-3214815	HDACs deacetylate histones	-10.7661415	-12.8981691	0	-9.8415034	-5.82686224	-9.03716643	-3.24544683	-15.1532621	0
R-HSA-3214858	RMTs methylate histone arginines	-5.6183791	-7.73158018	0	-7.62596783	-6.40043428	-8.41715258	-2.89611921	-8.97946265	0
GO:0006335	DNA replication-dependent chromatin assembly	-2.26205138	-4.68639981	0	-3.14696932	-2.64544768	-5.05935624	0	-4.31763717	0
Downregulated lgs										
GO	Description	BLCA	BRCA	COAD	ESCA	GBM	KIRC	LUAD	PRAD	
GO:0060341	regulation of cellular localization	0	-3.02547227	-3.02547227	0	0	0	0	0	0
GO:0007423	sensory organ development	0	-2.85392912	-2.85392912	0	0	0	0	0	0
GO:0042391	regulation of membrane potential	0	-4.93064526	-4.93064526	0	0	0	0	0	0
WP236	Adipogenesis	0	-4.11141647	-4.11141647	0	0	0	0	0	0
WP4239	Epithelial to mesenchymal transition in colorectal cancer	0	-3.86830061	-3.86830061	0	0	-3.25734219	0	0	0
GO:0040007	growth	0	-5.9012524	-5.9012524	0	0	-2.70415439	0	-3.0093592	0
GO:0030029	actin filament-based process	-2.04254675	-2.07623128	-2.07623128	0	0	0	0	-4.23424649	0
GO:0010942	positive regulation of cell death	-2.04254675	-2.07623128	-2.07623128	0	0	0	0	-3.0093592	0
GO:0090066	regulation of anatomical structure size	-2.69983122	-8.03809881	-8.03809881	0	-3.49212146	0	-2.43731317	-2.57100774	0
R-HSA-373076	Class A/1 (Rhodopsin-like receptors)	0	-2.36047679	-2.36047679	0	-3.89604303	0	-4.62297482	-2.15756482	0
GO:0042752	regulation of circadian rhythm	0	0	0	0	-2.77079124	0	0	0	0
GO:0009725	response to hormone	0	0	0	0	-3.1073114	-2.70415439	0	0	0
GO:0043269	regulation of ion transport	-2.69983122	0	0	0	-3.49212146	0	0	0	0
GO:0007193	adenylate cyclase-inhibiting G protein-coupled receptor signaling pathway	-2.66479428	0	0	0	-3.01063029	0	-3.40702042	0	0
GO:1903530	regulation of secretion by cell	-2.57842819	-2.62363875	-2.62363875	0	-3.2975469	0	0	0	0
GO:0008285	negative regulation of cell population proliferation	0	0	0	0	0	-2.15855955	0	0	0
GO:2000027	regulation of animal organ morphogenesis	0	0	0	0	0	-2.84414144	0	0	0
GO:0030162	regulation of proteolysis	0	0	0	0	0	-3.98141135	0	0	0
GO:0009611	response to wounding	0	0	0	0	0	0	0	0	0
WP2877	Vitamin D receptor pathway	0	0	0	0	0	0	0	-3.04033691	0

Supplementary Table 3. Cancer-specific deregulated genes

Each column contains the list of cancer-specific DE intronless genes.

Protein complex	Coding genes
CD40 signaling pathway	FANCL, FANCG, FANCF, FANCE, FANCC, APITD1, C17orf70, FANCM, FANCA, STRA13, FANCB, C1orf86
Multi-multicellular organism process	JUNB, FOSL1, FOSL2, BATF, MAFB, CEBPD, FOS, FOSB
Transcription preinitiation complex assembly	GTF2B, TAF13, HIST1H1C, NDN, TAF1L, TAF11, TAF6, TAF7, TAF9B, TAF12, AK6, TP53, TAF5L, TAF3, TAF7L, TAF8, HIST1H1D, GTF2F1, TAF4, TAF5, TBPL2, HMGB1, TAF10, TAF2, TBP
Electron transport chain	UQCRB, CYC1, UQCRCQ, UQCRHL
Wnt signaling pathways	WNT5A, SRF, FZD8, KLF5, AGRP, FOXD3, CEBPB, DDT3, RPA1, FOXG1, RPA4, SPI1, RBBP4, PAX6, NANOG, SFN, MYB, FZD7, CEBPA, SMAD3, POU5F1, LRP6, ESR1, P2RY1, WNT7A, DAXX, SALL4, SOX4, DVL1, FZD2, FOXO1, FZD9, HDAC1, JUN, EP300, RPA3, HSPB3, SDC4, KLF4, SMAD4, POU2F1, SOX2, MC4R, FFAR1, E2F4, TLE1, POMC, CDK2, PPARG, GNAQ, CHD7, WNT3A, CTNBB1, FZD1, WNT1
Chromatid sister cohesion	SMC6, SMC5, UBE2I, NDNL2, EID3, SUMO4, NSMCE1, NSMCE2, PIA54
Follicle-stimulating hormone signaling	GNAS, SLC9A3R1, ADRBK1, GAS1, ARRB1, SHH, ADRB2, ADRB1, PTCH1, ARRB2, CFTR
Dendrite morphogenesis	CUL7, CCDC8, OBSL1
tRNA modifications	CTU1, NFS1, MOCS3, CTU2, URM1
Protein localization to nucleolus	RPF2, EBNA1BP2, RRS1, NIFK
GPI anchor	PIGM, PIGX, PIGV
Hormone-mediated apoptosis	SST, SSTR5
Protein folding	DNAJC2, DNAJB6, HSP90AA1, HSPB8, ST13, HIST1H2AJ, ACTB, HSPA1B, UBB, BAG5, HSPH1, PTGES3, MAGEL2, HSPBP1, KDM1B, RPS27A, ACTL6A, BAG4, UBC, TRIM27, NR3C1, HSP90AB1, STIP1, DNAJB1, BAG2, BAG3, UBA52, BAG1
Light stimulus	GNB5, RGS11, PDE6H, PDE6C, GNAT2, RGS9, PDE6A, RGS9BP
Vasodilatation	LDHD, LDHAL6B, GOT1
Heart morphogenesis	NRARP, NOTCH1
tRNA and NcRNA processing	RPP14, RPP30, RPP21, RPP38, POP7, RPP40, RPP25, POP5, POP4, POP1
Splicing	SF3B5, SNIP1, SNRPA1, SF3A1, SNRPB, SNRPE, SNRPG, SART1, SF3A3, DHX16, LSM8, SF3A2, PHF5A, SF3B6, SNRPB2, RBMX2, LSM4, LSM5, SNRPF, SF3B3, SNRPD2, LSM7, SNRPD3, SF3B4, BUD13, SF3B1, LSM3, SF3B2
Habituation	NHLRC1, EPM2A
Vesicle transport	RALA, EXOC6B, EXOC4, EXOC3, EXOC6, EXOC2, EXOC5, EXOC7, RALB, EXOC8, EXOC1
Intracellular transport	TTC30B, IFT52, TTC30A
Regulation of feeding behavior	RXFP1, RXFP4, INSL5, RLN3
BMP signaling pathway	BMP4, GATA4, SOST, BMPR2, BMP7, GDF5, BMPR1B, NKX2-5, RGMA, HFE2, BMPR1A, BMP2, BAMBI, RGM8, NOG
Innate immune response	GALNT4, MUC15, MUC16, MUC4, MUC7, C1GALT1, MUC5AC, MUC17, MUC1, MUC13, MUC5B, MUC21
Blood coagulation system	PROCR, PROC, F2, CPB2, THBD
Translation and posttranslation processes	EARS2, YARS, MARS2
Beta-catenin destruction complex	FRAT2, GSK3B

Supplementary 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	Interactions shared	Percentage of total interactions in compared cancers
COAD,BRCA	145	44.6153846153846
KIRC,PRAD	277	28.6452947259566
KIRC,BRCA	132	21.3247172859451
KIRC,COAD	138	20.7518796992481
BLCA,KIRC	118	14.1656662665066
BRCA,PRAD	95	11.6421568627451
COAD,PRAD	95	10.9447004608295
BLCA,PRAD	95	9.3503937007874
BLCA,BRCA	43	7.47826086956522
BLCA,COAD	43	6.85805422647528
BLCA,GBM	14	1.64512338425382
PRAD,GBM	15	1.31233595800525
KIRC,GBM	9	0.910010111223458
BRCA,GBM	2	0.301659125188537
COAD,GBM	2	0.27972027972028
BLCA,ESCA	1	0.175131348511383

Supplementary Table 5. Identified PPI complexes. Protein complexes identified where cancer-specific IG-encoded proteins and their interactors play a key role.

Protein complex	Enriched functions	FDR	Cancers
CD40 signaling pathway	Regulation of protein monoubiquitination;Regulation of CD40 signaling pathway	0.018	BLCA
Multi-multicellular organism process	Response to corticosterone;Multi-multicellular organism process;Myeloid leukocyte differentiation;Positive regulation of transcription by RNA polymerase II	0.000014	BLCA
Electron transport chain	Mitochondrial electron transport, ubiquinol to cytochrome c;Pons development;Pyramidal neuron development	0.0067	BLCA
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	0.000051	BLCA,PRAD

Dendrite morphogenesis	Positive regulation of dendrite morphogenesis;Regulation of dendrite morphogenesis;Regulation of mitotic nuclear division	0.00057	BRCA
tRNA modifications	Protein urmylation;TRNA thio-modification;TRNA wobble position uridine thiolation	1.5E-10	BRCA
Protein localization to nucleolus	Ribosomal large subunit export from nucleus;Protein localization to nucleolus	0.0055	BRCA
Wnt signaling pathways	Canonical Wnt signaling pathway;Wnt signaling pathway;Cell-cell signaling by wnt	3.9E-19	KIRC,COAD,BRCA,BLCA,GBM
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	0.000065	BRCA
GPI anchor	Preassembly of GPI anchor in ER membrane;GPI anchor biosynthetic process;GPI anchor metabolic process	3.4E-08	BRCA
Hormone-mediated apoptosis	Hormone-mediated apoptotic signaling pathway;Somatostatin receptor signaling pathway;Somatostatin signaling pathway;Positive regulation of T cell anergy	0.022	COAD
Light stimulus	Wnt signaling pathway, calcium modulating pathway;Detection of light stimulus;Phototransduction	0.0011	ESCA
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	0.021	ESCA
Protein folding	Regulation of cellular response to heat;Chaperone cofactor-dependent protein refolding;Inclusion body assembly;de novo protein folding	4.1E-08	ESCA
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	0.0099	ESCA

	heart valve morphogenesis		
tRNA and NcRNA processing	TRNA 5'-leader removal;TRNA 5'-end processing;NcRNA 5'-end processing;Endonucleolytic cleavage involved in tRNA processing	6.90E-03	ESCA
Splicing	U2-type prespliceosome assembly;MRNA 3'-splice site recognition;Histone mRNA metabolic process;Spliceosomal snRNP assembly	0.00000094	GBM
Habituation	Habituation;Nonassociative learning;Regulation of glycogen (starch) synthase activity	0.012	GBM
Vesicle transport	Golgi to plasma membrane transport;Vesicle docking involved in exocytosis;Vesicle docking;Exocytic process	0.00000085	GBM
Intraciliary transport	Intraciliary anterograde transport;Negative regulation of myotube differentiation;Intraciliary transport	0.019	GBM
Regulation of feeding behavior	Positive regulation of feeding behavior;Regulation of feeding behavior;Positive regulation of behavior;Negative regulation of myotube differentiation	0.0005	KIRC
Transcription preinitiation complex assembly	RNA polymerase II preinitiation complex assembly;Transcription preinitiation complex assembly	2.90E-15	BLCA,LUAD,GBM
BMP signaling pathway	Pharyngeal system development;Endocardial cushion development;Positive regulation of bone mineralization;Negative regulation of BMP signaling pathway	8.8E-12	LUAD
Innate immune response	Innate immune response activating cell surface receptor signaling pathway;Innate immune response-activating signal transduction;Activation of innate immune response	8.9E-13	LUAD
Blood coagulation system	Positive regulation of blood coagulation;Positive regulation of hemostasis;Positive regulation of coagulation	6.3	LUAD
Translation and postranslation processes	Translation;Peptide biosynthetic process;Amide biosynthetic process	0.0013	PRAD

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	0.041	PRAD
----------------------------------	---	-------	------

References

- Ohki, R., Nemoto, J., Murasawa, H., Oda, E., Inazawa, J., Tanaka, N., Taniguchi, T. Reprimo, a new candidate mediator of the p53-mediated cell cycle arrest at the G2 phase. *The Journal of biological chemistry* 275(30), 22627–22630 (2000). <https://doi.org/10.1074/jbc.C000235200>
- Grzybowska, E. A. Human intronless genes: functional groups, associated diseases, evolution, and mRNA processing in absence of splicing. *Biochemical and biophysical research communications* 424(1), 1–6 (2012). <https://doi.org/10.1016/j.bbrc.2012.06.092>
- Liu, X. Y., Fan, Y. C., Gao, S., Zhao, J., Chen, L. Y., Li, F., Wang, K. Methylation of SOX1 and VIM promoters in serum as potential biomarkers for hepatocellular carcinoma. *Neoplasia* 64(5), 745–753 (2017). https://doi.org/10.4149/neo_2017_513
- Hung, M. S., Lin, Y. C., Mao, J. H., Kim, I. J., Xu, Z., Yang, C. T., Jablons, D. M., You, L. Functional polymorphism of the CK2alpha intronless gene plays oncogenic roles in lung cancer. *PloS one* 5(7), e11418 (2010). <https://doi.org/10.1371/journal.pone.0011418>
- Xu, W., Li, J. Y. SOX11 expression in mantle cell lymphoma. *Leukemia & lymphoma* 51(11), 1962–1967 (2010). <https://doi.org/10.3109/10428194.2010.514968>
- Bujko, M., Kober, P., Mikula, M., Ligaj, M., Ostrowski, J., Siedlecki, J. A. Expression changes of cell-cell adhesion-related genes in colorectal tumors. *Oncology letters* 9(6), 2463–2470 (2015). <https://doi.org/10.3892/ol.2015.3107>
- Amigo, J. D., Opazo, J. C., Jorquera, R., Wichmann, I. A., Garcia-Bloj, B. A., Alarcon, M. A., Owen, G. I., Corvalán, A. H. The Reprimo Gene Family: A Novel Gene Lineage in Gastric Cancer with Tumor Suppressive Properties. *International journal of molecular sciences* 19(7), 1862 (2018). <https://doi.org/10.3390/ijms19071862>
- Aviña-Padilla, K., Ramírez-Rafael, J. A., Herrera-Oropeza, G. E., Muley, V. Y., Valdivia, D. I., Díaz-Valenzuela, E., García-García, A., Varela-Echavarría, A., Hernández-Rosales, M. Evolutionary Perspective and Expression Analysis of Intronless Genes Highlight the Conservation of Their Regulatory Role. *Frontiers in genetics* 12, 654256 (2021). <https://doi.org/10.3389/fgene.2021.654256>

9. Wang, Y. Y., Yu, Z., Yi, S. H., Li, Z. J., Li, C. H., Xie, Z. Q., Li, F., Zang, M. R., Hao, M., Qiu, L. G. Expression of SOX11 mRNA in mantle cell lymphoma and its clinical significance 33(7), 556–560 (2012).
10. Yao, Z., Sun, B., Hong, Q., Yan, J., Mu, D., Li, J., Sheng, H., Guo, H. The role of tumor suppressor gene SOX11 in prostate cancer. *Tumour biology: the journal of the International Society for Oncodevelopmental Biology and Medicine* 36(8), 6133–6138 (2015). <https://doi.org/10.1007/s13277-015-3296-3>
11. Sander, B., Quintanilla-Martinez, L., Ott, G., Xerri, L., Kuzu, I., Chan, J. K., Swerdlow, S. H., Campo, E. Mantle cell lymphoma--a spectrum from indolent to aggressive disease. *Virchows Archiv : an international journal of pathology* 468(3), 245–257 (2016). <https://doi.org/10.1007/s00428-015-1840-6>
12. Croci, G. A., Hoster, E., Beà, S., Clot, G., Enjuanes, A., Scott, D. W., Cabeçadas, J., Velloza, L., Campo, E., Clasen-Linde, E., Goswami, R. S., Helgeland, L., Pileri, S., Rymkiewicz, G., Reinke, S., Dreyling, M., Klapper, W. Reproducibility of histologic prognostic parameters for mantle cell lymphoma: cytology, Ki67, p53 and SOX11. *Virchows Archiv : an international journal of pathology* 477(2), 259–267 (2020). <https://doi.org/10.1007/s00428-020-02750-7>
13. Sander B. Mantle cell lymphoma: recent insights into pathogenesis, clinical variability, and new diagnostic markers. *Seminars in diagnostic pathology* 28(3), 245–255 (2011). <https://doi.org/10.1053/j.semdp.2011.02.010>
14. Inamdar, A. A., Goy, A., Ayoub, N. M., Attia, C., Oton, L., Taruvai, V., Costales, M., Lin, Y. T., Pecora, A., Suh, K. S. 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>
15. 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 & development* 19(4), 472–488 (2005). <https://doi.org/10.1101/gad.1267105>
16. Wang, F., Dai, J., Daum, J. R., Niedzialkowska, E., Banerjee, B., Stukenberg, P. T., Gorbsky, G. J., Higgins, J. M. 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>
17. Jao, T. M., Fang, W. H., Ciou, S. C., Yu, S. L., Hung, Y. L., Weng, W. T., Lin, T. Y., Tsai, M. H., Yang, Y. C. PCDH10 exerts tumor-suppressor functions through modulation of EGFR/AKT axis in colorectal cancer. *Cancer letters* 499, 290–300 (2021). <https://doi.org/10.1016/j.canlet.2020.11.017>
18. An, G., Ng, A.Y., Meka, C.S., Luo, G.Z., Bright, S., Cazares, L., Wright, G., Veltri, R; Cloning and Characterization of *UROC28*, a Novel Gene Overexpressed in Prostate, Breast, and Bladder Cancers. *Cancer Research* 60 (24), 7014–7020 (2000).

19. Martin, A. M., Kulski, J. K., Witt, C., Pontarotti, P., Christiansen, F. T. Leukocyte Ig-like receptor complex (LRC) in mice and men. *Trends in immunology* 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)
20. Denny, J. C., Crawford, D. C., Ritchie, M. D., Bielinski, S. J., Basford, M. A., Bradford, Y., Chai, H. S., Bastarache, L., Zuvich, R., Peissig, P., Carrell, D., Ramirez, A. H., Pathak, J., Wilke, R. A., Rasmussen, L., Wang, X., Pacheco, J. A., Kho, A. N., Hayes, M. G., Weston, N., de Andrade, M. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. *American journal of human genetics* 89(4), 529–542 (2011). <https://doi.org/10.1016/j.ajhg.2011.09.008>
21. Pan, J., Silva, T. C., Gull, N., Yang, Q., Plummer, J. T., Chen, S., Daigo, K., Hamakubo, T., Gery, S., Ding, L. W., Jiang, Y. Y., Hu, S., Xu, L. Y., Li, E. M., Ding, Y., Klempner, S. J., Gayther, S. A., Berman, B. P., Koeffler, H. P., Lin, D. C. Lineage-Specific Epigenomic and Genomic Activation of Oncogene HNF4A Promotes Gastrointestinal Adenocarcinomas. *Cancer research* 80(13), 2722–2736 (2020). <https://doi.org/10.1158/0008-5472.CAN-20-0390>
22. Patrinos, G. P., Forero, D. A. Genome Plasticity in Health and Disease. Elsevier Gezondheidszorg (2020). <https://doi.org/10.1016/B978-0-12-817819-5.00009-7>
23. Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., Barabási, A. L. The human disease network. *Proceedings of the National Academy of Sciences of the United States of America* 104(21), 8685–8690 (2007).
24. 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 systems biology* 12(Suppl 2), 13 (2018). <https://doi.org/10.1186/s12918-018-0537-2>
25. Chen, K., Bao, Z., Gong, W., Tang, P., Yoshimura, T., Wang, J. M. Regulation of inflammation by members of the formyl-peptide receptor family. *Journal of autoimmunity* 85, 64–77 (2017). <https://doi.org/10.1016/j.jaut.2017.06.012>
26. Le, Y., Murphy, P. M., Wang, J. M. Formyl-peptide receptors revisited. *Trends in immunology* 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)
27. Liang, W., Chen, K., Gong, W., Yoshimura, T., Le, Y., Wang, Y., Wang, J. M. The Contribution of Chemoattractant GPCRs, Formylpeptide Receptors, to Inflammation and Cancer. *Frontiers in endocrinology* 11, 17 (2020). <https://doi.org/10.3389/fendo.2020.00017>
28. 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>

29. Dai, E., Zhu, Z., Wahed, S. *et al.* Epigenetic modulation of antitumor immunity for improved cancer immunotherapy. *Molecular Cancer* 20, 171 (2021). <https://doi.org/10.1186/s12943-021-01464-x>
30. Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T., Malta, T.M., Pagnotta, S.M., Castiglioni, I., Ceccarelli, M., Bontempi, G., Noushmehr, H. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Research* (2015). <http://doi.org/10.1093/nar/gkv1507>.
31. Ramos, M., Schiffer, L., Waldron, L. TCGAutils: TCGA utility functions for data management. R package version 1.14.0 (2021).
32. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., Marra, M. A. Circos: an information aesthetic for comparative genomics. *Genome research* 19(9), 1639–1645 (2009). <https://doi.org/10.1101/gr.092759.109>
33. Szklarczyk, D., et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acid Research* 49, D605-D612 (2021).
34. Hagberg, A., Schult, D., Swart, P. *Exploring Network Structure, Dynamics, and Function using NetworkX in Proceedings of the 7th Python in Science conference (SciPy 2008)*, G Varoquaux, T Vaught, J Millman (Eds.), pp. 11-15
35. 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>