The network structure of hematopoietic cancers

Arturo Kenzuke Nakamura-García[1] **and Jesús Espinal-Enríquez**[1*]

[1]National Institute of Genomic Medicine, Mexico

Hematopoietic cancers (HCs) are a heterogeneous group of malignancies that affect blood, bone marrow and lymphatic system. Despite arising from derivatives of hematopoietic stem cells, each disease has a unique set of characterizing genomic irregularities. Gene co-expression networks (GCNs) have been useful to analyze and integrate information of cancer transcriptomes. Here, we explored the co-expression landscape in HC, by inferring GCNs from four hematopoietic cancers (B and T-cell acute leukemia, -BALL, TALL-, acute myeloid leukemia -AML- and multiple myeloma -MM-) as well as non-cancer bone marrow. We characterized their structure (topological features) and function (enrichment analyses). We found that, as in other types of cancer, the highest co-expression interactions are intra-chromosomal, which is not the case for control GCNs. We also detected a highly co-expressed group of overexpressed pseudogenes in HC networks. The four GCNs present only a small fraction of common interactions, related to canonical functions, like immune response or erythrocyte differentiation. Those genes are differentially expressed in a unique fashion for each HC. For instance, cell cycle-associated genes are underexpressed in MM and AML, slightly overexpressed in BALL but highly overexpressed in TALL. With this approach, we are able to reveal cancer-specific features useful for detection of disease manifestations.

Hematopoietic cancers | Co-expression network | Pseudogenes |
Correspondence: ***jespinal@inmegen.gob.mx***

Introduction

Hematological cancers (HC) are a heterogeneous group of malignancies from immune cells or their precursors. They can be categorized into three major groups: leukemias, lymphomas and multiple myeloma (1). Leukemias are malignancies from precursors of leukocytes that accumulate in blood or in the bone marrow. They can arise from myeloid or lymphoid linage. Depending on the velocity of the disease progression, they can be acute, with a fast development of the disease, or chronic, if the diseases progress slowly. Lymphomas are tumors from lymphocytes that affect the lymph nodes, thymus or spleen, and can be broadly categorized into Hodgkin and Non-Hodgkin lymphomas. Multiple myeloma is a tumor of differentiated plasma cells, which unlike normal plasma cells, can proliferate and accumulate in the bone marrow. It is also characterized by the secretion of large amounts of a monoclonal immunoglobulin protein.

Results and Discussion

Cancer networks are topologically different to their normal counterparts. By looking at the architecture of the networks, we observe that the GCNs of the normal bone marrow are mainly composed of a single giant component, while the cancerous networks have several smaller components (Fig. 1 and Supplementary Material 1). As expected, several of these smaller components are made up of genes from the same chromosome. This result coincide with the previous GCNs analyzed from breast cancer (2–4), lung (5), and clear cell renal carcinoma (6, 7) suggesting a common regulatory mechanism affecting the entire system, which is manifested in the network topologies.

Another aspect that we consider to be crucial in the topological analysis of the GCNs of cancer phenotypes is the regulation of inter-chromosomal interactions. Here, we show that the loss of strong inter-chromosomal interactions is also present in hematological cancer networks. This phenomenon is more clearly appreciated in MM, in which approximately 80% of the strongest interactions (on the cut-off point of 100,000) are intra-chromosome (Fig. 2). TALL and BALL have a similar fraction of intra-chromosomal interactions between them, and AML has only a slight increase in these interactions in comparison with the normal bone marrow. One difference between the MM samples and the samples from leukemias is the ages of the patients in each group. The MM group is conformed by samples from adults, while our leukemias samples correspond mainly to pediatric patients. This could suggest that the loss of inter-chromosomal interactions could be associated with the age of the patients. Further studies including samples from pediatric cancer patients are needed to further explore this possibility.

A limitation of the studies involving gene co-expression lies on the fact that there are still many questions around the meaning of this phenomenon. Recent studies have shown that multiple DNA regulatory elements (such as transcription factors, CTCF binding sites, copy number alterations, among others) must influence the co-expression landscape (8, 9). Still, the question remains of whether the phenomenon is provoked by these molecular mechanisms or if they are a consequence of the co-expression profile. A plausible explanation about this, is that the cancer co-expression landscape could be modified by a mechanical principle affecting the transcriptional machinery that somehow force the genes to co-express in a small local vicinity.

Cancer networks show elevated interactions between pseudogenes. A feature that has not been explored in previous works of our group is the co-expression between different types of genes other than protein coding. Here, we did not applied any filter regarding RNA biotype and constructed net-

DRAFT

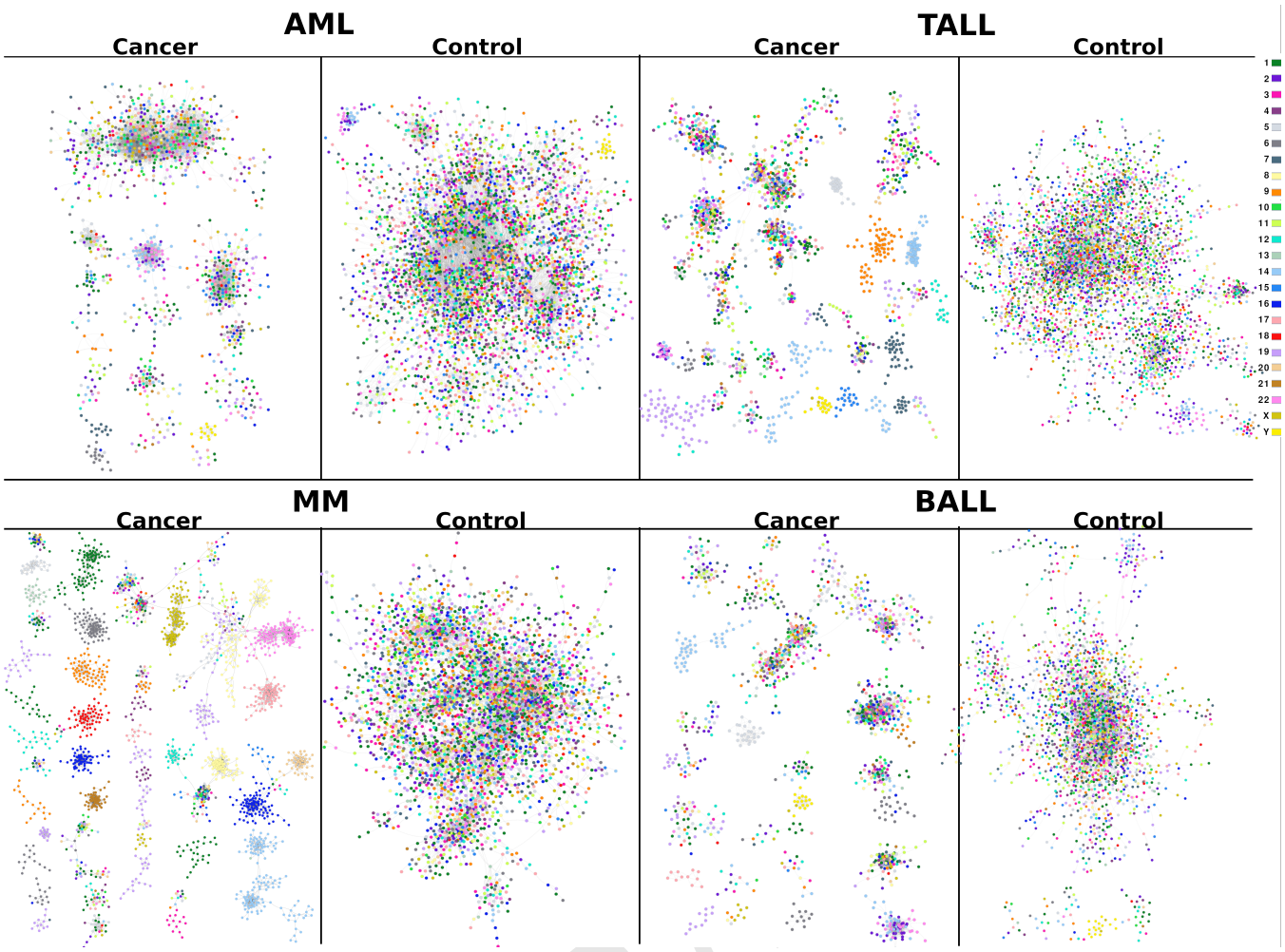

Fig. 1. MI networks for the for phenotypes and their controls, painted by chromosome.

works with all the available biotypes in the samples. For instance, in AML network, more than 50% of the interactions occur between pseudogenes (see 3). This network has the least amount of intra-chromosome interactions of the four HCs. The opposite seems to happen in the MM network, with the least amount of interactions between pseudogenes and the biggest fractions of intra-chromosomal edges. This could suggests a relationship between the co-expression of pseudogenes and the intra-chromosomal regulation. although further studies are needed to clarify if this is the case.

The co-expression pattern between pseudogenes in the cancer networks forms a cluster that is not seen in the topology of the normal BM networks (large black component of AML network in Fig. 4). These are pseudogenes of housekeeping genes, in particular of ribonucleoproteins and elongation factors. Previous reports regarding the expression of pseudogenes have revealed that their transcription is tissue specific, suggesting an active role in the tissues (10, 11).

A very compelling explanation regarding the pseudogenes function is centered on their role as competitive endogenous RNA (ceRNA). According to this, the transcripts of pseudogenes regulate their parental gene expression through competition for its regulatory elements, such as miRNAs (12, 13). However, our results show that most pseudogenes have very strong co-expression relationships between them, which are even stronger than the interactions with their respective parental genes, and that occur between pseudogenes with different parental genes. This suggests that these co-expression interactions are not a manifestation of the ceRNA capability of pseudogenes.

It could be possible that a dysregulation on the transcription of ribonucleoproteins is provoking their co-expression, and that this serves as a mechanism used by cancer cells to help sustain high rates of protein translation that contribute to their proliferation. A similar mechanism could be affecting the expression pattern of their pseudogenes, since they retain some characteristics of their parental genes, which could prompted their co-expression. However, the question would still remain as of what type of signals or mechanisms are indicating the transcriptional machinery to transcribe these genes at a similar rate.

Additionally, the mechanism regulating the protein coding genes and the pseudogenes should be similar, but not exactly the same, since otherwise one would detect interactions between pseudogenes and protein coding genes in the top interactions. Further research on the subject of the transcription regulation of ribonucleoproteins and their pseudogenes in cancer is necessary to propose a more robust hypothesis

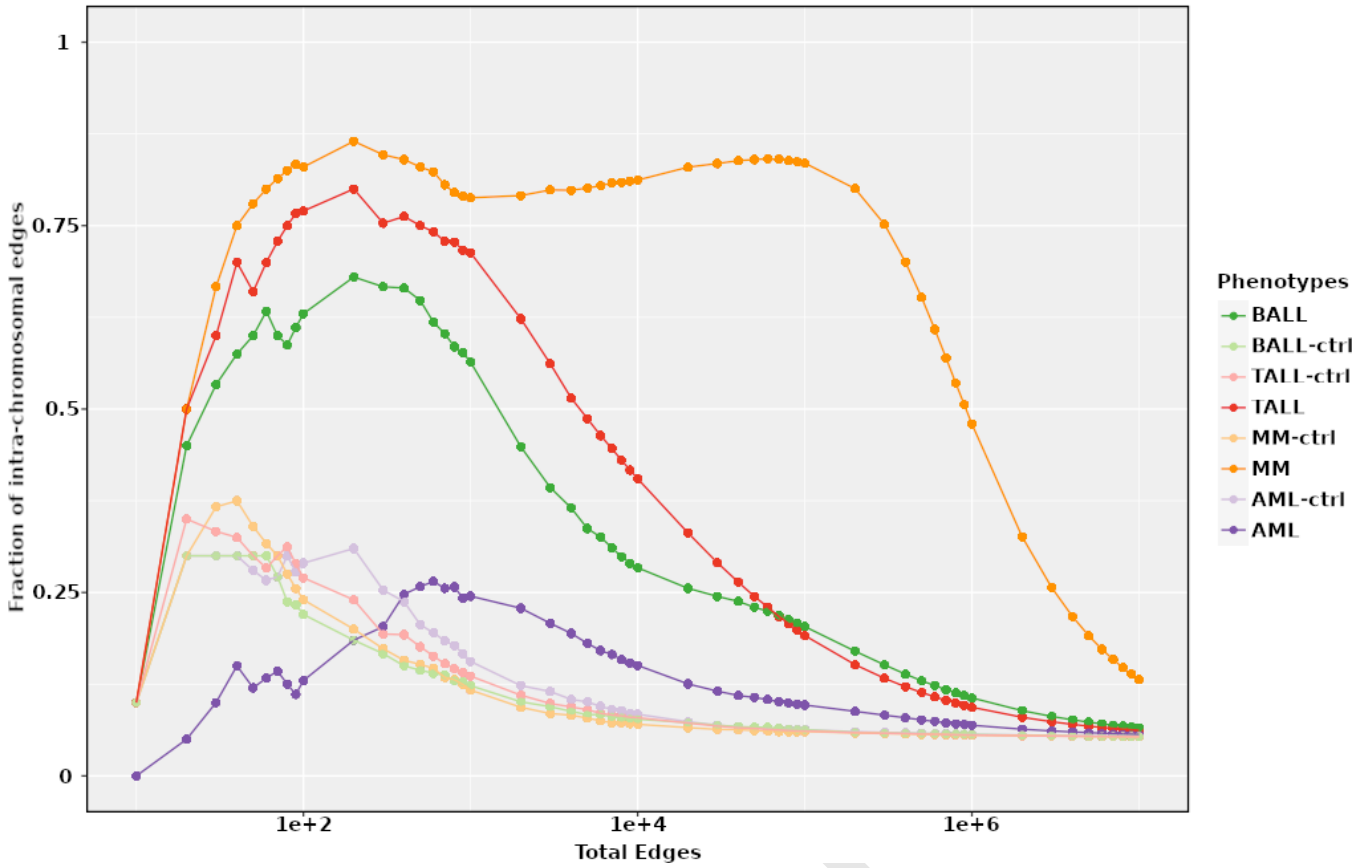

Fig. 2. Fraction of intrachromosomal (cis) interactions across different cut off values of the networks.

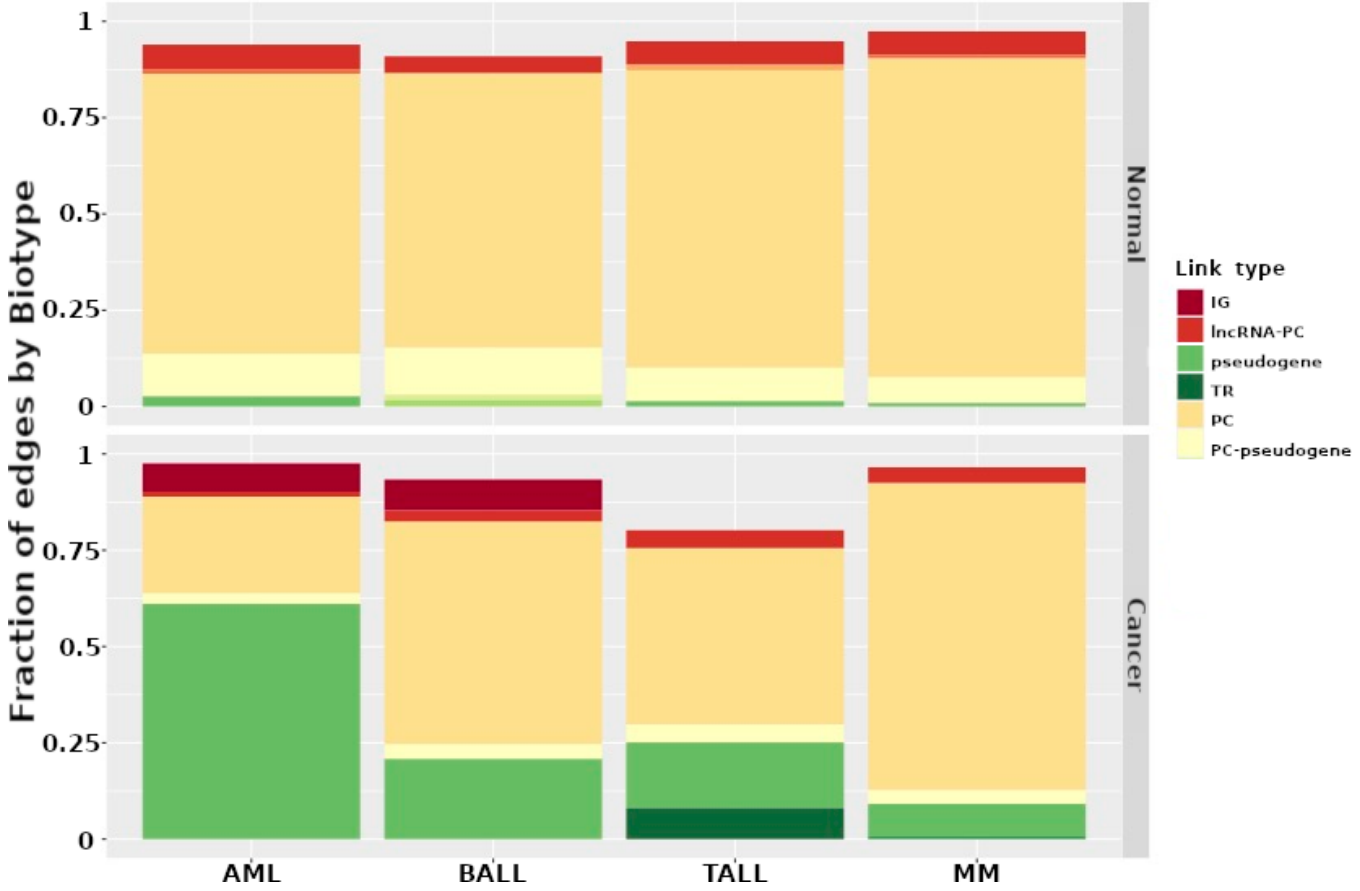

Fig. 3. Fraction of Top 5 interacting biotypes on HC co-expression networks

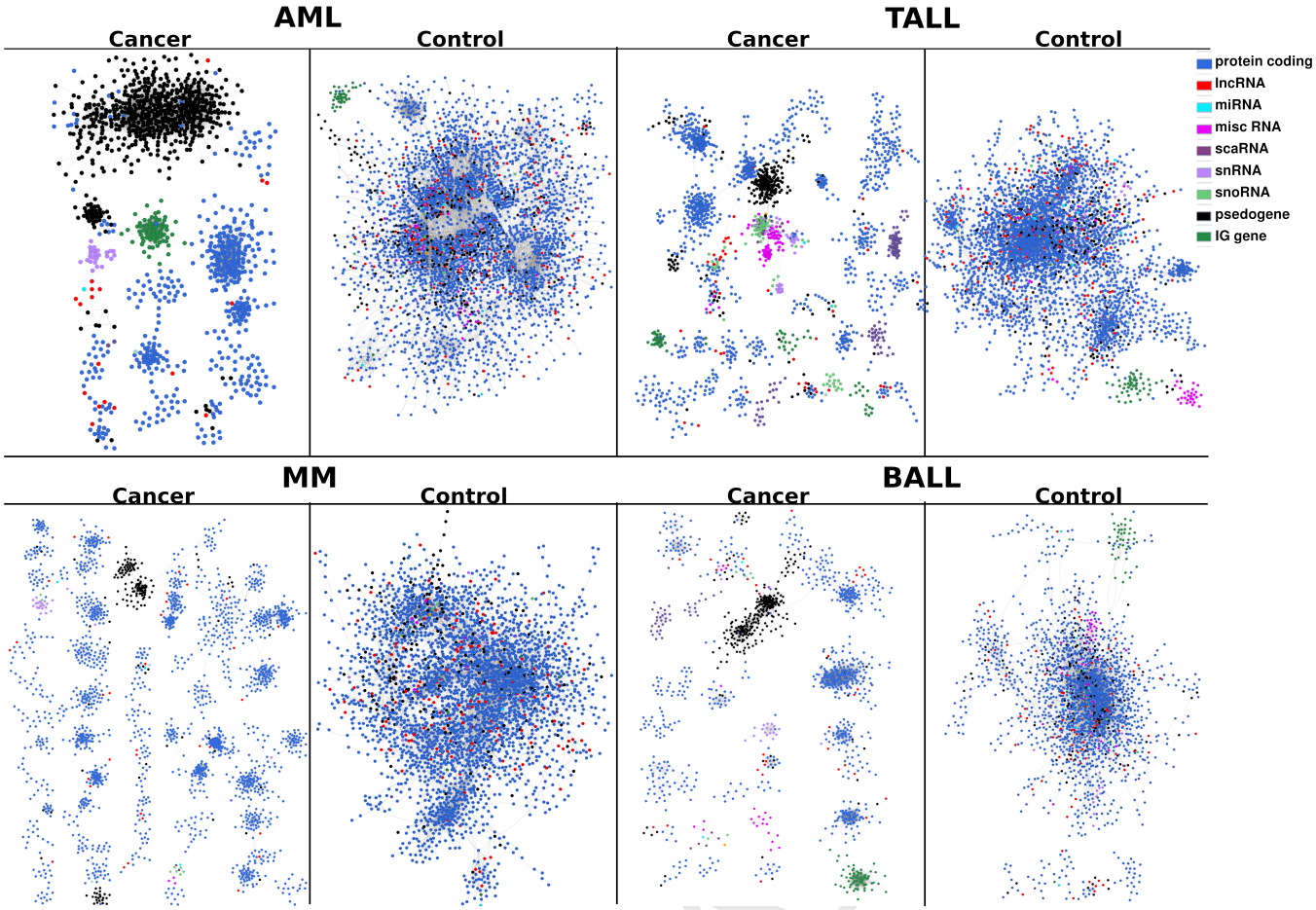

Fig. 4. MI networks for the four phenotypes and their controls, painted by the biotype.

regarding their co-expression patterns.

The co-expression between pseudogenes could be related to the phenomenon of the loss of inter-chromosomal co-expression, since our results suggest a preference by pseudogenes to co-express with other pseudogenes from different chromosomes in hematological cancer GCNs.

So far, with the GCNs shown here we observe that 1) the gene co-expression landscape in HCs (as in carcinomas) is topologically different from normal GCNs, which is mainly manifested in the the loss of inter-chromosomal co-expression and the smaller connected components in HCs; and 2) that the co-expression interactions between pseudogenes is increased in the cancer phenotypes, which is a feature that have not been explored by our group. As far as the authors are aware of, this has not been reported in the current literature.

Differential expression analysis. A differential expression analysis was carried out to detect DEGs in each cancer phenotype in comparison the normal bone marrow group. Fig- 5 shows largest components of the four GCNs with the genes colored by its differential expression value. It is worth noticing that in the four cases, network components are predominantly formed by genes with the same differential expression trend. Tables 1-4 shows the top five differentially expressed genes in BALL, TALL, MM, and AML, respectively. The complete differential expression analysis can be found in Supplementary Material 2.

On its own, this analysis revealed several differentially expressed genes. The top over-expressed and the top under-expressed genes are unique in each phenotype, except for DSC3, a member of the desmocollin subfamily that is over-expressed in both BALL and AML. BALL and TALL have members of the olfactory receptor family over-expressed, namely OR5P2 and OR10T2, respectively.

OR10R3P, a pseudogene of the aforementioned family is also in the top overexpressed genes in TALL. The olfactory receptor family genes encode for G-protein coupled receptors related with the sense of smell, and have been reported to be over-expressed in several cancer cell lines (14), although their exact role in cancer is still unknown.

On the top under-expressed genes of BALL, TALL and MM are many genes related with the immune system, and in AML this top is conformed by four small nucleolar RNAs and a pseudogene of a keratin protein. Notwithstanding, the biological insights that we could obtain from these mere list of DEGs is limited; however, integrated with the information of GCNs and functional enrichment analysis, we identified specific molecular processes that were exacerbated or diminished in HCs.

GCNs of hematological malignancies share several interactions. In order to identify the main similarities in the

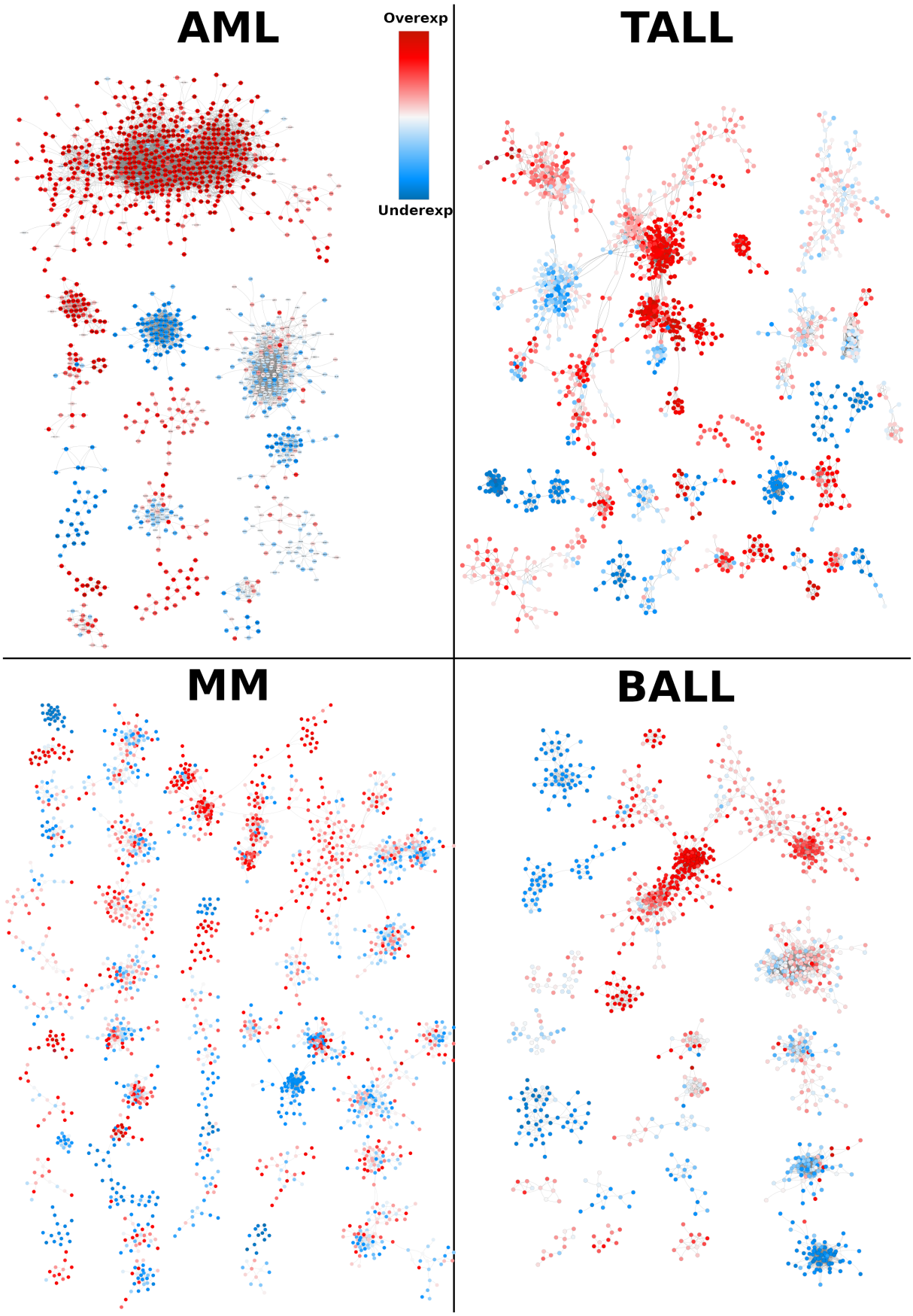

Fig. 5. MI networks for the four cancer phenotypes painted by the differential expression values. Red nodes represent overexpression meanwhile underexpressed genes are in blue.

architecture of the GCNs of the HCs, we inferred a network of the common interactions between the strongest 100,000 interactions of our four cancer networks. This network has 1,995 nodes and 4,305 edges.

We performed the community detection algorithm HiDeF (Hierarchical community Decoding Framework (15)) to identify robust structures across different scales, followed by a functional enrichment analysis using the clusterProfiler (16) package from Bioconductor in R (see Methods). This yielded a total of 206 hierarchical communities, from which 44 were enriched for at least one biological process (BP) from Gene Ontology (Supplementary Material 3)

The functional enrichment of these communities is represented in Fig. 6 in the form of a bipartite network of communities (circles) linked with their respective BPs (square diamonds). In that figure, communities are associated with those processes that resulted significantly enriched at the cut-off value ($p_{val} < 1e^{-10}$).

These results were combined with a differential expression analysis to determine the differential expression trend of those communities for each type of HC. This is shown in the pie plots on each community (Supplementary Material 4). Fig. 6 shows the differential expression trend for MM. This allowed us to visualize the communities of the intersection network, the processes that they are associated with, and to determine which BPs are enhanced or decreased in each HC in comparison to the normal bone marrow.

The largest component of the four GCNs is observed in the AML network. This is composed mostly of pseudogenes of ribonucleoproteins (RPs) and translation initiation factors. These genes are mainly over-expressed in every HCs. The large set of co-expressed ribonucleoprotein interchromosome genes, has been also reported in networks from other 15 cancer tissues (17).

In the present work, the communities detected in this cluster were enriched for cytosplasmatic translation and other processes related with RNA processing. Two communities in this cluster are also enriched for `positive regulation of signal transduction by p53 class mediator`. The genes responsible for this enrichment are ribosomal proteins that have been described to affect the p53 levels by negatively regulating Mdm2 and MdmX, and therefore promoting apoptosis (18).

We found two communities enriched for processes related with energy production, specifically, to oxidative phosphorylation. The genes in theses communities have all an over-expression trend in AML and BALL, while in MM and TALL there are some under-expressed genes.

Four communities in the intersection network (C23349, C23298, C23366 ,C23365) are enriched for `immunoglobulin production`. These communities are all conformed by immunoglobulin light chain genes that are under-expressed on the three leukemias and over-expressed in the case of multiple myeloma, which could be associated with the elevated secretion of monoclonal immunoglobulin proteins by myeloma cells.

Another community in our intersection network (C23272) is enriched for `erythrocyte homeostasis` and `erythrocyte differentiation`, although with a p-value of $1.34e^{-09}$, which is below our significance threshold. Notwithstanding, we consider this particular enrichment to be relevant, since anemia is one of the most common symptoms of hematological malignancies. Indeed, most genes in this community are under-expressed in every HC network, reaffirming that the production of erythrocytes is strongly impaired.

The functional enrichment analysis on each HC network shows similar biological processes present in the cancer networks. The intersection network of the four HCs allowed us to identify which processes were represented by the same interactions in the four GCNs. Nevertheless, this analysis does not account for the shared BPs that are represented by different interactions in each phenotype, which could bias our comparison between the phenotypes. To have a better sense of the expression trends of the BPs in each phenotype, we performed the same community detection and functional enrichment analysis workflow on the individual cancer networks and identified the processes that were commonly present in the four HCs. We integrated this information in a heatmap (Fig. 7) that shows the expression trends of these common processes in each phenotype.

This analysis shows a contrast in the differential expression trends among HCs. For example, several communities are enriched for processes related to cell cycle and regulation of the immune system, but they have a clear different expression trend among the four phenotypes. The latter suggests differences on their regulation in each case.

The main similarity in the figure are those BPs related with the assembly of ribonucleoprotein complexes, whose genes are mainly overexpressed. The processes related with cell division present a mixture of differential expression trend, which is also depicted in the four networks at the left part of the figure.

In the case of multiple myeloma, all protein-coding genes except for one (DCTN6) are underexpressed. We believe that this underexpression pattern must be retained on myeloma cells from their normal cell of origin, the long-lived plasma cell, which no longer divides after its differentiation from naive B cells (19–21).

Notwithstanding, this expression profile in MM should not to be understood as if the myeloma cells have stopped proliferating. MM is a progressive disease, which typically arises after monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM). The progression from these stages to MM is approximately 1% per year for MGUS and 10% per year for SMM (22). Given these relatively slow rates of progression, we could assume that the proliferation of malignant plasma cells occurs slowly and in a lower rate than the proliferation of normal cells inside the bone marrow, which in turn would explain the underexpression of genes related with cell cycle, and the differences of this in comparison with the acute leukemias, which are malignancies of rapid evolution.

Since the RNA-Seq samples for these analyses were per-

DRAFT

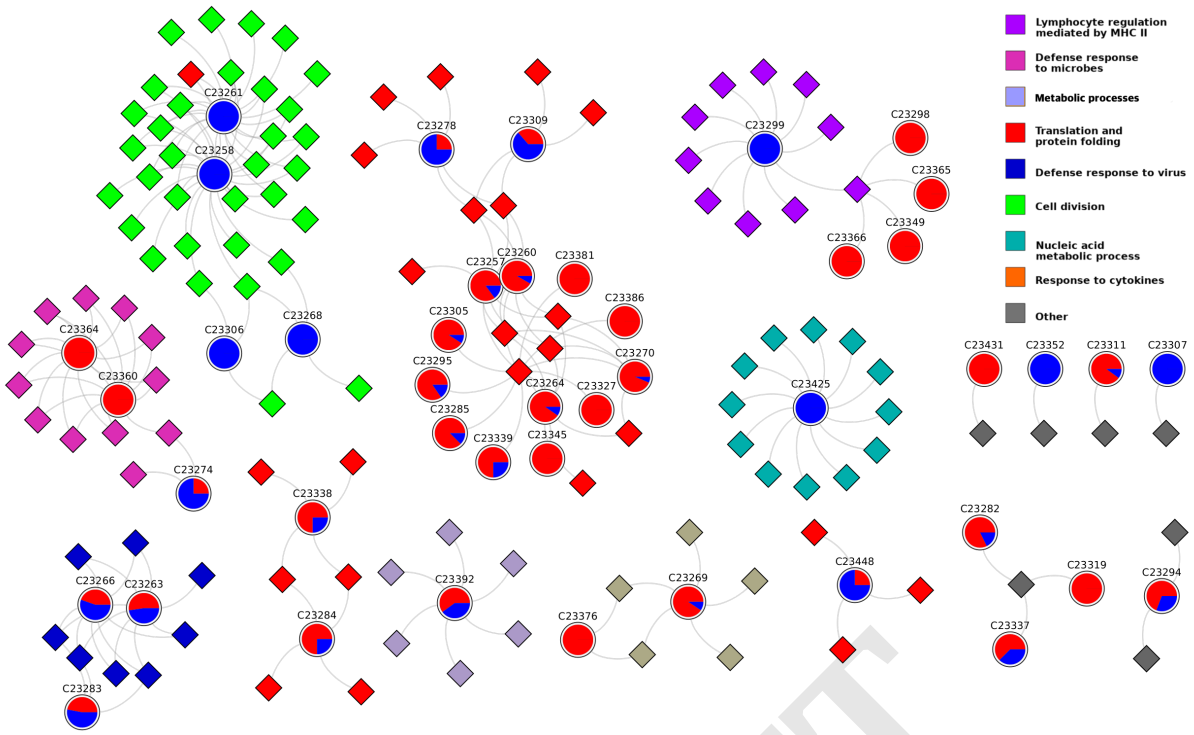

Fig. 6. Bipartite network of communities (circles) detected in the intersection network and their associated biological processes (diamonds). The pie plots in the communities represent the proportions of the expression trends of the genes in those communities. A completely blue circle means that every gene in the community has an under-expression trend, and the opposite for a completely red circle. In this figure, we depict the differential expression trend of the analysis on multiple myeloma.

formed in bulk, we cannot dissect the specific expression profile of the cells into the sample. Given that the BM has a relatively small amount of plasma cells (around 1% of total cells (23)), we believe that further studies involving myeloma and plasma cells could better asses the cell cycle regulation in the disease.

In AML we observed a similar underexpression trend in cell cycle processes, however in this case there are some overexpressed genes that have an active role in the regulation of cell cycle, like CKS1B that is crucial for the function of cyclin dependent kinases and for progression through cell cycle in cancer cells (24, 25).

We also detected the overexpression of PCLAF, a factor associated with PCNA (also overexpressed in AML), that is associated with DNA synthesis and the G1/S transition (26). On the other hand, we detected several underexpressed genes involved in the regulation of the mitotic spindle and chromosome segregation, namely KIFC1, CENPF, CENPI, ASPM (27–32). This suggests that the leukemic cells are actively proliferating, but the mechanisms related with the quality of cell division are downregulated, which could trigger genomic irregularities, like aneuploidies, which are a hallmark of cancer.

For BALL, the majority of the most significant differentially expressed genes are overexpressed and also related with cell proliferation. For example, MYBL2 is the most overexpressed gene in BALL enriched for cell cycle processes and is known to be ubiquitously expressed on proliferating cells (33). Members of the E2F family, namely E2F2 and E2F1, that activate genes for G1/S transition (34, 35) are also overexpressed. As in AML, PCLAF and PCNA are overexpressed, but several underexpressed genes are related with the kinetochore function and cytokinesis, such as DLGAP5 and ANLN.

The aforementioned results in BALL suggest (as in AML) the cancer cells proliferation and, at the same time, a remarkable downregulation of some processes associated with an adequate control of cell division.

In TALL, almost every gene related with cell cycle is overexpressed. The most relevant underexpressed genes are E2F8 (log2FoldChange = -0.9151, padj = $1.21e^{-06}$) and PIF1 (log2FoldChange = -0.6808, padj = $1.49e^{-04}$). The synergystic activity of E2F8 with E2F7 on the regulation of cell cycle is essential for embryonic development, and their knockdown promotes widespread apoptosis (36). Interestingly, E2F7 is overexpressed in the TALL network. Lack of more studies involving the axis of E2F8 and E2F7 limits the interpretation of this result.

Considering the over-expression of E2F1 (which is in turn target of E2F7/8 (36)), we could argue that the underexpression of E2F8 indirectly promotes the over-expression of E2F1, whose inhibition in chronic myeloid leukemia leads to cell cycle arrest (37). Additional studies of the transcriptional regulation of E2F genes, including the E2F7/8 axis, could help elucidate unknown mechanisms of cell cycle regulation in leukemia by this family of genes.

An important amount of common biological processes of the four cancer GCNs are related to the adaptive immune sys-

DRAFT

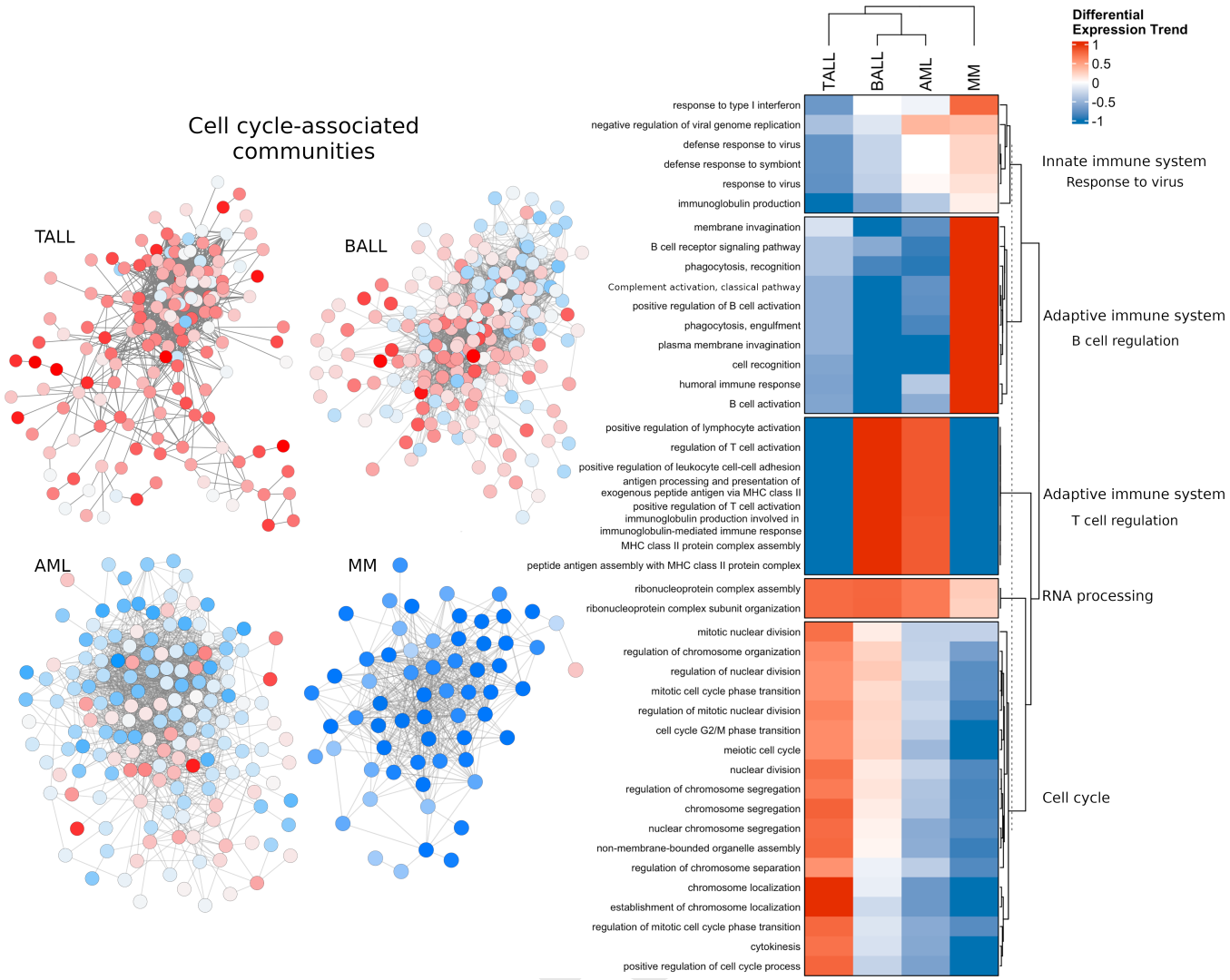

Fig. 7. Differential expression trend of the common biological processes in the four HC networks. A value of 0 means that half of the genes enriched for a BP are over-expressed and the other half under-expressed.

tem. This was expected given the origin of these malignancies. The processes of B cell activation are under-regulated in every phenotype except for MM. On the other hand, the processes related with T cell regulation are exacerbated in TALL and MM, but diminished in BALL and AML. We believe that this expression trend is related to the cells of origin of the malignancies. B cells and members of the myeloid lineage are professional antigen presenting cells (APCs) which activate T cells through MHC molecules. Hence, the over-expression of genes related with this processes (mainly HLA genes) could be a legacy from the normal cell of origin that are specialized in the presentation of antigens. The genes enriched for B cell regulation are mainly immunoglobulins, and their over-expression in MM is expected as its normal cell of origin, the plasma cell, produces antibodies and do not have the ability to present antigens to T cells (which explains the underexpression of HLA genes in this phenotype).

These shared BPs are very similar to those enriched on the intersection network. This indicates that hematological malignancies have a shared set of affected BPs, and that these processes are probably deregulated through different mechanisms in each HCs, given their unique expression pattern in each phenotype.

Multiple myeloma network analysis. After analyzing our four HCs networks, we focused our attention on the MM network, since its co-expression landscape shows the highest fractions of intra-chromosomal interactions, which is a trait that is maintained in every cut-off point of the network until the 100,000 interactions. From this point forward, the fractions starts to decrease (see Fig 1). This shows that in MM, as well as in several carcinomas, the strongest and more relevant interactions of the co-expression landscape are intra-chromosomal. These type of interactions are present in small proportions in the networks of healthy phenotypes, almost as in a randomly constructed co-expression network (73, 74), indicating that the co-expression regulation in cancer is importantly biased towards local and physically close distances between genes. The observed differences in this phenomenon among HCs could help us to elucidate what mechanism are being tempered in cancer to produce this co-expression land-

Table 1. Top and bottom five deferentially expressed genes on B-cell acute lymphoblastic leukemia.

Gene Symbol	Log2FC	p-value	Related function	References
IRX2	12.72	1.75e-66	Related to poor outcome on BALL	(38, 39)
EPHA7	11.84	3.47e-116	Increased expression of this gene is associated with multiple forms of carcinoma	
KLKP1	11.34	5.32e-120	Down-regulated in prostate cancer	(40)
OR5P2	11.12	1.86e-42	-	-
DSC3	11.03	7.05e-31	Its promoter has been reported to be hypermethylated in pediatric ALL	(41)
...	...	...	...	...
UNC5A	-6.49	1.04e-51	Transmembrane receptor that promotes cell death and is reported to be downregulated in some cancers	(42–44)
CCR2	-6.23	6.22e-76	Mediates monocyte chemotaxis, specially in inflammtory disease and in the response against tumors	
HK3	-6.01	1.16e-68	Upregulated in several tumors and correlated with immune infiltration in lung cancer	(45, 46)
CHI3L1	-6.01	3.58e-48	High levels of this gene has been correlated with poor survival on pedriatic ALL and other tumors	(47)
CEACAM3	-5.97	1.17e-60	Glycoprotein receptor on granulocyte lineage cells, mediates the recognition and phagocytosis of microbial pathogens	(48, 49)

Table 2. Top and bottom five differentially expressed genes on T-cell acute lymphoblastic leukemia.

Gene Symbol	Log2FC	p-value	Related function	References
PEX5L-AS2	15.21	5.89e-121	-	-
OR10T2	13.98	7.83e-71	-	-
RNU7-7P	13.09	8.90e-91	-	-
OR10R3P	12.51	6.98e-65	-	-
PCDH10	12.39	4.67e-201	The promoter of this gene has been reported to be hypermethylated in several types of cancers	(50, 51)
...	...	...	...	...
IGHA1	-6.92	1.09e-118	Low levels of IgA has been correlated with advanced Rai stages in patients of CLL	(52)
IGHA2	-6.75	2.34e-105	Low levels of IgA has been correlated with advanced Rai stages in patients of CLL	(52)
NRGN	-6.34	1.47e-99	-	-
TARM1	-6.29	3.02e-100	Expressed primarily in myeloid cells, is involved in T-cell activation. Lack of this receptor on dendritic cells impairs their antigen presenting ability	(53, 54)
RPH3A	-6.27	2.68e-155	-	-

DRAFT

scape.

Interestingly, the MM network has the least amount of interactions between pseudogenes (see Fig. 4). As in the others HC networks, these pseudogenes interact mainly through inter-chromosomal interactions and correspond to pseudogenes of housekeeping genes. This observation of the high fraction of cis-chromosomal edges and the low amount of interactions between pseudogenes in MM, and the opposite behavior of AML suggest that there is a connection between these two variables in the networks. Further research on this is necessary to clarify whether indeed there is a connection between them and what are their role in the cancerous phenotypes.

The cell cycle regulation among HCs seems to form a gradient of differential expression across them. In TALL, these genes are strongly overexpressed, while in BALL they are a little less over-expressed. In AML, genes are predominantly underexpressed, but it still has a relevant number of overexpressed genes related with proliferation. Finally, for MM, its genes are completely underexpressed. As we mentioned before, we believe that in MM these trend of under-expression in cell cycle related genes reflects the low proliferation rates of myeloma cells in comparison with the rates of proliferation in the normal bone marrow.

Table 3. Top and bottom five differentially expressed genes on Multiple myeloma.

Gene Symbol	Log2FC	p-value	Related function	References
MAGEC2	16.33	1.12e-99	Member of the Melanoma Antigen Gene (MAGE) protein family; expression of proteins of this family has been associated with poor prognosis in cancer patients	(55)
MAGEC1	15.09	2.99e-175	Member of MAGE protein family	(55)
LINC01518	14.08	9.07e-64	Its knockdown inhibits tumorigenicity in oesophageal squamous cell carcinoma cells by suppression of the PIK3CA/Akt pathway	(56)
MAGEA3	12.86	1.66e-66	Member of the MAGE protein family	(55)
MAGEA6	12.71	9.59e-66	Member of the MAGE protein family	(55)
...	...	...	...	...
PZP	-8.75	6.73e-191	Protease inhibitor downregulated in several types of cancer, associated with poor prognosis	(57, 58)
MIR3609	-8.55	3.86e-244	Downregulated in chemoresistant breast cancer cells and correlated with poor prognosis	(59)
RN7SL5P	-8.51	5.12e-180	-	-
CD40LG	-8.47	1.54e-106	Transmembrane protein located mainly in the T-cells membrane interacts with CD40 on B-cells, promoting their maturation. Its downregulation has been correlated with poor prognosis on breast cancer. Multiple myeloma cells transfected with CD40LG seems to induce tumor remission on mices	(60, 61)
RMRP	-8.43	1.83e-150	LncRNA overexpressed in various tumors, functioning as a sequestrator of tumor supressor miRNAs. Also functions as a inactivator of p53 activity	(62, 63)

The functional enrichment analysis on the MM network yielded several biological processes related with the disease. Besides the communities of underexpressed genes enriched for processes of cell cycle we also detected communities of metabolism and translation with mainly over-expressed genes. There are also communities related with the regulation of the immune system. We found that the communities enriched for `Immunoglobulin production` are composed of genes of the kappa light chain. On the other hand, a community formed by overexpressed genes of the immunoglobulin gamma heavy chain is also enriched for several processes related with B cell regulation (Supplementary Material). This type of kappa light chain and the gamma heavy chain conform the most common type of immunoglobulin in patients of MM, which typically produce an IgG kappa type immunoglobulin (75). Additionally, a community formed mainly by the genes of the human leukocyte antigen (HLA) corresponding to the class II of the major histocompatibility complex (MHC-II) is enriched for processes of T cell regulation and for immunoglobulin production, but these genes are underexpress in MM. Like we stressed before, this reflects the lack of antigen presenting function in myeloma cells.

Final considerations and future perspectives

In this work, we performed a global analysis of the transcriptional landscape from several hematopoietic cancers. Through the use of GCNs we were able to identify several gene interactions that are shared by the four HC networks, which translates in several biological processes common to these malignancies. Nevertheless, the genes in these common interactions have a unique differential expression profile that reflects differences in their regulation in each phenotype. It is necessary to further analyze the topology of the individual GCNs of HCs to understand how the circuits of normal hematopoietic cells are deregulated to promote these diseases.

An aspect that also needs to be further explored is the topology of normal bone marrow network. Here, the data set used was made up of 37 RNA-Seq samples, which is the total number of normal bone marrow samples present in the TCGA dataset. Given that we performed a normalization work-flow for each HC data set with the same normal group, we obtained a total of four networks for the normal bone marrow. However, our approach can be highly sensitive to the number of samples analyzed, which could represent a limitation to detect co-expression interactions with a high confidence. Further studies using a bigger group size could provide more robust networks to further compare the topology of the normal and cancerous phenotype and identify interactions dereg-

DRAFT

Table 4. Top and bottom five differentially expressed genes on Acute myeloid leukemia.

Gene Symbol	Log2FC	p-value	Related function	References
HMX3	9.74	2.40e-23	Usually silenced during normal hematopoiesis. Its expression seems to be related with myeloid differentiation arrest and with oncogenic ERK-signalling	(64)
DSC3	9.72	1.32e-14	Its promoter has been reported to be hypermethylated in pediatric ALL	(41)
PRAME	9.18	1.51e-79	Associated with inhibition of myeloid differentiation. Overexpression in Chronic myeloid leukemia has been associated with poor prognosis, however, the overexpression in pedriatic AML has been associated with good outcomes	(65, 66)
PCDH7	8.97	5.32e-07	Its overexpression inhibits homotypic cell-in-cell structures and promotes tumor growth in breast cancer cell lines	(67)
IGF2BP1	8.92	6.57e-39	This gene promotes stemness properties in leukemia cells through post-transcriptional regulation of HOXB4, MYB and ALDH1A1	(68)
...	...	...	...	...
SCARNA10	-9.54	1.59e-147	High levels of this lncRNA have been found in patients of hepatocellular carcinoma and breast cancer	(69, 70)
SCARNA5	-7.47	4.64e-124	-	-
SNORA74B	-7.06	1.83e-97	Reportedly over-expressed in gallbladder cancer	(71)
SNORA53	-7.05	2.05e-113	Downregulated upon RAS-induced senescence in human skin fibroblasts	(72)
KRT17P8	-7.05	1.42e-88	-	-

DRAFT

ulated in the disease. Nevertheless, our results show that our network theory approach to investigate cancer phenotypes is able to elucidate relevant clinical features. For example, the communities associated with erythrocyte development that are downregulated, evidence the anemia suffered by patients of these diseases. To clarify whether these biological features of HCs are a consequence of the deregulation of the normal co-expression landscape, additional research is necessary.

Our results also show other features shared by the co-expression landscape of HCs, like the loss of inter-chromosomal regulation and the co-expression between genes with similar functions. Furthermore, we identified a set of common biological processes that are deregulated in the four HC phenotypes. However, the way in which these common processes are altered in each phenotype must differ, given their distinct differential expression trends. Future research should be done regarding the co-expression regulation between pseudogenes of housekeeping genes, which appear to interact through trans-chromosome interactions and that could be related with the loss of inter-chromosomal co-expression that has been observed in every cancer network we have analyzed.

An important aspect of the GCNs of hematologic cancers should be taken into account: the effect of the ages of the patients whose samples are included in these type of studies. Multiple myeloma, the network with the more pronounced loss of inter-chromosomal co-expression, was inferred with all patients in advanced edges, meanwhile the rest of HCs were obtained from pediatric patients. This difference must to be broadly explored.

Future investigation should be done involving lymphomas samples, which were neglected here given the lack of an appropriate control group to compare these samples. Single-cell RNA-Seq data will certainly improve the accuracy of the gene co-expression landscape. To analyze other -omics data could add molecular interactions that may help to disentangle the molecular mechanisms that make different these hematopoietic malignancies. However, with this approach we emphasize the importance of this kind of analyses to better understand the co-expression landscape in cancer and its associated functional features.

Methods and Materials

The computational code to develop the present pipeline is available on GitHub. It can be found at https://github.com/AKNG97/Network_structure_of_hematopoietic_cancers. All data shown here can be reproduced using this code.

Data acquisition. The samples used in this worked were retrieved from https://portal.gdc.cancer.gov, The Cancer Genome Atlas consortium. The cancer samples included those categorized as "Recurrent Blood Derived

Cancer - Bone Marrow" and "Primary Blood Derived Cancer - Bone Marrow" of patients with a primary diagnosis of "acute myeloid leukemia, nos", "precursor b-cell lymphoblastic leukemia", "t lymphoblastic leukemia/lymphoma" and "multiple myeloma". Also, normal bone marrow samples used were those categorized as "bone marrow normal".
A complementary annotation file was obtained from BioMart in order to obtain the chromosome name, gene type, GC content, gene start and gene end. The gene length was inferred by the rest of the gene end minus the gene start.

Data pre-processing. Each cancer dataset was pre-processed along with the normal bone marrow dataset. Hence, we obtained four cancer expression matrices and four normal bone marrow matrices, one for each cancer phenotype. The first step in our workflow was to apply according filters to remove low count genes by means of the TCGAbiolinks package in R (76), as follows:

1. Genes on the 0.25 quantile
2. Genes with an expression value of 0 across the majority of the samples in the datasets
3. Genes with a mean expression less than 10 were removed.

After filtering the genes, we performed normalization procedures for several biases of the sequencing technology using EDASeq (77) and NOISeq (78) in R:

1. Normalization for the length transcript and the GC content of the genes.
2. Between-lane normalization followed by Trimmed Mean of M values (TMM).
3. Lastly, we apply the multidimensional noise reduction algorithm ARSyN to assure that the cancer and normal samples were separate and clustered correctly, PCA visual exploration was used to confirm this (Supplementary Material 5).

Data processing.

Differential expression analysis. We tested for differentially expressed genes in each cancer data set in comparison to the normal bone marrow data set using the DESeq2 (79) package in R, which is based on a model using the negative binomial distribution.

Networks construction. Despite mutual information (MI) is a good estimator of dependency between variables, it can be strongly affected by the number of samples in the analysis. Considering the small amount of samples in our control group, we constructed networks with both MI and Spearman correlation, which is less dependent on the number of samples, although it specifically detects monotonic relationships. To decide with which type of network (MI or Spearman correlation) we should continue our analysis, we identified the gene interactions that are shared between each network in different cut-off points (Supplementary Material 4).
Mutual information values were calculated by implementing the ARACNe algorithm (80). Spearman correlation values were calculated using the function rcorr from the package Hmisc in R. These co-expression values were sorted using their absolute values for further analyses.

Intersection network construction The intersection network was constructed from the common interactions between each of the four cancer GCNs using their respective 100,000 strongest interactions.

Community detection and functional enrichment analysis. In order to correlate topological structures inside our networks with biological processes, we first performed the community detection algorithm HiDeF through the CyCommunityDetection app in Cytoscape (81). This integrates the concept of persistent homology with existing algorithms for community detection in order to identify robust structures intrinsic to a dataset. We applied HiDeF with its default setting in Cytoscape using the Louvain algorithm. The communities detected were then enriched using the R package clusterProfiler to correlate the genes inside the communities to biological processes of Gene Ontology. We used the EnsemblID of each gene in this analysis and considered that a community was enriched if the correlation to a biological process had an adjusted p-value less than $1e^{-10}$).

Analysis of the common biological processes on the four HCs We calculated the percentage of up and down expressed genes on the communities enriched for each BP on the individual cancer networks. We rested the proportions and called the result "differential expression trend" for a BP on a specific phenotype. A value of 1 depicts that every gene enriched for a BP has an over-expression tendency, a value of -1 depicts that every gene has an under-expression tendency, and a value of 0 depicts that there are the same proportion of up and down expressed genes enriched for the process.

Acknowledgments

Arturo Kenzuke Nakamura-Garcia is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 806341 from CONACYT. Both authors thank to National Institute of Genomic Medicine (INMEGEN) where this work took place.

Supplementary material

1. Network files for the top-100,000 interactions in the five phenotypes.
2. Differential gene expression analysis for all hematopoietic cancers.
3. Community structure of the four HC networks. These files contain the name of all communities and the corresponding elements for each community.

DRAFT

4. Top. Bipartite networks of communities and enriched biological processes. Bottom. Shared interactions between MI and Spearman networks.

5. Quality control for data pre-processing.

Bibliography

1. Tak W. Mak, Mary E. Saunders, and Bradley D. Jett, editors. *Chapter 20 - Hematopoietic Cancers*, pages 553–585. Academic Cell, Boston, second edition edition, 2014. ISBN 978-0-12-385245-8. doi: https://doi.org/10.1016/B978-0-12-385245-8.00020-0.
2. Jesus Espinal-Enriquez, Cristóbal Fresno, Guillermo Anda-Jáuregui, and Enrique Hernández-Lemus. Rna-seq based genome-wide analysis reveals loss of inter-chromosomal regulation in breast cancer. *Scientific reports*, 7(1):1–19, 2017.
3. Diana García-Cortés, Guillermo de Anda-Jáuregui, Cristóbal Fresno, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Gene co-expression is distance-dependent in breast cancer. *Frontiers in oncology*, 10:1232, 2020.
4. Alfredo González-Espinoza, Jose Zamora-Fuentes, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Gene co-expression in breast cancer: A matter of distance. *Frontiers in oncology*, page 4743, 2021.
5. Sergio Daniel Andonegui-Elguera, José María Zamora-Fuentes, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Loss of long distance co-expression in lung cancer. *Frontiers in genetics*, 12, 2021.
6. Jose María Zamora-Fuentes, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Gene expression and co-expression networks are strongly altered through stages in clear cell renal carcinoma. *Frontiers in genetics*, page 1232, 2020.
7. Jose María Zamora-Fuentes, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Oncogenic role of mir-217 during clear cell renal carcinoma progression. *Frontiers in oncology*, 12:934711–934711, 2022.
8. Diogo M Ribeiro, Simone Rubinacci, Anna Ramisch, Robin J Hofmeister, Emmanouil T Dermitzakis, and Olivier Delaneau. The molecular basis, genetic control and pleiotropic effects of local gene co-expression. *Nature communications*, 12(1):1–13, 2021.
9. Diana García-Cortés, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Luminal a breast cancer co-expression network: Structural and functional alterations. *Frontiers in genetics*, 12:629475, 2021.
10. Deyou Zheng, Adam Frankish, Robert Baertsch, Philipp Kapranov, Alexandre Reymond, Siew Woh Choo, Yontao Lu, France Denoeud, Stylianos E Antonarakis, Michael Snyder, et al. Pseudogenes in the encode regions: consensus annotation, analysis of transcription, and evolution. *Genome research*, 17(6):839–851, 2007.
11. Shanker Kalyana-Sundaram, Chandan Kumar-Sinha, Sunita Shankar, Dan R Robinson, Yi-Mi Wu, Xuhong Cao, Irfan A Asangani, Vishal Kothari, John R Prensner, Robert J Lonigro, et al. Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell*, 149 (7):1622–1634, 2012.
12. Leonardo Salmena, Laura Poliseno, Yvonne Tay, Lev Kats, and Pier Paolo Pandolfi. A cerna hypothesis: the rosetta stone of a hidden rna language? *Cell*, 146(3):353–358, 2011.
13. Pavel Sumazin, Xuerui Yang, Hua-Sheng Chiu, Wei-Jen Chung, Archana Iyer, David Llobet-Navas, Presha Rajbhandari, Mukesh Bansal, Paolo Guarnieri, Jose Silva, et al. An extensive microrna-mediated network of rna-rna interactions regulates established oncogenic pathways in glioblastoma. *Cell*, 147(2):370–381, 2011.
14. Marco Ranzani, Vivek Iyer, Ximena Ibarra-Soria, Martin Del Castillo Velasco-Herrera, Mathew Garnett, Darren Logan, and David J Adams. Revisiting olfactory receptors as putative drivers of cancer. *Wellcome open research*, 2, 2017.
15. Fan Zheng, She Zhang, Christopher Churas, Dexter Pratt, Ivet Bahar, and Trey Ideker. Hidef: identifying persistent structures in multiscale 'omics data. *Genome biology*, 22(1): 1–15, 2021.
16. Tianzhi Wu, Erqiang Hu, Shuangbin Xu, Meijun Chen, Pingfan Guo, Zehan Dai, Tingze Feng, Lang Zhou, Wenli Tang, LI Zhan, et al. clusterprofiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, 2(3):100141, 2021.
17. Diana Garcia-Cortes, Enrique Hernandez-Lemus, and Jesus Espinal-Enriquez. Loss of long-range co-expression is a common trait in cancer. *bioRxiv*, 2022.
18. Lilyn Daftuar, Yan Zhu, Xavier Jacq, and Carol Prives. Ribosomal proteins rpl37, rps15 and rps20 regulate the mdm2-p53-mdmx network. *PloS one*, 8(7):e68667, 2013.
19. Michelle R Tourigny, Josie Ursini-Siegel, Hayyoung Lee, Kai-Michael Toellner, Adam F Cunningham, David S Franklin, Scott Ely, Meihong Chen, Xiao-Feng Qin, Yue Xiong, et al. Cdk inhibitor p18ink4c is required for the generation of functional plasma cells. *Immunity*, 17(2): 179–189, 2002.
20. Adriana Zingone and W Michael Kuehl. Pathogenesis of monoclonal gammopathy of undetermined significance and progression to multiple myeloma. In *Seminars in hematology*, volume 48, pages 4–12. Elsevier, 2011.
21. S Chen-Kiang. Biology of plasma cells. *Best practice & research Clinical haematology*, 18 (4):493–507, 2005.
22. María-Victoria Mateos and Ola Landgren. Mgus and smoldering multiple myeloma: diagnosis and epidemiology. *Plasma Cell Dyscrasias*, pages 3–12, 2016.
23. Heather A Minges Wols, Gregory H Underhill, Geoffrey S Kansas, and Pamela L Witte. The role of bone marrow-derived stromal cells in the maintenance of plasma cell longevity. *The Journal of Immunology*, 169(8):4213–4221, 2002.
24. Wenwen Shi, Qiudi Huang, Jiacui Xie, He Wang, Xiyong Yu, and Yi Zhou. Cks1b as drug resistance-inducing gene—a potential target to improve cancer therapy. *Frontiers in Oncology*, 10:582451, 2020.
25. Louise Westbrook, Marina Manuvakhova, Francis G Kern, Norman R Estes, Harish N Ramanathan, and Jaideep V Thottassery. Cks1 regulates cdk1 expression: a novel role during mitotic entry in breast cancer cells. *Cancer Research*, 67(23):11393–11401, 2007.
26. Xiaowei Liu, Yuanxia Cai, Cheng Cheng, Yaoyao Gu, Xiaoxiao Hu, Kai Chen, Yeming Wu, and Zhixiang Wu. Pclaf promotes neuroblastoma g1/s cell cycle progression via the e2f1/pttg1 axis. *Cell death & disease*, 13(2):1–12, 2022.
27. Zhenglin Xu, Qi Zhang, Frank Luh, Baiye Jin, and Xiyong Liu. Overexpression of the aspm gene is associated with aggressiveness and poor outcome in bladder cancer. *Oncology letters*, 17(2):1865–1876, 2019.
28. Na Ding, Rongxin Li, Wenhao Shi, and Cui He. Cenpi is overexpressed in colorectal cancer and regulates cell migration and invasion. *Gene*, 674:80–86, 2018.
29. Jingbo Sun, Jingzhan Huang, Jin Lan, Kun Zhou, Yuan Gao, Zhigao Song, Yunyao Deng, Lixin Liu, Ying Dong, and Xiaolong Liu. Overexpression of cenpf correlates with poor prognosis and tumor bone metastasis in breast cancer. *Cancer cell international*, 19(1):1–11, 2019.
30. Muhammad Shahid, Min Young Lee, Honit Piplani, Allen M Andres, Bo Zhou, Austin Yeon, Minjung Kim, Hyung L Kim, and Jayoung Kim. Centromere protein f (cenpf), a microtubule binding protein, modulates cancer metabolism by regulating pyruvate kinase m2 phosphorylation signaling. *Cell Cycle*, 17(24):2802–2818, 2018.
31. Hao Wu, Yingjuan Duan, Siming Gong, Qiang Zhu, Xuanyou Liu, and Zhenguo Liu. An integrative pan-cancer analysis of kinesin family member c1 (kifc1) in human tumors. *Biomedicines*, 10(3):637, 2022.
32. Yu-Xi Xiao and Wan-Xi Yang. Kifc1: a promising chemotherapy target for cancer treatment? *Oncotarget*, 7(30):48656, 2016.
33. Maddalena Frau, Sara Ladu, Diego F Calvisi, Maria M Simile, Piero Bonelli, Lucia Daino, Maria L Tomasi, Maria A Seddaiu, Francesco Feo, and Rosa M Pascale. Mybl2 expression is under genetic control and contributes to determine a hepatocellular carcinoma susceptible phenotype. *Journal of hepatology*, 55(1):111–119, 2011.
34. A Iglesias-Ara, O Zenarruzabeitia, L Buelta, J Merino, and AM Zubiaga. E2f1 and e2f2 prevent replicative stress and subsequent p53-dependent organ involution. *Cell Death & Differentiation*, 22(10):1577–1589, 2015.
35. Nicholas Dyson. The regulation of e2f by prb-family proteins. *Genes & development*, 12 (15):2245–2262, 1998.
36. Jing Li, Cong Ran, Edward Li, Faye Gordon, Grant Comstock, Hasan Siddiqui, Whitney Cleghorn, Hui-Zi Chen, Karl Kornacker, Chang-Gong Liu, et al. Synergistic function of e2f7 and e2f8 is essential for cell survival and embryonic development. *Developmental cell*, 14 (1):62–75, 2008.
37. Francesca Pellicano, Laura Park, Lisa EM Hopcroft, Mansi M Shah, Lorna Jackson, Mary T Scott, Cassie J Clarke, Amy Sinclair, Sheela A Abraham, Alan Hair, et al. hsa-mir183/egr1–mediated regulation of e2f1 is required for cml stem/progenitor cell survival. *Blood, The Journal of the American Society of Hematology*, 131(14):1532–1544, 2018.
38. Huining Kang, Carla S Wilson, Richard C Harvey, I-Ming Chen, Maurice H Murphy, Susan R Atlas, Edward J Bedrick, Meenakshi Devidas, Andrew J Carroll, Blaine W Robinson, et al. Gene expression profiles predictive of outcome and age in infant acute lymphoblastic leukemia: a children's oncology group study. *Blood, The Journal of the American Society of Hematology*, 119(8):1872–1881, 2012.
39. Maria E Figueroa, Shann-Ching Chen, Anna K Andersson, Letha A Phillips, Yushan Li, Jason Sotzen, Mondira Kundu, James R Downing, Ari Melnick, Charles G Mullighan, et al. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *The Journal of clinical investigation*, 123(7):3099–3111, 2013.
40. Aneel Kaushal, Stephen A Myers, Ying Dong, John Lai, Olivia L Tan, Loan T Bui, Melanie L Hunt, Matthew R Digby, Hemamali Samaratunga, Robert A Gardiner, et al. A novel transcript from the klkp1 gene is androgen regulated, down-regulated during prostate cancer progression and encodes the first non-serine protease identified from the human kallikrein gene locus. *The Prostate*, 68(4):381–399, 2008.
41. Zac Chatterton, Leah Morenos, Francoise Mechinaud, David M Ashley, Jeffrey M Craig, Alexandra Sexton-Oates, Minhee S Halemba, Mandy Parkinson-Bates, Jane Ng, Debra Morrison, et al. Epigenetic deregulation in pediatric acute lymphoblastic leukemia. *epigenetics*, 9(3):459–467, 2014.
42. Megan E Williams, Xiaowei Lu, William L McKenna, Raesha Washington, Adam Boyette, Phyllis Strickland, Allison Dillon, Zaven Kaprielian, Marc Tessier-Lavigne, and Lindsay Hinck. Unc5a promotes neuronal apoptosis during spinal cord development independent of netrin-1. *Nature neuroscience*, 9(8):996–998, 2006.
43. Karine Thiébault, Laetitia Mazelin, Laurent Pays, Fabien Llambi, Marie-Odile Joly, Jean-Yves Scoazec, Jean-Christophe Saurin, Giovanni Romeo, and Patrick Mehlen. The netrin-1 receptors unc5h are putative tumor suppressors controlling cell death commitment. *Proceedings of the National Academy of Sciences*, 100(7):4173–4178, 2003.
44. FengFei Huang, Manel Ben Aissa, Georges Lévesque, and Madeleine Carreau. Fancc localizes with unc5a at neurite outgrowth and promotes neuritogenesis. *BMC Research Notes*, 11(1):1–7, 2018.
45. Kristina Seiler, Magali Humbert, Petra Minder, Iris Mashimo, Anna M Schläfli, Deborah Krauer, Elena A Federzoni, Bich Vu, James J Moresco, John R Yates, et al. Hexokinase 3 enhances myeloid cell survival via non-glycolytic functions. *Cell death & disease*, 13(5): 1–14, 2022.
46. Zhan Tuo, Xin Zheng, Yan Zong, Jie Li, Chunyan Zou, Yi Lv, and Jun Liu. Hk3 is correlated with immune infiltrates and predicts response to immunotherapy in non-small cell lung cancer. *Clinical and translational medicine*, 10(1):319–330, 2020.
47. Alinne Ayulieth Rivas-Alarcón, Yazmin Gómez-Gómez, Jorge Organista-Nava, Marco Antonio Jiménez-López, Ana Berta Rivera-Ramírez, Eloisa Ibarra-Sierra, Mónica Virginia Saavedra-Herrera, Berenice Illades-Aguiar, and Marco Antonio Leyva-Vázquez. Plasma levels of ykl-40 as a prognostic factor in childhood acute lymphoblastic leukemia. *Molecular and Clinical Oncology*, 15(2):1–8, 2021.
48. Stefan Pils, Dave T Gerrard, Axel Meyer, and Christof R Hauck. Ceacam3: an innate immune receptor directed against human-restricted bacterial pathogens. *International Journal of Medical Microbiology*, 298(7-8):553–560, 2008.
49. Thomas Streichert, Alireza Ebrahimnejad, Stefanie Ganzer, Raid Flayeh, Christoph Wagener, and Jens Brümmer. The microbial receptor ceacam3 is linked to the calprotectin complex in granulocytes. *Biochemical and biophysical research communications*, 289(1): 191–197, 2001.

DRAFT

50. Jianming Ying, Zifen Gao, Hongyu Li, Gopesh Srivastava, Paul G Murray, Hwee Koon Goh, Chai Yen Lim, Yajun Wang, Teresa Marafioti, David Y Mason, et al. Frequent epigenetic silencing of protocadherin 10 by methylation in multiple haematologic malignancies. *British journal of haematology*, 136(6):829–832, 2007.
51. Gopeshwar Narayan, Allen J Freddy, Dongxu Xie, Hema Liyanage, Lorraine Clark, Sergey Kisselev, Ji Un Kang, Subhadra V Nandula, Catherine McGuinn, Shivakumar Subramaniyam, et al. Promoter methylation-mediated inactivation of pcdh10 in acute lymphoblastic leukemia contributes to chemotherapy resistance. *Genes, Chromosomes and Cancer*, 50(12):1043–1053, 2011.
52. Ganchimeg Ishdorj, Erin Streu, Pascal Lambert, Harbhajan S Dhaliwal, Salaheddin M Mahmud, Spencer B Gibson, Versha Banerji, Aaron J Marshall, and James B Johnston. Iga levels at diagnosis predict for infections, time to treatment, and survival in chronic lymphocytic leukemia. *Blood advances*, 3(14):2188–2198, 2019.
53. Valeria Radjabova, Piero Mastroeni, Karsten Skjødt, Paola Zaccone, Bernard de Bono, Jane C Goodall, Edwin R Chilvers, Jatinder K Juss, Des C Jones, John Trowsdale, et al. Tarm1 is a novel leukocyte receptor complex–encoded itam receptor that costimulates proinflammatory cytokine secretion by macrophages and neutrophils. *The Journal of Immunology*, 195(7):3149–3159, 2015.
54. Rikio Yabe, Soo-Hyun Chung, Masanori A Murayama, Sachiko Kubo, Kenji Shimizu, Yukiko Akahori, Takumi Maruhashi, Akimasa Seno, Tomonori Kaifu, Shinobu Saijo, et al. Tarm1 contributes to development of arthritis by activating dendritic cells through recognition of collagens. *Nature communications*, 12(1):1–12, 2021.
55. Jenny L Weon and Patrick Ryan Potts. The mage protein family and cancer. *Current opinion in cell biology*, 37:1–8, 2015.
56. Donghong Zhang, Haifeng Zhang, Xiaolong Wang, Baoli Hu, Feng Zhang, Haitao Wei, and Li Li. Linc01518 knockdown inhibits tumorigenicity by suppression of pik3ca/akt pathway in oesophageal squamous cell carcinoma. *Artificial Cells, Nanomedicine, and Biotechnology*, 47(1):4284–4292, 2019.
57. Minhua Wu, Hui Lan, Zhongwei Ye, and Yonghui Wang. Hypermethylation of the pzp gene is associated with hepatocellular carcinoma cell proliferation, invasion and migration. *FEBS Open bio*, 11(3):826–832, 2021.
58. Rohit Kumar, Ekaterina Kuligina, Anna Sokolenko, Quadir Siddiqui, Nilesh Gardi, Sudeep Gupta, Ashok K Varma, and Syed K Hasan. Genetic ablation of pregnancy zone protein promotes breast cancer progression by activating tgf-β/smad signaling. *Breast Cancer Research and Treatment*, 185(2):317–330, 2021.
59. Duolu Li, Xinru Wang, Meng Yang, Quancheng Kan, and Zhenfeng Duan. mir3609 sensitizes breast cancer cells to adriamycin by blocking the programmed death-ligand 1 immune checkpoint. *Experimental Cell Research*, 380(1):20–28, 2019.
60. Manqiu Yuan, Jianying Pei, Ruihao Li, Lirong Tian, Xin He, and Yanping Li. Cd40lg as a prognostic molecular marker regulates tumor microenvironment through immune process in breast cancer. *International Journal of General Medicine*, 14:8833, 2021.
61. Young Kim, John W Strehl, Elisabeth Sievers, Marcus Gorschlueter, Peter Brossart, and Ingo GH Schmidt-Wolf. Cd40l-transfected myeloma cells transfer prolonged immunity in vivo. *in vivo*, 24(1):45–48, 2010.
62. Qian Hao and Xiang Zhou. The emerging role of long noncoding rna rmrp in cancer development and targeted therapy. *Cancer Biology & Medicine*, 19(2):140, 2022.
63. Yajie Chen, Qian Hao, Shanshan Wang, Mingming Cao, Yingdan Huang, Xiaoling Weng, Jieqiong Wang, Zhen Zhang, Xianghuo He, Hua Lu, et al. Inactivation of the tumor suppressor p53 by long noncoding rna rmrp. *Proceedings of the National Academy of Sciences*, 118(29):e2026813118, 2021.
64. Stefan Nagel, Claudia Pommerenke, Corinna Meyer, Roderick AF MacLeod, and Hans G Drexler. Aberrant expression of nkl homeobox genes hmx2 and hmx3 interferes with cell differentiation in acute myeloid leukemia. *Plos one*, 15(10):e0240120, 2020.
65. Vivian G Oehler, Katherine A Guthrie, Carrie L Cummings, Kathleen Sabo, Brent L Wood, Ted Gooley, Taimei Yang, Mirjam T Epping, Yaping Shou, Era Pogosova-Agadjanyan, et al. The preferentially expressed antigen in melanoma (prame) inhibits myeloid differentiation in normal hematopoietic and leukemic progenitor cells. *Blood, The Journal of the American Society of Hematology*, 114(15):3299–3308, 2009.
66. Daniel Steinbach, Johann Hermann, Susanne Viehmann, Felix Zintl, and Bernd Gruhn. Clinical implications of prame gene expression in childhood acute myeloid leukemia. *Cancer genetics and cytogenetics*, 133(2):118–123, 2002.
67. Chenxi Wang, Ang Chen, Banzhan Ruan, Zubiao Niu, Yan Su, Hongquan Qin, You Zheng, Bo Zhang, Lihua Gao, Zhaolie Chen, et al. Pcdh7 inhibits the formation of homotypic cell-in-cell structure. *Frontiers in cell and developmental biology*, 8:329, 2020.
68. Irina A Elcheva, Tyler Wood, Kathryn Chiarolanzio, Bryan Chim, Madeline Wong, Vikash Singh, Chethana P Gowda, Qingli Lu, Markus Hafner, Sinisa Dovat, et al. Rna-binding protein igf2bp1 maintains leukemia stem cell properties by regulating hoxb4, myb, and aldh1a1. *Leukemia*, 34(5):1354–1363, 2020.
69. Yongfeng Jia, Lin Shi, Fen Yun, Xia Liu, Yongxia Chen, Minjie Wang, Chen Chen, Yanni Ren, Yulong Bao, and Li Wang. Transcriptome sequencing profiles reveal lncrnas may involve in breast cancer (er/pr positive type) by interaction with ras associated genes. *Pathology-Research and Practice*, 215(6):152405, 2019.
70. Yawei Han, Wenna Jiang, Yu Wang, Meng Zhao, Yueguo Li, and Li Ren. Serum long non-coding rna scarna10 serves as a potential diagnostic biomarker for hepatocellular carcinoma. *BMC cancer*, 22(1):1–8, 2022.
71. Yiyu Qin, Li Meng, Yang Fu, Zhiwei Quan, Mingzhe Ma, Mingzhe Weng, Zhengdong Zhang, Cuixiang Gao, Xinghua Shi, and Koulan Han. Snora74b gene silencing inhibits gallbladder cancer cells by inducing phlpp and suppressing akt/mtor signaling. *Oncotarget*, 8(12): 19980, 2017.
72. Mary McMahon, Adrian Contreras, Mikael Holm, Tamayo Uechi, Craig M Forester, Xiaming Pang, Cody Jackson, Meredith E Calvert, Bin Chen, David A Quigley, et al. A single h/aca small nucleolar rna mediates tumor suppression downstream of oncogenic ras. *Elife*, 8: e48847, 2019.
73. Rodrigo Dorantes-Gilardi, Diana García-Cortés, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. Multilayer approach reveals organizational principles disrupted in breast cancer co-expression networks. *Applied Network Science*, 5(1):1–23, 2020.
74. Rodrigo Dorantes-Gilardi, Diana García-Cortés, Enrique Hernández-Lemus, and Jesús Espinal-Enríquez. k-core genes underpin structural features of breast cancer. *Scientific Reports*, 11(1):1–17, 2021.
75. Taran K Nandra, Amarpreet Devi, and John R Jones. Multiple myeloma: what a non-haematologist should know. *Clinical Medicine*, 22(3):230, 2022.
76. Antonio Colaprico, Tiago C Silva, Catharina Olsen, Luciano Garofano, Claudia Cava, Davide Garolini, Thais S Sabedot, Tathiane M Malta, Stefano M Pagnotta, Isabella Castiglioni, et al. Tcgabiolinks: an r/bioconductor package for integrative analysis of tcga data. *Nucleic acids research*, 44(8):e71–e71, 2016.
77. Davide Risso, Katja Schwartz, Gavin Sherlock, and Sandrine Dudoit. Gc-content normalization for rna-seq data. *BMC bioinformatics*, 12(1):1–17, 2011.
78. Sonia Tarazona, Fernando García-Alcalde, Joaquín Dopazo, Alberto Ferrer, and Ana Conesa. Differential expression in rna-seq: a matter of depth. *Genome research*, 21(12): 2213–2223, 2011.
79. Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12):1–21, 2014.
80. Adam A Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In *BMC bioinformatics*, volume 7, pages 1–15. BioMed Central, 2006.
81. Akshat Singhal, Song Cao, Christopher Churas, Dexter Pratt, Santo Fortunato, Fan Zheng, and Trey Ideker. Multiscale community detection in cytoscape. *PLoS computational biology*, 16(10):e1008239, 2020.
82. David Amar, Hershel Safer, and Ron Shamir. Dissection of regulatory networks that are altered in disease via differential co-expression. *PLoS Comput Biol*, 9(3):e1002955, 2013.
83. Valeria Fionda. Networks in biology. In Shoba Ranganathan, Michael Gribskov, Kenta Nakai, and Christian Schönbach, editors, *Encyclopedia of Bioinformatics and Computational Biology*, pages 915 – 921. Academic Press, Oxford, 2019. ISBN 978-0-12-811432-2. doi: https://doi.org/10.1016/B978-0-12-809633-8.20420-2.
84. Tadeo Enrique Velazquez-Caldelas, Sergio Antonio Alcalá-Corona, Jesús Espinal-Enríquez, and Enrique Hernandez-Lemus. Unveiling the link between inflammation and adaptive immunity in breast cancer. *Frontiers in immunology*, 10:56, 2019.
85. Sergio Antonio Alcalá-Corona, Jesús Espinal-Enríquez, Guillermo de Anda-Jáuregui, and Enrique Hernández-Lemus. The hierarchical modular structure of her2+ breast cancer network. *Frontiers in physiology*, 9:1423, 2018.
86. Melanie R Hassler and Gerda Egger. Epigenomics of cancer–emerging new concepts. *Biochimie*, 94(11):2219–2230, 2012.
87. Guillermo de Anda-Jáuregui, Tadeo E Velázquez-Caldelas, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Transcriptional network architecture of breast cancer molecular subtypes. *Frontiers in physiology*, 7:568, 2016.
88. Hugo Tovar, Rodrigo García-Herrera, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Transcriptional master regulator analysis in breast cancer genetic networks. *Computational biology and chemistry*, 59:67–77, 2015.
89. Sergio Antonio Alcalá-Corona, Guillermo de Anda-Jáuregui, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. Network modularity in breast cancer molecular subtypes. *Frontiers in physiology*, 8:915, 2017.
90. Enrique Hernández-Lemus, Helena Reyes-Gopar, Jesús Espinal-Enríquez, and Soledad Ochoa. The many faces of gene regulation in cancer: A computational oncogenomics outlook. *Genes*, 10(11):865, 2019.
91. Paolo Tieri, Lorenzo Farina, Manuela Petti, Laura Astolfi, Paola Paci, and Filippo Castiglione. Network inference and reconstruction in bioinformatics. *Encyclopedia of Bioinformatics and Computational Biology*, 2019.
92. Sipko van Dam, Urmo Vosa, Adriaan van der Graaf, Lude Franke, and Joao Pedro de Magalhaes. Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in bioinformatics*, 19(4):575–592, 2018.
93. W Du and O Elemento. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene*, 34(25):3215–3225, 2015.
94. Adam A Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In *BMC bioinformatics*, volume 7, page S7. Springer, 2006.
95. Shi Hao Tan, Fatima Carla Bertulfo, and Takaomi Sanda. Leukemia-initiating cells in t-cell acute lymphoblastic leukemia. *Frontiers in oncology*, 7:218, 2017.
96. Daniel L Morris, David W Kastner, Sabrina Johnson, Marie-Paule Strub, Yi He, Christopher KE Bleck, Duck-Yeon Lee, and Nico Tjandra. Humanin induces conformational changes in the apoptosis regulator bax and sequesters it into fibers, preventing mitochondrial outer-membrane permeabilization. *Journal of Biological Chemistry*, 294(50):19055–19065, 2019.
97. Alan G Porter and Saravanakumar Dhakshinamoorthy. Apoptosis initiated by dependence receptors: a new paradigm for cell death? *Bioessays*, 26(6):656–664, 2004.

DRAFT