Check for updates

The Role of Copy Number Variants in Gene Co-Expression Patterns for Luminal B Breast Tumors

Candelario Hernández-Gómez[1], Enrique Hernández-Lemus[1,2]* and Jesús Espinal-Enríquez[1,2]*

[1]Computational Genomics Division, National Institute of Genomic Medicine, Mexico City, Mexico, [2]Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Mexico City, Mexico

Gene co-expression networks have become a usual approach to integrate the vast amounts of information coming from gene expression studies in cancer cohorts. The reprogramming of the gene regulatory control and the molecular pathways depending on such control are central to the characterization of the disease, aiming to unveil the consequences for cancer prognosis and therapeutics. There is, however, a multitude of factors which have been associated with anomalous control of gene expression in cancer. In the particular case of co-expression patterns, we have previously documented a phenomenon of loss of long distance co-expression in several cancer types, including breast cancer. Of the many potential factors that may contribute to this phenomenology, copy number variants (CNVs) have been often discussed. However, no systematic assessment of the role that CNVs may play in shaping gene co-expression patterns in breast cancer has been performed to date. For this reason we have decided to develop such analysis. In this study, we focus on using probabilistic modeling techniques to evaluate to what extent CNVs affect the phenomenon of long/short range co-expression in Luminal B breast tumors. We analyzed the co-expression patterns in chromosome 8, since it is known to be affected by amplifications/deletions during cancer development. We found that the CNVs pattern in chromosome 8 of Luminal B network does not alter the co-expression patterns significantly, which means that the co-expression program in this cancer phenotype is not determined by CNV structure. Additionally, we found that region 8q24.3 is highly dense in interactions, as well as region p21.3. The most connected genes in this network belong to those cytobands and are associated with several manifestations of cancer in different tissues. Interestingly, among the most connected genes, we found MAF1 and POLR3D, which may constitute an axis of regulation of gene transcription, in particular for non-coding RNA species. We believe that by advancing on our knowledge of the molecular mechanisms behind gene regulation in cancer, we will be better equipped, not only to understand tumor biology, but also to broaden the scope of diagnostic, prognostic and therapeutic interventions to ultimately benefit oncologic patients.

Keywords: breast cancer, co-expression networks, conditional mutual information, copy number variants, luminal B breast cancer

1 INTRODUCTION

It is difficult to exaggerate the negative impact that cancer has, not only as a global health burden, but also as a phenomenon with enormous societal and economic consequences. Last year, cancer was the cause of death for 10 million people, mainly in poor and developing countries. Of all tumor types, breast cancer is the malignant neoplasm with the largest incidence worldwide (Siegel et al., 2020). Many survivors of the disease (five or more years after diagnosis) still have to live with physical and psychological problems that persist over time (Stein et al., 2008) and drastically reduce their quality of life and productivity.

Breast cancer is also a highly complex and heterogeneous disease, both from the molecular and from the clinical/phenotypic standpoints. It is known that breast cancer diagnosis, response to treatment, relapse, and outcome are largely dependent on the molecular features that have been associated with the so-called breast cancer subtypes (Liu et al., 2007; Kittaneh et al., 2013; de Anda-Jáuregui et al., 2016): Luminal A, the most common, estrogen and progesterone receptor positive but, in general, epidermal growth factor 2 receptor negative; Luminal B subtype is positive for estrogen receptor and epidermal growth and negative for progesterone; HER2+, these tumors are negative for estrogen and progesterone and positive for epidermal growth factor 2; finally the Basal subtype tumors, which are mostly (around 80%) *triple negative*, i.e., negative for estrogen, progesterone, and epidermal growth factor 2. These classifications are useful to determine the origin, evolution and treatment to be followed in each case, although each patient is unique and each subtype has peculiarities.

Of the estrogen-positive subtypes, Luminal B is often the most aggressive and although luminal breast tumors are susceptible to be treated with targeted therapy (a fact that commonly, but not always, is associated with better outcomes), in some cases present the worst prognosis for patients. This is so, due to several molecular and functional features that have been related to higher proliferation rates and pharmacological resistance (Cheang et al., 2009; Tran and Bedard, 2011; Creighton, 2012; Ades et al., 2014), or even due to metabolic alterations (Serrano-Carbajal et al., 2020).

Luminal B breast cancer subtype is hormone-receptor positive (estrogen-receptor and/or progesterone-receptor positive), and either HER2 positive or HER2 negative. This subtype presents high levels of Ki-67. In general, Luminal B tumors grow slightly faster than those from Luminal A subtype. Additionally, the prognosis is commonly worse (Li et al., 2016).

Luminal B breast tumors are characterized by a lower expression of estrogen receptor, and low expression of progesterone receptor (Harbeck et al., 2013). It is also defined by aggressive clinical behavior; its prognosis is similar to that of non-luminal cancers (Tran and Bedard, 2011). Bone metastasis appears more often in Luminal B patients than in non-luminal ones. However, recurrence or metastasis in this subtype have a better prognosis after treatment than non-luminal tumors. It has also been shown that Luminal B subtype presents high metabolic deregulation (Li et al., 2016; Serrano-Carbajal et al., 2020).

Luminal B subtype tumors accounted for nearly 40% of all breast cancers (Metzger-Filho et al., 2013). Therefore, understanding the molecular basis of the luminal B subtype is a matter of current concern.

One of the most actively investigated genomic regions in the manifestation of breast cancer is cytoband 8q24.3 (Wokolorczyk et al., 2008; Dorantes-Gilardi et al., 2021). This region results particularly relevant to study in the present context, since 8q24 has been repeatedly reported to harbor multiple variants associated with the incidence of breast cancer and other type of neoplasms (Tong et al., 2020). Indeed, genomic variants in 8q24 have been ascertained to be associated with risk of breast cancer on systematic reviews and meta-analyses (Wang et al., 2020). Particularly interesting is the fact that larger genomic alterations (including copy number variants, CNVs) have been linked to breast cancer onset (Jia et al., 2019), often via disruption of healthy breast cells transcriptional programmes (Ibragimova et al., 2020). Indeed, such effects have been actually linked via clinical and pathological features to breast tumors of the Luminal B subtype (or related: ER+, PR- and HER2-) giving rise to basal-like and endocrine resistant phenotypes (Liu et al., 2018).

It has been argued that some of these genomic alterations have relevant consequences for transcriptional regulation anomalies associated with cancer. Gene regulatory programmes are known to be altered, a fact that has been linked with the onset and development of tumor phenotypes (de Anda-Jáuregui et al., 2016; Hernández-Lemus et al., 2019). In this regard, our group has thoroughly described how the more relevant gene-gene co-expression interactions in several cancer types often occur between genes from the same chromosome (*cis-*), even in proximal chromosomal locations. Conversely, inter-chromosome (*trans-*) interactions (or even long distance intra-chromosome co-expression interactions) are comparatively less abundant and have lower values of diverse statistical dependency measures (Espinal-Enríquez et al., 2017; de Anda-Jáuregui et al., 2019; Dorantes-Gilardi et al., 2020; García-Cortés et al., 2020; Zamora-Fuentes et al., 2020; Andonegui-Elguera et al., 2021; Dorantes-Gilardi et al., 2021; García-Cortés et al., 2021).

Since a large number of molecular players and processes are known to be involved in (normal and) anomalous transcriptional regulatory patterns, establishing which, among the multitude of potential causes of this phenomenon in tumors are actually more relevant given the available experimental evidence becomes desirable. Particularly important has been the discussion on the effects that CNVs have on gene expression patterns in several diseases (Stranger et al., 2007; Henrichsen et al., 2009), including breast cancer (Kumaran et al., 2017; Ohshima et al., 2017; Safonov et al., 2017; Sun et al., 2018; Shao et al., 2019).

With this in mind, we have decided to investigate the effect that CNVs may have on the phenomenon of distance-associated gene co-expression in breast cancer. For the reasons already discussed we have chosen to perform a detailed study centered in the chromosome 8 and the 8q24 region in Luminal B breast tumors. In this study, we implemented a probabilistic modeling approach to evaluate how CNVs affect long and short range co-expression profiles in Luminal B breast tumors. We analyzed the co-expression patterns in chromosome 8, since it is known that

amplifications and deletions in that chromosome may influence cancer development. We found that CNV signatures in chromosome 8 of Luminal B network do not significantly alter the co-expression patterns, hence the co-expression program in this tumor phenotype is not determined by CNV structure. Additionally, we found that the 8q24.3 region is highly dense in co-expression interactions, as well as p21.3. The most connected genes in this network belong to those cytobands and have been associated with several tumor types. Interestingly, among the most connected genes, we found MAF1 and POLR3D, which may constitute an axis of regulation of gene transcription, in particular for non-coding RNA species. We believe that by advancing our knowledge of the molecular mechanisms behind gene regulation in cancer, we will be improving, not only to understand tumor biology, but also the scope of diagnostic, prognostic and therapeutic interventions to ultimately benefit oncologic patients.

This article is organized as follows: **Section 2** explains what conditional mutual information (CMI) is and why it is appropriate in the present investigation; the Kolmogorov-Smirnov method is also presented to quantify the difference between two probability distributions. In **section 3**, the method used to build the networks is explained. It is also established (via conditional mutual information distributions), that CNVs are not significantly associated with the structural features (in particular distance-dependent co-expression patterns) of the analyzed co-expression networks. The results are analyzed in **section 4**. There, the genes with the highest connectivity as well as their possible relationship in the development of cancer are identified and discussed. Finally, in **section 5** the conclusions and some consequences of this work are presented.

2 METHODS

2.1 Data Acquisition
The complete collection of The Cancer Genome Atlas (TCGA) breast RNA-Seq samples was downloaded in January 2019 from the GDC repository https://portal.gdc.cancer.gov/repository. This collection included 113 solid tissue, normal samples and 1,102 primary tumor samples. From these samples, 192 corresponded to Luminal B tumors. Data acquisition was carried out by using the TCGABiolinks R package (Colaprico et al., 2016).

2.2 Data Integration
An integrity check was carried out in raw expression files using gene annotations from BioMart. Only protein coding genes belonging to conventional chromosomes (1, 2, . . . , 22, X and Y) were kept. The CNVs of the micro-RNAs were *masked* from the 8q24.3 region and excluded from the analysis. **Supplementary Materials S1, S2** contains gene expression and genetic information of chromosome 8 genes. Pre-processing and quality control of the gene expression samples were performed as in (Espinal-Enríquez et al., 2017). In brief, we used for NOIseq R library for quality control (Tarazona et al., 2011, 2015). For batch

effect removal, normalization, transcript length and GC content correction, EDASeq library was implemented (Risso et al., 2011). Finally, for multi-dimensional noise reduction we used ARSyN R library (Nueda et al., 2012).

2.3 Conditional Mutual Information Measures
We considered information theoretical measures of statistical dependency as follows: Let X, Y and Z denote discrete random variables having the following features:

1. Finite alphabets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, respectively
2. Joint probability mass distributions $p(X, Y, Z)$, and partial-joint probability mass distributions $p(X, Y)$, $p(X, Z)$, etc.,
3. Marginal probability mass distributions p (X), p (Y) and p (Z)

Let also $\hat{X}$, $\hat{Y}$ and $\hat{Z}$ denote additional discrete random variables defined on X, Y and Z respectively, the associated probability mass distributions will be $\widehat{p(X)}$, $\widehat{p(Y)}$ and $\widehat{p(Z)}$, their joint probability mass distribution $p(\hat{X}, \hat{Y}, \hat{Z})$ defined on $\mathcal{J}$, the joint probability sampling space; $\mathcal{J} = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. For particular realizations, we have $p(x) = P(X = x)$, $\widehat{p(y)} = P(\hat{Y} = y)$, etc., It is possible to define the Conditional Mutual Information (CMI) function $I(X; Y|Z)$ as follows:

$$I(X; Y|Z) = \sum_{z \in \mathcal{Z}} \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p_{X,Y,Z}(x, y, z) \log \frac{p_Z(z) p_{X,Y,Z}(x, y, z)}{p_{X,Z}(x, z) p_{Y,Z}(y, z)}$$

$$(1)$$

Formally $I(X; Y|Z)$ is a measure representing the expected value of the mutual information of two random variables X and Y given the value of a third random value Z. Thus $I(X; Y|Z)$ represents the expected value (w.r.t. Z) of the Kullback-Liebler divergence from the conditional joint distribution $P(X, Y|Z)$ to the product of the conditional marginals $P(X|Z)$ and $P(Y|Z)$.

CMI calculations were performed with R infotheo library (Meyer and Meyer, 2009).

2.4 Assessment of the Impact of Copy Number Variant in Gene Co-Expression Programs
In order to ascertain to what extent the CNVs are able to influence the gene co-expression programs, we performed Kolmogorov-Smirnov (KS) tests to evaluate the differences among the diverse CMI distributions. The KS statistic between 2 distributions is defined as:

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)| (2)$$

here, $F_{1,n}(x)$ and $F_{2,m}(x)$ are the empirical distribution functions of the first and the second sets. Statistical significance of the KS tests is asymptotically given as follows.

The null hypothesis is rejected (at significance level α), whenever

$$D_{n,m} > c(\alpha)\sqrt{\frac{n+m}{n \cdot m}} \qquad (3)$$

with $c(\alpha) = \sqrt{-\ln\left(\frac{\alpha}{2}\right) \cdot \frac{1}{2}}$

KS tests were performed with the ks. test library in R.

Circos plots visualizations were made using the BioCircos JavaScript library (Cui et al., 2016).

3 IMPLEMENTATION

It has been argued that mutual information (MI) is a reliable measure to establish links between genes in co-expression networks (Margolin et al., 2006a,b; Hernández-Lemus and Siqueiros-García, 2013; Lachmann et al., 2016). Given that there is generally a sufficient amount of data available to reconstruct the probability distributions associated with the expression of genes, it is possible to use it without major restrictions when measuring the statistical dependency structures between them (Hernández-Lemus and Rangel-Escareño, 2011). It is also a measure that takes into account the non-linear contributions of interdependence in the series, which makes it more appropriate in the context of the complex regulatory patterns of gene expression.

The adoption of MI-based network deconvolution methods has opened the entrance of the tenets of information theory in the analysis of biomolecular networks. However, when, as in the present work, it is necessary to evaluate the mutual information between two random variables (corresponding to each of the individual gene expression profiles in the present case), given a third, potentially influential feature, one relevant alternative is the evaluation of the Conditional Mutual Information measures. Indeed CMI has already been used in the construction of regulatory networks in a different but related context to the one presented here (Liang and Wang, 2008; Zhang et al., 2012).

An underlying problem when applying MI and CMI to establish the dependency between variables is choosing the method for reconstruction of the associated probability distributions. There are two main ways to do this, using the k-nearest neighbor non-parametric method (Kraskov et al., 2004) and through kernel density estimation (Terrell and Scott, 1992). In this work we have chosen the second method, in particular with a Gaussian kernel estimation. The implementation used is the one corresponding to the *R* package Infotheo.

The role of CNVs has been pointed out as a possible element causing the loss of co-expression with the distance that is observed in cancer and that has been reported previously (García-Cortés et al., 2021). Ideally, a non-cancerous person has two copies of each gene, however many disorders of genetic origin are caused by the deletion, repetition or insertion of DNA segments, sometimes as long as the arm of a chromosome or a complete chromosome. Given that in the past, they have been found to be responsible for genetic diseases, to think that they may also have a relevant role in the loss of co-expression is an appealing idea.

To systematically evaluate the contribution of CNVs in gene co-expression in cancer, it is convenient to use an area of the genome that is particularly active in its manifestation and to calculate the influence of CNVs both in that region and in the surrounding *areas*. As previously mentioned, 8q24.3 has been identified as particularly active and connected in co-expression networks, mainly in breast cancer tissues. Thus, we chose the Luminal B breast cancer subtype, given the poor prognosis compared with the Luminal A subtype. Furthermore, gene expression patterns have shown altered metabolic pathways even more evident than in Basal-like or HER2+ phenotypes (Serrano-Carbajal et al., 2020). We took the CNV values of 183 coding genes from the 8q24.3 region (**Supplementary Materials S3, S4**), we take them one by one as conditional and calculate the CMI of 442 genes, including not only those of this region but of the entire chromosome 8. **Supplementary Materials S3–S6** show the copy number alteration map of all samples for chromosome 8.

We repeated the above calculation for the same genes with control tissue samples assuming a number of CNVs equal to two in all samples as conditional. From the obtained values we construct co-expression networks. For each of the 183 conditional layers there are 97,461 links. The question behind this analysis is: How important is the difference obtained between the different conditional layers? For control tissue networks the question becomes trivial, since all the conditionals are essentially the same, but for those corresponding to Luminal B this difference may be relevant due to the already documented anomalous CNV structure in the region.

To quantify the differences between each of the 183 conditional layers obtained, we calculated the difference between the distributions of CMI values for all possible pairs of them (16,653 comparisons) using the Kolmogorov-Smirnov test. In essence, this test compares the cumulative functions of two distributions and it establishes as a measure, the D statistic, i.e. the maximum vertical difference between them (**Supplementary material S5**). In **Figure 1** we show graphically the workflow for this project.

4 RESULTS AND DISCUSSION

4.1 Contribution of Copy Number Variants to the Co-Expression Program is Marginal

Figure 2 shows a heatmap with the values of the D statistic of the Kolmogorov-Smirnov test for each different pair of the 183 distributions of CMI values obtained. The largest of them, which occurs between the layers corresponding to the CNVs of the COL22A1 gene (ENSG00000169436) and (ENSG0000016943), is less than 0.06. This result indicates that the contribution of 8q24.3 CNVs in the co-expression networks of genes on chromosome 8 in the luminal B subtype of cancer is marginal.

This result acquires relevance since copy number alterations are a mesoscopic phenomena, which can affect large parts of the genome. On the other hand, changes in the co-expression landscape are considered a microscopic event, since those changes may affect specific genes and their regulatory relationships. Hence, with this result we provide evidence that
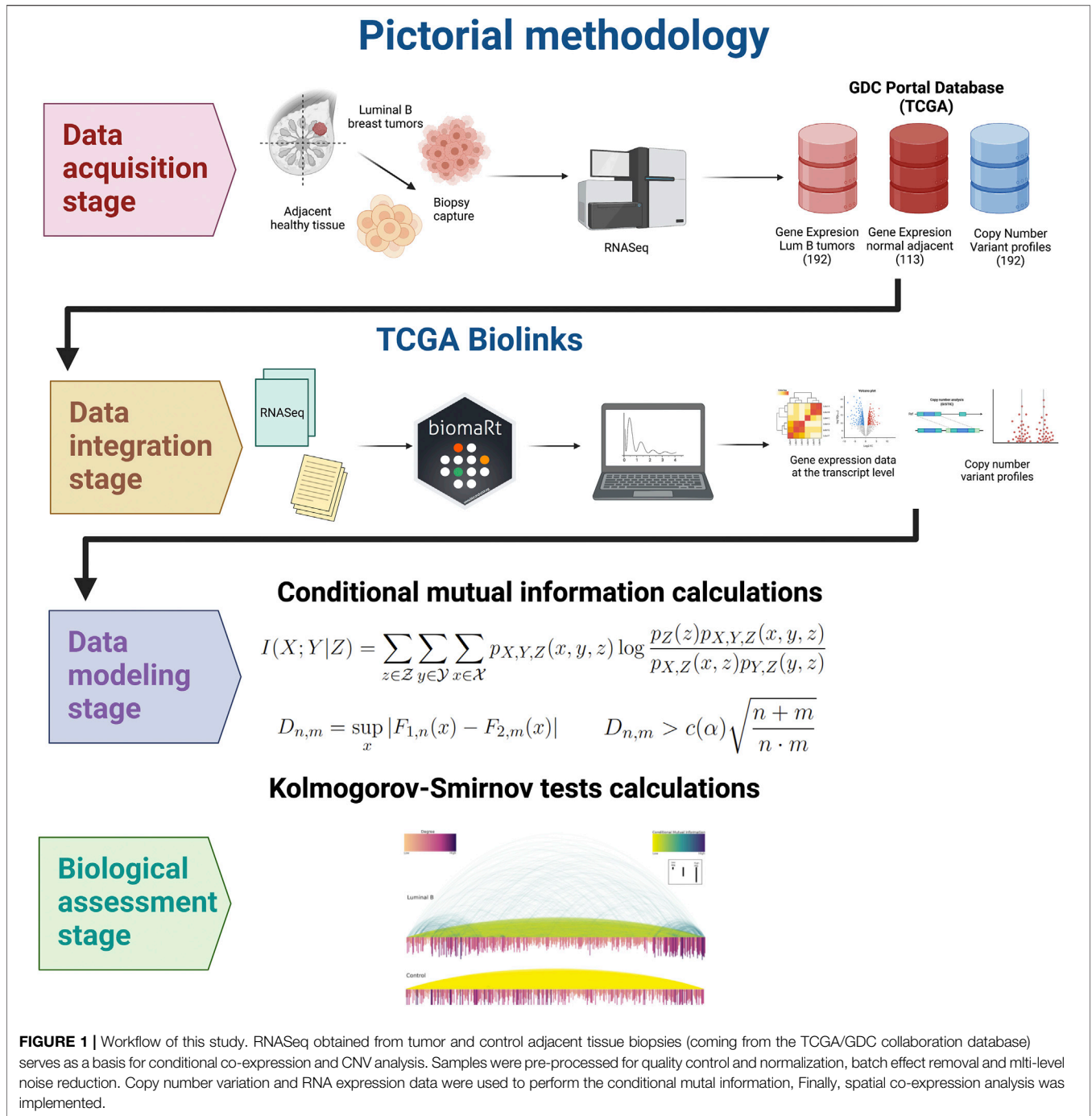
FIGURE 1 | Workflow of this study. RNASeq obtained from tumor and control adjacent tissue biopsies (coming from the TCGA/GDC collaboration database) serves as a basis for conditional co-expression and CNV analysis. Samples were pre-processed for quality control and normalization, batch effect removal and mlti-level noise reduction. Copy number variation and RNA expression data were used to perform the conditional mutal information, Finally, spatial co-expression analysis was implemented.
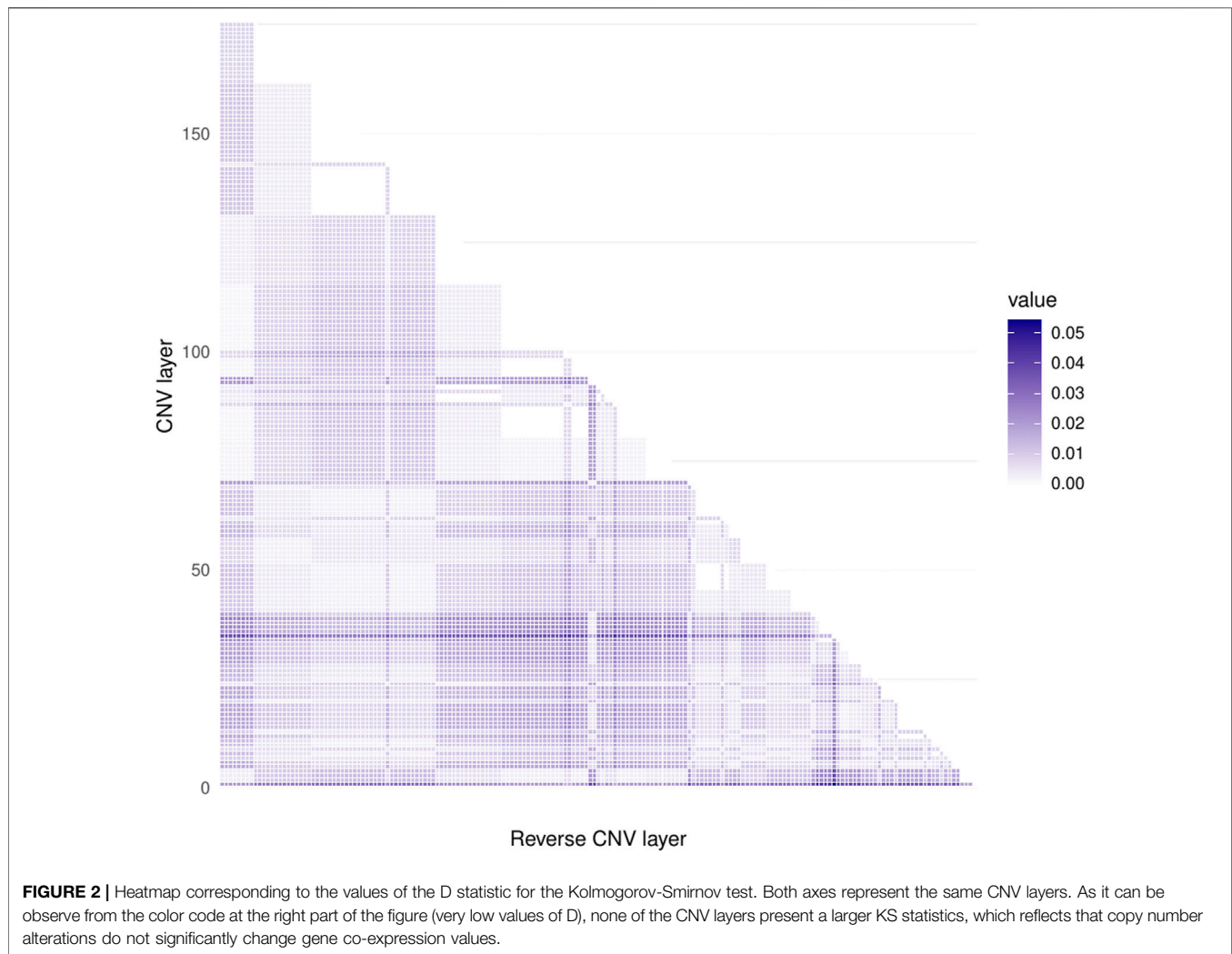
a large-scale event such as amplifications/deletions of large portions of chromosome 8 do not significantly alter the gene co-expression program in cancer.

4.2 Conditional Mutual Information is Higher in Luminal B Than in Controls

Since all the layers corresponding to Luminal B samples have essentially the same distribution of CMI, we take the first, whose conditionals are the CNVs of the COL22A1 gene, as representative of the entire family of networks and proceed to analyze it. **Figure 3** shows the difference between CMI from the Luminal B subtype network and the one from the control network. As it can be clearly observed, the control CMI values are lower than the cancer counterpart.

In (Espinal-Enríquez et al., 2017), we observed that mutual information values from control gene co-expression networks are higher than in the cancer-derived network. However, the calculation was taken over all the gene-gene co-expression interactions, i.e. intra and inter-chromosomal gene

FIGURE 2 | Heatmap corresponding to the values of the D statistic for the Kolmogorov-Smirnov test. Both axes represent the same CNV layers. As it can be observe from the color code at the right part of the figure (very low values of D), none of the CNV layers present a larger KS statistics, which reflects that copy number alterations do not significantly change gene co-expression values.

correlations. In this case, the CMI is taken between chromosome 8 genes only, i.e., just intra-chromosome interactions.

Taking into account that the loss of long-range co-expression phenomenon has been previously reported in networks form different cancer types (Zamora-Fuentes et al., 2020; Andonegui-Elguera et al., 2021), and also for breast carcinoma (Espinal-Enríquez et al., 2017; Alcalá-Corona et al., 2018; de Anda-Jáuregui et al., 2019; Dorantes-Gilardi et al., 2020), and in particular for breast cancer subtype networks (de Anda-Jáuregui et al., 2019; García-Cortés et al., 2020; Dorantes-Gilardi et al., 2021), the finding of a higher average co-expression in intra-chromosome 8 of Luminal B breast cancer compared with the control one, reinforces the fact that this phenomenon is a common trait in several types of cancer.

4.3 Cancer Network *Grows* in Small Local Regions, but not in Control

Figure 4 corresponds to Luminal B subtype interactions. Top-left to bottom-right shows the Top-500, 1,000, 2,000, and 3,000 interactions in chromosome 8. As it can be appreciated,

interactions appear local in the first place, markedly in cytobands q24.3 (yellow arcs) and p11.23. Subsequently, inter-cytoband and inter-arm interactions occur, but in a small fraction (top-right). For the top-2,000 edges, it is clear that several interactions appear from q24.3. Finally, for the top-3,000 edges, 8q24.3 region is strongly connected with the rest of chromosome 8, and more inter-arm and inter-cytoband edges appear.

On the other hand, the **Figure 5** shows the existence of a non-localized connectivity for the control network; since for the first 500 highest connections, interactions occur between genes from different cytobands, or even from different arms, much more often than in the cancer phenotype. The same effect is shown in the top-1,000 and 2,000 interaction circos plots. It is worth noticing the small number of intra-cytoband interactions, even in the top-3,000 edges (830 out of 3,000, dark blue edges).

We believe that this pattern where the highest correlations appear between physically close genes, is in agreement with other phenomena observed in breast and other cancers in which the loss of long-distance co-expression is evident, such as kidney (Zamora-Fuentes et al., 2020) or lung (Andonegui-Elguera et al., 2021).
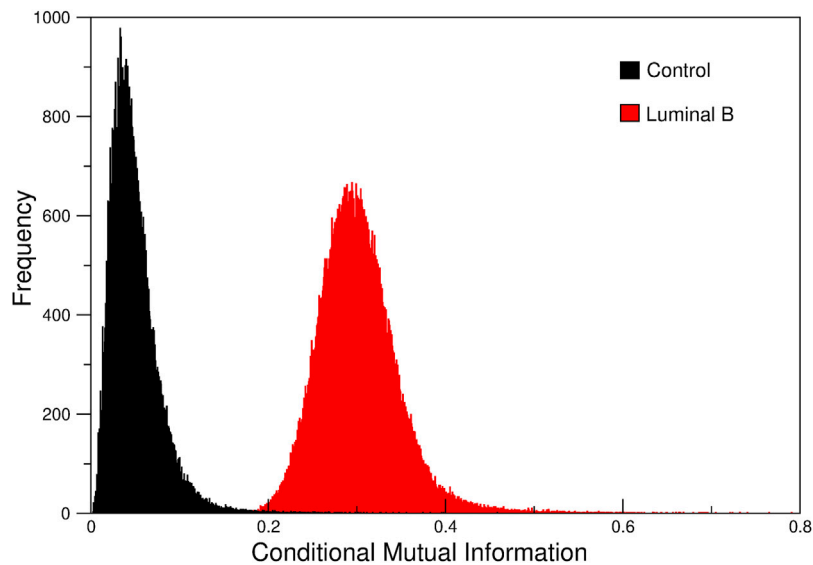
FIGURE 3 | Conditional Mutual Information for normal adjacent tissue (black) and Luminal B tumors (red). We can notice that Luminal B tumors present distinctively higher values of CMI than Controls.

4.4 Extreme Regions Exhibit a High Connectivity Pattern in Luminal B Network

We decided to analyze the topology of the resulting networks, but we only kept the highest interactions for this purpose. We set 0.35 as a threshold for CMI in the Luminal B network. This threshold resulted in a network with 11,449 edges and 420 genes. For comparison, we conserved the same number of edges in the control network. Taking into account the highest CMI values in the Luminal B phenotype, a very intriguing phenomenon appears: the first (8p11.21–23) and last (8q24.3) codifying regions of the chromosome 8 exhibit strong connections. Additionally, this strength decrease towards the centromere. Conversely, in the control network, the strength of interactions does not depend on the location of genes. This remarkable difference between the breast cancer and the non-tumor adjacent phenotype can be appreciated in **Figure 6**

Regarding the Luminal B chromosome 8 network, the genes of the extreme regions also constitute the most connected nodes of the network. The case of cytoband q24.3 is the most emblematic one. Cytoband q24.3 is the one with more intra-cytoband edges (1,640 out of 11,449). It is also the cytoband with more genes in the network (79 out of 420). Regarding the inter-cytoband edges, the large majority of interactions from any cytoband also correspond to those from q24.3. That is the reason for which the average degree of region q24.3 is the highest of the network (52). The non-extreme region p11.23 is an exception that has strong interactions.

It is important to mention that in region p11.23, there are some crucial genes in terms of cell maintenance, as well as for cancer development. For instance, four genes, namely BAG4, LSM1, ASH2L and BRF2 serve as housekeeping genes. On the other hand, FGFR, a keystone gene in cancer development, has been observed to be amplified in luminal B breast tumors (Erber et al., 2020; Voutsadakis, 2020; Amina et al., 2021).
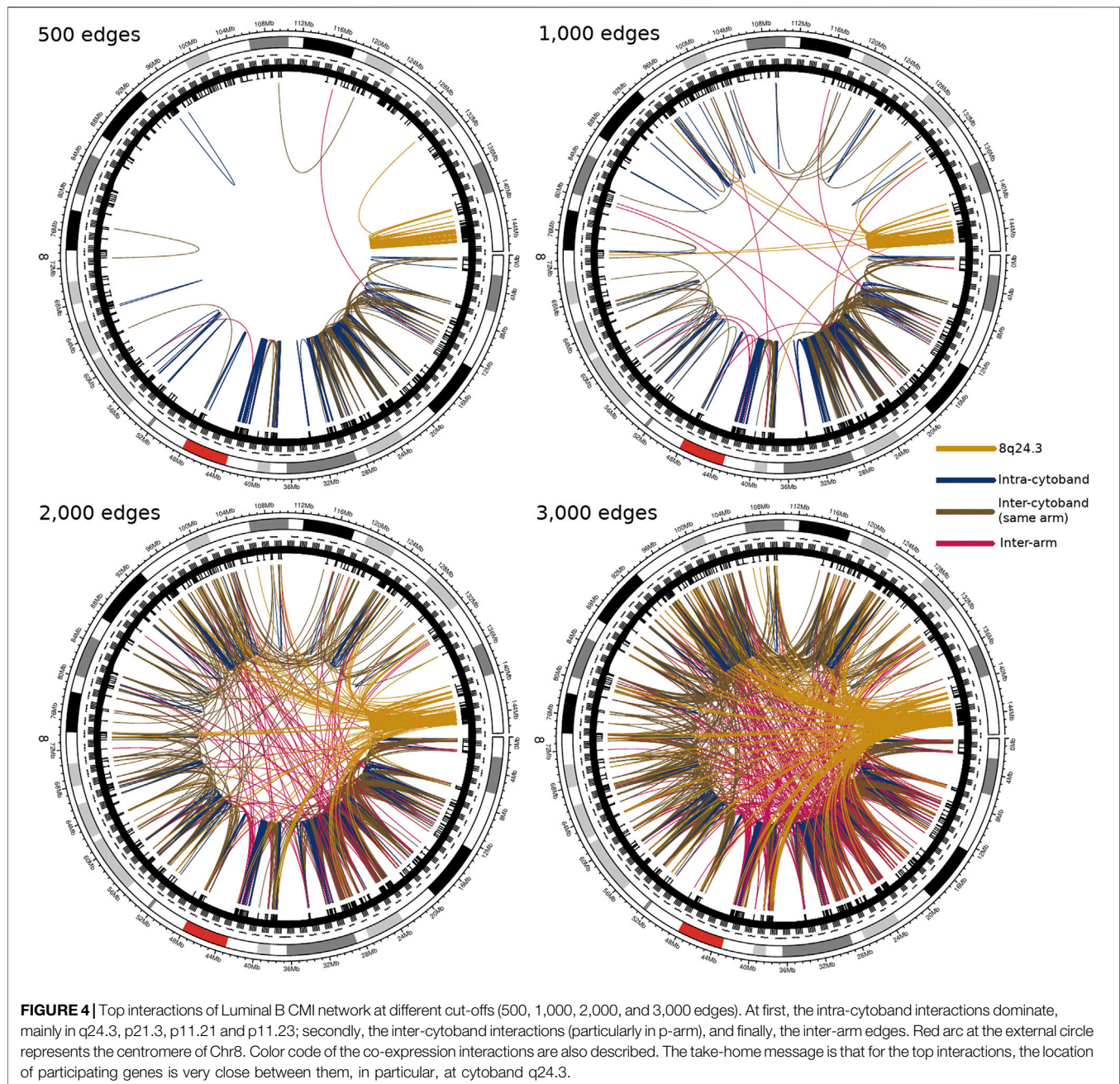
4.5 Highly Connected Genes and Their Possible Relationship With Cancer

In **Table 1** we show the most connected genes in the Luminal B network. As it can be observed, most of the genes belong to region q24.3. However, three genes, KCTD9, POL3RD, and ATP6V1B2 belong to the region p21.3. The gene with the highest Betweenness Centrality (BC) is KCTD9. Below we will provide a brief summary of what it is known of those genes in Luminal B breast cancer or other carcinomas.

ZNF7 (Zinc Finger Protein 7) has been indicated as a biomarker of survival in glioblastoma Esteve-Codina et al. (2021) and Burkitt's lymphoma (Gallego and Lazo, 1994). ZNF7 is the most connected gene in the Luminal B breast cancer co-expression network, and it is slighlty overexpressed (**Table 1**).

Many genes on chromosome 8 have been associated with mental illnesses, mainly schizophrenia, and their mutations are presumed to be involved in the development of our mental abilities. This is the case of the coding genes for the KCTD (Potassium Channel Tetramerization Domain) proteins. However, it has recently been indicated (Angrisani et al., 2021) that the 25 members of this family are potentially involved in a second fundamental activity: 13 of them have an exclusive pro-tumor function, 5 an exclusive anti-tumor function, 5 a pro/anti-tumor role, and 2 with a function not yet determined. KCTD9 has a pro-tumor function not yet reported in the literature, but inferred through databases. However, in this case, KCTD9 is underexpressed (−1.16) which may implicate a dual role in the phenotype.

MAF1 gene is known for its regulatory effect on the polymerase III, although it has also been associated with cancer, given that it activates the expression of the PTEN protein, which is an important tumor suppressor (Zhang et al.,

FIGURE 4 | Top interactions of Luminal B CMI network at different cut-offs (500, 1,000, 2,000, and 3,000 edges). At first, the intra-cytoband interactions dominate, mainly in q24.3, p21.3, p11.21 and p11.23; secondly, the inter-cytoband interactions (particularly in p-arm), and finally, the inter-arm edges. Red arc at the external circle represents the centromere of Chr8. Color code of the co-expression interactions are also described. The take-home message is that for the top interactions, the location of participating genes is very close between them, in particular, at cytoband q24.3.

2018). Interestingly POLR3D is also one of the most connected genes in the Luminal B breast cancer network. POLR3D gene encodes an RNA polymerase 3 subunit D, and synthesizes small RNAs, such as 5S rRNA and tRNAs, which, when inhibited by TRIPLIDE (TPL) influence the control of colorectal cancer (Liang et al., 2019). POLR3D is also inhibited by miR-320 (Ramassone et al., 2018).
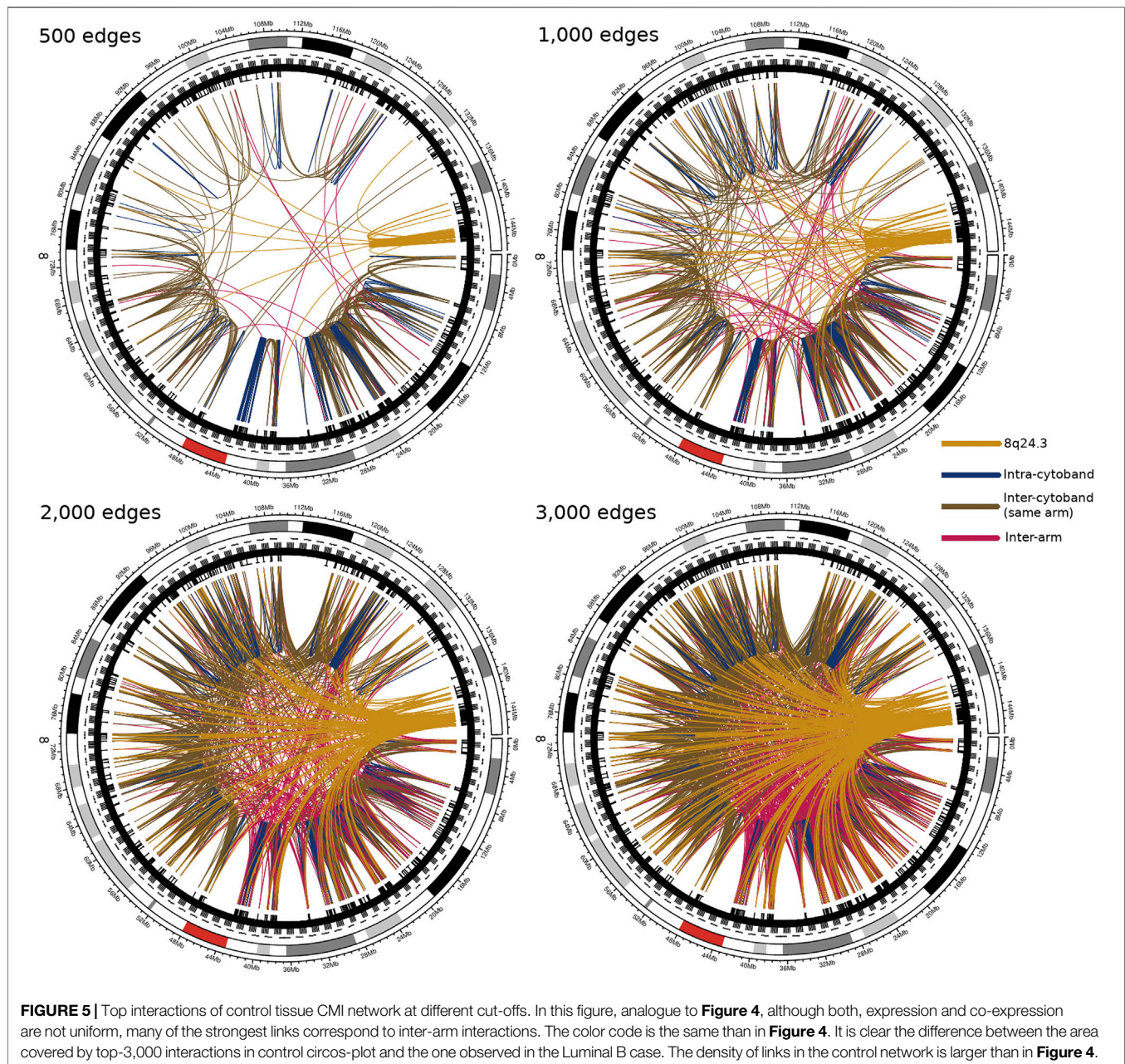
The fact that MAF1 and POLR3D genes were highly connected could be an indicator of alterations in the transcriptional regulation program. Since MAF1 regulates polymerase III, and POLR3D encodes an RNA polymerase III subunit, the emergence of an axis of transcriptional regulation of

non-coding RNA species results appealing. The latter may imply that in the Luminal B subtype, transcriptional regulation mediated by non-coding RNAs could affect the gene regulatory program. The latter coincides with the fact that POLR3D abnormal activity is characteristic of cancer cells (White, 2004).

ADCK5 has been indicated as an intermediary in the growth and metastasis of lung cancer. This gene promotes invasion and migration of lung cancer cells through the ADCK5-SOX9-PTTG1 pathway Qiu et al. (2020).

Regarding PTK2, this gene in association with KCNMA1 gene has been reported as a tumor suppressor in gastric cancer Ma

FIGURE 5 | Top interactions of control tissue CMI network at different cut-offs. In this figure, analogue to **Figure 4**, although both, expression and co-expression are not uniform, many of the strongest links correspond to inter-arm interactions. The color code is the same than in **Figure 4**. It is clear the difference between the area covered by top-3,000 interactions in control circos-plot and the one observed in the Luminal B case. The density of links in the control network is larger than in **Figure 4**.

et al. (2017). Additionally, the SMARCE1 gene regulates metastasis in breast cancer through its interaction with HIF1A and PTK2 Sethuraman et al. (2016). It has also been associated with hepatocarcinoma (Okamoto et al., 2003). Additionally, PTK2/FAK is considered a driver of radio-resistance in HPV-negative head and neck cancer (Skinner et al., 2016).
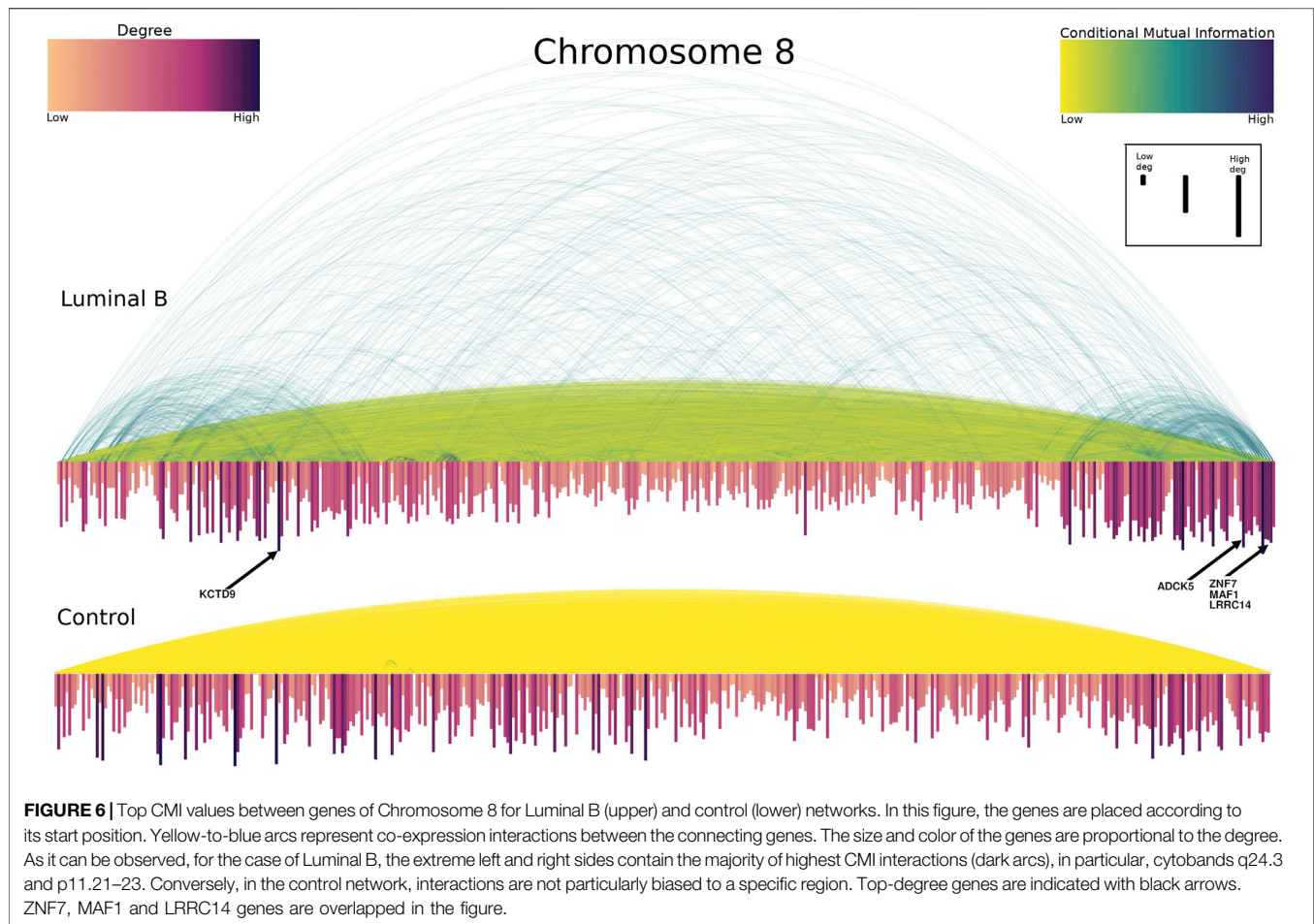
Recently, the FAM83H gene (also located on chromosome 8, specifically in the 8q24.3 region) has been related to Zinc Finger Proteins (ZNFs) genes, specifically ZNF16. Gallbladder cancer is highly associated with the expression of these two genes Ahn et al. (2020).

ATP6V1B2 gene has been identified as a possible biomarker (from controlled-to-aggressive growth with invasion of muscle

tissue) for bladder cancer. Specifically, it is underexpressed in the early stages of the disease and over-expressed in the advanced stages Fang et al. (2013). Additionally, it has been associated with follicular lymphoma, activating autophagic flux and mTOR pathway (Wang et al., 2019).

ZNF707 has recently been Kim et al. (2020) identified as highly overexpressed in the Japanese population. High expression has also been shown in kidney and luminal B cancer patients Machnik et al. (2019) which suggests that ZNF707 could be involved in the development of cancer in general, regardless of the tissue.

In brief, we can establish that the most connected genes in the chromosome 8 network for Luminal B breast cancer are

FIGURE 6 | Top CMI values between genes of Chromosome 8 for Luminal B (upper) and control (lower) networks. In this figure, the genes are placed according to its start position. Yellow-to-blue arcs represent co-expression interactions between the connecting genes. The size and color of the genes are proportional to the degree. As it can be observed, for the case of Luminal B, the extreme left and right sides contain the majority of highest CMI interactions (dark arcs), in particular, cytobands q24.3 and p11.21–23. Conversely, in the control network, interactions are not particularly biased to a specific region. Top-degree genes are indicated with black arrows. ZNF7, MAF1 and LRRC14 genes are overlapped in the figure.

TABLE 1 | The most connected genes belong to q24.3 and p21.2.

Gene	Band	Degree	BC	Log$_2$FC
ZNF7	q24.3	109	0.00544	0.47
KCTD9	p21.2	106	0.00656	−1.16
MAF1	q24.3	105	0.00562	0.07
LRRC14	q24.3	102	0.00512	0.34
ADCK5	q24.3	101	0.00523	0.93
PTK2	q24.3	99	0.00516	−0.03
POLR3D	p21.3	97	0.00508	−0.36
ZNF16	q24.3	97	0.00518	0.38
ATP6V1B2	p21.3	96	0.00579	−0.15
ZNF707	q24.3	96	0.00410	0.99

significantly related to the oncogenic phenotype. Particularly interesting is the case of MAF1, which in turn regulates POL3RD expression, suggesting an axis of non-coding RNA regulation.

As it can be also observed in **Table 1**, the differential expression in the large majority of hub genes is not significant. However, the co-expression patterns of these genes are importantly different than in the control case. We have observed previously, in clear cell renal carcinoma progression,

that the gene expression patterns do not change significantly over progression stages, however, the co-expression networks are clearly different between those stages (Zamora-Fuentes et al., 2020). Therefore, slightly different gene expression patterns may implicate a dramatical alteration in the co-expression landscape. The differential expression of all Chr8 genes is provided in **Supplementary Material S6**.

5 CONCLUSION

The understanding of the intricate relationship between copy number alterations, which can be seen as a mesoscopic dimension, with the regulation of the gene co-expression program, which can be understood as a microscopic phenomenon, is a highly promissory pathway for intense research in the near future. In this work, we have shown, for the case of Chromosome 8 in Luminal B breast cancer, that the Copy Number Alteration scenario does not influence, in a relevant manner, the Conditional Mutual Information program. That is independent on which region is analyzed. As a summary of findings, we can establish the following:

- CNVs do not influence Conditional Mutual Information (as observed in **Figure 2**).
- Co-expression program in the chromosome 8 for Luminal B breast cancer shows localized hotspots regions in certain cytobands.
- The large majority of Chromosome 8 gene co-expression interactions shows low CMI values, meanwhile the extreme parts of the chromosome show higher values.
- Cytoband q24.3 has the highest values of MI, is the most dense in terms of interactions, and its genes have the highest degree.
- In the control phenotype there is a homogeneous CMI distribution regarding the location of genes in the top interactions, contrary to the case of Luminal B network.
- Taking into account a *growing* of the networks from highest to lowest CMI values, in the case of Luminal B network, the top CMI values appear between intra-cytoband genes; after that, between inter-cytoband and same-arm genes; finally between inter-arm genes. In the case of the control network, there is no clear localization pattern.
- Genes such as ZNF7, KCTD9, MAF1, or POLR3D have the highest degree centralities. Those genes have been reported to have influence in cancer.
- MAF1 and POLR3D could form an axis of non-coding RNA regulation, which can be a possible complex for future research.

Further steps towards a whole understanding of how copy number alterations may affect the co-expression program in breast cancer must include the analysis of the conditional mutual information of all chromosomes in Luminal B breast cancer. A similar study in the other molecular subtypes is also needed. The influence of the progression stage must be also taken into account. Finally, this analysis over other cancer tissues will provide a solid and robust landscape of the role of copy number alterations in the rise and development of cancer.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ **Supplementary Material**.

AUTHOR CONTRIBUTIONS

CH-G organized data, developed code, performed calculations, analyzed data, drafted the manuscript; EH-L co-devised the project, designed the methodological approach, analyzed data, discussed results, co-supervised the project. JE-E co-devised and coordinated the project, contributed to the methodological strategy, analyzed data and integrated the biological results, discussed results, co-supervised the project; All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.806607/full#supplementary-material

Supplementary Material S1 | Gene expression data. This table includes the normalized RNA-seq expression of all Luminal B samples.

Supplementary Material S2 | Gene information. Name, ENSEMBL-id, GC content, biotype, gene description, cytoband location, base pair and base end of each gene of chromosome 8 is provided.

Supplementary Material S3 | CNV data. This table contains the copy number alteration information for all Luminal B breast cancer samples.

Supplementary Material S4 | Gene information data from copy number regions.

Supplementary Material S5–S8 | Chromosome 8 copy number alteration map of 140 Luminal B samples. For these figures we used the IGV platform (https://software.broadinstitute.org/software/igv/).

Supplementary Material S9 | D-statistic heatmap and p-values. In the first figure, the layers with the largest discrepancies, all less than 0.05, appear in red. The corresponding p-values appear in the second heatmap; these should be read in conjunction with the D-values. Most of these numbers are low and appear in shades of white, the blue dots correspond to *p*-values equal to one, indicating that the data are the same in both distributions. Like adjacency matrices, these matrices are symmetric.

Supplementary Material S10 | Differential gene expression of Chr8. Log2 fold change and *p*-values are provided.

REFERENCES

Ades, F., Zardavas, D., Bozovic-Spasojevic, I., Pugliano, L., Fumagalli, D., De Azambuja, E., et al. (2014). Luminal B Breast Cancer: Molecular Characterization, Clinical Management, and Future Perspectives. *J. Clin. Oncol.* 32, 2794–2803. doi:10.1200/jco.2013.54.1870

Ahn, S. W., Ahn, A.-R., Ha, S. H., Hussein, U. K, Do Yang, J., Kim, K. M., et al. (2020). Expression of Fam83h and Znf16 Are Associated with Shorter Survival of Patients with Gallbladder Carcinoma. *Diagn. Pathol.* 15, 1–14. doi:10.1186/s13000-020-00985-1

Alcalá-Corona, S. A., de Anda-Jáuregui, G., Espinal-Enriquez, J., Tovar, H., and Hernández-Lemus, E. (2018). "Network Modularity and Hierarchical Structure in Breast Cancer Molecular Subtypes," in International Conference on Complex Systems, Cambridge, MA (Springer), 352–358. doi:10.1007/978-3-319-96661-8_36

Amina, B., Lynda, A. K., Sonia, S., Adel, B., Jelloul, B. H., Miloud, M., et al. (2021). Fibroblast Growth Factor Receptor 1 Protein (Fgfr1) as Potential Prognostic and Predictive Marker in Patients with Luminal B Breast Cancers Overexpressing Human Epidermal Receptor 2 Protein (Her2). *Indian J. Pathol. Microbiol.* 64, 254. doi:10.4103/IJPM.IJPM_87_20

Andonegui-Elguera, S. D., Zamora-Fuentes, J. M., Espinal-Enríquez, J., and Hernández-Lemus, E. (2021). Loss of Long Distance Co-expression in Lung Cancer. *Front. Genet.* 12, 625741. doi:10.3389/fgene.2021.625741

Angrisani, A., Di Fiore, A., De Smaele, E., and Moretti, M. (2021). The Emerging Role of the Kctd Proteins in Cancer. *Cell Commun. Signal.* 19, 1–17. doi:10.1186/s12964-021-00737-8

Cheang, M. C., Chia, S. K., Voduc, D., Gao, D., Leung, S., Snider, J., et al. (2009). Ki67 index, Her2 Status, and Prognosis of Patients with Luminal B Breast Cancer. *J. Natl. Cancer Inst.* 101, 736–750. doi:10.1093/jnci/djp082

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., et al. (2016). Tcgabiolinks: an R/bioconductor Package for Integrative Analysis of Tcga Data. *Nucleic Acids Res.* 44, e71. doi:10.1093/nar/gkv1507

Creighton, C. J. (2012). The Molecular Profile of Luminal B Breast Cancer. *Biol. Targets Ther.* 6, 289. doi:10.2147/btt.s29923

Cui, Y., Chen, X., Luo, H., Fan, Z., Luo, J., He, S., et al. (2016). Biocircos. Js: an Interactive Circos Javascript Library for Biological Data Visualization on Web Applications. *Bioinformatics* 32, 1740–1742. doi:10.1093/bioinformatics/btw041

de Anda-Jáuregui, G., Velázquez-Caldelas, T. E., Espinal-Enríquez, J., and Hernández-Lemus, E. (2016). Transcriptional Network Architecture of Breast Cancer Molecular Subtypes. *Front. Physiol.* 7, 568. doi:10.3389/fphys.2016.00568

de Anda-Jáuregui, G., Espinal-Enriquez, J., and Hernández-Lemus, E. (2019). Spatial Organization of the Gene Regulatory Program: an Information Theoretical Approach to Breast Cancer Transcriptomics. *Entropy* 21, 195. doi:10.3390/e21020195

Dorantes-Gilardi, R., García-Cortés, D., Hernández-Lemus, E., and Espinal-Enríquez, J. (2020). Multilayer Approach Reveals Organizational Principles Disrupted in Breast Cancer Co-expression Networks. *Appl. Netw. Sci.* 5, 1–23. doi:10.1007/s41109-020-00291-1

Dorantes-Gilardi, R., García-Cortés, D., Hernández-Lemus, E., and Espinal-Enríquez, J. (2021). K-Core Genes Underpin Structural Features of Breast Cancer. *Sci. Rep.* 11, 1–17. doi:10.1038/s41598-021-95313-y

Erber, R., Rübner, M., Davenport, S., Hauke, S., Beckmann, M. W., Hartmann, A., et al. (2020). Impact of Fibroblast Growth Factor Receptor 1 (Fgfr1) Amplification on the Prognosis of Breast Cancer Patients. *Breast Cancer Res. Treat.* 184, 311–324. doi:10.1007/s10549-020-05865-2

Espinal-Enríquez, J., Fresno, C., Anda-Jáuregui, G., and Hernández-Lemus, E. (2017). Rna-seq Based Genome-wide Analysis Reveals Loss of Inter-chromosomal Regulation in Breast Cancer. *Sci. Rep.* 7, 1–19. doi:10.1038/s41598-017-01314-1

Esteve-Codina, A., Alameda, F., Carrato, C., Pineda, E., Arpí, O., Martinez-García, M., et al. (2021). Rna Sequencing and Immunohistochemistry Reveal Zfn7 as a Stronger Marker of Survival Than Molecular Subtypes in G-Cimp–Negative Glioblastoma. *Clin. Cancer Res.* 27, 645–655. doi:10.1158/1078-0432.ccr-20-2141

Fang, Z., Zang, W., Chen, R., Ye, B., Wang, X., Yi, S., et al. (2013). Gene Expression Profile and Enrichment Pathways in Different Stages of Bladder Cancer. *Genet. Mol. Res.* 12, 1479–1489. doi:10.4238/2013.may.6.1

Gallego, M. I., and Lazo, P. A. (1994). The Human Zinc-finger Protein-7 Gene Is Located 90 Kb 3 Prime of Myc and Is Not Expressed in Burkitt Lymphoma Cell Lines. *Int. J. Cancer* 58, 855–859. doi:10.1002/ijc.2910580618

García-Cortés, D., de Anda-Jáuregui, G., Fresno, C., Hernández-Lemus, E., and Espinal-Enríquez, J. (2020). Gene Co-expression Is Distance-dependent in Breast Cancer. *Front. Oncol.* 10, 1232. doi:10.3389/fonc.2020.01232

García-Cortés, D., Hernández-Lemus, E., and Espinal-Enríquez, J. (2021). Luminal a Breast Cancer Co-expression Network: Structural and Functional Alterations. *Front. Genet.* 12, 514. doi:10.3389/fgene.2021.629475

Harbeck, N., Thomssen, C., and Gnant, M. (2013). St. Gallen 2013: Brief Preliminary Summary of the Consensus Discussion. *Breast care* 8, 102–109. doi:10.1159/000351193

Henrichsen, C. N., Chaignat, E., and Reymond, A. (2009). Copy Number Variants, Diseases and Gene Expression. *Hum. Mol. Genet.* 18, R1–R8. doi:10.1093/hmg/ddp011

Hernández-Lemus, E., and Rangel-Escareño, C. (2011). The Role of Information Theory in Gene Regulatory Network Inference. *Inf. Theor. New Res.*, 109–144.

Hernández-Lemus, E., and Siqueiros-García, J. M. (2013). Information Theoretical Methods for Complex Network Structure Reconstruction. *Complex Adaptive Syst. Model.* 1, 1–22. doi:10.1186/2194-3206-1-8

Hernández-Lemus, E., Reyes-Gopar, H., Espinal-Enríquez, J., and Ochoa, S. (2019). The many Faces of Gene Regulation in Cancer: A Computational Oncogenomics Outlook. *Genes* 10, 865. doi:10.3390/genes10110865

Ibragimova, M., Tsyganov, M., and Litviakov, N. (2020). Transcriptome of Breast Tumor with Different Amplification Statuses of Long Arm of Chromosome 8. *Med. Genet.* 19, 31–32. doi:10.20538/1682-0363-2020-3-22-28

Jia, H., Li, S., Han, Y., Liu, F., Li, W., and Fu, L. (2019). Research Advances of Chromosomal 8q24 Aberrance in Breast Cancer. *Chin. J. Clin. Oncol.* 46, 150–153. doi:10.3969/j.issn.1000-8179.2019.03.157

Kim, J., Chung, J.-Y., Hwang, J. R., Lee, Y.-Y., Kim, T.-J., Lee, J.-W., et al. (2020). Identification of Candidate Genes Associated with Susceptibility to Ovarian clear Cell Adenocarcinoma Using Cis-Eqtl Analysis. *J. Clin. Med.* 9, 1137. doi:10.3390/jcm9041137

Kittaneh, M., Montero, A. J., and Glück, S. (2013). Molecular Profiling for Breast Cancer: a Comprehensive Review. *Biomarkers Cancer* 5, BIC.S9455. doi:10.4137/bic.s9455

Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating Mutual Information. *Phys. Rev. E* 69, 066138. doi:10.1103/PhysRevE.69.066138

Kumaran, M., Cass, C. E., Graham, K., Mackey, J. R., Hubaux, R., Lam, W., et al. (2017). Germline Copy Number Variations Are Associated with Breast Cancer Risk and Prognosis. *Sci. Rep.* 7, 1–15. doi:10.1038/s41598-017-14799-7

Lachmann, A., Giorgi, F. M., Lopez, G., and Califano, A. (2016). Aracne-ap: Gene Network Reverse Engineering through Adaptive Partitioning Inference of Mutual Information. *Bioinformatics* 32, 2233–2235. doi:10.1093/bioinformatics/btw216

Li, Z.-h., Hu, P.-h., Tu, J.-h., and Yu, N.-s. (2016). Luminal B Breast Cancer: Patterns of Recurrence and Clinical Outcome. *Oncotarget* 7, 65024. doi:10.18632/oncotarget.11344

Liang, K.-C., and Wang, X. (2008). Gene Regulatory Network Reconstruction Using Conditional Mutual Information. *EURASIP J. Bioinformatics Syst. Biol.* 2008, 1–14. doi:10.1155/2008/253894

Liang, X., Xie, R., Su, J., Ye, B., Wei, S., Liang, Z., et al. (2019). Inhibition of Rna Polymerase Iii Transcription by Triptolide Attenuates Colorectal Tumorigenesis. *J. Exp. Clin. Cancer Res.* 38, 1–13. doi:10.1186/s13046-019-1232-x

Liu, R., Wang, X., Chen, G. Y., Dalerba, P., Gurney, A., Hoey, T., et al. (2007). The Prognostic Role of a Gene Signature from Tumorigenic Breast-Cancer Cells. *New Engl. J. Med.* 356, 217–226. doi:10.1056/nejmoa063994

Liu, X., Ma, D., Jiang, Y., Liu, Y., Yu, K., and Shao, Z. (2018). Abstract P2-06-03: Clinicopathologic Characteristics and Genomic Essence Analyses Revealed Estrogen Receptor Positive, Progesterone Receptor Negative and Human Epidermal Growth Factor Receptor 2 Negative Breast Cancer to Be More Basal-like and Endocrine Resistant. *Cancer Res.* 78, 6711–6910. doi:10.1158/1538-7445.sabcs17-p2-06-03

Ma, G., Liu, H., Hua, Q., Wang, M., Du, M., Lin, Y., et al. (2017). Kcnma1 Cooperating with Ptk2 Is a Novel Tumor Suppressor in Gastric Cancer and Is Associated with Disease Outcome. *Mol. Cancer* 16, 1–10. doi:10.1186/s12943-017-0613-z

Machnik, M., Cylwa, R., Kiełczewski, K., Biecek, P., Liloglou, T., Mackiewicz, A., et al. (2019). The Expression Signature of Cancer-Associated Krab-Znf Factors Identified in Tcga Pan-Cancer Transcriptomic Data. *Mol. Oncol.* 13, 701–724. doi:10.1002/1878-0261.12407

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., et al. (2006a). Aracne: an Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* 7, 1–15. Springer. doi:10.1186/1471-2105-7-S1-S7

Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., and Califano, A. (2006b). Reverse Engineering Cellular Networks. *Nat. Protoc.* 1, 662–671. doi:10.1038/nprot.2006.106

Metzger-Filho, O., Sun, Z., Viale, G., Price, K. N., Crivellari, D., Snyder, R. D., et al. (2013). Patterns of Recurrence and Outcome According to Breast Cancer Subtypes in Lymph Node–Negative Disease: Results from International Breast Cancer Study Group Trials Viii and Ix. *J. Clin. Oncol.* 31, 3083. doi:10.1200/jco.2012.46.1574

Meyer, P. E., and Meyer, M. P. E. (2009). *Package 'infotheo'*. Princeton, NJ, USA: Citeseer. R Package Version.

Nueda, M. J., Ferrer, A., and Conesa, A. (2012). Arsyn: a Method for the Identification and Removal of Systematic Noise in Multifactorial Time Course Microarray Experiments. *Biostatistics* 13, 553–566. doi:10.1093/biostatistics/kxr042

Ohshima, K., Hatakeyama, K., Nagashima, T., Watanabe, Y., Kanto, K., Doi, Y., et al. (2017). Integrated Analysis of Gene Expression and Copy Number Identified Potential Cancer Driver Genes with Amplification-dependent Overexpression in 1,454 Solid Tumors. *Sci. Rep.* 7, 1–13. doi:10.1038/s41598-017-00219-3

Okamoto, H., Yasui, K., Zhao, C., Arii, S., and Inazawa, J. (2003). Ptk2 and Eif3s3 Genes May Be Amplification Targets at 8q23–Q24 and Are Associated with Large Hepatocellular Carcinomas. *Hepatology* 38, 1242–1249. doi:10.1053/jhep.2003.50457

Qiu, M., Li, G., Wang, P., Li, X., Lai, F., Luo, R., et al. (2020). Aarf Domain Containing Kinase 5 Gene Promotes Invasion and Migration of Lung Cancer Cells through Adck5-Sox9-Pttg1 Pathway. *Exp. Cel. Res.* 392, 112002. doi:10.1016/j.yexcr.2020.112002

Ramassone, A., Pagotto, S., Veronese, A., and Visone, R. (2018). Epigenetics and Micrornas in Cancer. *Int. J. Mol. Sci.* 19, 459. doi:10.3390/ijms19020459

Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). Gc-content Normalization for Rna-Seq Data. *BMC Bioinformatics* 12, 1–17. doi:10.1186/1471-2105-12-480

Safonov, A., Jiang, T., Bianchini, G., Győrffy, B., Karn, T., Hatzis, C., et al. (2017). Immune Gene Expression Is Associated with Genomic Aberrations in Breast Cancer. *Cancer Res.* 77, 3317–3324. doi:10.1158/0008-5472.can-16-3478

Serrano-Carbajal, E. A., Espinal-Enríquez, J., and Hernández-Lemus, E. (2020). Targeting Metabolic Deregulation Landscapes in Breast Cancer Subtypes. *Front. Oncol.* 10, 97. doi:10.3389/fonc.2020.00097

Sethuraman, A., Brown, M., Seagroves, T. N., Wu, Z.-H., Pfeffer, L. M., and Fan, M. (2016). Smarce1 Regulates Metastatic Potential of Breast Cancer Cells through the Hif1a/ptk2 Pathway. *Breast Cancer Res.* 18, 1–15. doi:10.1186/s13058-016-0738-9

Shao, X., Lv, N., Liao, J., Long, J., Xue, R., Ai, N., et al. (2019). Copy Number Variation Is Highly Correlated with Differential Gene Expression: a Pan-Cancer Study. *BMC Med. Genet.* 20, 1–14. doi:10.1186/s12881-019-0909-5

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA: A Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Skinner, H. D., Giri, U., Yang, L., Woo, S. H., Story, M. D., Pickering, C. R., et al. (2016). Proteomic Profiling Identifies Ptk2/fak as a Driver of Radioresistance in Hpv-Negative Head and Neck Cancer. *Clin. Cancer Res.* 22, 4643–4650. doi:10.1158/1078-0432.ccr-15-2785

Stein, K. D., Syrjala, K. L., and Andrykowski, M. A. (2008). Physical and Psychological Long-Term and Late Effects of Cancer. *Cancer* 112, 2577–2592. doi:10.1002/cncr.23448

Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., et al. (2007). Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. *Science* 315, 848–853. doi:10.1126/science.1136678

Sun, W., Bunn, P., Jin, C., Little, P., Zhabotynsky, V., Perou, C. M., et al. (2018). The Association between Copy Number Aberration, Dna Methylation and Gene Expression in Tumor Samples. *Nucleic Acids Res.* 46, 3009–3018. doi:10.1093/nar/gky131

Tarazona, S., García, F., Ferrer, A., Dopazo, J., and Conesa, A. (2011). Noiseq: a Rna-Seq Differential Expression Method Robust for Sequencing Depth Biases. *Embnet. J.* 17, 18–19. doi:10.14806/ej.17.b.265

Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A. D., Nueda, M. J., Ferrer, A., et al. (2015). Data Quality Aware Analysis of Differential Expression in Rna-Seq with Noiseq R/bioc Package. *Nucleic Acids Res.* 43, e140. doi:10.1093/nar/gkv711

Terrell, G. R., and Scott, D. W. (1992). Variable Kernel Density Estimation. *Ann. Stat.* 20, 1236–1265. doi:10.1214/aos/1176348768

Tong, Y., Tang, Y., Li, S., Zhao, F., Ying, J., Qu, Y., et al. (2020). Cumulative Evidence of Relationships between Multiple Variants in 8q24 Region and Cancer Incidence. *Medicine* 99, e20716. doi:10.1097/MD.0000000000020716

Tran, B., and Bedard, P. L. (2011). Luminal-b Breast Cancer and Novel Therapeutic Targets. *Breast Cancer Res.* 13, 1–10. doi:10.1186/bcr2904

Voutsadakis, I. A. (2020). 8p11. 23 Amplification in Breast Cancer: Molecular Characteristics, Prognosis and Targeted Therapy. *J. Clin. Med.* 9, 3079. doi:10.3390/jcm9103079

Wang, F., Gatica, D., Ying, Z. X., Peterson, L. F., Kim, P., Bernard, D., et al. (2019). Follicular Lymphoma–Associated Mutations in Vacuolar Atpase Atp6v1b2 Activate Autophagic Flux and Mtor. *J. Clin. Invest.* 129, 1626–1640. doi:10.1172/jci98288

Wang, X., He, X., Guo, H., and Tong, Y. (2020). Variants in the 8q24 Region Associated with Risk of Breast Cancer: Systematic Research Synopsis and Meta-Analysis. *Medicine* 99, e19217. doi:10.1097/MD.0000000000019217

White, R. J. (2004). Rna Polymerase Iii Transcription and Cancer. *oncogene* 23, 3208–3216. doi:10.1038/sj.onc.1207547

Wokolorczyk, D., Gliniewicz, B., Sikorski, A., Zlowocka, E., Masojc, B., Dkebniak, T., et al. (2008). A Range of Cancers Is Associated with the Rs6983267 Marker on Chromosome 8. *Cancer Res.* 68, 9982–9986. doi:10.1158/0008-5472.CAN-08-1838

Zamora-Fuentes, J. M., Hernández-Lemus, E., and Espinal-Enríquez, J. (2020). Gene Expression and Co-expression Networks Are Strongly Altered through Stages in clear Cell Renal Carcinoma. *Front. Genet.* 11, 578679. doi:10.3389/fgene.2020.578679

Zhang, X., Zhao, X.-M., He, K., Lu, L., Cao, Y., Liu, J., et al. (2012). Inferring Gene Regulatory Networks from Gene Expression Data by Path Consistency Algorithm Based on Conditional Mutual Information. *Bioinformatics* 28, 98–104. doi:10.1093/bioinformatics/btr626

Zhang, S., Li, X., Wang, H.-Y., and Zheng, X. S. (2018). Beyond Regulation of Pol Iii: Role of Maf1 in Growth, Metabolism, Aging and Cancer. *Biochim. Biophys. Acta (BBA)-Gene Regul. Mech.* 1861, 338–343. doi:10.1016/j.bbagrm.2018.01.019