Loss of trans regulation in breast cancer molecular subtypes

Diana García-Cortés[1], Guillermo de Anda-Jáuregui[1], Cristobal Fresno[1], Enrique Hernández-Lemus[1,2], Jesús Espinal-Enriquez[1,2,*]

[1]National Insitute of Genomic Medicine, Mexico City, Mexico
[2] Centro de Ciencias de la Complejidad (C3). Universidad Nacional Autónoma de México, Ciudad de México, México

Abstract

Breast cancer is a complex heterogeneous disease. A clear example is given by the four molecular subtypes: Luminal A, Luminal B, HER2-Enriched and Basal-like. These subtypes give way to different therapeutic approaches to deal with different prognosis. Despite these differences, common hallmark features of cancer can be found, which its origin is traced back to the intricate relationships governing regulatory programs. In our recent work, by constructing RNA-Seq normal tissue and breast cancer gene regulatory networks, we have observed the phenomenon of loss of inter-chromosomal regulation. Our results showed that *cis*- regulation in breast cancer tissue occurs mostly between neighbour genes. On the contrast, in non-cancerous tissue, gene-gene regulation appears along the whole genome. Here, we extend the aforementioned approach, in order to observe into what extent the loss of *trans*- regulation occurs in the different intrinsic breast cancer subtypes. A collection of 780 RNA-Seq The Cancer Genome Atlas breast cancer samples were classified using PAM50 algorithm. Differential expression analysis was performed between each subtype and additional 101 normal tissue samples. Gene regulatory networks were inferred for each of the four subtypes and the normal tissue. Circos plots visualization was used to contrast the *cis/trans* regulation proportion. Finally, power-law regression analyses were fitted to explain the statistical relationship between genes and the distance between genes. Inter and intra-chromosome relationships occur approximately in the same proportion in the healthy network. Meanwhile, the four subtypes present a loss of *trans*- regulation. The decrease of *trans*- regulations exhibits different patterns among subtypes. Additionally, the strength of *cis*- regulatory interactions decays exponentially with the distance in the four subtypes. But, in the non-cancerous phenotype, distance does not influence the strength of the interactions. With this kind of approach, we have been able to integrate gene regulation and physical distance to elaborate a more comprehen-

sive landscape in cancer genomics. Here, we opened the possibility to analyse in a complementary fashion the regulatory program of molecular subtypes of breast cancer. This effort may be complemented with copy number alterations, micro-RNAs or Hi-C data with the aim of providing a multi-omics-based framework to elaborate more specific questions in the era of personalized medicine.

Introduction

Genomic instability is a highly relevant hallmark of cancer [1, 2], in particular for breast cancer. Along with this, molecular heterogeneity adds a more complicated landscape to understand the disease. Furthermore, epigenetic marks may influence either, directly or indirectly, the global gene expression patterns and utterly alter protein function triggering the oncogenic phenotype [3, 4].

Next Generation Sequencing has provided us an unprecedented tool to have a broader and deeper landscape of the molecular origins of cancer. However, at the same time, the astronomic amount of data requires the development of a theoretical framework to deal with the massive information and be able to understand, interpret and provide explanations of said results [5, 6]. In these terms, network theory has proven to be a robust tool to analyze and understand the context of global genetic regulation in cancer [7, 8].

In a previous work [9], by means of a gene regulatory network inferred from gene expression data, collected by The Cancer Genome Atlas (TCGA) consortium, we have demonstrated that inter-chromosome (*trans*-) gene regulatory interactions are less abundant and weaker in breast cancer tissue compared to the non-cancerous one.

We also observed important differences regarding the *cis*- (intra-chromosome) regulations: the most intense relationships in the cancerous phenotype occur between genes that are physically close in terms of their chromosome position. This is not observed in the healthy network tissue, i.e., correlation strength is not dependent on the chromosomal location or the chromosome where the genes belong to.

Based on this previous work, in which we provide a physical context of the genomic regulation in breast cancer, here we went one step further by classifying into the molecular subtypes said samples and observe the topological differences among groups. We have found that each subtype network presents a distinctive regulatory program.

Network size of the largest components (lc) are different. The lc for the healthy network is an order of magnitude bigger than any subtype lc. Additionally, there is a notable distinction between lc in tumour. Luminal B subtype lc's contains 2,387 genes, meanwhile the other three subtypes have less than 400

gene *lc*.

Regarding inter-chromosomal gene interactions, it is worth to mention that the most notorious loss of *trans*- regulation occurs in the Basal subtype, followed by HER2-Enriched, i.e., non-Luminal subtypes. Luminal A and B present a less severe loss of *trans*- regulation.

With respect to the intensity of intra-chromosomal (*cis*-) gene interactions, all breast cancer subtypes present a strong dependency of gene-pair physical proximity. The overall correlation strength behaviour follows a power-law decay dependent on gene-pair distance, although the parameters are different for each subtype. The power-law decay on network subtype curves shows higher mutual information values on the closest gene-pairs, when compared to the healthy network curve, suggesting that this phenomenon involves not only the loss of inter-chromosomal regulation but also gain of *short-range* correlation.

In the context of these results, we argue that patterns of genome instability distinguish each molecular subtype. The loss of long-range correlations may be related to a global loss of functional coupling and gain of short-range interactions could implicate local specialization.

To our knowledge, there is not a previous study that analyse statistical dependencies between genes and global gene expression in terms of physical proximity for breast cancer molecular subtypes. Studies like the one presented here may shed some light to the spatial rearrangements that occur in cancer, and distinguish (dys)regulations in molecular subtypes. This may have important implications for both, the basic understanding of cancer biology, and future developments of novel therapeutic interventions.

Results

Network structures differ between subtypes and compared to normal samples

Figure 1 shows the healthy network structure as well as the four network subtypes, using the 0.01% highest interactions between genes (11,675 regulatory interactions in all cases). In these visualizations, genes are coloured according to the chromosome in which they belong to. As it can be observed, in the healthy network, the largest component (*lc*) comprises almost all the genes, in contrary to any tumour subtype network, where the vast majority of genes in any component, belongs to a single chromosome.

The case of Luminal B subtype network is slightly different: even when the *lc* contains 2,387 genes (almost the half of the whole size), co-regulated gene-pair relationships occur mostly between intra-chromosome genes. Table 1 shows

the proportion of $\frac{trans-}{cis-}$ interactions in the *lc* for each network under study, as well as other structural features. *Cis-* interactions also prevail in the small components.

Table 1: **Topological features of the gene regulatory networks inferred**

			Breast cancer subtypes			
			Luminal			
Feature		**Healthy**	**B**	**A**	**HER2-Enriched**	**Basal-like**
Number of Genes		10,353	5,846	4,379	5,207	5,382
Chrom. Interactions	**Total interactions**	11,675	11,675	11,675	11,675	11,675
	Intra (*cis-*)	1,874	10,336	10,747	11,302	11,438
	Inter (*trans-*)	9,801	1,339	928	373	237
Largest component	**Number of genes**	7,384	2,387	337	311	130
	$trans-/cis-$	$\frac{6504}{9678}$	$\frac{280}{6870}$	$\frac{66}{1707}$	$\frac{33}{1190}$	$\frac{0}{450}$

Notice the parameters differences between the healthy network compared to all subtypes.

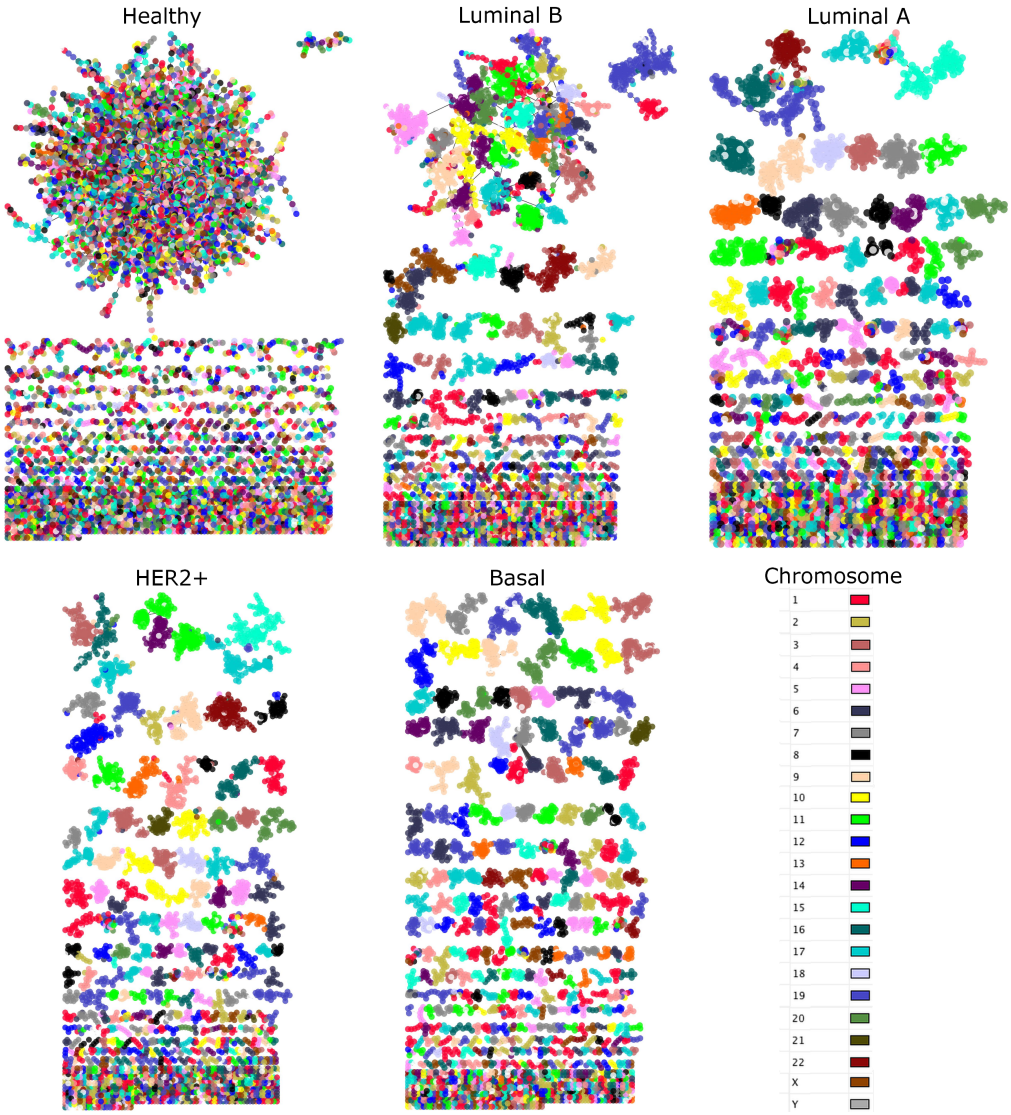

Figure 1: Networks per subtype. In this representation we observe the 0.01% strongest interactions per each network (healthy and the four subtypes). Nodes are colored according to the chromosome each gene belongs to. Notice the evident difference between the combination of colors in the largest component in the healthy network, meanwhile the breast cancer subtype networks have almost the same color in the largest component. An interesting case is luminal B subtype; despite it has a largest component with several genes from different chromosomes, genes are grouped into the component according to their corresponding chromosome.

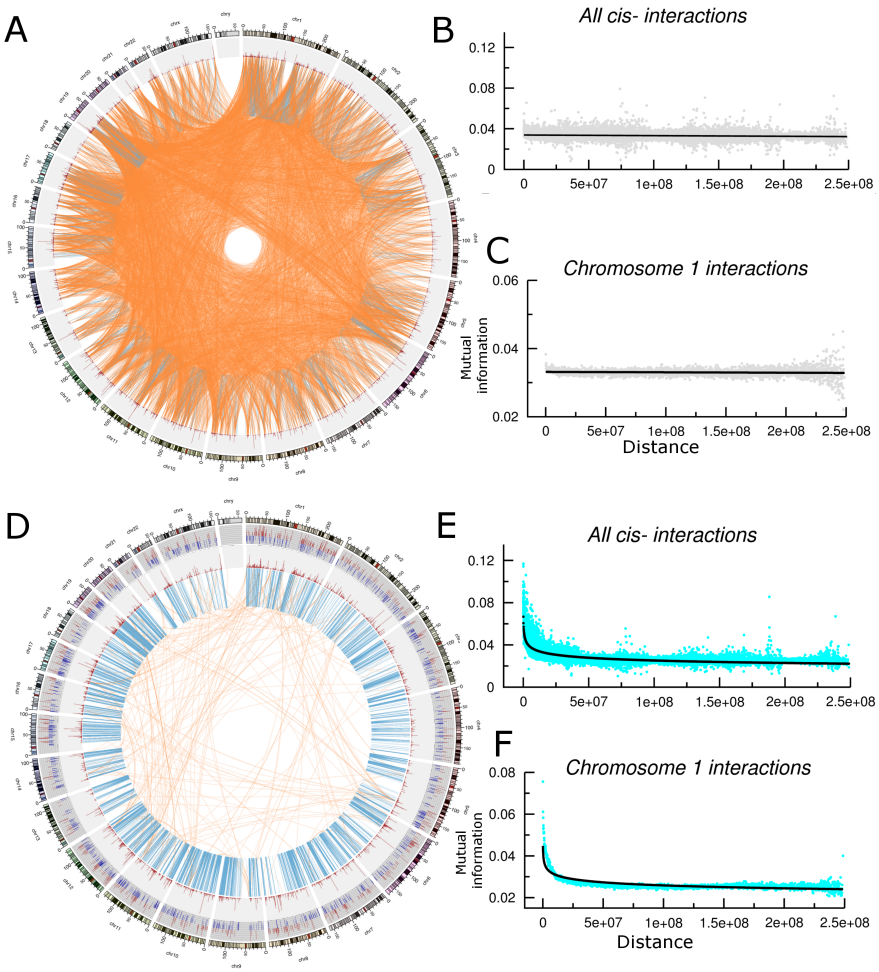

Figure 2: **Abundance and strength of gene regulation in cancer and healthy networks.** The circos plots show the dramatic difference in *cis-/trans-* regulatory patterns between healthy network (A), and basal subtype network (D). Both plots contain the 0.01% highest mutual information values for both networks. Orange lines join inter-chromosomal genes, meanwhile the blue lines takes account for intra-chromosomal couples. External circles indicate the chromosome band. The next circle (red histogram) indicates the number of neighbors that each gene has. In the lower circos plot for basal subtype, is also presented the differential expression value of the appearing genes in the network: blue/red outer histogram takes account for under/overexpressed genes. B) The plot shows the average mututal information value of the whole-genome *cis-* interactions, depending on the distance between couples of genes for the healhty network. The black line is the linear fitting of said distribution. The value of the slope is $6.17e^{-12}$. C) plots *cis-* interactions in chromosome 1 only. E) and F) are analogous to B) and C) plots, but the curve fitting correspond to an exponential distribution with an exponent of -0.14 for E) and -0.07 for F). With this comparison it is shown how the strength of correlation between intra-chromosomal genes decay exponentially with the distance.

Strongest interactions occur between physically close genes

Differences regarding chromosome composition of network components can be thus related to gene-pair physical proximity. Figure 2 presents a `circos plot` representation of the networks depicted in Figure 1. In this representation, blue inner links represent *cis*- interactions, meanwhile orange inner links account for *trans*- regulations. As it can be observed, there is a remarkable distinction in blue/orange proportion not only when we compare the healthy circos plots to the four subtypes, but also comparing cancer subtypes.

Figure 3 shows a hyper-surface plots for the healthy (A) and basal (B) networks. There, the mutual information values between gene pairs are placed depending on the distance (left axis at the plane), i.e., ; at the same time, gene-pairs are placed according to the rank of MI values. Green lines are a guide-to-the-eye to observe the different MI values depending on the distance in both phenotypes. Basal subtype follows a power law decay, meanwhile the healthy network MI values does not depend on the distance.

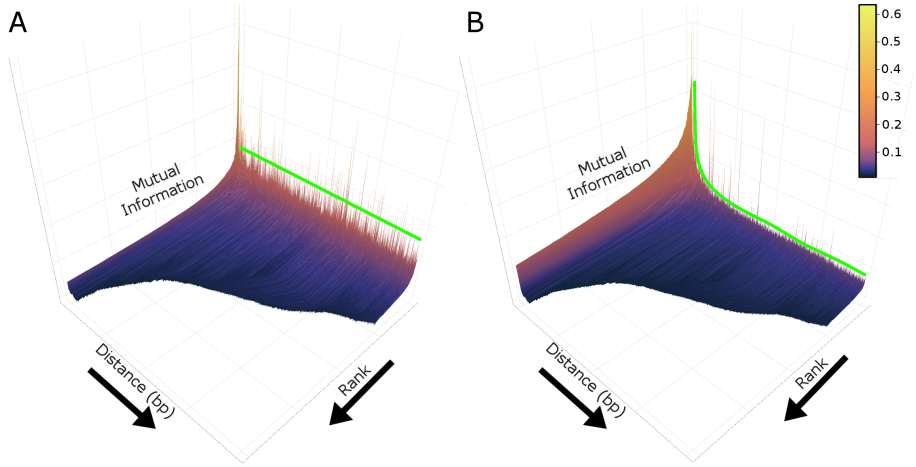

Figure 3: **Strength of *cis*- interactions decay exponentially in cancer.** This hyper-surface plots show the mutual information values depending on the distance for chromosome 1 in healthy (A) and basal (B) networks. Left axis at the plane represents the distance between gene-pairs. Right axis at the plane is a sorting of the mutual information values per distance. Vertical axis is the MI value. Notice that in basal network, the MI values decay strongly (wall at the right in B), meanwhile for healthy network this effect does not occur. Green lines are a guide-to-the-eye to observe said effect. Is can also be noticed that MI values in basal are higher than healthy network at the closest distances: orange color at the left wall, representing MI values at the closest distances, is more abundant in basal than in healthy phenotype.

Loss of *trans*- regulation is more evident in non-luminal subtypes

It is also remarkable the fact that the cancer circos plots with more *trans*- interactions are those for luminal subtypes, meanwhile HER2+ and basal figures are barely populated with orange links(Supplementary Figure 1). Table 1 shows the proportions of all networks. There, it can be clearly observed that the strongest relationships for each phenotype are different, in particular between healthy and cancer networks.

Strength of *cis*- regulations decays exponentially in cancer

Results observed in Figure 2 and Table 1, were obtained only with the 11,675 strongest relationships (the 0.01% highest values). Then, said results may be biased due to the measured subset of links. To overcome this problem, we calculated the whole set of interactions, to observe whether the strength of correlations is dependent on the distance between genes. Figure 2B, C, E, and F shows the mean value of mutual information between gene-pairs with a determined distance in Chromosome 1, for healthy and basal subtype. The other subtype plots are presented in Supplementary Figure 1.

As it is observed, the instensity of intra-cromosomal interactions in the healhty network does not change with proximity (black lines in plots at the right of Figure 2). The fitted line is practically horizontal ($m = 6.17e^{-12}$, see Methods section). Meanwhile, for any subtype, the mean value of mutual information follows a clear power law decay relative to the distance of gene pairs. Figure 2E and F show this for the basal subtype with an exponent of -0.14. Supplementary material 2 contains this information for all chromosomes in the four subtypes.

Concomitant with the exponential decay of correlation strength in *cis*- interactions in subtype networks, it can be observed a clear increase in the strength of the closest gene-pairs for all subytypes. Figure 3 shows the distribution of mutual information values for the basal network subtype (black) and the healthy network (red). If we focus on the left part of the curves, gene-pairs with close proximity in the basal network have mutual information values higher than the healthy one. However, the decay reaches a lower plateau in the former than the latter. Apparently, the effect on *cis*- regulation during cancer is an interplay between loss of *long-range* and gain of *short range* gene-pair correlation. This effect is more clearly observed in Figure 4, where the average mutual information values in chromosome 1 of healthy and basal networks are depicted. It is noticeable the lower MI values in basal interactions at long distances, and at the same time, this subtype presents higher values at shorter physical distances ($<\approx 7.5 \times 10^7$), compared to the invariant MI values in the healthy phenotype.

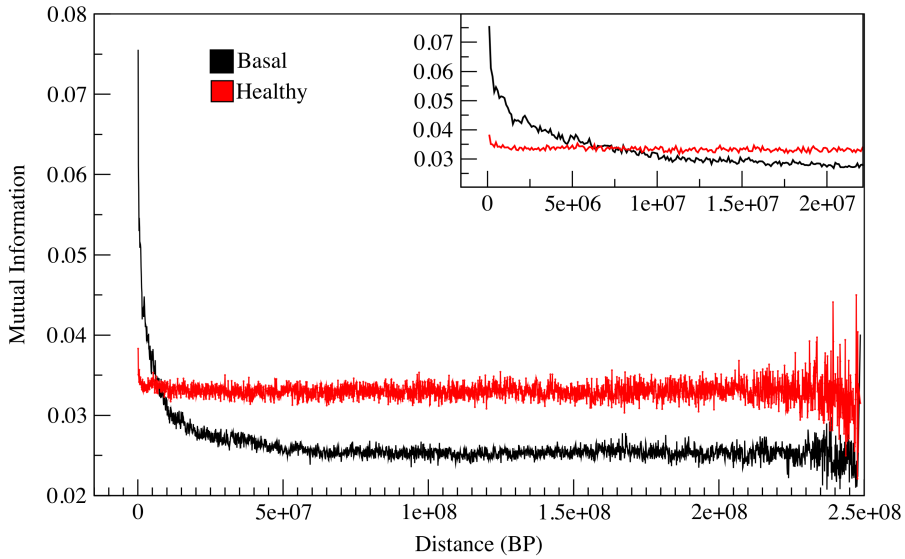

Figure 4: **Interplay between loss of long-range and gain of short-range strength of interactions in breast cancer.** This plot shows the average mutual information values of gene-pairs according to their physical distances in chromosome 1. These curves are the same than Figure 2C and E. Red line represents the distribution of MI values and black takes account for basal subtype. Notice that at long distances ($> 10^7$ bp), basal MI values are lower, but at shorter distances ($<\approx 7.5 \times 10^6$ bp), MI values for basal are consistently higher than the healthy ones. The upper-right inset is a zoom of the main figure, showing the point in which both curves intersect.

Discussion

Breast cancer is, at its core, defined by changes in the cellular regulatory program, which lead to the clinical and pathological characteristics of the disease [10]. We have shown, along with others, that this regulatory program may be reconstructed through the statistical inference of relationships from high throughput gene expression experiments [11, 12] and modeled as transcriptional networks.

Previously, we have described a marked difference in the abundance of inter (*cis*-) and intra chromosomal (*tran*-) interactions found in cancer and non-cancer breast tissue [9]. We also reported a pattern displayed by the *cis*- interactions, in which phisically close genes in terms of chromosome position had the strongest relationships. This result provides physical context to genomic regulation in breast cancer. Following that context, a major question to be answered with this work was whether the loss of trans regulation exhibited different patterns

between the breast cancer molecular subtypes. The heterogeneity of breast cancer is another feature that may be recovered through the analysis of transcriptional networks [13, 14].

Our results show that loss of *trans*- regulation is a common feature of breast cancer molecular subtypes when compared to the healthy breast tissue network. However, each subtype ehxhibits a unique pattern of *trans*- regulation loss. Differences are illustrated in figure 1: the healthy network has a giant component filled with both *cis*- and *trans*- interactions. Meanwhile, the molecular subtypes present different levels of network cohesion, with smaller components comprising mostly intrachromosomal interactions. The luminal B subtype in particular displays an intermediate behavior, as it is the only subtype that has a large connected component, but with more gene-pair interactions belonging to the same chromosome, reminiscent to the concept of network modularity that has been previously studied by our group in the context of molecular subtypes [14].

In table 1, we show the proportion of *cis*- and *trans*- interactions in each molecular subtype, along with the healthy network. It may be seen that each molecular subtype has a different *cis/trans* connection proportion: the basal subtype has the highest proportion, followed by the HER2+ subtype, luminal A, and finally the luminal B subtype. Luminal subtypes usually have a better prognosis than the non-luminal subtypes. Therefore, the proportion of *cis/trans* connections may be related to the physiopathological features involved in this prognosis difference, such as chromosomal instability.

The topic of chromosomal instability in breast cancer subtypes has been mostly explored from an inmunohistochemical perspective [15]. In [16], an association between genomic instability and breast cancer subtypes was established based on a genomic instability gene signature. Other studies have shown particular chromosomal instability features that characterize specific subtypes: for instance, [17] shows a distinct pattern of DNA amplification in the luminal B subtype, not unlike the distinct trans regulation pattern found in our work.

Strenght of *cis*- gene-pair interactions and its correlation to physical proximity in terms of gene chromosomal location was also explored. Although there have been described instances of correlation between gene distances and their coexpression in prokaryotes [18], this phenomenon has been less characterized in eukaryotes [19], and has not been described as a general feature of cancer.

Figure 2, panels B, C, E, F, figure 4, and Supplementary Figure 1 depict mutual information value and distance correlation. We have found that in healthy breast tissue, there is no evident link between physical proximity and the intensity of gene-pair interaction, inside a given chromosome. However, in all molecular subtypes of breast cancer, there is a marked trend: average MI value decay with respect to gene distance, following a power law distribution. To our knowledge, this is the first time that an information theory derived approach

has been used to associate the gene regulatory program to the physical distance between genes.

Comparing the aformentioned power law decays curves for the cancer molecular subtypes with the behavior of the correlation in the healthy newtork, it can be argued that what we describe as loss of *trans*- regulation is not an isolated feature, as it also involves an increase in mutual information values between the closest genes. *Short-range* interactions are favored over distant relationships.

The loss of *long-range* correlations and the increment of strength in physically close interactions may be related to a global `loss of functional coupling` and a `local specialization`, respectively. Although experimental evidence is needed to confirm the extent to which each feature contributes to the breast cancer phenotype and their corresponding implications on prognosis, they may be associated to the well-known hallmark of genomic instability.

Chromosomic instabilities leading to genome re-arrangements may alter the patterns of gene regulation and trigger a global reconstitution of the regulatory program. Such is the case of promoter translocation, in which the promoter region for one gene is relocated so that it is able to activate or repress new genes that were not previously present on its repertoire [20]. Patterns of genomic copy number variation have also been correlated with gene expression and patient survival [21, 22]. Chromotripsis, a phenomenon that involves DNA breaks affecting various chromosomes has been observed in breast cancer, with differences in breakage patterns related to inmunohistochemical profiles in breast cancer [15].

All of the aformentioned events, along with epigenetic deregulations [23, 4, 24] influence global gene expression patterns promoting the oncogenic phenotype with pathological and clinical implications. We consider that our findings provide further evidence to the role of genomic instability and its effect in transcriptional deregulation, and the heterogeneity of breast cancer manifestations.

Conclusion

In this work we provide novel insights into the global behavior of whole genome regulatory programs. We show that a healthy regulatory program involves genomic regulation independent of distance or chromosomal location, whereas each clinical manifestation of breast cancer involves a specific divergence of the global and orchestrated gene regulatory program that defines the healthy physiological state.

By integrating the complex gene expression patterns characteristic of a different breast cancer molecular subtypes into well-defined regulatory programs

represented accurately in the form of gene regulatory networks, we were able to unveil one possible way in which chromosomal instability leads to system level regulation. Global gene regulation may be deemed ultimately responsible for the differential physiology of breast cancer subtypes. This may, under deeper scrutiny provide further advances towards individualized cancer therapeutics.

Methods

Databases

A collection of The Cancer Genome Atlas (TCGS) breast invasive carcinoma datasets were used in this work [25]. Briefly, 780 tumour and 101 normal IlluminaHiSeq RNASeq samples were adquired and pre-processed until log_2 normalized gene expression values were obtained as described in [9].

Data processing

The tumour log_2 normalized expression values were classified using PAM50 algorithm into the respective intrinsic breast cancer subtypes (Normal-like, Luminal A, Luminal B, Basal and HER2-Enriched) using the Permutation-Based Confidence for Molecular Classification [26] as implemented in the *pbcmc* R package [27]. Tumour samples with a non-reliable breast cancer subtype call, were removed from the analysis as described in Table 2.

Table 2: **Breast samples used in this work**

		Breast cancer subtypes			
	Healthy	**Luminal B**	**Luminal A**	**HER2-Enriched**	**Basal-like**
Number of samples	101	58	163	72	142

Notice that out of the 780 original breast cancer samples only 345 samples have been reliably classified into the molecular subtypes.

Multidimensional Principal Component Analysis (PCA) over the gene expression values showed a blurred overlapped pattern among the different breast cancer subtypes. Hence, multidimensional noise reduction using ARSyN R implementation was used [28]. Finally, PCA visual exploration showed that the noisy pattern was removed, thus breast cancer subtypes clustered without overlap.

Differential expression analysis

A linear effect model was adjusted in a gene by gene basis using *limma* R package [29] to find differentially expressed genes between each breast cancer

subtype with respect to the normal samples. P-values were adjusted for multiple comparisons using the False Discovery Rate (FDR) [30]. Thus, selected gene candidates cope with a $FDR < 1 \times 10^{-5}$ and a $|log_2(Fold_change)| > 1$. Visual inspection using heatmaps confirmed the appropriate clustering of the samples according to the experimental design.

Network construction

Gene regulatory network deconvolution from experimental data has been extensively used to unveil co-regulatory interactions between genes by looking out for patterns in their experimentally-measured mRNA expression levels. A number of correlation measures have been used to deconvolute transcriptional interaction networks based on the inference of the corresponding statistical dependency structure in the associated gene expression patterns [31, 32, 33, 34, 35]. It has long been known that the maximum likelihood estimator of statistical dependency is mutual information (MI) [34, 35, 11, 36]. ARACNE [37] is the flagship algorithm used to quantify the degree of statistical dependence between pairs of genes.

In a nutshell, the algorithm calculates the *Mutual Information (MI)* –a non-parametric measure that captures non-linear dependencies between variables [38]– in a relatively fast implementation. The method associates a MI value to each significance value (p-value) based on permutation analysis, as a function of the sample size. In order to make comparable networks, only the top 0.01% MI were kept for every tested network. [1]

Competing interests

The authors declare that they have no competing interests.

Author's contributions

DGC performed computational analyses, developed and implemented programming code, contributed to the writing of the manuscript. GDJ performed computational analyses, developed and implemented programming code, contributed to the writing of the manuscript. CF performed pre-processing and low-level data analysis, developed and implemented programming code, contributed to the writing of the manuscript. EHL devised the project's strategy and methodological approach, contributed to the theoretical and modeling analysis, co-supervised the project, contributed and supervised the writing of the manuscript. JEE designed the overall project, co-supervised the project, took the lead in the biological analyses, drafted the manuscript. All authors read and approved the final version of the manuscript.

[1] *** Supplementary materials for this work are available under request.**

Acknowledgements

DGC is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM). This work is part of her PhD Thesis. This work was supported by CONACYT (grant no.285544/2016, JEE), as well as by federal funding from the National Institute of Genomic Medicine (Mexico). Additional support has been granted by the National Laboratory of Complexity Sciences (grant no. 232647/2014 CONACYT). EHL is a recipient of the 2016 Marcos Moshinsky Fellowship in the Physical Sciences.

References

[1] Hanahan, D., Weinberg, R.A.: Hallmarks of cancer: the next generation. cell **144**(5), 646–674 (2011)

[2] Negrini, S., Gorgoulis, V.G., Halazonetis, T.D.: Genomic instability—an evolving hallmark of cancer. Nature reviews Molecular cell biology **11**(3), 220 (2010)

[3] Lustberg, M., Ramaswamy, B.: Epigenetic targeting in breast cancer: therapeutic impact and future direction. Drug news & perspectives **22**(7), 369 (2009)

[4] Montenegro, M., Gonzalez-Guerrero, R., Sanchez-del-Campo, L., Piñero-Madrona, A., Cabezas-Herrera, J., Rodriguez-Lopez, J.: Targeting the epigenetics of the dna damage response in breast cancer. Cell death & disease **7**(4), 2180 (2016)

[5] Espinal-Enríquez, J., Mejía-Pedroza, R.A., Hernández-Lemus, E.: Computational approaches in precision medicine. In: Progress and Challenges in Precision Medicine, pp. 233–250. Elsevier, ??? (2017). doi:10.1016/b978-0-12-809411-2.00013-1

[6] Hood, L., Rowen, L.: The human genome project: big science transforms biology and medicine. Genome medicine **5**(9), 79 (2013)

[7] Davidson, E., Levin, M.: Gene regulatory networks. Proceedings of the National Academy of Sciences **102**(14), 4935–4935 (2005)

[8] Davidson, E.H., Erwin, D.H.: Gene regulatory networks and the evolution of animal body plans. Science **311**(5762), 796–800 (2006)

[9] Espinal-Enriquez, J., Fresno, C., Anda-Jáuregui, G., Hernandez-Lemus, E.: Rna-seq based genome-wide analysis reveals loss of inter-chromosomal regulation in breast cancer. Scientific Reports **7**(1), 1760 (2017)

[10] Li, R., Campos, J., Iida, J.: A gene regulatory program in human breast cancer. Genetics

201(4), 1341–1348 (2015). doi:10.1534/genetics.115.180125. http://www.genetics.org/content/201/4/1341.full.pdf

[11] Margolin, A.A., Wang, K., Lim, W.K., Kustagi, M., Nemenman, I., Califano, A.: Reverse engineering cellular networks. Nature protocols **1**(2), 662 (2006)

[12] Tovar, H., García-Herrera, R., Espinal-Enríquez, J., Hernández-Lemus, E.: Transcriptional master regulator analysis in breast cancer genetic networks. Computational biology and chemistry **59**, 67–77 (2015)

[13] de Anda-Jáuregui, G., Velázquez-Caldelas, T.E., Espinal-Enríquez, J., Hernández-Lemus, E.: Transcriptional network architecture of breast cancer molecular subtypes. Frontiers in physiology **7**, 568 (2016)

[14] Alcalá-Corona, S.A., de Anda-Jáuregui, G., Espinal-Enríquez, J., Hernández-Lemus, E.: Network modularity in breast cancer molecular subtypes. Frontiers in physiology **8**, 915 (2017)

[15] Przybytkowski, E., Lenkiewicz, E., Barrett, M.T., Klein, K., Nabavi, S., Greenwood, C.M., Basik, M.: Chromosome-breakage genomic instability and chromothripsis in breast cancer. BMC Genomics **15**(1), 579 (2014). doi:10.1186/1471-2164-15-579

[16] Habermann, J.K., Doering, J., Hautaniemi, S., Roblick, U.J., Bündgen, N.K., Nicorici, D., Kronenwett, U., Rathnagiriswaran, S., Mettu, R.K.R., Ma, Y., Krüger, S., Bruch, H.-P., Auer, G., Guo, N.L., Ried, T.: The gene expression signature of genomic instability in breast cancer is an independent predictor of clinical outcome. International journal of cancer **124**, 1552–1564 (2009). doi:10.1002/ijc.24017

[17] Melchor, L., Honrado, E., Garcia, M., Alvarez, S., Palacios, J., Osorio, A., Nathanson, K., Benitez, J.: Distinct genomic aberration patterns are found in familial breast cancer associated with different immunohistochemical subtypes. Oncogene **27**(22), 3165 (2008)

[18] Pannier, L., Merino, E., Marchal, K., Collado-Vides, J.: Effect of genomic distance on coexpression of coregulated genes in e. coli. PloS one **12**(4), 0174887 (2017)

[19] Michalak, P.: Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics **91**(3), 243–248 (2008)

[20] Zheng, J.: Oncogenic chromosomal translocations and human cancer. Oncology reports **30**(5), 2011–2019 (2013)

[21] Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.-L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., *et al.*: Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer cell **10**(6), 529–541 (2006)

[22] Hicks, J., Krasnitz, A., Lakshmi, B., Navin, N.E., Riggs, M., Leibu, E., Esposito, D., Alexander, J., Troge, J., Grubor, V., *et al.*: Novel patterns of genome rearrangement and their association with survival in breast cancer. Genome research **16**(12), 1465–1479 (2006)

[23] Lustberg, M., Ramaswamy, B.: Epigenetic targeting in breast cancer: therapeutic impact and future direction. Drug news & perspectives **22**(7), 369 (2009)

[24] Longacre, M., Snyder, N.A., Housman, G., Leary, M., Lapinska, K., Heerboth, S., Willbanks, A., Sarkar, S.: A comparative analysis of genetic and epigenetic events of breast and ovarian cancer related to tumorigenesis. International journal of molecular sciences **17**(5), 759 (2016)

[25] Tomczak, K., Czerwińska, P., Wiznerowicz, M.: The cancer genome atlas (tcga): an immeasurable source of knowledge. Contemporary oncology **19**(1A), 68 (2015)

[26] Fresno, C., González, G.A., Merino, G.A., Flesia, A.G., Podhajcer, O.L., Llera, A.S., Fernández, E.A.: A novel non-parametric method for uncertainty evaluation of correlation-based molecular signatures: its application on pam50 algorithm. Bioinformatics **33**(5), 693–700 (2016)

[27] Fresno, C., González, G.A., Llera, A.S., Fernández, E.A.: pbcmc: Permutation-based confidence for molecular classification. R package version **1**(2) (2016)

[28] Nueda, M.J., Ferrer, A., Conesa, A.: Arsyn: a method for the identification and removal of systematic noise in multifactorial time course microarray experiments. Biostatistics **13**(3), 553–566 (2012)

[29] Smyth, G.K.: Limma: linear models for microarray data. In: Bioinformatics and Computational Biology Solutions Using R and Bioconductor, pp. 397–420. Springer, ??? (2005)

[30] Storey, J.D.: False discovery rate. In: International Encyclopedia of Statistical Science, pp. 504–508. Springer, ??? (2011)

[31] Usaite, R., Jewett, M.C., Oliveira, A.P., Yates, J.R., Olsson, L., Nielsen, J.: Reconstruction of the yeast snf1 kinase regulatory network reveals its role as a global energy regulator. Molecular systems biology **5**(1), 319 (2009)

[32] Friedman, N., Linial, M., Nachman, I., Pe'er, D.: Using bayesian networks to analyze expression data. Journal of computational biology **7**(3-4), 601–620 (2000)

[33] Wang, Z., Hou, Z., Xin, H.: Internal noise stochastic resonance of synthetic gene network. Chemical physics letters **401**(1-3), 307–311 (2005)

[34] Emmert-Streib, F., Glazko, G., De Matos Simoes, R., *et al.*: Statistical inference and reverse engineering of gene regulatory networks from observational expression data. Frontiers in genetics **3**, 8 (2012)

[35] Hernández-Lemus, E., Rangel-Escareño, C.: The role of information theory in gene regulatory network inference. Information Theory: New Research, 109–144 (2011)

[36] Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., Califano, A.: Reverse engineering of regulatory networks in human b cells. Nature genetics **37**(4), 382 (2005)

[37] Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., Califano, A.: Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In: BMC Bioinformatics, vol. 7, p. 7 (2006). BioMed Central

[38] Hernández-Lemus, E., Siqueiros-García, J.M.: Information theoretical methods for complex network structure reconstruction. Complex Adaptive Systems Modeling **1**(1), 8 (2013)