2022 19th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE). Mexico City, Mexico. November 9-11, 2022

Genetic Network of Breast Cancer Metastasis in Lymph Nodes via Information Theory Algorithms

Matías Alvarado
Departamento de Computación
Centro de Investigación y de Estudios Avanzados del IPN
México City, México
matias@cs.cinvestav.mx
ORCID 0000-0003-2853-7451

Ivan Valdespin*
Departamento de Computación
Centro de Investigación y de Estudios Avanzados del IPN
México City, México
ivan.valdespin@cinvestav.mx

Moises León**
Departamento de Computación
Centro de Investigación y de Estudios Avanzados del IPN
México City, México
moises.leon@cinvestav.mx

Sergio A. Alcalá-Corona***
Departamento de Computación
Centro de Investigación y de Estudios Avanzados del IPN
México City, México
sergio.alcala@cinvestav.mx
ORCID:0000-0001-8643-9214

***Abstract*—Breast cancer is a public health challenge because its high rate of growth around the world in recent decades. Breast cancer induces metastasis on diverse organs, quite hard to bound and inhibit. Metastasis is the main whatfor to cancer-associated deaths in patients. This paper introduces the matrix of gene expressions, algorithmically inferred, with data on biopsis of (GEO database) breast cancer metastasis in lymph nodes. The matrix grants for generating the gene network (GN) with each node per gene and the weighted links between genes the interaction intensity. On the base of GN structural analysis we identify the importance rank of each gene through the set of biopsis. This result is by applying information-theory-based algorithms for ranking the gene importance –in breast cancer metastasis. A brief and basic analysis focus on the higher rank genes-nodes says that, both the gene-node degree and the correlation values with the other genes support the every gene relevance.**

***Index Terms*—Information Theory Algorithms, Breast Cancer, Gene Network, Genes Relevance, Metastasis.**

I. Introduction

Any cancer primary tumor not bounded nor inhibit by the immune system dangerous grows and proliferates [1], [2]. Cancer cells (CC) that drop off the primary tumor then being incorporated in the blood flow in the angiogenesis process [3], [4], eventually may invade tissues in other organs. In adittion, exosomes from primary tumor containing DNA, microRNAs, proteins, and other molecules also may travel and arrive to distant tissues for supporting the CC invasion. The CC and exosomes modify the micro-environment of the host tissues during the premetastatic niche (PMN). PMN is the preliminary process to cancer metastasis in a tissue derived by the presence of CC (seed) and exosomes (soil) coming from a cancer primary tumor [5], [6]. We use data on breast cancer first metastasis in lymph nodes. Breast cancer is one of the most vigorous for making metastasis in diverse organs' tissues. Bone, liver, lung, lymph nodes, are the most frequent organs for metastasis occurrence [7], [8].

Usually, to study cancer and metastasis is by clinical experiments, *in vitro* and *in situ*, in oncology laboratories [3], [9], [10]. This practice is money expensive and highly time consuming. Last decades ago, the alternatives of mathematical modeling and algorithmic simulation, *in silico*, advantages the understanding of cancer and metastasis diseases [11], [12], [13], [14], [15], [16] [17]; as well, mathematical and computing tools have strongly diminished the cost of cancer trials and studies. Current mathematical models have improved the precision of research in specific cancer research issues; computer algorithms allow simulating diverse experimental conditions and making several realizations for statistical analysis. Of course, algorithmic results should be validated and compared with real data coming from experiments essays and clinical studies.

Genetic analysis supports to identify the key role of set of genes in biological processes [18]: The gene expression profiling (GEP) is a basic gene measure unit useful for quantify every gene relationships. Gene interactions shape the gene network of correlations between pairs of genes; the level of expression determines the specific anatomic and functional configuration over each organ in the body. Thousands of genes constitute the human genome; around 22000 genes shape the GN of interaction between pairs of genes with expression in the metastasis process.

In this work, we focus and introduce the GN inferred

Thanks to Consejo Nacional de Ciencia y Tecnología de México grants: Master Scholarships CVU No. *1147565 and **1144833; ***Postdoctoral Researcher grant in Project A1-S-20037.

978-1-6654-5508-4/22/$31.00 ©2022 IEEE2022

with data from both breast cancer primary tumor and once of metastasis in lymph nodes from that tumor. We use information-theory-based algorithms [19] for GN inferring; the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNE) [20] multicore version [21] is efficient supporting our purpose. The information is from biopsies in the Gene Expression Omnibus (GEO) database https://www.ncbi.nlm.nih.gov/geo/. Data sets with accession number GSE32489 correspond to the GEP of formalin-fixed paraffin-embedded (FFPE) breast cancer metastases, autopsy lymph nodes and tissues (platform GPL8432 Illumina HumanRef-8 WG-DASL v3.0).

II. Gene Expression Matrix

To generate the GN of breast cancer that metastasizes in lymph nodes data are processed and sorted for applying ARACNE. Samples are from the same type of tissue.

A. Transcription per Million Measure

GEP is measured in transcripts per million (TPM). TPM is the normalization for RNA-seq and indicates the degree of expression of every gene. One TPM is equal to $\frac{1}{1{,}000{,}000}$ RNA molecules in the RNA-seq sample from the gen. Thus, the higher the TPM value the more expressed is the gen.

In this analysis the need information is about the function of every gene and the scientific name gene symbol; using that symbols a dictionary is created. This is the dictionary associated to the genes in the dataset, without depend on the particular Illumina nomenclature. Dictionaries have miscellaneous information about genes they manipulate, such as the nucleotide sequence, the specie which comes the tissue, the protein product, and other references of Ilumina datasets. To identify each of the genes the Illumina HumanRef-8 WG-DASL v3.0 platform dictionaries are used [22]. A dictionary with the scientific standard gene identifier and the symbol of genes in this project was required and done.

The downloaded GEO database file of samples *Series Matrix File(s)* is in a txt format; the information is on title, serial number, type of sample tissue, normalized method, and the each gene TPM. Samples of GSE32489 data set are:

- 5 negative control samples
- 6 positive control samples
- 90 lymph node metastasis samples

90 out of 120 samples in the dataset correspond to breast cancer with metastasis in lymph node (the 30 not cancer samples were removed).With these data each being identified with gene symbol the expression matrix are created.

B. The information theory algorithms

The Table I shows an example of the matrix of expression: the first column with the gene symbol ID, the first row with the name of each sample per column, and in the next entries the TPM of each gene per sample.

A fundamental information theory (IT) concept for quantifying information relationships is the entropy $S(t)$ of discrete variable t [23]. The antecedent of IT entropy is the thermodynamics entropy, that is the waste of energy not useful for work [24]. Entropy is formally defined as follows:

ID_REF	GSM803917	GSM803918	GSM803919
EEF1A1	10.65796523	11.44461781	9.285907809
SLC35E2	7.047777431	6.554849504	7.130089085
RPS28	10.74388693	10.06837991	9.821975287
IPO13	6.698028587	8.252343047	6.242786112
AFAP1	6.868091137	6.194919819	8.044367827
GGTLC1	6.521264145	6.389951163	8.322217225
CDT1	6.801231704	8.549550323	8.849987275
TRPV1	6.548158108	6.846182417	6.840801471
LPP	7.306225172	9.137877398	6.142776208
CCNE2	6.37355111	6.39454826	7.52573492
HNRNPAB	6.261411622	8.546198504	7.036285894
LOH12CR1	7.495838805	6.973083944	8.796533412
SNIP1	6.255716184	6.473306895	6.210033094
COL17A1	6.356384366	7.135407628	7.04436924
BCL6B	6.229631032	6.676598297	6.006150766

TABLE I: Part of expression matrix of GSE32489 data set

$$S(t) = -\kappa \sum_i p_i(t) log p_i(t)$$

with $p_i(t)$ the probability of each (discrete) state i of the system; in this case, the system emerges from the interaction between the information of pairs of genes in the samples. κ is the Boltzman constant, first proposed by the genius Ludwig Botzman in th 19th century, and wrote in his thumb's surface.

As well, $S(t)$ is used to estimate the information relationships between variables x, y. The IT equation on mutual information (MI) is, $I(x, y) = S(x) + S(y) - S(x, y)$.

From the values of transcriptional interactions between the GEPs, whose relationships are derived from irreducible statistical dependencies -inexplicable through others- we could inferred the GN. GEPs values joined to the predefined list of gene regulators (e.g., transcription factors) are the variables that input ARACNE; output are a *.sif* and *.sort* files with the MI of pairs of genes, that constitute the matrix of expression. In the *sort* file the data are in MI descending order; in the *sif* file the first elements are the more connected with the other genes and so on. From these ordered information the weighted GN is generated.

With the input information of $N = 24,143$ genes from $M = 90$ samples, the ARACNE-Multicore [21] generates 291,405,595 connections between the genes; the file size is 10.3 GB that use 104 minutes of computer processing[1]. The run includes the following steps[20]: MI threshold estimation, and the reconstruction of the network regarding the choice threshold.

To estimate the GEP between pairs of genes, g_i, g_j, with i, j positive integer numbers, the MI formula is used:

$$I(g_i, g_j) = I(g_i, g_j) = S(g_i) + S(g_j) - S(g_i, g_j).$$

a

With N genes and M samples the JPD (Joint Probablilit distribution) of the system [20] is quantified by:

$$P(\{g_i\}_1^N) =$$

[1]The used device is with RAM 16.0 GiB, processor 11th Gen Intel® Core™ i7-11800H @ 2.30GHz × 16, graphics board NVIDIA GA106M [GeForce RTX 3060 Mobile/Max-Q], disk 1.0 TB, OS Ubuntu 22.04 LTS of 64 bits.

$\frac{1}{M}[-\sum_i^N \phi_i(g_i) - \sum_{i,j}^N \phi_{ij}(g_i g_j) - \sum_{i,j,k}^N \phi_{ijk}(g_i g_j g_k)]$;

ϕ is the potential of interaction. Sets of genes stay in interaction if and only if their emerging potential is not null; henceforth, this probability is not null and added in the potential term, and is algorithmically identified.

Again, from IT, the data processing inequality (DPI) states that if pair of individuals, for us the genes g_1 y g_3 interact only through another gene g_2, then $I(g1, g3) \leq min\{I(g1, g2), I(g2, g3)\}$. Thus, the smallest MI values come only from indirect interactions [20].

Observe that with M samples and N genes, the computational complexity of ARCANE is $O(N^3 + N^3M^2)$ [20]. The term N^3 refers to DPI analysis, and N^3M^2 to MI of pairs of genes. Observe that the greater the N and M, the amount of operations is higher computational resources and time consuming. A GN pruning is desired but maintaining the GN centralities without essential changes. This is the reconstruction step. To the reconstruction of the matrix of expression, the MI values should exceed a threshold I0. The DPI is applied to a subset of triplets that exceed the threshold. This pruning eliminates a lot of interactions: the less meaningful are pruned and not appear in the final GN.

III. Results

As said before, the full GN is too heavy to computational deal with. GN is pruned, without loss the essential structural centralities [18]. Connections with MI minor or equal to 0.4 are pruned and the result is a GN with 109,244 connections (2.4 MB file size), a manageable GN for analysis; the GN links between genes-nodes with weights upper this threshold are meaningful and determine the GN topology [2] The pruned weighted GN is shown in Figure 1.

A. GN topology

To learned the GN centralities the structural analysis allows for identifying the nodes-genes with the highest degree (number of links), alongside the links with the higher MI pairs of genes –associated to their weighted links.

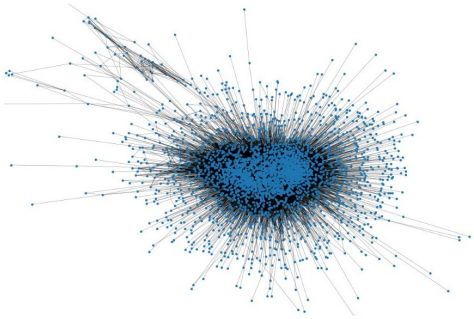

Fig. 1: GN of breast cancer metastasis in lymph nodes with genes of high MI.

The GN is fully connected in a single body without islands [26], [27], and with the next centralities:

- Average degree(K): $60.99 \approx 61$
- Clustering coefficient (C): 0.52
- Average shortest path length (L): 2.746

It is plausible say that GN topology is small-world [28], [29]: If N is the number of nodes, the network is small-world whenever $L \approx Log_{10}(N)$; that is, the diameter being compared with $Log_{10}(N)$ results of similar size. In this case:

$$L = 2.74$$

$$Log_{10}(N) = Log_{10}(24143)$$

$$Log_{10}(N) = 4.38$$

.

The GN degree distribution is scale-free, so a huge amount of nodes has few connections, and, too few nodes have a huge amount of connections; formally a scale-free network degree distribution follows a power-law distribution with negative sign exponent. $p_k \ K^{-\gamma}$, $\gamma > 0$, see Figure 2.

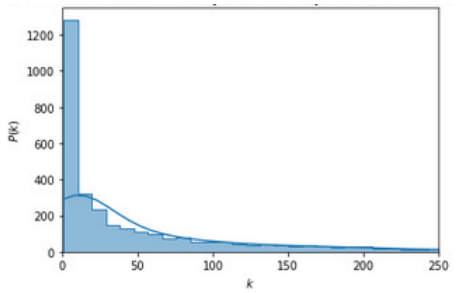

Fig. 2: Degree distribution of genes with mutual information greater than 0.4

B. Relevant Genes

The relevant genes in the breast cancer first metastasis in lymph nodes corresponds to the 15 pairs with the highest weight of mutual correlation.

Correlation	Symbol1	Symbol2
0.689759	LHFPL3	CYorf15B
0.687776	FGA	SALL3
0.687739	TMEM34	CYorf15B
0.681159	LHFPL3	TMEM34
0.676752	FGA	TMEM34
0.675167	FGA	CYorf15B
0.670746	SALL3	TMEM34
0.669893	FGA	P2RX7
0.669491	TMEM34	SLC23A1
0.66921	SALL3	CYorf15B
0.668622	CYorf15B	SLC23A1
0.667756	SALL3	P2RX7
0.666312	FGA	PSD3
0.6657	FGA	LHFPL3
0.661615	TMEM34	P2RX7

TABLE II: Gene pairs with the highest correlation

As a very preliminary analysis on the pair of genes with the highest correlation follows:

- The gene LHFPL3 is member of the lipoma HMGIC fusion partner (LHFP) gene family, a subset of the

[2]The GN analysis was done with the Python NETWORKX library [25].

superfamily of tetraspan transmembrane protein encoding genes. Mutations in one LHFP-like gene result in deafness in humans and mice; the LHFP-like gene is fused to a high-mobility group gene in a translocation-associated lipoma. [30]

- CYHYorf15B chromosome Y open reading frame [31], hence is a portion of a DNA sequence that not include a stop codon.[32]

Up to our knowledge not scientific reference was found that explains the correlation between these two genes. However, due to the studies about cancer and as explained in [17] and [6] we may thing that the first gene is being affected by the second; and probably they are generating or promoting the formation of cancer cells.

The 15 genes with the highest number of connections with other nodes are shown below.

Gen	Num of conections
LHFPL3	1458
TMEM34	1391
SALL3	1352
Cyrof15B	1352
FGA	1348
P2RX7	1090
EFNB2	1079
SLC23A1	1037
PLK1	946
LOC401296	921
LRDD	912
PRSS35	897
C12orf39	866
CARM1	773
CDH26	750

TABLE III: Gene-Nodes with the highest degrees in the GN of breast cancer first metastasis in lymph node metastasis

The gene LHFPL3 is the highest degree, whereas the gene CYorf15B is the fifth of node degree in the GN. This could confirm, even preliminary, our hypothesis that the big number of connections the genes have, it determines the breast cancer first metastasis in lymph nodes.

Other interesting thing from the analysis of the second element in Table II, is the high correlation between the fourth and the 6th element of Table III:

- SALL3: This gene encodes a sal-like C2H2-type zinc-finger protein; it belongs to a family of evolutionarily conserved genes found in diverse species like Drosophila, C. elegans, and vertebrates. Mutations in some of these genes are associated with congenital disorders in human, suggesting their importance in embryonic development. It is suggested that silencing this gene it may result in acceleration of DNA methylation, so may have a role in oncogenesis.[33].
- FGA: The protein encoded by this gene is the alpha component of fibrinogen, a blood-borne glycoprotein composed of three pairs of nonidentical polypeptide chains. Following vascular injury, fibrinogen is cleaved by thrombin to form fibrin, which is the most abundant component of blood clots. In addition, various cleavage products of fibrinogen and fibrin regulate cell adhesion and spreading, display vasoconstrictor and chemotactic activities, and are mitogens for several cell types. Mutations in this gene lead to several disorders, including dysfibrinogenemia, hypofibrinogenemia, afibrinogenemia, and renal amyloidosis. Alternative splicing results in two isoforms that vary in the carboxy-terminus[34].

A second hypothesis can raise about the correlation between these two genes and the frequent connections that each one presents: probably, they are factors to promote the metastasis in lymph node tissues; [13] the pathway transport to promote the PMN is through the blood.

IV. Discussion and Future Work

With information theory algorithms a GN of metastatic lymph node tumors derived from primary breast cancer is generated and analyzed. The distribution of probabilities of gene-node degree support for identifying every gene importance. The GN topology is scale-free degree distribution and small-world. Hence, there are few nodes with big degree and a huge number of genes with low degree; as well, the path between each pair of nodes is quite short. May this identification be a preliminary step to advance for understanding the Pre-Metastatic Niche (PMN)? PMN is the complex process prior to cancer metastasis, given the interaction of molecules and cells that transform the micro-environment in organ tissues [35]; all these interactions induce the remodeling of extracellular matrix, increased vascular permeability, angiogenesis, immunosuppression; also, the PMN recruits bone marrow-derived cells (BMDC) a principal soil in the cancer cells colonization that invaded organ tissues [36]. The PMN involves interactions in and from the micro-environments of cancer primary tumor, that promotes cancer springs in secondary organs and tissues, previous to the proper cancer metastasis.

V. Conclusions

We identified, in a preliminary manner, the main genes in the GN of metastasis in lymph node derived from breast cancer. For every node-gene, both its level of correlations with each other gene joined to its degree, could indicate its relevance in the metastasis. The GN topology is scale-free degree distribution and small-world. The LHFPL3 gene is the highest degree and weight connection in the network; the CYorf15B gene is the 5th more connected and with high weight connections. We guess that both are involved in the lymph node metastasis from breast cancer.

References

[1] Douglas Hanahan and Robert Weinberg. “The Hallmarks of Cancer”. In: *Cell* 100 (Feb. 2000), pp. 57–70. DOI: 10.1016/S0092-8674(00)81683-9.

[2] Paul K. Newton, Jeremy Mason, Kelly Bethel, Lyudmila A. Bazhenova, Jorge Nieva, and Peter Kuhn. “A Stochastic Markov Chain Model to Describe Lung Cancer Growth and Metastasis”. In: *PLOS ONE* 7.4 (Apr. 2012). DOI: 10.1371/journal.pone.0034637. URL: https://doi.org/10.1371/journal.pone.0034637.

[3] Douglas Hanahan and Robert A Weinberg. “Hallmarks of cancer: the next generation”. In: *cell* 144.5 (2011), pp. 646–674.

[4] Douglas Hanahan and Lisa M. Coussens. “Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment”. In: *Cancer Cell* 21.3 (2012), pp. 309–322. ISSN: 1535-6108. DOI: https://doi.org/10.1016/j.ccr.2012.02.022. URL: https://www.sciencedirect.com/science/article/pii/S1535610812000827.

[5] Yang Liu and Xuetao Cao. “Characteristics and significance of the pre-metastatic niche”. In: *Cancer cell* 30.5 (2016), pp. 668–681.

[6] Héctor Peinado, Haiying Zhang, Irina R Matei, Bruno Costa-Silva, Ayuko Hoshino, Goncalo Rodrigues, Bethan Psaila, Rosandra N Kaplan, Jacqueline F Bromberg, Yibin Kang, et al. “Pre-metastatic niches: organ-specific homes for metastases”. In: *Nature Reviews Cancer* 17.5 (2017), pp. 302–317.

[7] Paul K. Newton, Jeremy Mason, Kelly Bethel, Lyudmila Bazhenova, Jorge Nieva, Larry Norton, and Peter Kuhn. “Spreaders and Sponges Define Metastasis in Lung Cancer: A Markov Chain Monte Carlo Mathematical Model”. In: *Cancer Research* 73.9 (Apr. 2013), pp. 2760–2769. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-12-4488. eprint: https://aacrjournals.org/cancerres/article-pdf/73/9/2760/2700408/2760.pdf. URL: https://doi.org/10.1158/0008-5472.CAN-12-4488.

[8] Paul Newton, Jeremy Mason, Neethi Venkatappa, Maxine Jochelson, Brian Hurt, Jorge Nieva, Larry Norton, and Peter Kuhn. “Spatiotemporal progression of metastatic breast cancer: a Markov chain model highlighting the role of early metastatic sites”. In: *npj Breast Cancer* 1 (Oct. 2015), p. 15018. DOI: 10.1038/npjbcancer.2015.18.

[9] Yousef Ahmed Fouad and Carmen Aanei. “Revisiting the hallmarks of cancer”. In: *American journal of cancer research* 7.5 (2017), p. 1016.

[10] Jeffrey West and Paul K. Newton. “Cellular interactions constrain tumor growth”. In: *Proceedings of the National Academy of Sciences* 116.6 (2019), pp. 1918–1923. DOI: 10.1073/pnas.1804150116. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1804150116. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1804150116.

[11] Sabrina L Spencer, Matthew J Berryman, José A García, and Derek Abbott. “An ordinary differential equation model for the multistep transformation to cancer”. In: *Journal of Theoretical Biology* 231.4 (2004), pp. 515–524.

[12] Grazziela P Figueredo, Peer-Olaf Siebers, Markus R Owen, Jenna Reps, and Uwe Aickelin. “Comparing stochastic differential equations and agent-based modelling and simulation for early-stage cancer”. In: *PloS one* 9.4 (2014), e95150.

[13] Georgina Cosma, Giovanni Acampora, David Brown, Robert C Rees, Masood Khan, and A Graham Pockley. “Prediction of pathological stage in patients with prostate cancer: a neuro-fuzzy model”. In: *PloS one* 11.6 (2016), e0155856.

[14] Colin G Cess and Stacey D Finley. “Multi-scale modeling of macrophage—T cell interactions within the tumor microenvironment”. In: *PLoS computational biology* 16.12 (2020), e1008519.

[15] Sophie Bekisz and Liesbet Geris. “Cancer modeling: From mechanistic to data-driven approaches, and from fundamental insights to clinical applications”. In: *Journal of Computational Science* 46 (2020), p. 101198.

[16] Mehran Akbarpour Ghazani, Mohsen Saghafian, Peyman Jalali, and Madjid Soltani. “Mathematical simulation and prediction of tumor volume using RBF artificial neural network at different circumstances in the tumor microenvironment”. In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 235.11 (2021), pp. 1335–1355.

[17] Alfonso Rojas-Domínguez, Renato Arroyo-Duarte, Fernando Rincón-Vieyra, and Matías Alvarado-Mentado. “Modeling cancer immunoediting in tumor microenvironment with system characterization through the ising-model Hamiltonian”. In: *BMC bioinformatics* 23.1 (2022), pp. 1–25. DOI: https://doi.org/10.1186/s12859-022-04731-w. URL: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-022-04731-w.

[18] Guillermo de Anda-Jáuregui, Sergio Alcalá Corona, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. “Functional and transcriptional connectivity of communities in breast cancer co-expression networks”. In: *Applied Network Science* 4 (May 2019). DOI: 10.1007/s41109-019-0129-0.

[19] Ilya Nemenman. “Information theory, multivariate dependence, and genetic network inference”. In: *arXiv:q-bio/0406015* (July 2004).

[20] Adam Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla-Favera, and Andrea Califano. “ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context”. In: *BMC bioinformatics* 7 Suppl 1 (Feb. 2006), S7. DOI: 10.1186/1471-2105-7-S1-S7.

[21] @Josemaz Github. *ARACNE-Multicore*. URL: https://github.com/josemaz/aracne-multicore. (accessed: 06.2022).

[22] Illumina Inc. GEO. *Platform GPL8432*. URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL8432. (accessed: 06.2022).

[23] Paul Newton, Jeremy Mason, Brian Hurt, Kelly Bethel, Lyudmila Bazhenova, Jorge Nieva, and Peter Kuhn. “Entropy, complexity, and Markov diagrams for random walk cancer models”. In: *Scientific reports* 4 (Dec. 2014), p. 7558. DOI: 10.1038/srep07558.

[24] C. E. Shannon. “A mathematical theory of communication”. In: *The Bell System Technical Journal* 27.4 (1948), pp. 623–656. DOI: 10.1002/j.1538-7305.1948.tb00917.x.

[25] Dan Aric Hagberg and Pieter. *NETWORKX Network Analysis Python*. URL: https://networkx.org/. (accessed: 07.2022).

[26] Sergio Alcalá Corona, Guillermo de Anda-Jáuregui, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. “Network Modularity in Breast Cancer Molecular Subtypes”. In: *Frontiers in Physiology* 8 (Nov. 2017). DOI: 10.3389/fphys.2017.00915.

[27] Sergio Antonio Alcalá-Corona, Santiago Sandoval-Motta, Jesús Espinal-Enríquez, and Enrique Hernández-Lemus. “Modularity in Biological Networks”. In: *Frontiers in Genetics* 12 (2021). ISSN: 1664-8021. DOI: 10.3389/fgene.2021.701331. URL: https://www.frontiersin.org/articles/10.3389/fgene.2021.701331.

[28] Albert-László Barabási. *Network Science. The Barabási- Albert model*. London: Network Science, 2014. URL: https://barabasi.com/f/622.pdf.

[29] M E J Newman. “The Structure and Function of Complex Networks *”. In: *Society for Industrial and Applied Mathematics* 45 (2 2003), pp. 167–256. URL: http://www.siam.org/journals/sirev/45-2/42480.html.

[30] SciLifeLab. *The Human Protein Atlas*. URL: https://www.proteinatlas.org/ENSG00000187416-LHFPL3. (accessed: julio 2022).

[31] National Institutes of Health-National Center for Biotecnology Information. *CYHYorf15B*. URL: https://www.ncbi.nlm.nih.gov/gene/100271730. (accessed: julio 2022).

[32] National Institutes of Health-National Human Genome Research Institute. *Open Reading Frame*. URL: https://www.genome.gov/genetics-glossary/Open-Reading-Frame. (accessed: julio 2022).

[33] National Institutes of Health-National Center for Biotecnology Information. *SALL3 spalt like transcription factor 3 [Homo sapiens (human)]*. URL: https://www.ncbi.nlm.nih.gov/gene/27164. (accessed: julio 2022).

[34] Wikipedia. *Fibrinogen alpha chain*. URL: https://en.wikipedia.org/wiki/Fibrinogen_alpha_chain. (accessed: julio 2022).

[35] Qi Dong, Xue Liu, Ke Cheng, Jiahao Sheng, Jing Kong, and Tingjiao Liu. “Pre-metastatic Niche Formation in Different Organs Induced by Tumor Extracellular Vesicles”. In: *Frontiers in Cell and Developmental Biology* 9 (2021).

[36] Danielle Harmer, Carolyne Falank, and Michaela R Reagan. “Interleukin-6 interweaves the bone marrow microenvironment, bone loss, and multiple myeloma”. In: *Frontiers in endocrinology* (2019), p. 788.