

RESEARCH ARTICLE

Bayesian networks established functional differences between breast cancer subtypes

Lucía Trilla-Fuertes¹✉, Angelo Gámez-Pozo^{1,2}✉, Jorge M. Arevalillo¹, Rocío López-Vacas², Elena López-Camacho¹, Guillermo Prado-Vázquez^{1,2}, Andrea Zapater-Moros^{1,2}, Mariana Díaz-Almirón⁴, María Ferrer-Gómez², Hilario Navarro³, Paolo Nanni⁵, Pilar Zamora⁶, Enrique Espinosa^{6,7}, Paloma Maín⁸, Juan Ángel Fresno Vara^{1,2,7*}

1 Biomedica Molecular Medicine SL, Madrid, Spain, **2** Molecular Oncology & Pathology Lab, Institute of Medical and Molecular Genetics-INGEMM, La Paz University Hospital-IdiPAZ, Madrid, Spain, **3** Operational Research and Numerical Analysis, National Distance Education University (UNED), Madrid, Spain, **4** Biostatistics Unit, La Paz University Hospital-IdiPAZ, Madrid, Spain, **5** Functional Genomics Centre Zurich, University of Zurich/ETH Zurich, Zurich, Switzerland, **6** Medical Oncology Service, La Paz University Hospital-IdiPAZ, Madrid, Spain, **7** Biomedical Research Networking Center on Oncology-CIBERONC, ISCIII, Madrid, Spain, **8** Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Madrid, Spain

✉ These authors contributed equally to this work.

* juanangel.fresno@salud.madrid.org



OPEN ACCESS

Citation: Trilla-Fuertes L, Gámez-Pozo A, Arevalillo JM, López-Vacas R, López-Camacho E, Prado-Vázquez G, et al. (2020) Bayesian networks established functional differences between breast cancer subtypes. PLoS ONE 15(6): e0234752. <https://doi.org/10.1371/journal.pone.0234752>

Editor: Elda Tagliabue, Fondazione IRCCS Istituto Nazionale dei Tumori, ITALY

Received: March 23, 2020

Accepted: June 1, 2020

Published: June 11, 2020

Copyright: © 2020 Trilla-Fuertes et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All the mass spectrometry raw data files acquired in this study may be downloaded from Chorus (<http://chorusproject.org>) under the project name Breast Cancer Proteomics.

Funding: This study was supported by Instituto de Salud Carlos III, Spanish Economy and Competitiveness Ministry, Spain and co-funded by the FEDER program, “Una forma de hacer Europa” (PI07/1302). LT-F is supported by the Spanish Economy and Competitiveness Ministry (DI-15-

Abstract

Breast cancer is a heterogeneous disease. In clinical practice, tumors are classified as hormonal receptor positive, Her2 positive and triple negative tumors. In previous works, our group defined a new hormonal receptor positive subgroup, the TN-like subtype, which had a prognosis and a molecular profile more similar to triple negative tumors. In this study, proteomics and Bayesian networks were used to characterize protein relationships in 96 breast tumor samples. Components obtained by these methods had a clear functional structure. The analysis of these components suggested differences in processes such as mitochondrial function or extracellular matrix between breast cancer subtypes, including our new defined subtype TN-like. In addition, one of the components, mainly related with extracellular matrix processes, had prognostic value in this cohort. Functional approaches allow to build hypotheses about regulatory mechanisms and to establish new relationships among proteins in the breast cancer context.

Introduction

Breast cancer is one of the most prevalent cancers in the world [1]. In clinical practice, breast cancer is classified according to the expression of hormonal receptors (estrogen or progesterone) and Her2, into positive hormonal receptor (ER+), HER2+ and triple negative (TNBC). In previous studies, our group defined a new ER+ molecular subgroup, named TN-like, with a molecular profile and a prognosis more similar to TNBC tumors [2]. We denoted the remaining ER+ tumors as ER-true. We also found significant molecular differences among breast cancer subtypes. For instance, differences related with metabolism of glucose were described between ER-true, TN-like and TNBC tumors [2, 3].

07614). AZ-M is supported by Consejería de Educación e Investigación de la Comunidad de Madrid (IND2018/BMD-9262). GP-V is supported by Consejería de Educación, Juventud y Deporte of Comunidad de Madrid (IND2017/BMD7783). EL-C is supported by the Spanish Economy and Competitiveness Ministry (PTQ2018-009760). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests: JAFV, EE and AG-P are shareholders in Biomedica Molecular Medicine SL, and LT-F, EL-C, AZ-M, and GP-V are currently employees of the company. The other authors declare that they have no competing interests. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

Proteomics supplies complementary information to genomics experiments. Proteomics has been used to find differences between subtypes on a proteomic level in sporadic and hereditary breast tumors [4]. In addition, proteomics coupled with super-SILAC has been also used to define molecular signatures that are differentially expressed between breast cancer subtypes [5].

Proteomics provides useful information about biological process effectors and may quantify thousands of proteins. Undirected probabilistic graphical models (PGM), based on a Bayesian approach, allow characterizing differences between tumor samples at functional level [2, 3, 6, 7]. In this study we explored the utility of Bayesian networks in the molecular characterization of breast cancer. The main feature of targeted Bayesian networks is that they provide a hierarchical structure and targeted relationships between proteins.

Bayesian networks (BN) have been previously used to inference protein signaling networks using phase-reverse protein array data from a breast cancer cell line [8]. In this study, the authors also experimentally tested some of the relationships by an inhibition approach. Balandayuthapani et al. also applied BN to phase-reverse protein array data, in this case from a panel of ovarian and breast cancer cell lines. Their model was capable to distinguish between both cell line types [9]. Most recently, BN has been used to determine genes related to bone metastasis development in breast cancer [10]. Also related to gene expression, BN inference leaded to the identification of TRIB1 as a regulator of cell cycle progression and survival in triple negative cancer cells [11]. BN have been also used to suggest therapeutic targets in breast cancer. In the study of Vundavilli et al., applying BN to gene expression data, the resulting network was used to rank different interventions in order to achieve an apoptosis induction [12]. Beretta et al. used BN to study the inference of signaling downstream of tyrosine kinase receptors, comparing predictions about inhibition of several nodes with experimental data [13]. BN was applied even to rank treatments in triple negative breast cancer datasets [14]. Finally, BN have demonstrated its utility in making associations between clinical data in breast cancer patients [15] or between lifestyle factors in breast cancer survivors [16].

In this work, we aim to explore if BN can be applied to proteomics expression data and if that the results provided by these analyses provide useful biological and clinical information. For this, we used mass-spectrometry proteomics data and Bayesian networks to characterize protein relationships in a cohort of breast cancer tumor paraffin samples. These networks maintained a functional structure and some of them showed prognostic value. This approach also reflected previously described protein-protein interactions and it could be used to propose new hypotheses and mechanisms of regulation of these proteins.

Materials and methods

Ethics statement

Written informed consent had been obtained for the participants on the study. The approval of the study was obtained from Hospital Doce de Octubre and Hospital Universitario La Paz Ethics Committees.

Samples

One hundred and six FFPE samples from patients with breast cancer were recovered from I +12 Biobank and from IdiPAZ Biobank, both integrated in the Spanish Hospital Biobank Network. The histopathological characteristics were reviewed by a pathologist to confirm tumor content. Samples had to include no less than half of tumor cells. The endorsement of the study was obtained by Hospital Doce de Octubre and Hospital Universitario La Paz Ethics Committees. These samples were utilized in previous studies [2, 3, 17].

Protein preparation

Proteins were extracted from formalin-fixed paraffin-embedded (FFPE) samples as previously described [18]. Briefly, FFPE sections were deparaffinized in xylene and washed twice with absolute ethanol. Protein extracts from FFPE samples were set up in 2% SDS buffer using a protocol based on heat-induced antigen retrieval. Protein concentration was quantified using the MicroBCA Protein Assay Kit (Pierce-Thermo Scientific). Protein extracts (10 µg) were processed with trypsin (1:50) and SDS was removed from digested lysates using Detergent Removal Spin Columns (Pierce). Peptide samples were additionally desalted using ZipTips (Millipore), dried, and resolubilized in 15 µL of a 0.1% formic acid and 3% acetonitrile solution before mass-spectrometry (MS) experiments.

Label-free proteomics

Samples were analyzed on a LTQ-Orbitrap Velos hybrid mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to NanoLC-Ultra system (Eksigent Technologies, Dublin, CA, USA) as described previously [2, 3]. Briefly, after separation, peptides were eluted with a gradient of 5 to 30% acetonitrile in 95 minutes. The mass spectrometer was operated in data-dependent mode (DDA), followed by CID (collision-induced dissociation) fragmentation on the twenty most intense signals per cycle. The acquired raw MS data were processed by MaxQuant (version 1.2.7.4) [19], followed by protein identification using the integrated Andromeda search engine [20]. Briefly, spectra were searched against a forward UniProtKB/Swiss-Prot database for human, concatenated to a reversed decoyed fasta database (NCBI taxonomy ID 9606, release date 2011-12-13). The maximum false discovery rate (FDR) was set to 0.01 for peptides and 0.05 for proteins. Label free quantification was calculated on the basis of the normalized intensities (LFQ intensity). Quantifiable proteins were defined as those detected in at least 75% of samples in at least one type of sample (either ER+ or TNBC samples) showing two or more unique peptides. Only quantifiable proteins were considered for subsequent analyses. Protein expression data were log2 transformed and missing values were replaced using data imputation for label-free data, as explained in [21], using default values. Finally, protein expression values were z-score transformed. All the mass spectrometry raw data files acquired in this study may be downloaded from Chorus (<http://chorusproject.org>) under the project name Breast Cancer Proteomics.

Network construction

PGM are graph-based representations of joint probability distributions where nodes represent random variables and edges (directed or undirected) represent stochastic dependencies among the variables. In particular, we have used a type of PGM called Bayesian networks (BN) [13]. With these models, the dependences between the variables in our data are specified by a directed acyclic graph (DAG). The obtained networks will indicate causality i.e. if protein A and B are connected and protein A changes its expression value, protein B changes its expression value as well [22].

Firstly, we find the BN that best explains our data [23]. There are different algorithms to learn a DAG from data but we have selected the well-known PC algorithm (named as its inventors Peter Spirtes and Clark Glymour), a constraint-based structure learning algorithm [24] based on conditional independence tests. The PC algorithm was shown to be consistent in high-dimensional settings [25]. Moreover, an order-independent version of the PC algorithm, called PC-stable, was proposed in [26]. All these procedures are implemented in R within packages pcalg [25] and graph [27]. We used protein expression data without other a priori information.

In this way, our data are represented by a large graph that can be partitioned into several connected components. Then, we focused on finding suitable subgraphs that give us a much clearer understanding of the interrelations therein.

STRING v11 (<https://string-db.org/>) was used to check if some of the protein relations obtained in the DAG analysis were previously described.

Gene ontology analyses

Protein to Gene Symbol conversion was performed using Uniprot (www.uniprot.org) and DAVID (www.david.ncifcrf.gov) [28]. Gene Ontology Analysis was also done in DAVID selecting only “Homo sapiens” background and GOTERM-FAT (<http://geneontology.org/>), Biocarta (<http://doi.org/10.1089/152791601750294344>) and KEGG (<https://www.genome.jp/kegg/kegg1.html>) databases.

Component activity measurements

Component activities were calculated as previously described [2, 3]. Briefly, activity measurement was calculated by the mean expression of all the proteins of each component related with the established major component function.

Statistical analyses

Network visualization was performed using Cytoscape software [29]. Statistical comparison between tumor groups were done in GraphPad Prism v6 using a non-parametric Mann-Whitney test. Prognostic signatures were developed using R v3.2.4 and BRB Array Tools, developed by Dr. Richard Simon and BRB Array Tools Development Team [30]. Briefly, functional node activities were ranked according their p-values in a Kaplan-Meier analysis. Then, a Cox regression including a leave-one-out validation using 1,000 random permutations was used to validate the prognostic capability. P-values under 0.05 were considered statistically significant.

Results

Patient characteristics

Clinical characteristics of this patient cohort have been previously described [2, 3, 31]. Briefly, one hundred and six patients were enrolled into the study. They all had node positive disease, Her2 negative and all had received adjuvant chemotherapy and hormonal therapy in the case of ER+ tumors. Among ER+ tumors, 50 patients were characterized as ER-true and 21 were defined as TN-like (S1 Table) [2].

Mass spectrometry analysis

Proteomics analyses from these samples have been previously described [2]. In summary, one hundred and two FFPE samples had enough protein to perform the MS analyses. After MS workflow, 96 samples provided useful protein expression data. After quality criteria, 1,095 proteins presented at least two unique peptides and detectable expression in at least 75% of the samples in at least one type of sample (either ER+ or TNBC).

Directed networks

Using proteomics data, directed acyclic graphs (DAG) were performed. Altogether, it was possible to establish 789 edges of which 662 were guided and 127 are undetermined. These edges

Table 1. Characteristics of the components obtained from DAG.

Number of nodes	1	2	3	4	5	6	8	9	10	11	12	13	15	17	18	23	464
Number of components	188	62	13	18	6	4	2	1	1	1	1	1	1	1	1	1	1

Number of nodes = number of proteins contained in each component, Number of components = directed components obtained.

<https://doi.org/10.1371/journal.pone.0234752.t001>

formed 303 components formed by different number of nodes or proteins. An overview of the number of nodes (proteins) included in each component is provided in [Table 1](#).

We characterized components from DAG analysis. Components including less than 9 nodes were dismissed because they were little informative. All components were named with the number of nodes included by the DAG analysis.

Afterwards, components were interrogated for biological function. Characteristics about all components are supplied in [Table 2](#) and [S1 File](#).

Component activity measurements

Component activities were calculated for each node. There were significant differences between ER-true, TN-like and TNBC tumors in the component activity for component 23: mitochondria, component 17: RNA binding, component 13: extracellular matrix, and component 10: extracellular exosome ([Fig 1](#)).

Component 13: Extracellular matrix

Component 13 activity showed prognosis value in our series, splitting our population into a high and a low risk group and it can be used as a distant-metastasis free survival (DMFS) predictor ($p = 0.045$, HR = 0.35, 30–70%) ([Fig 2](#)). Interestingly, the predictor classified all the TN-like tumors and most of TNBC tumors into the high-risk group ([Table 3](#)).

Component 13 contains thirteen proteins mainly related to extracellular matrix ([Fig 3](#)). The five proteins related by gene ontology analysis to extracellular matrix were OGN, BGN, LUM, CMA1, and DCN, all of them, with the exception of CMA1, belonged to small leucine-rich proteoglycan family of proteins.

Additionally, the existence of described connections was checked using STRING v11 ([Fig 4](#)). Interestingly, some of the connections suggested by BN analysis in Component 13, (NDUFB2 and NDUFB4; CPA3 and CMA1; and BGN, LUM, and DCN) were connected both in the DAG graph and in the STRING network, meaning that these interactions have been previously described.

Table 2. Features of components obtained by DAG analysis.

Component	Number of nodes	Main function
Component 23	23	Membrane and mitochondria
Component 18	18	Cytoskeleton
Component 17	17	RNA binding
Component 15	15	Extracellular exosome
Component 13	13	Extracellular matrix
Component 12	12	RNA binding and translation
Component 11	11	Proliferation and ophoxos
Component 10	10	Extracellular exosome
Component 9	9	RNA binding and splicing

<https://doi.org/10.1371/journal.pone.0234752.t002>

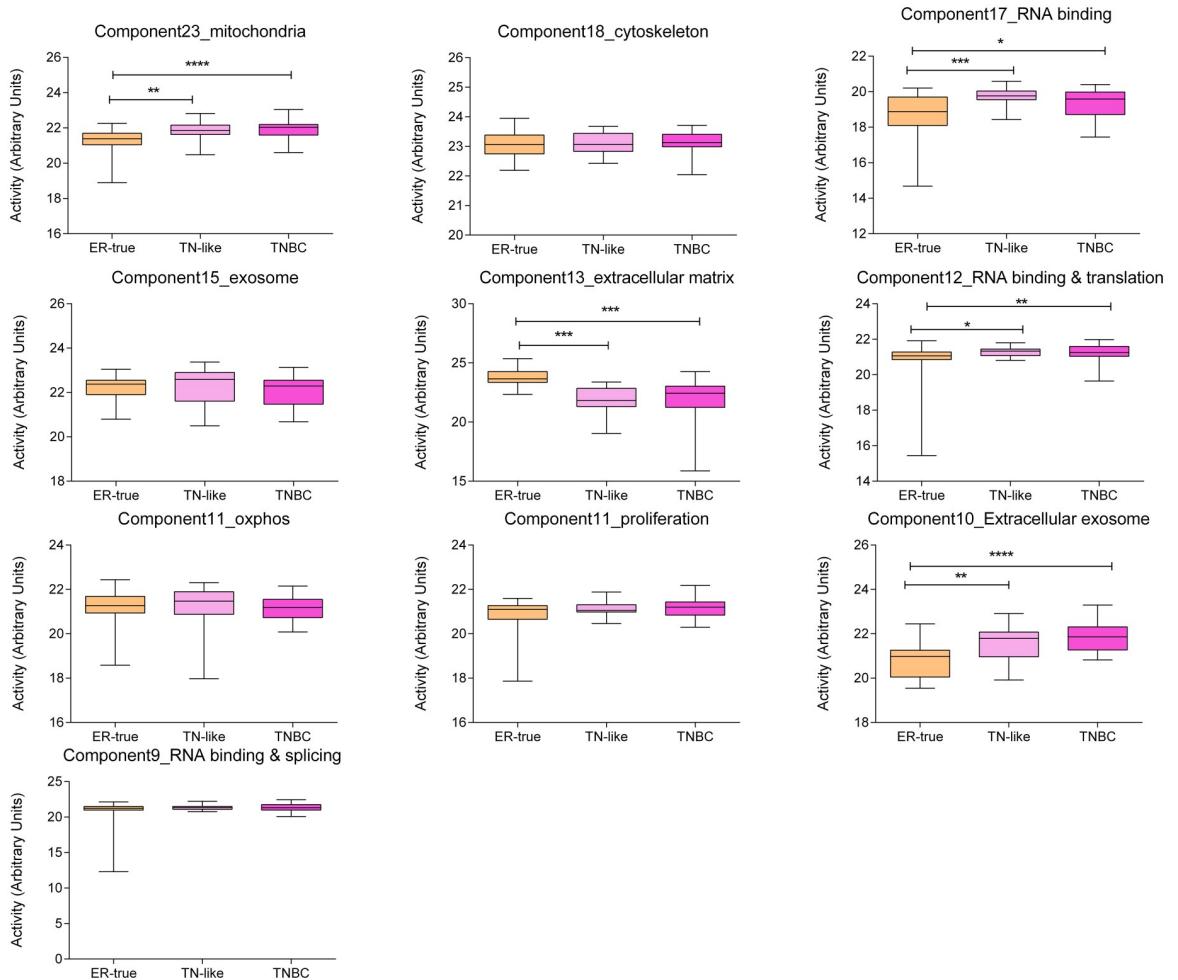


Fig 1. Component activity measurements for ER-true, TN-like and TNBC respectively.

<https://doi.org/10.1371/journal.pone.0234752.g001>

Discussion

In this study, we used proteomics and DAG to characterize relationships between proteins in breast cancer tumor samples. Unlike other approaches, such as Genes2FANS [32], our DAG method supplies directed relationships between proteins and a hierarchical structure. Traditionally, protein-protein interaction (PPI) networks, such as STRING, are based in relationships described in the literature. However, we built a directed network, i.e. a graph formed by edges with a direction, using protein expression data without other a priori information, so it was possible to propose new hypotheses about protein interactions. We used probabilistic graphical models (PGM) because they offer a way to relate many random variables with a complex dependency structure.

As it has been previously mentioned, arrows in directed networks indicate causality between two proteins. This approach allows making hypotheses about causal relationships between proteins and proposes a hierarchical structure. In some cases, an experimental relationship between two proteins connected in the directed network had been previously described. For instance, in component 18, it has been widely described that PIP binds AZGP1

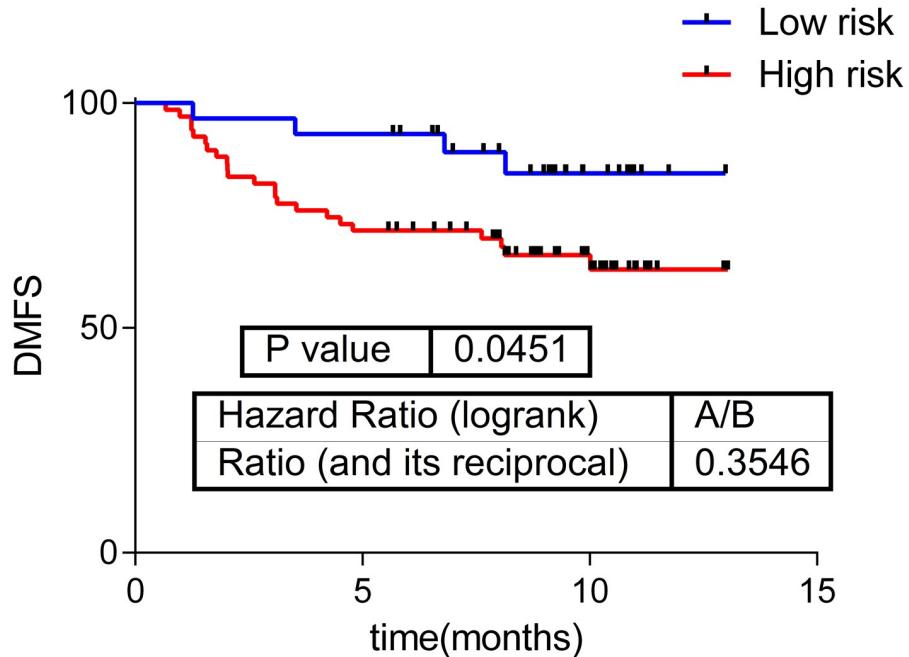


Fig 2. Component 13 activity prognostic value in the whole cohort.

<https://doi.org/10.1371/journal.pone.0234752.g002>

in breast cancer [33]. Another example is component 11 which related COX5A, MT-CO2 and COX6C, all proteins of mitochondrial complex IV [34].

We demonstrated in previous works that non-directed graphs provided functional information [2, 3]. Interestingly, a functional structure also appeared in this type of networks. Component activities suggested differences in functions such as extracellular matrix and mitochondria. Differences in mitochondria between these subtypes have been previously described using non-directed PGMs [2, 3].

On the other hand, component 13, composed by thirteen proteins, five of them related to extracellular matrix, had prognostic value in our series. Of these five proteins, four of them belonged to the small leucine-rich proteoglycan (SLRP) family (lumican, biglycan, osteoglycin, and decorin). Biglycan (BGN) could promote migration in breast cancer [35]. It has been widely described the anti-metastatic role of decorin (DCN1) in breast cancer [36–38]. Lumican (LUM) significantly attenuated cell functional processes, including proliferation, migration and invasion [39]. In the same study, it was described that lumican modulates matrix effectors in MCF7 and MDAMB231 cells. Finally, osteoglycin (OGN) has been suggested as a biomarker of ECM in TNBC [40]. On the other hand, chymase 1 (CMA1) is secreted by mast cells and may play a role in angiogenesis [41]. It is also involved in extracellular matrix degradation. Higher levels of this protein were observed in Luminal subtype [42]. DCN, OGN, BGN, and

Table 3. Number of patients classified by the DMFS predictor into a low or a high-risk group.

Subtype	Number in low-risk group	Number in high-risk group
ER-true	26	24
TN-like	0	21
TNBC	3	22

<https://doi.org/10.1371/journal.pone.0234752.t003>

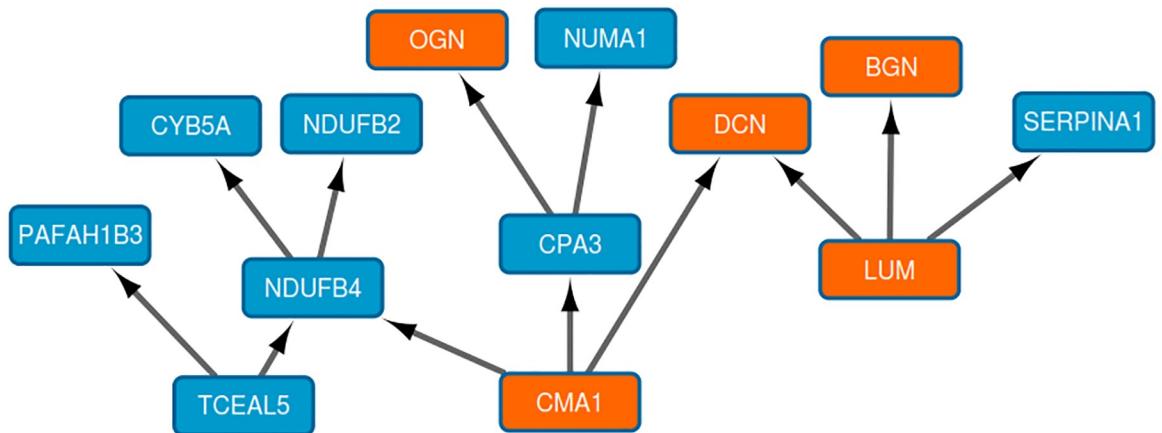


Fig 3. Component 13. Orange nodes: Proteins related to extracellular matrix ontology.

<https://doi.org/10.1371/journal.pone.0234752.g003>

LUM appeared also interconnected in the STRING network, so the DAG graph reflected protein interactions previously described. NDUFB2 and NDUFB4 were also connected in both networks.

Other proteins included in this component related to cancer are NUMA1, SERPINA1, and PAFAH1B3. Nuclear mitotic apparatus (NUMA1) is a structural component of the nuclear matrix. The encoded protein interacts with microtubules and plays a role in formation and organization of the mitotic spindle during cell division. It also modulates p-53 mediated transcription in cancer cells [43]. Serpin family A member 1 (SERPINA1) is a direct estrogen

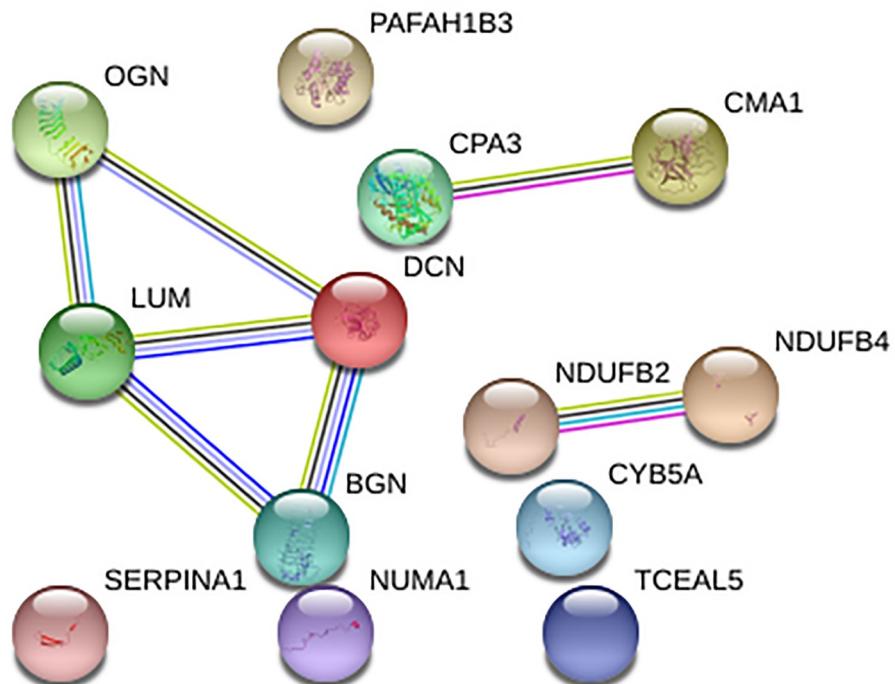


Fig 4. Network built with the proteins from Component13 using STRING.

<https://doi.org/10.1371/journal.pone.0234752.g004>

receptor target and a predictor of survival in breast cancer patients [44]. Platelet activating factor acetylhydrolase 1b catalytic subunit 3 (PAFAH1B3) encodes an acetylhydrolase that catalyzes the removal of an acetyl group from the glycerol backbone of platelet-activating factor. A study identified PAFAH1B3 as a key metabolic driver of breast cancer pathogenicity that is upregulated in primary human breast tumors and correlated with poor prognosis [45]. This enzyme may be dysregulated across many cancer types [46].

In previous studies we have used functional node activities from non-directed network to develop prognostic predictors [7]. Now, this approach is also validated in directed networks.

We used a mathematical method as DAG analysis and applied them to proteomics data of breast cancer tumors in order to infer causal relationships between these proteins. This method supplied some known relationships but also proposed new ones. Additionally, it associated proteins with a similar function. Therefore, it seems that it is a good approach to propose new hypotheses about mechanisms of action. Moreover, it was possible to associate the results obtained by DAG analysis with prognosis and built a prognostic signature. As far we know, this is the first time that this type of analysis is applied to clinical data and is associated with clinical outcome.

Our study has some limitations. Proteomics provides complementary information to other techniques such as genomics. However, an improvement in the number of detected proteins is still necessary. On the other hand, breast cancer clinical scenario is far more complex, and stratified analyses (by molecular or clinical subtypes, for example) could provide more complex and insightful information. Finally, all these mathematical approaches (and others), despite being useful by themselves, should be combined to obtain more information about the clinical scenario analyzed, as long as it seems that different analyses provide different and complementary information from the same data.

To sum up, in this study, we used proteomics and directed networks to characterize relationships between proteins in breast cancer tumors. This approach reflected some previously described interactions and it could be used to propose new hypotheses and mechanisms of action.

Supporting information

S1 File. Components obtained by DAG analysis.
(DOCX)

S1 Table. Patient characteristics.
(DOCX)

Author Contributions

Conceptualization: Lucía Trilla-Fuertes, Angelo Gámez-Pozo, Pilar Zamora, Enrique Espinosa, Paloma Maín, Juan Ángel Fresno Vara.

Data curation: Pilar Zamora, Enrique Espinosa.

Formal analysis: Lucía Trilla-Fuertes, Angelo Gámez-Pozo, Jorge M. Arevalillo, Elena López-Camacho, Guillermo Prado-Vázquez, Andrea Zapater-Moros, Mariana Díaz-Almirón, María Ferrer-Gómez, Hilario Navarro, Paloma Maín.

Methodology: Rocío López-Vacas, Paolo Nanni.

Supervision: Enrique Espinosa, Juan Ángel Fresno Vara.

Writing – original draft: Lucía Trilla-Fuertes.

Writing – review & editing: Angelo Gámez-Pozo, Enrique Espinosa, Paloma Maín, Juan Ángel Fresno Vara.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015; 136(5): E359–86. Epub 2014/10/09. <https://doi.org/10.1002/ijc.29210> PMID: 25220842.
2. Gámez-Pozo A, Trilla-Fuertes L, Berges-Soria J, Selevsek N, López-Vacas R, Díaz-Almirón M, et al. Functional proteomics outlines the complexity of breast cancer molecular subtypes. *Scientific Reports*. 2017; 7(1):10100. <https://doi.org/10.1038/s41598-017-10493-w> PMID: 28855612
3. Gámez-Pozo A, Berges-Soria J, Arevalillo JM, Nanni P, López-Vacas R, Navarro H, et al. Combined label-free quantitative proteomics and microRNA expression analysis of breast cancer unravel molecular differences with clinical implications. *Cancer Res*; 2015. p. 2243–53. <https://doi.org/10.1158/0008-5472.CAN-14-1937> PMID: 25883093
4. Waldemarson S, Kurbasic E, Krogh M, Cifani P, Berggård T, Borg Å, et al. Proteomic analysis of breast tumors confirms the mRNA intrinsic molecular subtypes using different classifiers: a large-scale analysis of fresh frozen tissue samples. *Breast Cancer Res*. 2016; 18(1):69. Epub 2016/06/29. <https://doi.org/10.1186/s13058-016-0732-2> PMID: 27357824.
5. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 2016; 13(9):731–40. Epub 2016/06/27. <https://doi.org/10.1038/nmeth.3901> PMID: 27348712.
6. Trilla-Fuertes L, Gámez-Pozo A, Arevalillo JM, Díaz-Almirón M, Prado-Vázquez G, Zapater-Moros A, et al. Molecular characterization of breast cancer cell response to metabolic drugs. *Oncotarget*. 2018; 9 (11):9645–60. Epub 2018/01/08. <https://doi.org/10.18632/oncotarget.24047> PMID: 29515760.
7. de Velasco G, Trilla-Fuertes L, Gamez-Pozo A, Urbanowicz M, Ruiz-Ares G, Sepúlveda JM, et al. Urothelial cancer proteomics provides both prognostic and functional information. *Sci Rep*. 2017; 7 (1):15819. Epub 2017/11/17. <https://doi.org/10.1038/s41598-017-15920-6> PMID: 29150671.
8. Hill SM, Lu Y, Molina J, Heiser LM, Spellman PT, Speed TP, et al. Bayesian inference of signaling network topology in a cancer cell line. *Bioinformatics*. 2012; 28(21):2804–10. Epub 2012/08/24. <https://doi.org/10.1093/bioinformatics/bts514> PMID: 22923301.
9. Baladandayuthapani V, Talluri R, Ji Y, Coombes KR, Lu Y, Hennessy BT, et al. BAYESIAN SPARSE GRAPHICAL MODELS FOR CLASSIFICATION WITH APPLICATION TO PROTEIN EXPRESSION DATA. *Ann Appl Stat*. 2014; 8(3):1443–68. <https://doi.org/10.1214/14-AOAS722> PMID: 26246866.
10. Park SB, Chung CK, Gonzalez E, Yoo C. Causal Inference Network of Genes Related with Bone Metastasis of Breast Cancer and Osteoblasts Using Causal Bayesian Networks. *J Bone Metab*. 2018; 25 (4):251–66. Epub 2018/11/30. <https://doi.org/10.11005/jbm.2018.25.4.251> PMID: 30574470.
11. Gendelman R, Xing H, Mirzoeva OK, Sarde P, Curtis C, Feiler HS, et al. Bayesian Network Inference Modeling Identifies TRIB1 as a Novel Regulator of Cell-Cycle Progression and Survival in Cancer Cells. *Cancer Res*. 2017; 77(7):1575–85. Epub 2017/01/13. <https://doi.org/10.1158/0008-5472.CAN-16-0512> PMID: 28087598.
12. Vundavilli H, Datta A, Sima C, Hua J, Lopes R, Bittner M. Bayesian Inference Identifies Combination Therapeutic Targets in Breast Cancer. *IEEE Trans Biomed Eng*. 2019; 66(9):2684–92. Epub 2019/01/23. <https://doi.org/10.1109/TBME.2019.2894980> PMID: 30676941.
13. Beretta S, Castelli M, Gonçalves I, Henriques R, Ramazotti D. Learning the structure of Bayesian networks: A quantitative assessment of the effect of different algorithmic schemes Hindawi 2018.
14. Chen H, Lu W, Zhang Y, Zhu X, Zhou J, Chen Y. A Bayesian network meta-analysis of the efficacy of targeted therapies and chemotherapy for treatment of triple-negative breast cancer. *Cancer Med*. 2019; 8(1):383–99. Epub 2018/12/07. <https://doi.org/10.1002/cam4.1892> PMID: 30525293.
15. Soto-Ferrari M, Prieto D, Munene G. A Bayesian network and heuristic approach for systematic characterization of radiotherapy receipt after breast-conservation surgery. *BMC Med Inform Decis Mak*. 2017; 17(1):93. Epub 2017/06/28. <https://doi.org/10.1186/s12911-017-0479-4> PMID: 28659177.
16. Xu S, Thompson W, Kerr J, Godbole S, Sears DD, Patterson R, et al. Modeling interrelationships between health behaviors in overweight breast cancer survivors: Applying Bayesian networks. *PLoS One*. 2018; 13(9):e0202923. Epub 2018/09/04. <https://doi.org/10.1371/journal.pone.0202923> PMID: 30180192.
17. Gámez-Pozo A, Trilla-Fuertes L, Prado-Vázquez G, Chiva C, López-Vacas R, Nanni P, et al. Prediction of adjuvant chemotherapy response in triple negative breast cancer with discovery and targeted

- proteomics. *PLoS One.* 2017; 12(6):e0178296. Epub 2017/06/08. <https://doi.org/10.1371/journal.pone.0178296> PMID: 28594844.
- 18. Gámez-Pozo A, Ferrer NI, Ciruelos E, López-Vacas R, Martínez FG, Espinosa E, et al. Shotgun proteomics of archival triple-negative breast cancer samples. *Proteomics Clin Appl.* 2013; 7(3–4):283–91. <https://doi.org/10.1002/prca.201200048> PMID: 23436753.
 - 19. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol.* 2008; 26(12):1367–72. Epub 2008/11/30. <https://doi.org/10.1038/nbt.1511> PMID: 19029910.
 - 20. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res.* 2011; 10(4):1794–805. Epub 2011/02/22. <https://doi.org/10.1021/pr101065j> PMID: 21254760.
 - 21. Deeb SJ, D'Souza RC, Cox J, Schmidt-Suprian M, Mann M. Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles. *Mol Cell Proteomics.* 2012; 11(5):77–89. Epub 2012/03/21. <https://doi.org/10.1074/mcp.M111.015362> PMID: 22442255.
 - 22. Needham CJ, Bradford JR, Bulpitt AJ, Westhead DR. A primer on learning in Bayesian networks for computational biology. *PLoS Comput Biol.* 2007; 3(8):e129. <https://doi.org/10.1371/journal.pcbi.0030129> PMID: 17784779.
 - 23. Neapolitan R, Xue D, Jiang X. Modeling the altered expression levels of genes on signaling pathways in tumors as causal bayesian networks. *Cancer Inform.* 2014; 13:77–84. Epub 2014/05/25. <https://doi.org/10.4137/CIN.S13578> PMID: 24932098.
 - 24. Spirtes P, Glymour C, Scheines R. *Causation, Prediction, and Search.* Adaptive Computation and Machine Learning. 2nd ed. The MIT Press 2000.
 - 25. Kalisch M, Maechler M, Colombo D, Maathuis MH, Bluehlmann P. *Causal Inference Using Graphical Models with the R Package pcalg.* *Journal of Statistical Software* 2012. p. 1–26.
 - 26. Colombo D, Maathuis M. Order-independent constraint-based causal structure. *arXiv:1211.3295v2;* 2013.
 - 27. Gentleman R, Whalen E, Huber W, Falcon S. *graph:* A package to handle graph data structures. R package version 1.54.0.
 - 28. Huang dW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009; 4(1):44–57. <https://doi.org/10.1038/nprot.2008.211> PMID: 19131956.
 - 29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003; 13(11):2498–504. <https://doi.org/10.1101/gr.1239303> PMID: 14597658.
 - 30. Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol.* 2005; 23(29):7332–41. Epub 2005/09/06. <https://doi.org/10.1200/JCO.2005.02.8712> PMID: 16145063.
 - 31. Trilla-Fuertes L, Gámez-Pozo A, Díaz-Almirón M, Prado-Vázquez G, Zapater-Moros A, López-Vacas R, et al. Computational metabolism predicts risk of distant relapse-free survival in breast cancer patients. *Future Oncology.* 2019; 30:3483–90. <https://doi.org/10.2217/fon-2018-0698> PMID: 31580166.
 - 32. Dannenfelser R, Clark NR, Ma'ayan A. Genes2FANS: connecting genes through functional association networks. *BMC Bioinformatics.* 2012; 13:156. <https://doi.org/10.1186/1471-2105-13-156> PMID: 22748121.
 - 33. Debily MA, Marhomy SE, Boulanger V, Eveno E, Mariage-Samson R, Camarca A, et al. A functional and regulatory network associated with PIP expression in human breast cancer. *PLoS One.* 2009; 4(3):e4696. Epub 2009/03/05. <https://doi.org/10.1371/journal.pone.0004696> PMID: 19262752.
 - 34. Lazarou M, Smith SM, Thorburn DR, Ryan MT, McKenzie M. Assembly of nuclear DNA-encoded subunits into mitochondrial complex IV, and their preferential integration into supercomplex forms in patient mitochondria. *FEBS J.* 2009; 276(22):6701–13. Epub 2009/10/16. <https://doi.org/10.1111/j.1742-4658.2009.07384.x> PMID: 19843159.
 - 35. Chen A, Wang L, Liu S, Wang Y, Liu Y, Wang M, et al. Attraction and Compaction of Migratory Breast Cancer Cells by Bone Matrix Proteins through Tumor-Osteocyte Interactions. *Sci Rep.* 2018; 8(1):5420. Epub 2018/04/03. <https://doi.org/10.1038/s41598-018-23833-1> PMID: 29615735.
 - 36. Reed CC, Waterhouse A, Kirby S, Kay P, Owens RT, McQuillan DJ, et al. Decorin prevents metastatic spreading of breast cancer. *Oncogene.* 2005; 24(6):1104–10. <https://doi.org/10.1038/sj.onc.1208329> PMID: 15690056.
 - 37. Oda G, Sato T, Ishikawa T, Kawachi H, Nakagawa T, Kuwayama T, et al. Significance of stromal decorin expression during the progression of breast cancer. *Oncol Rep.* 2012; 28(6):2003–8. Epub 2012/09/17. <https://doi.org/10.3892/or.2012.2040> PMID: 22992875.

38. Goldoni S, Seidler DG, Heath J, Fassan M, Baffa R, Thakur ML, et al. An antimetastatic role for decorin in breast cancer. *Am J Pathol.* 2008; 173(3):844–55. Epub 2008/08/07. <https://doi.org/10.2353/ajpath.2008.080275> PMID: 18688028.
39. Karamanou K, Franchi M, Piperigkou Z, Perreau C, Maquart FX, Vynios DH, et al. Lumican effectively regulates the estrogen receptors-associated functional properties of breast cancer cells, expression of matrix effectors and epithelial-to-mesenchymal transition. *Sci Rep.* 2017; 7:45138. Epub 2017/03/23. <https://doi.org/10.1038/srep45138> PMID: 28332606.
40. Moriggi M, Giussani M, Torretta E, Capitanio D, Sandri M, Leone R, et al. ECM Remodeling in Breast Cancer with Different Grade: Contribution of 2D-DIGE Proteomics. *Proteomics.* 2018; 18(24):e1800278. <https://doi.org/10.1002/pmic.201800278> PMID: 30353998.
41. de Souza DA Junior, Santana AC, da Silva EZ, Oliver C, Jamur MC. The Role of Mast Cell Specific Chymases and Tryptases in Tumor Angiogenesis. *Biomed Res Int.* 2015; 2015:142359. Epub 2015/06/04. <https://doi.org/10.1155/2015/142359> PMID: 26146612.
42. Glajcar A, Szpor J, Pacek A, Tyrak KE, Chan F, Streb J, et al. The relationship between breast cancer molecular subtypes and mast cell populations in tumor microenvironment. *Virchows Arch.* 2017; 470(5):505–15. Epub 2017/03/18. <https://doi.org/10.1007/s00428-017-2103-5> PMID: 28315938.
43. Endo A, Moyori A, Kobayashi A, Wong RW. Nuclear mitotic apparatus protein, NuMA, modulates p53-mediated transcription in cancer cells. *Cell Death Dis.* 2013; 4:e713. Epub 2013/07/04. <https://doi.org/10.1038/cddis.2013.239> PMID: 23828576.
44. Chan HJ, Li H, Liu Z, Yuan YC, Mortimer J, Chen S. SERPINA1 is a direct estrogen receptor target gene and a predictor of survival in breast cancer patients. *Oncotarget.* 2015; 6(28):25815–27. <https://doi.org/10.18632/oncotarget.4441> PMID: 26158350.
45. Mulvihill MM, Benjamin DI, Ji X, Le Scolan E, Louie SM, Shieh A, et al. Metabolic profiling reveals PAFAH1B3 as a critical driver of breast cancer pathogenicity. *Chem Biol.* 2014; 21(7):831–40. Epub 2014/06/19. <https://doi.org/10.1016/j.chembiol.2014.05.008> PMID: 24954006.
46. Kohnz RA, Mulvihill MM, Chang JW, Hsu KL, Sorrentino A, Cravatt BF, et al. Activity-Based Protein Profiling of Oncogene-Driven Changes in Metabolism Reveals Broad Dysregulation of PAFAH1B2 and 1B3 in Cancer. *ACS Chem Biol.* 2015; 10(7):1624–30. Epub 2015/05/07. <https://doi.org/10.1021/acscchembio.5b00053> PMID: 25945974.