bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Probabilistic graphical models and computational metabolic models applied to the analysis of metabolomics data in breast cancer

Lucía Trilla-Fuertes[1¶]; Angelo Gámez-Pozo[1,2¶]; Jorge M Arevalillo[3]; Guillermo Prado-Vázquez[1]; Andrea Zapater-Moros[1,2]; Mariana Díaz-Almirón[4]; Hilario Navarro[3]; Paloma Maín[5]; Enrique Espinosa[6,7]; Pilar Zamora[6,7]; and Juan Ángel Fresno Vara[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]Department of Statistics, Operational Research and Numerical Analysis, National University of Distance Education (UNED).

[4]Biostatistics Unit, La Paz University Hospital-IdiPAZ, Madrid, Spain

[5]Department of Statistics and Operations Research, Faculty of Mathematics, Complutense University of Madrid, Madrid, Spain.

[6]Medical Oncology Service, La Paz University Hospital-IdiPAZ, Madrid, Spain

[7]Biomedical Research Networking Center on Oncology-CIBERONC, ISCIII, Madrid, Spain

¶ These authors contributed equally to this work.

[*] Corresponding Author: juanangel.fresno@salud.madrid.org (JAFV)

Short title: Computational models applied to metabolomics data in breast cancer

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Metabolomics has great potential in the development of new biomarkers in cancer. In this study, metabolomics and gene expression data from breast cancer tumor samples were analyzed, using (1) probabilistic graphical models to define associations using quantitative data without other *a priori* information; and (2) Flux Balance Analysis and flux activities to characterize differences in metabolic pathways. A metabolite network was built through the use of probabilistic graphical models. Interestingly, the metabolites were organized into metabolic pathways in this network, thus it was possible to establish differences between breast cancer subtypes at the metabolic pathway level. Additionally, the lipid metabolism node had prognostic value. A second network associating gene expression with metabolites was built. Associations were established between the biological functions of genes and the metabolites included in each node. A third network combined flux activities from Flux Balance Analysis and metabolomics data, showing coherence between the metabolic pathways of the flux activities and the metabolites in each branch. In this study, probabilistic graphical models were valuable for the functional analysis of metabolomics data from a functional point of view, allowing new hypotheses in metabolomics and associating metabolomics data with the patient's clinical outcome.

Author summary

Metabolomics is a promising technique to describe new biomarkers in cancer. In this study we proposed computational methods to manage this type of data and associate it with gene expression data. We also employed a metabolic computational model to compare predictions from this model with metabolomics measurements. Finally, we built predictors of relapse based on the integration of those high-dimensional data in breast cancer patients.

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Breast cancer is one of the most common malignancies, with 266,120 estimated new cases and 40,920 estimated deaths in the United States in 2018 [1]. In clinical practice, the expression of hormonal receptors and HER2 allows the classification of this disease into three groups: hormonal receptor-positive (ER+), HER2+ and triple negative (TNBC).

Metabolomics is the most recent -omics. It consists of measuring the entire set of metabolites present in a biological sample [2]. The most common techniques in metabolomics experiments are mass spectrometry-related methods, which are based on the mass/charge relationships of each metabolite or its fragments [3]. Metabolomics is a promising tool for the development of new biomarkers [4].

We used two different methods to merge metabolomics and gene expression data in breast cancer. In previous studies, we used probabilistic graphical models (PGMs) to study differences between breast tumor subtypes and to characterize muscle-invasive bladder cancer at a functional level using proteomics data [5-7]. Flux Balance Analysis (FBA), however, is a method that has been widely used to study biochemical networks [8]. FBA predicts the growth rate or the rate of production of a given metabolite [9], and it has previously been used to characterize breast cancer cell responses against drugs targeting metabolism [10]. In this study, flux activities were proposed as a feasible method to compare flux patterns in metabolic pathways.

In the present study, metabolomics and gene expression data from 67 fresh tissue samples [11] were analyzed through PGMs and FBA. Our aim was to find associations between metabolomics and gene expression data.

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

Patient characteristics

The data used in this study are from the previous work of Terunuma et al. [11]. A total of 67 paired normal and tumor fresh tissue samples from patients with breast cancer were studied. We only selected samples from tumor tissues for the present analyses.

This cohort included 67 patients, 33 ER+ and 34 ER- (of which 14 were TNBC). The median follow-up was 50 months, and 31 deaths had occurred during this time. No significant differences regarding overall survival were observed between patients with ER+ or ER- tumors. Patient characteristics are shown in Table 1.

	n (%)	ER+	ER-
Number of patients	67	33	34
Age (years)			
Median	51	57	48
Range	30–93	34–93	30–75
TNM stage			
I	6 (9%)	4 (12%)	2 (6%)
II	2 (3%)	1 (3%)	1 (3%)
IIA	23 (35%)	12 (37%)	11 (32%)
IIB	21 (31%)	7 (21%)	14 (41%)
IIIA	9 (13%)	5 (15%)	4 (12%)
IIIB	6 (9%)	4 (12%)	2 (6%)
N category			
pN0	37 (55%)	17 (52%)	20 (59%)
pN1	24 (35%)	13 (39%)	11 (32%)

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pN2	5 (8%)	3 (9%)	2 (6%)
Missing	1 (2%)	0 (0%)	1 (3%)
Grade			
G1	8 (12%)	8 (24%)	0 (0%)
G2	20 (30%)	14 (43%)	6 (18%)
G3	29 (43%)	7 (21%)	22 (64%)
Missing	10 (15%)	4 (12%)	6 (18%)
Neoadjuvant therapy			
Yes	6 (9%)	2	4 (12%)
No	50 (75%)	26	24 (70%)
Missing	11 (16%)	5	6 (18%)

Table 1: Patient characteristics.

Analysis of metabolomics data

An overall survival predictor using metabolomics data was built. This signature included five metabolites: glutamine, 2-hydroxypalmitate, deoxycarnitine, butyrylcarnitine and glycerophosphorylcholine (p-value =0.003, hazard ratio [HR] = 0.342, cut-off = 50:50) (Fig 1). A multivariate analysis showed that the predictor provided additional prognostic information to that of the clinical data (S1 Table).

Fig 1: Predictive signature built using metabolomics data.

Metabolomics data, including 237 metabolites, were analyzed through PGM. The resulting network was built assigning a main metabolic pathway to each node using IMPaLA. IMPaLA is a tool that allows ontology analyses based on metabolic pathways instead of genes. Strikingly, this network had a functional structure, grouping the metabolites into metabolic pathways. The network had five nodes, each with a different overrepresented metabolic pathway (Fig 2).

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Fig 2: Probabilistic graphical model from metabolomics data.

The activity of each node was calculated as previously described [6, 7, 10, 12]. Significant differences were found between ER+ and ER- tumors in lipid metabolism and purine metabolism ($p<0.05$) (S1 Fig).

The lipid metabolism node had prognostic value in this cohort ($p = 0.045$, HR = 0.479, cut-off = 50:50) (Fig 3). Differences remained when stratified by the expression of hormonal receptors. However, a multivariate analysis did not show that the predictor supplied additional prognostic information to that of the clinical data (S2 Table).

Fig 3: Predictor based on lipid metabolism node activity.

Analyses combining gene expression with metabolomics data

A network combining metabolomics and gene expression data was built. Although most metabolites were grouped in the same node, some were integrated into gene nodes (Fig 4).

Fig 4: A. Network associating genes (red) and metabolites (blue). B. Metabolite and gene network functionally characterized.

This combined network was then functionally characterized. The resulting network had eleven functional nodes and a twelfth node that grouped the metabolites (Fig 4).

Once the main functions were assigned, a literature review was performed to study the relationship between metabolites included in the gene nodes and the main function of each node. A relationship with functional nodes had been previously described for 4 of 20 metabolites: succinate, cytidine, histamine and 1,2-propanediol. The relationships between metabolites and their node function are shown in Table 2.

Metabolite	Node	Described relationship	Reference
Succinate	Immune	Increases immune response, induces IL-1b production,	PMID: 28109906

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

	response	promotes adaptive immune response.	
Cytidine	Immune response	5-aza-2'-deoxycytidine potentiates antitumor immune response, role in innate immune response	PMID: 23865062, PMID: 24559534
Histamine	Angiogenesis	Histamine promotes angiogenesis by enhancing VEGF production	PMID: 23225320
1,2-propanediol (prev.X-4796)	Angiogenesis	Modulates the immune system through S1P, which promotes angiogenesis and proliferation. 14C-sulfoquinovosyl acylpropanediol is an antiangiogenic drug	PMID: 21632869, PMID: 29543539

Table 2: Previously described relationships between metabolites included in gene nodes and the function of these nodes.

Flux Balance Analysis and flux activities

FBA and flux activities were calculated as previously described [10]. No significant differences were found in the tumor growth rate between ER+ and ER- tumors (S2 Fig).

Flux activities showed significant differences between ER+ and ER- in glycerophospholipid metabolism, phosphatidyl inositol metabolism, urea cycle, propanoate metabolism, pyrimidine catabolism and reactive oxygen species (ROS) detoxification (S3 Fig).

A predictor for overall survival was built with flux activities of glutamate metabolism and alanine and aspartate metabolism (p-value = 0.024, HR = 0.411, cut-off = 50:50) (Fig 5). A multivariate analysis showed that the predictor provided prognostic information independent from clinical data (S3 Table).

Fig 5: OS predictor based on glutamate metabolism and alanine and aspartate metabolism flux activities.

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

PGM analysis combining flux activities with metabolomics data

Using flux activities and metabolomics data, a new network was built. Interestingly, this network combined both types of data; however, flux activities appeared at the periphery of the network (Fig 6).

Fig 6: A. Network combining flux activities (purple) and metabolite (pink) expression data. B. Division in branches of the network formed by flux activities and metabolomics data.

The resulting network was split into several branches to study the relationship of the metabolites to the flux activities included in each branch (Fig 6). Coherence between both types of data was shown, associating flux activities and metabolites related to these flux activities in the same branch. For instance, branch 1 includes glycolysis flux activity and three metabolites previously related to glycolysis (S4 Table). Regarding vitamins and cofactors, it was not possible make comparisons because the IMPaLA label for this category is “Vitamin and co-factor metabolism” and Recon2 labels differentiate between the various vitamins, labeling them as “Vitamin B6 metabolism”, “Vitamin A metabolism”, etc.

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

Metabolomics is attracting considerable interest as a technique for finding new biomarkers in cancer. In this study, a new analytical workflow for the management and study of metabolomics data was proposed. This workflow allowed global metabolic characterization, beyond analyses based on unique metabolites.

Genomics and metabolomics data from Terunuma et al. have previously been used by The Cancer Genome Atlas Consortium to correlate gene expression data with metabolomics data [11, 13]. Based on this dataset, we applied PGMs for the first time in metabolomics data from tumor samples and also in metabolomics data combined with gene expression data and flux activities, with the aim of confirming known associations and finding new ones.

First, we evaluated whether metabolomics data were related to overall survival in patients with breast cancer. An overall survival predictive signature was built that included the expression values of glutamine, deoxycarnitine, butyrylcarnitine, glycerophosphorylcholine and 2-hydroxypalmitate [14]. The first three of these metabolites has previously been related to survival in breast cancer [15, 16]. However, to our knowledge, this is the first report associating 2-hydroxypalmitate with cancer survival. Additionally, in the previous study by Terunuma et al., 2-hydroxyglutarate was associated with a poor prognosis in patients with breast cancer [11]. 2-hydroxyglutarate is a glutamine intermediate in the tricarboxylic acid cycle, involved in the conversion of glutamine into lactate, a process known as glutaminolysis [14]. These results highlight the relevance of glutamine metabolism in breast cancer prognoses.

A metabolite network using metabolomics data was built using PGM. IMPaLA assigned a dominant metabolic function to each resulting node. In previous studies, we demonstrated that PGMs are useful for functionally characterizing gene or protein networks [6, 7, 12].

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

However, to our knowledge, this is the first time a PGM has been applied to metabolomics data from tumor samples. Just as observed in genes or proteins, metabolites are grouped into metabolic pathways, allowing the characterization of differences in metabolic pathways between ER+ and ER- tumors. For example, both lipid metabolism and purine metabolism node activities were higher in ER- tumors. ER- tumors usually overexpress genes related to lipid metabolism. [17]. Moreover, the activity of the lipid metabolism node had prognostic value. No relationship between purine metabolism and breast cancer has previously been defined.

On the other hand, the network combining gene expression data and metabolomics data grouped most of the metabolites into an isolated node. Yet, some metabolites were included in gene nodes. We found that four of the twenty metabolites showed a previously reported relationship with the main function of the gene node in which they were included. Succinate and cytidine were located in the immune response node. Succinate acts as an inflammation activation signal, inducing IL-1β cytokine production through hypoxia-inducible factor 1 [18]. In addition, succinate increases dendritic cell capability to act as antigen-presenting cells, prompting an adaptive immune response [19]. Regarding cytidine, Wachowska et al. described that 5-aza-2'-deoxycytidine modulates the levels of major histocompatibility complex class I molecules in tumor cells, induces P1A antigen and has immunomodulatory activity when combined with photodynamic therapy [20].

Both histamine and 1,2-propanediol appeared to be related to the angiogenesis node. Histamine is known to promote angiogenesis through vascular epithelial growth factor [21]. On the other hand, sulfoquinovosyl acylpropanediol, an 1,2-propanediol derivate, inhibits angiogenesis in murine models with pulmonary carcinoma [22].

FBA was used to model metabolism using gene expression data. Although FBA-predicted biomass did not show significant differences between ER+ and ER- tumors, differences in flux

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

activities were shown between both subtypes. Some of these activities were also related to prognosis. One of these flux activities is “Glutamine metabolism”, which agrees with the results obtained from the metabolomics data, including glutamine in the metabolite, a signature capable of predicting overall survival. With the aim of associating metabolomics and FBA results, flux activities and metabolomics data were combined to form a new network. As opposed to gene and metabolite data, metabolomics data and flux activities combined well in the network. Interestingly, flux activities are dead-end nodes, perhaps due to the fact that they are by definition a final summary of each pathway. IMPaLA assigned a main metabolic pathway to resulting branches; thus, it was possible to know how many metabolites were related to flux activity in each branch. In most cases with available information, there was coherence between metabolites included in the branch and its flux activity. This validates FBA and flux activities, both based on gene expression, as a method of simulating metabolism.

Our study has some limitations. The limited number of samples leads us to consider the results as preliminary, and validation in an independent cohort is needed. Additionally, our results are difficult to place in the current clinical landscape, given that tumors in the original series had not been assessed for HER2 expression. On the other hand, evolving techniques currently allow the detection of more metabolites, which would permit a more thorough analysis.

In conclusion, PGMs reveal their utility in the analysis of metabolomics data from a functional point of view, not only metabolomics data alone, but also in combination with flux or gene expression data. Therefore, PGM is postulated as a method to propose new hypotheses in the metabolomics field. We also found that it is possible to associate metabolomics data with clinical outcomes and to build prognostic signatures based on metabolomics data.

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Materials and methods

Patients included in the study

Metabolomics and gene expression data from 67 fresh tumor tissue samples originally analyzed by Terunuma et al. [11] were included in this study.

Preprocessing of gene expression and metabolomics data

For the metabolomics data, log2 was calculated. As quality criteria, data were filtered to include detectable measurements in at least 75% of the samples. Missing values were imputed to a normal distribution using Perseus software [23]. After quality control, 237 metabolites were considered for subsequent analyses.

In terms of gene expression data, the 2000 most variable genes, i.e., those genes with the highest standard deviation, were chosen to build the PGM.

Probabilistic graphical models and gene ontology analyses

As previously described [6, 7, 10, 12], PGMs compatible with high dimensional data were used, using correlations as associative criteria. The *grapHD* package [24] and R v3.2.5 [25] were employed to build the network. A majority function was assigned to each node using gene ontology analyses. In the case of genes, gene ontology analyses were performed using the DAVID web tool with “homo sapiens” as background and GOTERM, KEGG and Biocarta selected as categories [26]. In the case of metabolites, the Integrated Molecular Pathway Level Analysis (IMPaLA) web tool was used [27].

Node activities were calculated, as previously described [6, 7, 10, 12], as the mean of the expression/quantity of genes/metabolites of each node that are related to the main node function/metabolic pathway.

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Flux Balance Analysis and flux activities

FBA was calculated using the human metabolic reconstruction Recon2 [28]. As the objective function, the biomass reaction proposed in the Recon2 was used. FBA was performed using the COBRA Toolbox [29] available for MATLAB. Gene-Protein-Reaction rules were solved as described in previous studies [7, 10], using a modification of the Barker *et al.* algorithm [30], which were incorporated into the model by the E-flux method [31].

Flux activities were previously proposed as a measurement to compare differences at the flux pathway level [10]. Briefly, they were calculated as the sum of the fluxes of the reactions included in each pathway defined in Recon2.

Statistical analyses

The statistical analyses were performed with GraphPad Prism v6, and the network analyses were performed using Cytoscape software [32]. Predictor signatures were built with the BRB Array Tool from Dr. Richard Simon's team. All p-values are two-sided and are considered statistically significant under 0.05.

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Funding statement

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" (PI15/01310). LT-F is supported by the Spanish Economy and Competitiveness Ministry (DI-15-07614). GP-V is supported by Conserjería de Educación, Juventud y Deporte of Comunidad de Madrid (IND2017/BMD7783). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

All the authors have directly participated in the preparation of this manuscript and have approved the final version submitted. JMA, MD-A, HN and PM contributed the directed graphical models. LT-F, AG-P, G-PV, and AZ-M performed the statistical analyses, the graphical model interpretation and the ontology analyses. LT-F, AG-P, JAFV, PZ and EE conceived of the study and participated in its design and interpretation. MD-A and LT-F performed the Flux Balance Analysis. LT-F drafted the manuscript. AG-P, JAFV, and EE supported the manuscript drafting. AG-P and JAFV coordinated the study.

Competing interests

JAFV, EE and AG-P are shareholders in Biomedica Molecular Medicine SL. LT-F and GP-V are employees of Biomedica Molecular Medicine SL. The other authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7-30. Epub 2018/01/04. doi: 10.3322/caac.21442. PubMed PMID: 29313949.
2. Fiehn O. Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol. 2002;48(1-2):155-71. PubMed PMID: 11860207.
3. Emwas AH. The strengths and weaknesses of NMR spectroscopy and mass spectrometry with particular focus on metabolomics research. Methods Mol Biol. 2015;1277:161-93. doi: 10.1007/978-1-4939-2377-9_13. PubMed PMID: 25677154.
4. Gowda GA, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D. Metabolomics-based methods for early disease diagnostics. Expert Rev Mol Diagn. 2008;8(5):617-33. doi: 10.1586/14737159.8.5.617. PubMed PMID: 18785810; PubMed Central PMCID: PMCPMC3890417.
5. Sánchez-Navarro I, Gámez-Pozo A, Pinto A, Hardisson D, Madero R, López R, et al. An 8-gene qRT-PCR-based gene expression score that has prognostic value in early breast cancer. BMC Cancer. 2010;10:336. Epub 2010/06/28. doi: 10.1186/1471-2407-10-336. PubMed PMID: 20584321; PubMed Central PMCID: PMCPMC2906483.
6. 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.
7. 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. doi: 10.1038/s41598-017-10493-w.
8. Varma A, Palsson BO. Parametric sensitivity of stoichiometric flux balance models applied to wild-type Escherichia coli metabolism. Biotechnol Bioeng. 1995;45(1):69-79. doi: 10.1002/bit.260450110. PubMed PMID: 18623053.
9. Orth J, Thiele I, Palsson B. What is flux balance analysis? : Nat Biotechnol; 2010. p. 245-8.
10. 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. doi: 10.18632/oncotarget.24047. PubMed PMID: 29515760; PubMed Central PMCID: PMCPMC5839391.
11. Terunuma A, Putluri N, Mishra P, Mathé EA, Dorsey TH, Yi M, et al. MYC-driven accumulation of 2-hydroxyglutarate is associated with breast cancer prognosis. J Clin Invest. 2014;124(1):398-412. Epub 2013/12/09. doi: 10.1172/JCI71180. PubMed PMID: 24316975; PubMed Central PMCID: PMCPMC3871244.
12. 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. doi: 10.1038/s41598-017-15920-6. PubMed PMID: 29150671; PubMed Central PMCID: PMCPMC5694001.
13. Peng X, Chen Z, Farshidfar F, Xu X, Lorenzi PL, Wang Y, et al. Molecular Characterization and Clinical Relevance of Metabolic Expression Subtypes in Human Cancers. Cell Rep. 2018;23(1):255-69.e4. doi: 10.1016/j.celrep.2018.03.077. PubMed PMID: 29617665; PubMed Central PMCID: PMCPMC5916795.
14. DeBerardinis RJ, Mancuso A, Daikhin E, Nissim I, Yudkoff M, Wehrli S, et al. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. Proc Natl Acad Sci U S A. 2007;104(49):19345-50. doi: 10.1073/pnas.0709747104. PubMed PMID: 18032601; PubMed Central PMCID: PMCPMC2148292.
15. Bhowmik SK, Ramirez-Peña E, Arnold JM, Putluri V, Sphyris N, Michailidis G, et al. EMT-induced metabolite signature identifies poor clinical outcome. Oncotarget. 2015;6(40):42651-

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

60. doi: 10.18632/oncotarget.4765. PubMed PMID: 26315396; PubMed Central PMCID: PMCPMC4767460.

16. Cao MD, Sitter B, Bathen TF, Bofin A, Lønning PE, Lundgren S, et al. Predicting long-term survival and treatment response in breast cancer patients receiving neoadjuvant chemotherapy by MR metabolic profiling. NMR Biomed. 2012;25(2):369-78. Epub 2011/08/08. doi: 10.1002/nbm.1762. PubMed PMID: 21823183.

17. Wang J, Shidfar A, Ivancic D, Ranjan M, Liu L, Choi MR, et al. Overexpression of lipid metabolism genes and PBX1 in the contralateral breasts of women with estrogen receptor-negative breast cancer. Int J Cancer. 2017;140(11):2484-97. Epub 2017/03/21. doi: 10.1002/ijc.30680. PubMed PMID: 28263391.

18. Tannahill GM, Curtis AM, Adamik J, Palsson-McDermott EM, McGettrick AF, Goel G, et al. Succinate is an inflammatory signal that induces IL-1β through HIF-1α. Nature. 2013;496(7444):238-42. Epub 2013/03/24. doi: 10.1038/nature11986. PubMed PMID: 23535595; PubMed Central PMCID: PMCPMC4031686.

19. Jiang S, Yan W. Succinate in the cancer-immune cycle. Cancer Lett. 2017;390:45-7. Epub 2017/01/18. doi: 10.1016/j.canlet.2017.01.019. PubMed PMID: 28109906.

20. Wachowska M, Gabrysiak M, Muchowicz A, Bednarek W, Barankiewicz J, Rygiel T, et al. 5-Aza-2'-deoxycytidine potentiates antitumour immune response induced by photodynamic therapy. Eur J Cancer. 2014;50(7):1370-81. Epub 2014/02/18. doi: 10.1016/j.ejca.2014.01.017. PubMed PMID: 24559534; PubMed Central PMCID: PMCPMC4136636.

21. Lu Q, Wang C, Pan R, Gao X, Wei Z, Xia Y, et al. Histamine synergistically promotes bFGF-induced angiogenesis by enhancing VEGF production via H1 receptor. J Cell Biochem. 2013;114(5):1009-19. doi: 10.1002/jcb.24440. PubMed PMID: 23225320.

22. Ruike T, Kanai Y, Iwabata K, Matsumoto Y, Murata H, Ishima M, et al. Distribution and metabolism of. Xenobiotica. 2018:1-45. Epub 2018/03/15. doi: 10.1080/00498254.2018.1448949. PubMed PMID: 29543539.

23. 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. doi: 10.1038/nmeth.3901. PubMed PMID: 27348712.

24. Abreu G, Edwards D, Labouriau R. High-Dimensional Graphical Model Search with the gRapHD R Package Journal of Statistical Software2010. p. 1-18.

25. R Core Team. R: A language and environment for statistical computing. Vienna,Austria. R Foundation for Stattistical Computing, 2013.

26. 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. doi: 10.1038/nprot.2008.211. PubMed PMID: 19131956.

27. Cavill R, Kamburov A, Ellis JK, Athersuch TJ, Blagrove MS, Herwig R, et al. Consensus-phenotype integration of transcriptomic and metabolomic data implies a role for metabolism in the chemosensitivity of tumour cells. PLoS Comput Biol. 2011;7(3):e1001113. Epub 2011/03/31. doi: 10.1371/journal.pcbi.1001113. PubMed PMID: 21483477; PubMed Central PMCID: PMCPMC3068923.

28. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. Nat Biotechnol. 2013;31(5):419-25. Epub 2013/03/03. doi: 10.1038/nbt.2488. PubMed PMID: 23455439; PubMed Central PMCID: PMCPMC3856361.

29. Schellenberger J, Que R, Fleming R, Thiele I, Orth J, Feist A, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols; 2011. p. 1290-307.

30. Barker BE, Sadagopan N, Wang Y, Smallbone K, Myers CR, Xi H, et al. A robust and efficient method for estimating enzyme complex abundance and metabolic flux from expression data. Comput Biol Chem. 2015;59 Pt B:98-112. Epub 2015/09/01. doi:

bioRxiv preprint doi: https://doi.org/10.1101/370221; this version posted July 16, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

10.1016/j.compbiolchem.2015.08.002. PubMed PMID: 26381164; PubMed Central PMCID: PMCPMC4684447.

31. Colijn C, Brandes A, Zucker J, Lun D, Weiner B, Farhat M, et al. Interpreting expression data with metabolic flux models: Predicting Mycobacterium tuberculosis mycolic acid production. PLOS Comput Bio; 2009.

32. 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. doi: 10.1101/gr.1239303. PubMed PMID: 14597658; PubMed Central PMCID: PMCPMC403769.

Supporting information

S1 Table: Multivariate Cox regression model comparing OS predictor based on metabolomics data. T = tumor stage, N = lymph node status, G = tumor grade.

S2 Table: Multivariate Cox analysis comparing predictor based on node activity of lipid metabolism. T = tumor stage, N = lymph node status, G = tumor grade.

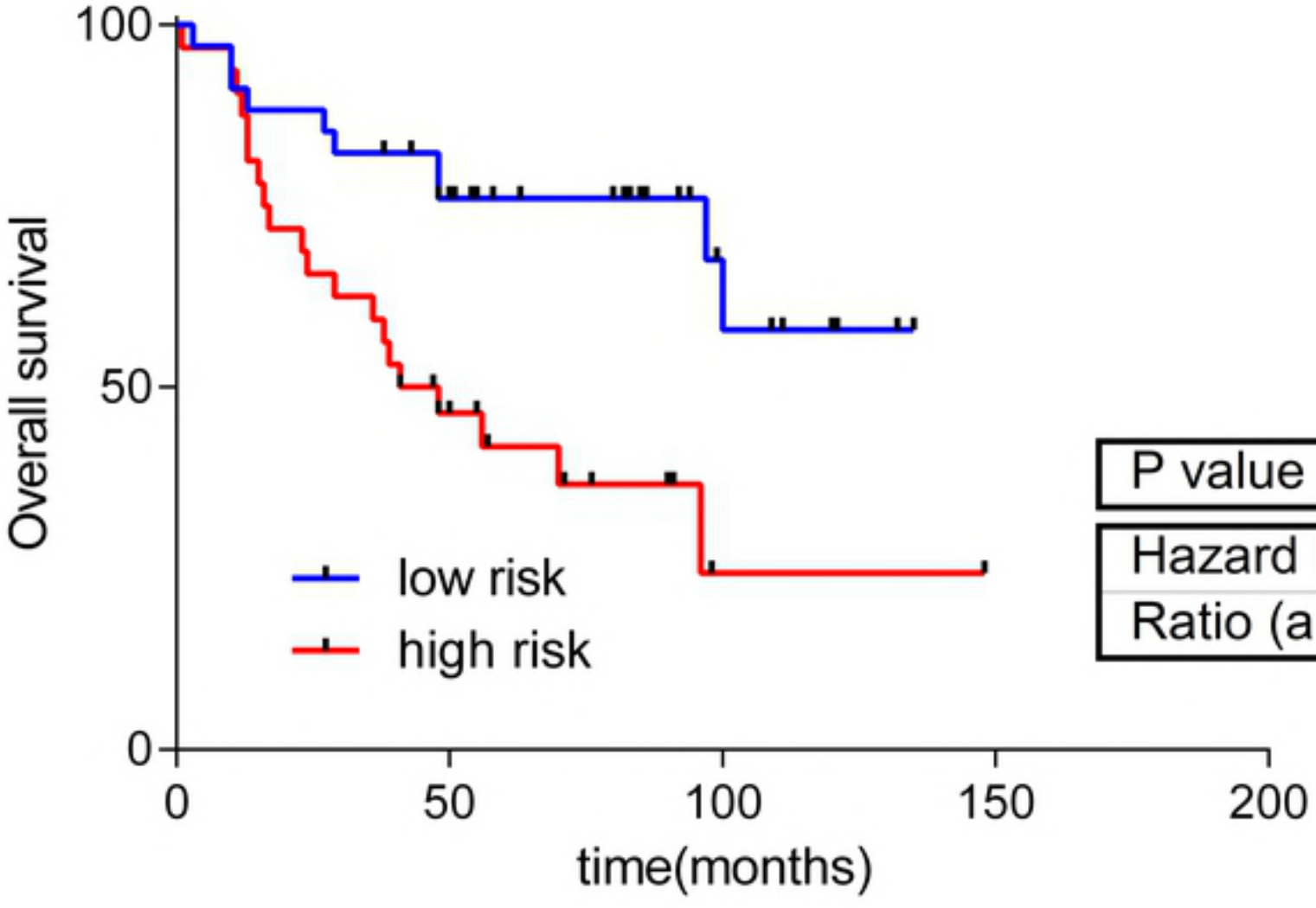
S3 Table: Multivariate Cox regression comparing predictor based on flux activities. T = tumor stage, N = lymph node status, G = tumor grade.

S4 Table: Metabolites associated with flux activity of each network branch.

S1 Fig: Node activities from the metabolic network.

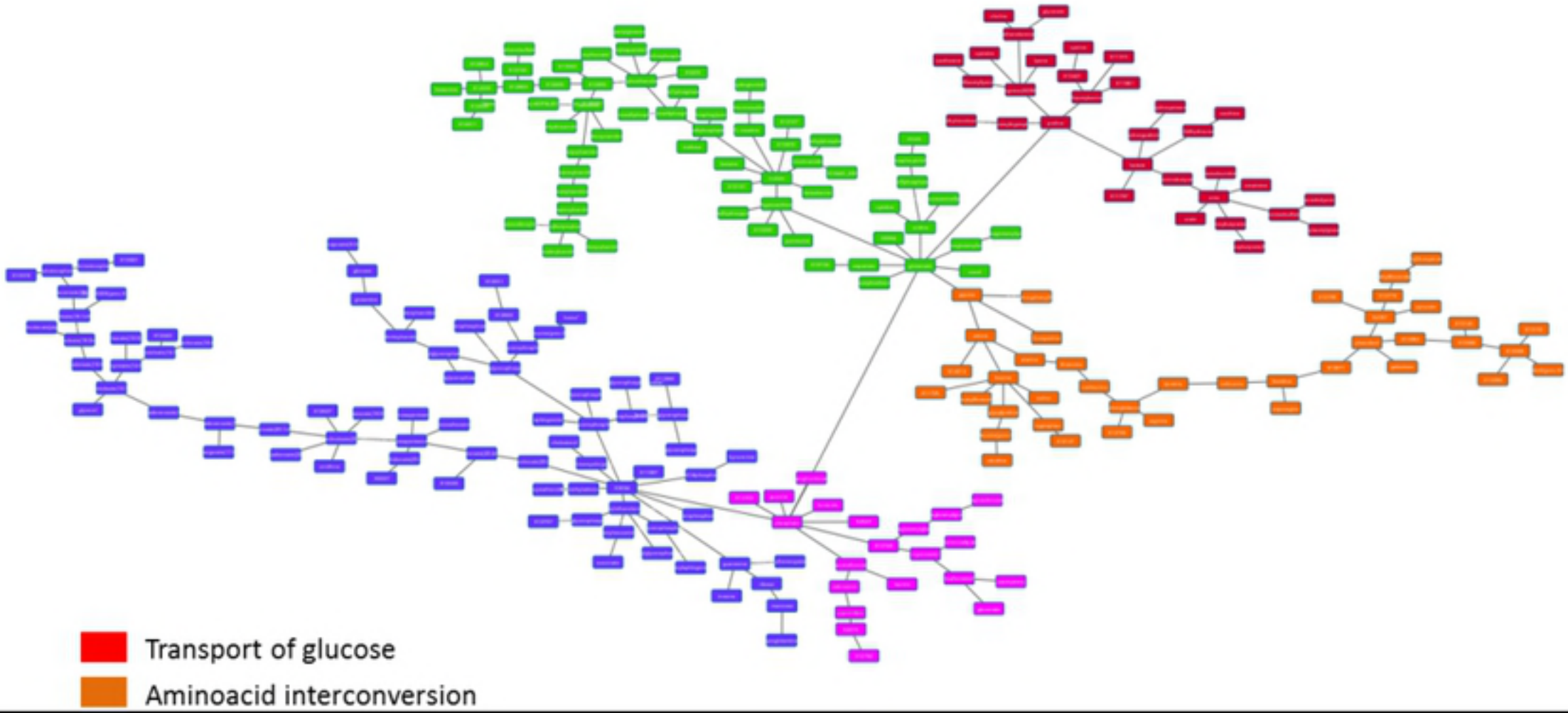
S2 Fig: Tumor growth rate predicted using FBA for ER+ and ER- tumors.

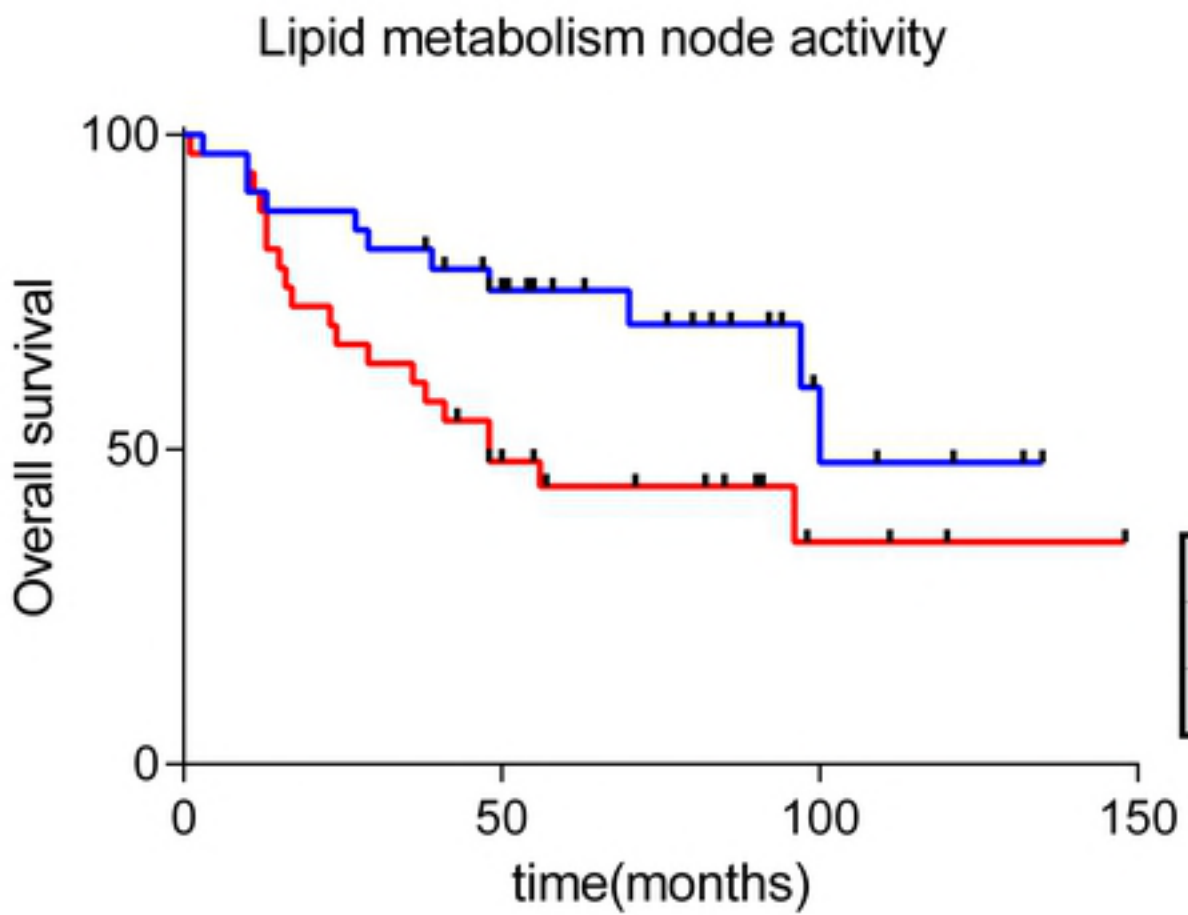
S3 Fig: Flux activities were significantly different between ER+ and ER-.



P value	0.0030

Hazard Ratio (logrank)	A/B
Ratio (and its reciprocal)	0.3422

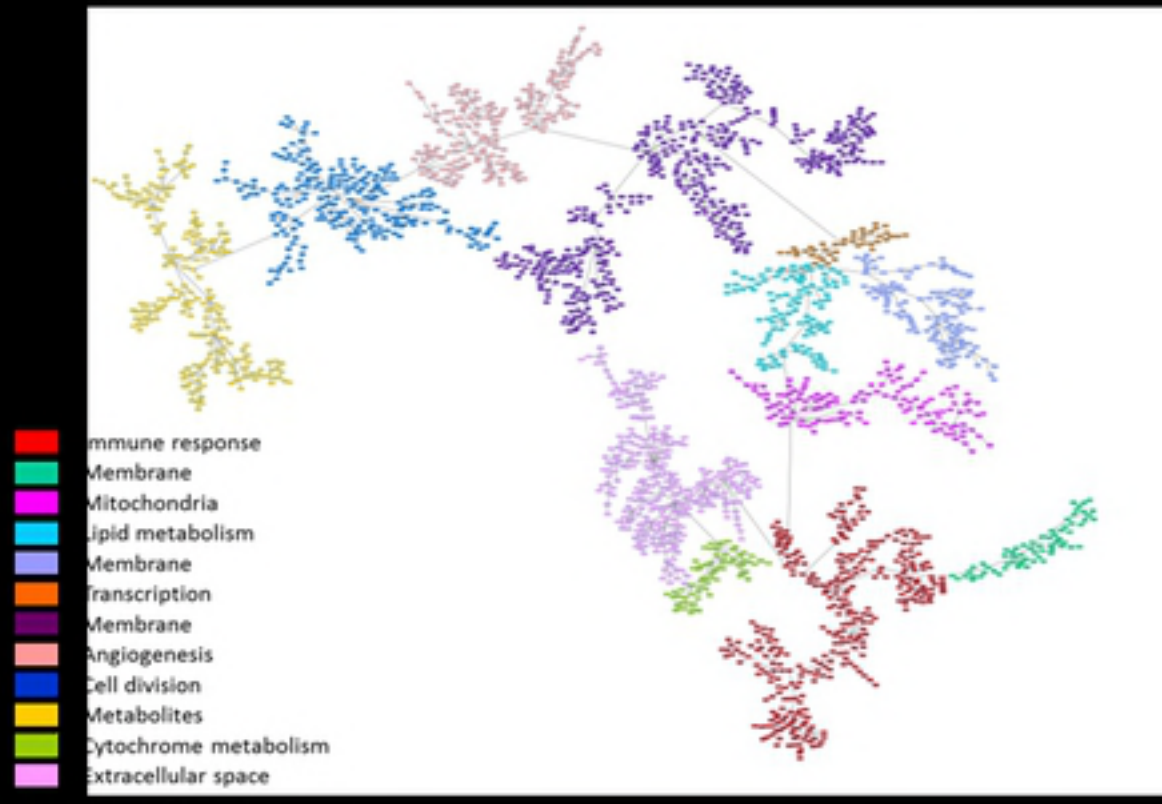

Transport of glucose
Aminoacid interconversion

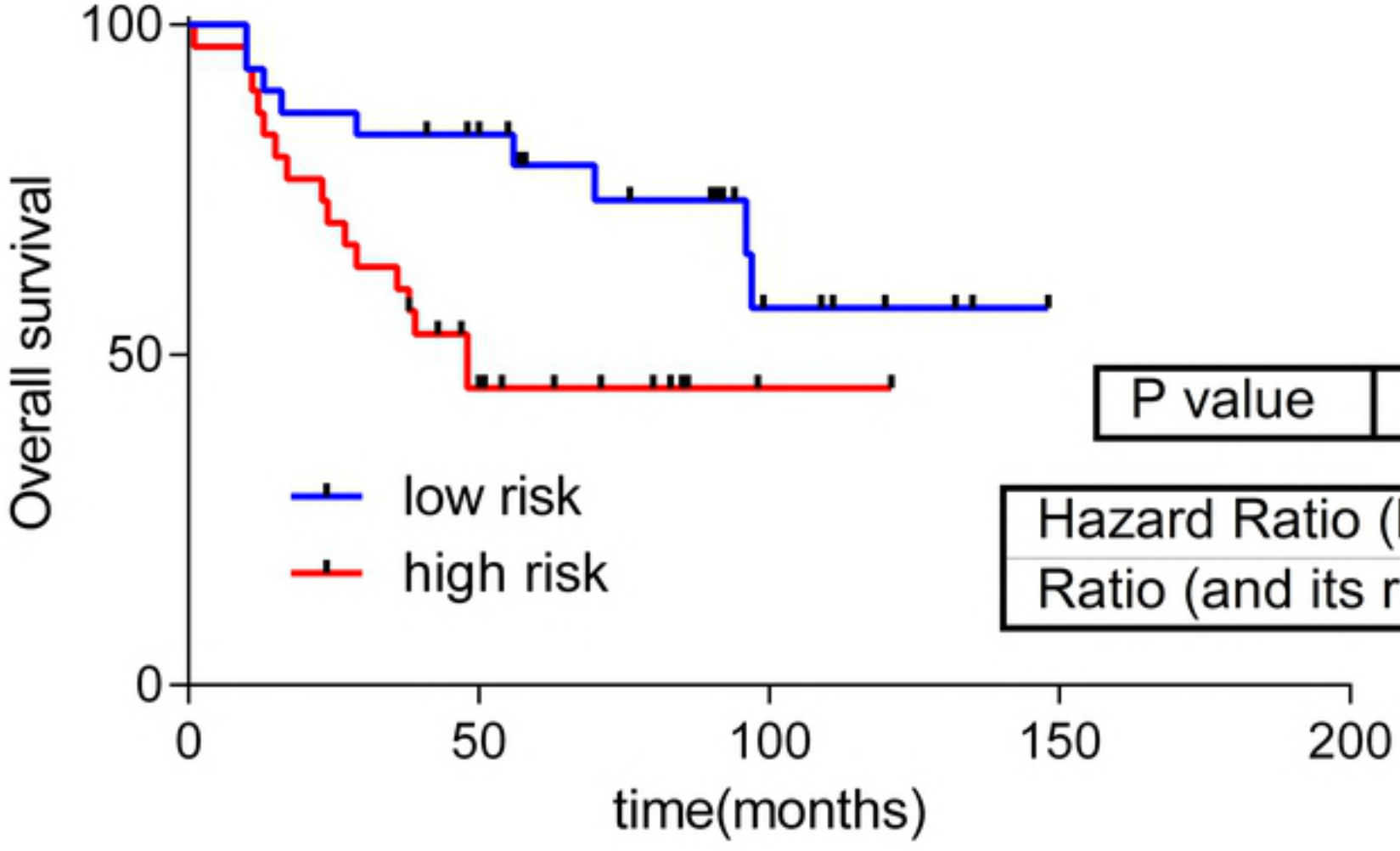


P value	0.0452

Hazard Ratio (logrank)	A/B
Ratio (and its reciprocal)	0.4796
95% CI of ratio	0.2321 to 0.9775

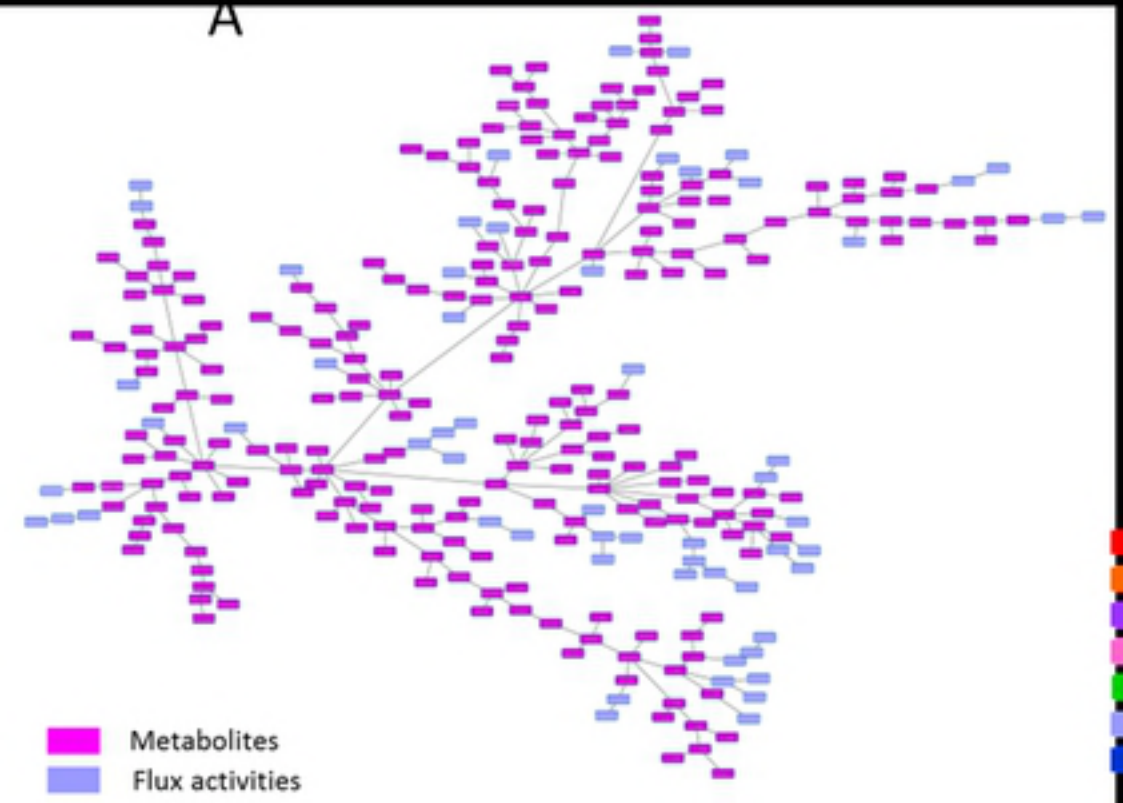
Genes
Metabolites

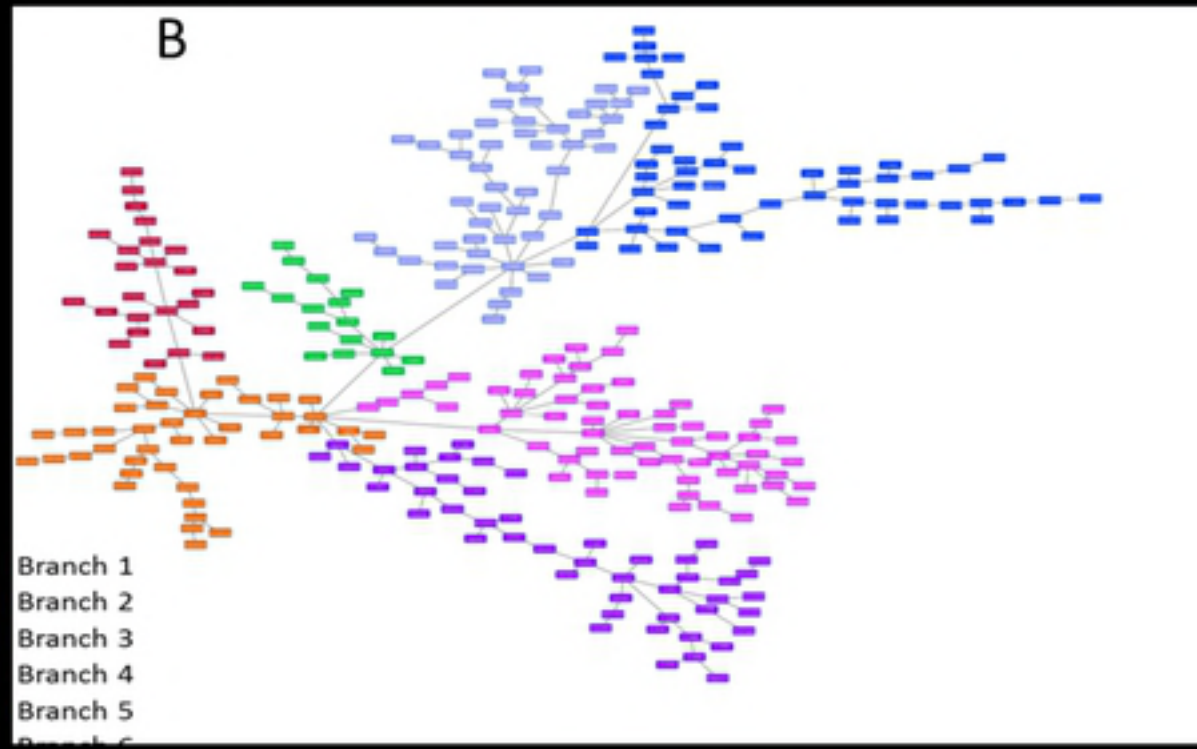

mmune response
Membrane
Mitochondria
ipid metabolism
Membrane
Transcription
Membrane
Angiogenesis
Cell division
Metabolites
Cytochrome metabolism
Extracellular space



P value	0.0243

Hazard Ratio (logrank)	A/B
Ratio (and its reciprocal)	0.4110


A
Metabolites
Flux activities


B
Branch 1
Branch 2
Branch 3
Branch 4
Branch 5