Computational metabolomics hints at the relevance of glutamine metabolism in breast cancer

Lucía Trilla-Fuertes[1¶]; Angelo Gámez-Pozo[1,2¶]; Elena López-Camacho[2]; Guillermo Prado-Vázquez[1]; Andrea Zapater-Moros[1,2]; Rocío López-Vacas[2] ; Jorge M Arevalillo[3]; 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)

Running title: Relevance of glutamine in breast cancer

Abstract

Metabolomics has a 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. On the one hand, both analyses highlighted the importance of glutamine in breast cancer. Moreover, cell experiments showed that treating breast cancer cells with drugs targeting glutamine metabolism significantly affects cell viability. On the other hand, these computational methods suggested some hypotheses and have demonstrated their utility in the analysis of metabolomics data and in associating metabolomics with patient's clinical outcome.

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 (Siegel et al, 2018). 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, a technique focused in the holistic study of the metabolites present in a biological system, is considered the most recent -omics. It consists of measuring the entire set of metabolites present in a biological sample (Fiehn, 2002). 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 (Emwas, 2015). Recent advances in this technique allow the measurement of thousands of metabolites from minimal amounts of biological samples (Emwas, 2015; Fuhrer & Zamboni, 2015). Therefore, metabolomics is a promising tool for the development of new biomarkers (Gowda et al, 2008).

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 (Gámez-Pozo et al, 2015; Gámez-Pozo et al, 2017; Sánchez-Navarro et al, 2010). Flux Balance Analysis (FBA), however, is a method that has been widely used to study biochemical networks (Varma & Palsson, 1995). FBA predicts the growth rate or the rate of production of a given metabolite (Orth et al, 2010), and it has previously been used to characterize breast cancer cell responses against drugs targeting metabolism (Trilla-Fuertes et al, 2018). In this study, flux activities were proposed as a feasible method to compare flux patterns in metabolic pathways.

Glutamine has a relevant role in tumor metabolism. The entrance of glutamine in the tricarboxylic acid cycle (TCA) generates lactate, a process known as glutaminolysis. The

57 metabolism of glutamine serves to maintain the availability of non-essential aminoacids and to
58 maintain TCA intermediates while NADH is generating (DeBerardinis et al, 2007). Glutamine is
59 necessary to cellular proliferation and its metabolism is regulated by the levels of *MYC*
60 oncogene (Eagle et al, 1956; Wise et al, 2008).

61 In the present study, metabolomics and gene expression data from 67 fresh tissue samples
62 (Terunuma et al, 2014) were analyzed through PGMs and FBA. Our aim was to find
63 associations between metabolomics and gene expression data and the characterization of
64 breast cancer from a metabolomics point of view.

Results

Patient characteristics

With the aim of study the relationships between metabolomics, gene expression, and FBA results, metabolomics and gene expression data , analyzed by mass-spectrometry and microarrays GeneChip Human Gene 1.0 ST (Affymetrix) respectively and published by Terunuma et al., were analyzed (Terunuma et al, 2014).

A total of 67 tumor fresh tissue samples from patients with breast cancer were studied. This patient cohort comprises 33 ER+ and 34 ER- (of which 14 were also TNBC). The median follow-up was 50 months, and 31 deaths occurred during this time. No significant differences regarding overall survival (OS) were observed between patients with ER+ or ER- tumors. Patient characteristics are shown in Table 1.

Analysis of metabolomics data

After Kaplan-Meier analysis, 29 metabolites were found related to OS ($p<0.05$) (Sup Table 1).

Then, an OS predictor using this metabolomics data was built. This metabolite-based signature included five metabolites: glutamine, 2-hydroxypalmitate, deoxycarnitine, butyrylcarnitine and glycerophosphorylcholine (p-value =0.003, hazard ratio [HR] = 0.34, 50:50%) (Fig 1). A multivariate analysis showed that the predictor provided additional prognostic information to clinical data (S1 Table).

Metabolomics data without using any a priori information were analyzed through PGM. Metabolomics database, including information about 536 metabolites, was reduced to 237 metabolites due to quality criteria. A main metabolic pathway was assigned to each functional node of the resulting network 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 as it has been previously shown

89 for gene and protein PGMs (de Velasco et al, 2017; Gámez-Pozo et al, 2015; Gámez-Pozo et al,
90 2017; Trilla-Fuertes et al, 2018). Five functional nodes were defined, each with a different
91 overrepresented metabolic pathway (Fig 2).

92 The activity of each functional node was calculated as previously described and comparisons
93 between ER+ and ER- were done (de Velasco et al, 2017; Gámez-Pozo et al, 2015; Gámez-Pozo
94 et al, 2017; Trilla-Fuertes et al, 2018). Significant differences were found between ER+ and ER-
95 tumors regarding lipid and purine metabolism ($p<0.05$) (S1 Fig).

96 Moreover, the lipid metabolism functional node showed prognostic value in this cohort (p
97 =0.045, HR = 0.48, 50:50%) (Fig 3). However, a multivariate analysis showed that the predictor
98 do not add additional prognostic information to that provided by clinical features (S2 Table).

99 Analyses combining gene expression with metabolomics data

100 On the other hand, a network combining metabolomics and gene expression data was built.
101 Due to the differences between both kinds of data, most of the metabolites were grouped
102 together. However, some metabolites were integrated into gene branches (Fig 4).

103 This combined network was then functionally characterized based on the majority function of
104 the genes contained in each branch. The resulting network had eleven functional nodes and a
105 twelfth branch that include the majority of the metabolites (Fig 4).

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

Flux Balance Analysis and flux activities

FBA and flux activities were calculated as previously described (Trilla-Fuertes et al, 2018). Briefly, gene expression data from 67 tumor samples were used, GPR rules were solved and the normalized values were introduced into the metabolic model using modified E-flux algorithm (Colijn et al, 2009; Gámez-Pozo et al, 2017). Finally, FBA was calculated using a biomass reaction representative of tumor growth. 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).

In addition, the combination of glutamate metabolism (the pathway that includes the glutamine) and alanine and aspartate metabolism flux activities showed prognostic value in this cohort (p-value = 0.024, HR = 0.41, 50:50%) (Fig 5). A multivariate analysis showed that this flux activity-based predictor provides prognostic information independently from clinical data (S3 Table).

PGM analysis combining flux activities with metabolomics data

Flux activities were calculated for each metabolic pathway defined in the Recon2. Then, using flux activities and metabolomics data, a new PGM was built to study association between both types of data. Interestingly, both types of data appeared mixed in the network; with the peculiarity that flux activities appeared usually at the periphery of the network (Fig 6).

The resulting network was split into several branches to study the relationship of the metabolites with the flux activities included in each branch (Fig 6). Coherence between both types of data was shown by the PGM, associating flux activities and metabolites related to these flux activities in the same branch. For instance, branch 1 includes glycolysis flux activity

136 and three metabolites previously related to glycolysis (S4 Table). Regarding vitamins and
137 cofactors, it was not possible make comparisons because the IMPaLA label for this category is
138 "Vitamin and co-factor metabolism" and Recon2 labels differentiate between the various
139 vitamins, labeling them as "Vitamin B6 metabolism", "Vitamin A metabolism", etc.

140 Cell viability assays using drugs targeting glutamine metabolism

141 As the computational analyses pointed out the relevance of glutamine and its metabolism in
142 breast cancer, cell viability assays employing two drugs targeting glutaminolysis
143 (aminooxyacetic acid [AOA], and L-Glutamic acid γ-(p-nitroanilide) hydrochloride [GPNA]) were
144 performed. Dose-response curves of these two drugs confirm that targeting glutamine
145 metabolism affected cell viability (Figure 7).

146 In AOA-treated cell lines, the IC_{50} did not show any differential response between breast
147 cancer subtypes. However, in GPNA-treated cells, IC_{50} for ER+ cell lines were lower than IC_{50} for
148 TNBC cells (Table 3).

149

150 # Discussion

151 Metabolomics is attracting considerable interest as a technique for finding new biomarkers in
152 cancer. In this study, a new analytical workflow for the management and study of
153 metabolomics data was proposed. This workflow allowed global metabolic characterization,
154 beyond analyses based on unique metabolites. On the other hand, this workflow pointed out
155 the relevance of glutamine metabolism in breast cancer, a hypothesis that was confirmed by
156 cellular experiments.

157 Genomics and metabolomics data from Terunuma et al. have previously been used by The
158 Cancer Genome Atlas Consortium to correlate gene expression data with metabolomics data
159 (Peng et al, 2018; Terunuma et al, 2014). Based on this dataset, we applied PGMs for the first
160 time in metabolomics data from tumor samples and also in metabolomics data combined with
161 gene expression data and flux activities, with the aim of confirming known associations and
162 finding new ones.

163 First, we evaluated whether metabolomics data were related to OS in patients with breast
164 cancer. An OS predictive signature was built that included the expression values of glutamine,
165 deoxycarnitine, butyrylcarnitine, glycerophosphorylcholine and 2-hydroxypalmitate
166 (DeBerardinis et al, 2007). The first four metabolites have been previously related to survival in
167 breast cancer (Bhowmik et al, 2015; Cao et al, 2012). However, to our knowledge, this is the
168 first report associating 2-hydroxypalmitate with cancer survival. Additionally, in the previous
169 study by Terunuma et al., 2-hydroxyglutarate was associated with poor prognosis in patients
170 with breast cancer (Terunuma et al, 2014). 2-hydroxyglutarate is a glutamine intermediate in
171 the tricarboxylic acid cycle, involved in the conversion of glutamine into lactate, a process
172 known as glutaminolysis (DeBerardinis et al, 2007). The negative sign in the predictor of
173 glutamine points a protective effect (the more glutamine the better prognosis). An increased
174 presence of glutamine could indicate that it has not been introduced into the Krebs cycle and

175 transformed into lactate, a fact associated with a more aggressive phenotype and a worst
176 prognosis.

177 A metabolite network using metabolomics data was built using PGMs. PGMs are based on
178 expression data, or quantification data in the case of metabolomics, and they do not need any
179 additional information. The output of this analysis is a network that reflects the correlations
180 between genes, proteins or metabolites. On the other hand, IMPaLA assigned a dominant
181 metabolic function to each resulting node. In previous studies, we demonstrated that PGMs
182 are useful for functionally characterizing gene or protein networks (de Velasco et al, 2017;
183 Gámez-Pozo et al, 2015; Gámez-Pozo et al, 2017). However, to our knowledge, this is the first
184 time a PGM has been applied to metabolomics data from tumor samples. Just as observed in
185 genes or proteins, metabolites are grouped into metabolic pathways, allowing the
186 characterization of differences in metabolic pathways between ER+ and ER- tumors. For
187 example, both lipid metabolism and purine metabolism node activities were higher in ER-
188 tumors. Although there has not been described a relationship between lipids and breast
189 cancer subtypes, it was described that ER- tumors usually overexpress genes related to lipid
190 metabolism (Wang et al, 2017). Moreover, the activity of the lipid metabolism node had
191 prognostic value. No relationship between purine metabolism and breast cancer has
192 previously been defined.

193 On the other hand, the network combining gene expression data and metabolomics data
194 grouped most of the metabolites into an isolated branch. Yet, some metabolites were included
195 into gene branches. We found that four out of twenty metabolites showed a previously
196 reported relationship with the main function of the gene functional node in which they were
197 included. Succinate and cytidine were located in the immune response functional node.
198 Succinate acts as an inflammation activation signal, inducing IL-1β cytokine production through
199 hypoxia-inducible factor 1 (Tannahill et al, 2013). In addition, succinate increases dendritic cell

200 capability to act as antigen-presenting cells, prompting an adaptive immune response (Jiang &
201 Yan, 2017). Regarding cytidine, Wachowska et al. described that 5-aza-2'-deoxycytidine
202 modulates the levels of major histocompatibility complex class I molecules in tumor cells,
203 induces P1A antigen and has immunomodulatory activity when combined with photodynamic
204 therapy (Wachowska et al, 2014).

205 Both histamine and 1,2-propanediol appeared to be related to the angiogenesis functional
206 node. Histamine is known to promote angiogenesis through vascular epithelial growth factor
207 (Lu et al, 2013). On the other hand, sulfoquinovosyl acylpropanediol, an 1,2-propanediol
208 derivate, inhibits angiogenesis in murine models with pulmonary carcinoma (Ruike et al, 2018).

209 The remaining sixteen metabolites require an in-depth study to establish associations with
210 their respective functional nodes. These results support the potential of PGMs as a tool to
211 generate hypotheses without the need of *a priori* knowledge.

212 FBA was used to model metabolism using gene expression data. Although FBA-predicted
213 biomass did not show significant differences between ER+ and ER- tumors, differences in flux
214 activities were shown between both subtypes. Some of these activities were also related to
215 prognosis, such as "Glutamine metabolism", which agrees with the results obtained from the
216 metabolomics data, including glutamine in the metabolite signature capable of predicting OS.
217 These results highlighted the relevance of glutamine metabolism in breast cancer, suggesting
218 the utility of drugs targeting this pathway such as GPNA, which it has already been described
219 as affecting proliferation in lung cancer cells (Hassanein et al, 2013). Strikingly, cell viability
220 experiments using GPNA and AOA, two drugs targeting glutamine metabolism, showed a
221 decreased in cell viability, confirming the relevant role of this process in breast cancer. Despite
222 the highest levels of glutamine-related enzymes described in TNBC and HER2 tumors
223 comparing with luminal tumors, dose-response curves did not show any differential response
224 between ER+ and TNBC breast cancer cells in the case of AOA (Cao et al, 2014; Kanaan et al,

225 2014). In addition, ER+ cells seem to be more sensible to GPNA. On the other hand, AOA has
226 been successfully tested in ER+ and ER- breast cancer xenograft models (Korangath et al,
227 2015).

228 With the aim of associating metabolomics and FBA results, flux activities and metabolomics
229 data were combined to form a new network. As opposed to gene and metabolite data,
230 metabolomics data and flux activities combined well in the network. Interestingly, flux
231 activities are dead-end nodes, perhaps due to the fact that they are by definition a final
232 summary of each pathway. IMPaLA assigned a main metabolic pathway to resulting branches;
233 thus, it was possible to know how many metabolites were related to flux activity in each
234 branch. In most cases with available information, there was coherence between metabolites
235 included in the branch and its flux activity. This validates FBA and flux activities, both based on
236 gene expression, as a method of simulating metabolism.

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

242 Metabolomics is postulated as a booming technique for the biomarker search in cancer.
243 Additionally, PGMs reveal their utility in the analysis of metabolomics data from a functional
244 point of view, not only metabolomics data alone, but also in combination with flux or gene
245 expression data. Therefore, PGM is postulated as a method to propose new hypotheses in the
246 metabolomics field. We also found that it is possible to associate metabolomics data with
247 clinical outcomes and to build prognostic signatures based on metabolomics data. Finally,
248 these computational analyses suggested a main role of glutamine metabolism in breast cancer,
249 a fact that was experimentally validated.

250 # Materials and methods

251 ## Patients included in the study

252 Metabolomics and gene expression data from 67 fresh tumor tissue samples originally
253 analyzed by Terunuma et al. (Terunuma et al, 2014) were included in this study.

254 ## Preprocessing of gene expression and metabolomics data

255 Metabolomics data contains information about 536 metabolites. Log2 was calculated. As
256 quality criteria, data were filtered to include detectable measurements in at least 75% of the
257 samples. Missing values were imputed to a normal distribution using Perseus software
258 (Tyanova et al, 2016). After quality control, 237 metabolites were considered for subsequent
259 analyses.

260 In terms of gene expression data, the 2,000 most variable genes, i.e., those genes with the
261 highest standard deviation, were chosen to build the PGM. This data was from an Affymetrix
262 array and they are available in Gene Expression Omnibus Database under the identifier
263 GSE37751.

264 ## Probabilistic graphical models and gene ontology analyses

265 As previously described (de Velasco et al, 2017; Gámez-Pozo et al, 2015; Gámez-Pozo et al,
266 2017; Trilla-Fuertes et al, 2018), PGMs compatible with high dimensional data were used,
267 using correlation as associative criteria. PGMs were built using metabolomics, gene expression
268 or flux activity data without any a priori information. The *grapHD* package (Abreu et al, 2010)
269 and R v3.2.5 were employed to build the PGMs. The management of the resulting network
270 was done using Cytoscape software (Shannon et al, 2003). The resulting networks were divided
271 into branches and ontology analyses were done to assign a majority function/metabolic
272 pathway to each branch, defining in this way different functional nodes in the networks. In the
273 case of genes, gene ontology analyses were performed using the DAVID web tool with "homo

274 sapiens" as background and GOTERM, KEGG and Biocarta selected as categories (Huang et al,
275 2009). In the case of metabolites, the Integrated Molecular Pathway Level Analysis (IMPaLA)
276 web tool was used to assign a main metabolic pathway to each branch (Cavill et al, 2011).

277 Once the functional structure was defined, functional node activities were calculated in order
278 to make comparisons between groups, as previously described (de Velasco et al, 2017; Gámez-
279 Pozo et al, 2015; Gámez-Pozo et al, 2017; Trilla-Fuertes et al, 2018). Briefly, each functional
280 node activity was calculated as the mean of the expression/quantity of genes/metabolites of
281 each node that are related to the main node function/metabolic pathway.

282 Flux Balance Analysis and flux activities

283 FBA is a method that allows the estimation of tumor growth rate. FBA was performed using
284 the library COBRA Toolbox v2.0 (Schellenberger et al, 2011) available for MATLAB. FBA was
285 calculated using the whole human metabolic reconstruction Recon2 (Thiele et al, 2013). This
286 metabolic reconstruction includes 2,191 genes collected into the Gene Protein Reaction rules
287 (GPRs), 5,063 metabolites and 7,440 reactions. GPRs represent the relationships between
288 genes and metabolic reactions and they are included into the model as Boolean expressions.
289 GPRs were solved as described in previous studies (Gámez-Pozo et al, 2017; Trilla-Fuertes et al,
290 2018), using a modification of the Barker *et al.* algorithm (Barker et al, 2015), which were
291 incorporated into the model by a modified E-flux method (Colijn et al, 2009; Trilla-Fuertes et
292 al, 2018). Briefly, the "OR" operators were solved as the sum and the "AND" operators were
293 solved as the minimum. Then, the GPR data were normalized using the Max-min function and
294 introduced into the model as the reaction bounds. As the objective function, the biomass
295 reaction proposed in the Recon2 was used as representative of tumor growth. This biomass
296 reaction was based on experimental measurements of leukemia cells. The 7,440 reactions are
297 grouped into 101 metabolic pathways.

298 Flux activities were previously proposed as a measurement to compare differences at the
299 metabolic pathway level (Trilla-Fuertes et al, 2018). Briefly, they were calculated as the sum of
300 the fluxes of the reactions included in each pathway defined in Recon2.

301 Statistical analyses

302 The statistical analyses were performed with GraphPad Prism v6. Predictor signatures were
303 built with the BRB Array Tool from Dr. Richard Simon's team (Simon, 2005). All p-values are
304 two-sided and are considered statistically significant under 0.05.

305 Cell culture and reagents

306 Breast cancer cell lines (MCF7, T47D and CAMA1 [ER+], and MDAB231, MDAMB468 and
307 HCC1143 [TNBC]) were cultured in RPMI-1640 medium with phenol red, supplemented with
308 10% heat-inactivated fetal bovine serum, 100 mg/mL penicillin and 100 mg/mL streptomycin.
309 Cell lines were cultured at 37°C in a humidified atmosphere with 5% (v/v) CO2 in the air. The
310 MCF7, T47D and MDA-MB-231 cell lines were kindly provided by Dr. Nuria Vilaboa (La Paz
311 University Hospital, previously obtained from ATCC in January 2014). The MDAMB468, CAMA1
312 and HCC1143 cell lines were obtained from ATCC (July 2014). Cell lines were routinely
313 monitored and authenticated, tested for Mycoplasma and frozen, and passaged for fewer than
314 6 months before experiments. The AOA (Sigma Aldrich C13408) and GPNA (Sigma Aldrich
315 G6133) were obtained from Sigma-Aldrich (St. Louis, MO, USA).

316 Cell viability assays

317 Dose-response curves were designed for AOA and GPNA. As the preparation of GPNA needs an
318 acid medium, HEPES (50mM) was added to buffer the medium. About 5,000 cells were seeded
319 in each well in 96-well plates and after 24 hours, drugs were added. After an incubation of 72
320 hours, cell viability was determined using CellTiter 96 AQueous One Solution Cell Proliferation
321 Assay (Promega) kit and absorbance was quantified on a microplate reader (TECAN). As a
322 control untreated cells were used and all the experiments were performed by triplicate.

Acknowledgements

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 and EL-C are supported by Consejería 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. EL-C did the cell viability experiments. LT-F drafted the manuscript. AG-P, JAFV, and EE supported the manuscript drafting. AG-P and JAFV coordinated the study.

Conflict of interest

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.

Data Availability Section

Not applicable.

References

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

Abreu G, Edwards D, Labouriau R. (2010) High-Dimensional Graphical Model Search with the gRapHD R Package Journal of Statistical Software, Vol. 37, pp. 1-18.

Barker BE, Sadagopan N, Wang Y, Smallbone K, Myers CR, Xi H, Locasale JW, Gu Z (2015) A robust and efficient method for estimating enzyme complex abundance and metabolic flux from expression data. *Comput Biol Chem* **59 Pt B:** 98-112

Bhowmik SK, Ramirez-Peña E, Arnold JM, Putluri V, Sphyris N, Michailidis G, Putluri N, Ambs S, Sreekumar A, Mani SA (2015) EMT-induced metabolite signature identifies poor clinical outcome. *Oncotarget* **6:** 42651-42660

Cao MD, Lamichhane S, Lundgren S, Bofin A, Fjøsne H, Giskeødegård GF, Bathen TF (2014) Metabolic characterization of triple negative breast cancer. *BMC Cancer* **14:** 941

Cao MD, Sitter B, Bathen TF, Bofin A, Lønning PE, Lundgren S, Gribbestad IS (2012) Predicting long-term survival and treatment response in breast cancer patients receiving neoadjuvant chemotherapy by MR metabolic profiling. *NMR Biomed* **25:** 369-378

Cavill R, Kamburov A, Ellis JK, Athersuch TJ, Blagrove MS, Herwig R, Ebbels TM, Keun HC (2011) Consensus-phenotype integration of transcriptomic and metabolomic data implies a role for metabolism in the chemosensitivity of tumour cells. *PLoS Comput Biol* **7:** e1001113

Colijn C, Brandes A, Zucker J, Lun D, Weiner B, Farhat M, Cheng T, Moody B, Murray M, Galagan J. (2009) Interpreting expression data with metabolic flux models: Predicting Mycobacterium tuberculosis mycolic acid production. PLOS Comput Bio, Vol. 5.

de Velasco G, Trilla-Fuertes L, Gamez-Pozo A, Urbanowicz M, Ruiz-Ares G, Sepúlveda JM, Prado-Vazquez G, Arevalillo JM, Zapater-Moros A, Navarro H, Lopez-Vacas R, Manneh R, Otero I, Villacampa F, Paramio JM, Vara JAF, Castellano D (2017) Urothelial cancer proteomics provides both prognostic and functional information. *Sci Rep* **7:** 15819

DeBerardinis RJ, Mancuso A, Daikhin E, Nissim I, Yudkoff M, Wehrli S, Thompson CB (2007) 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* **104:** 19345-19350

Eagle H, Oyama VI, LEVY M, Horton CL, Fleischman R (1956) The growth response of mammalian cells in tissue culture to L-glutamine and L-glutamic acid. *J Biol Chem* **218:** 607-616

386
387 Emwas AH (2015) The strengths and weaknesses of NMR spectroscopy and mass spectrometry
388 with particular focus on metabolomics research. *Methods Mol Biol* **1277:** 161-193

389
390 Fiehn O (2002) Metabolomics--the link between genotypes and phenotypes. *Plant Mol Biol* **48:**
391 155-171

392
393 Fuhrer T, Zamboni N (2015) High-throughput discovery metabolomics. *Curr Opin Biotechnol* **31:**
394 73-78

395
396 Gowda GA, Zhang S, Gu H, Asiago V, Shanaiah N, Raftery D (2008) Metabolomics-based
397 methods for early disease diagnostics. *Expert Rev Mol Diagn* **8:** 617-633

398
399 Gámez-Pozo A, Berges-Soria J, Arevalillo JM, Nanni P, López-Vacas R, Navarro H, Grossmann J,
400 Castaneda C, Main P, Díaz-Almirón M, Espinosa E, Ciruelos E, Fresno-Vara JA. (2015) Combined
401 label-free quantitative proteomics and microRNA expression analysis of breast cancer unravel
402 molecular differences with clinical implications. Cancer Res, Vol. 75, pp. 2243-2253.

403
404 Gámez-Pozo A, Trilla-Fuertes L, Berges-Soria J, Selevsek N, López-Vacas R, Díaz-Almirón M,
405 Nanni P, Arevalillo JM, Navarro H, Grossmann J, Gayá Moreno F, Gómez Rioja R, Prado-
406 Vázquez G, Zapater-Moros A, Main P, Feliú J, Martínez del Prado P, Zamora P, Ciruelos E,
407 Espinosa E et al (2017) Functional proteomics outlines the complexity of breast cancer
408 molecular subtypes. *Scientific Reports* **7:** 10100

409
410 Hassanein M, Hoeksema MD, Shiota M, Qian J, Harris BK, Chen H, Clark JE, Alborn WE,
411 Eisenberg R, Massion PP (2013) SLC1A5 mediates glutamine transport required for lung cancer
412 cell growth and survival. *Clin Cancer Res* **19:** 560-570

413
414 Huang dW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene
415 lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44-57

416
417 Jiang S, Yan W (2017) Succinate in the cancer-immune cycle. *Cancer Lett* **390:** 45-47

418
419 Kanaan YM, Sampey BP, Beyene D, Esnakula AK, Naab TJ, Ricks-Santi LJ, Dasi S, Day A,
420 Blackman KW, Frederick W, Copeland RL, Gabrielson E, Dewitty RL (2014) Metabolic profile of
421 triple-negative breast cancer in African-American women reveals potential biomarkers of
422 aggressive disease. *Cancer Genomics Proteomics* **11:** 279-294

423
424 Korangath P, Teo WW, Sadik H, Han L, Mori N, Huijts CM, Wildes F, Bharti S, Zhang Z, Santa-
425 Maria CA, Tsai H, Dang CV, Stearns V, Bhujwalla ZM, Sukumar S (2015) Targeting Glutamine
426 Metabolism in Breast Cancer with Aminooxyacetate. *Clin Cancer Res* **21:** 3263-3273

427

428 Lu Q, Wang C, Pan R, Gao X, Wei Z, Xia Y, Dai Y (2013) Histamine synergistically promotes
429 bFGF-induced angiogenesis by enhancing VEGF production via H1 receptor. *J Cell Biochem* **114:**
430 1009-1019

431
432 Orth J, Thiele I, Palsson B. (2010) What is flux balance analysis? Nat Biotechnol, Vol. 28, pp.
433 245-248.

434
435 Peng X, Chen Z, Farshidfar F, Xu X, Lorenzi PL, Wang Y, Cheng F, Tan L, Mojumdar K, Du D, Ge Z,
436 Li J, Thomas GV, Birsoy K, Liu L, Zhang H, Zhao Z, Marchand C, Weinstein JN, Bathe OF et al
437 (2018) Molecular Characterization and Clinical Relevance of Metabolic Expression Subtypes in
438 Human Cancers. *Cell Rep* **23:** 255-269.e254

439
440 Ruike T, Kanai Y, Iwabata K, Matsumoto Y, Murata H, Ishima M, Ohta K, Oshige M, Katsura S,
441 Kuramochi K, Kamisuki S, Sahara H, Miura M, Sugawara F, Sakaguchi K (2018) Distribution and
442 metabolism of. *Xenobiotica***:** 1-45

443
444 Schellenberger J, Que R, Fleming R, Thiele I, Orth J, Feist A, Zielinski D, Bordbar A, Lewis N,
445 Rahmanian S, Kang J, Hyduke D, Palsson B. (2011) Quantitative prediction of cellular
446 metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols, Vol. 6,
447 pp. 1290-1307.

448
449 Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T
450 (2003) Cytoscape: a software environment for integrated models of biomolecular interaction
451 networks. *Genome Res* **13:** 2498-2504

452
453 Siegel RL, Miller KD, Jemal A (2018) Cancer statistics, 2018. *CA Cancer J Clin* **68:** 7-30

454
455 Simon R (2005) Roadmap for developing and validating therapeutically relevant genomic
456 classifiers. *J Clin Oncol* **23:** 7332-7341

457
458 Sánchez-Navarro I, Gámez-Pozo A, Pinto A, Hardisson D, Madero R, López R, San José B,
459 Zamora P, Redondo A, Feliu J, Cejas P, González Barón M, Angel Fresno Vara J, Espinosa E
460 (2010) An 8-gene qRT-PCR-based gene expression score that has prognostic value in early
461 breast cancer. *BMC Cancer* **10:** 336

462
463 Tannahill GM, Curtis AM, Adamik J, Palsson-McDermott EM, McGettrick AF, Goel G, Frezza C,
464 Bernard NJ, Kelly B, Foley NH, Zheng L, Gardet A, Tong Z, Jany SS, Corr SC, Haneklaus M,
465 Caffrey BE, Pierce K, Walmsley S, Beasley FC et al (2013) Succinate is an inflammatory signal
466 that induces IL-1β through HIF-1α. *Nature* **496:** 238-242

467
468 Terunuma A, Putluri N, Mishra P, Mathé EA, Dorsey TH, Yi M, Wallace TA, Issaq HJ, Zhou M,
469 Killian JK, Stevenson HS, Karoly ED, Chan K, Samanta S, Prieto D, Hsu TY, Kurley SJ, Putluri V,
470 Sonavane R, Edelman DC et al (2014) MYC-driven accumulation of 2-hydroxyglutarate is
471 associated with breast cancer prognosis. *J Clin Invest* **124:** 398-412

472
473 Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML,
474 Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bölling C, Bordel S, Chavali AK, Dobson P,
475 Dunn WB, Endler L, Hala D, Hucka M et al (2013) A community-driven global reconstruction of
476 human metabolism. *Nat Biotechnol* **31:** 419-425

477
478 Trilla-Fuertes L, Gámez-Pozo A, Arevalillo JM, Díaz-Almirón M, Prado-Vázquez G, Zapater-
479 Moros A, Navarro H, Aras-López R, Dapía I, López-Vacas R, Nanni P, Llorente-Armijo S, Arias P,
480 Borobia AM, Maín P, Feliú J, Espinosa E, Fresno Vara J (2018) Molecular characterization of
481 breast cancer cell response to metabolic drugs. *Oncotarget* **9:** 9645-9660

482
483 Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J (2016) The Perseus
484 computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13:**
485 731-740

486
487 Varma A, Palsson BO (1995) Parametric sensitivity of stoichiometric flux balance models
488 applied to wild-type Escherichia coli metabolism. *Biotechnol Bioeng* **45:** 69-79

489
490 Wachowska M, Gabrysiak M, Muchowicz A, Bednarek W, Barankiewicz J, Rygiel T, Boon L,
491 Mroz P, Hamblin MR, Golab J (2014) 5-Aza-2'-deoxycytidine potentiates antitumour immune
492 response induced by photodynamic therapy. *Eur J Cancer* **50:** 1370-1381

493
494 Wang J, Shidfar A, Ivancic D, Ranjan M, Liu L, Choi MR, Parimi V, Gursel DB, Sullivan ME, Najor
495 MS, Abukhdeir AM, Scholtens D, Khan SA (2017) Overexpression of lipid metabolism genes and
496 PBX1 in the contralateral breasts of women with estrogen receptor-negative breast cancer. *Int*
497 *J Cancer* **140:** 2484-2497

498
499 Wise DR, DeBerardinis RJ, Mancuso A, Sayed N, Zhang XY, Pfeiffer HK, Nissim I, Daikhin E,
500 Yudkoff M, McMahon SB, Thompson CB (2008) Myc regulates a transcriptional program that
501 stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc Natl Acad Sci U*
502 *S A* **105:** 18782-18787

503

504

505 # Figure legends

506 **Fig 1: Predictive signature built using metabolomics data.**

507 **Fig 2: Probabilistic graphical model from metabolomics data.**

508 **Fig 3: Predictor based on lipid metabolism node activity.**

509 **Fig 4: A. Network associating genes (red) and metabolites (blue). B. Metabolite and gene**

510 **network functionally characterized.**

511 **Fig 5: OS predictor based on glutamate metabolism and alanine and aspartate metabolism**

512 **flux activities.**

513 **Fig 6: Probabilistic graphical model combining flux activities and metabolomics data.** A.

514 Network combining flux activities (purple) and metabolite (pink) expression data. B. Division in

515 branches of the network formed by flux activities and metabolomics data.

516 **Figure 7: Dose-response curves using two drugs targeting glutamine metabolism in breast**

517 **cancer cell lines.** A. Dose-response curve for AOA (0-6 Mm). B. Dose-response curve for GPNA

518 (0-4 Mm).

519 # Tables and their legends

	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%)
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 (6%)	4 (12%)
No	50 (75%)	26 (79%)	24 (70%)
Missing	11 (16%)	5 (15%)	6 (18%)

520 **Table 1: Patient characteristics.**

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

521 **Table 2: Previously described relationships between metabolites included in gene nodes and**

522 **the function of these nodes.**

Cell Line	Subtype	AOA IC_{50} (mM)	GPNA IC_{50} (mM)
T47D	ER+	2.05	0.48
MCF7	ER+	3.89	0.69
CAMA1	ER+	2.90	1.10
MDAMB231	TNBC	0.64	1.73
MDAMB468	TNBC	2.29	2.50
HCC1143	TNBC	4.22	2.59

523 Table 3: IC_{50} calculated for each drug in each breast cancer cell line.

524 # Supporting information

525 **S1 Table: Multivariate Cox regression model comparing OS predictor based on metabolomics**

526 **data.** T = tumor stage, N = lymph node status, G = tumor grade.

527 **S2 Table: Multivariate Cox analysis comparing predictor based on node activity of lipid**

528 **metabolism.** T = tumor stage, N = lymph node status, G = tumor grade.

529 **S3 Table: Multivariate Cox regression comparing predictor based on flux activities.** T = tumor

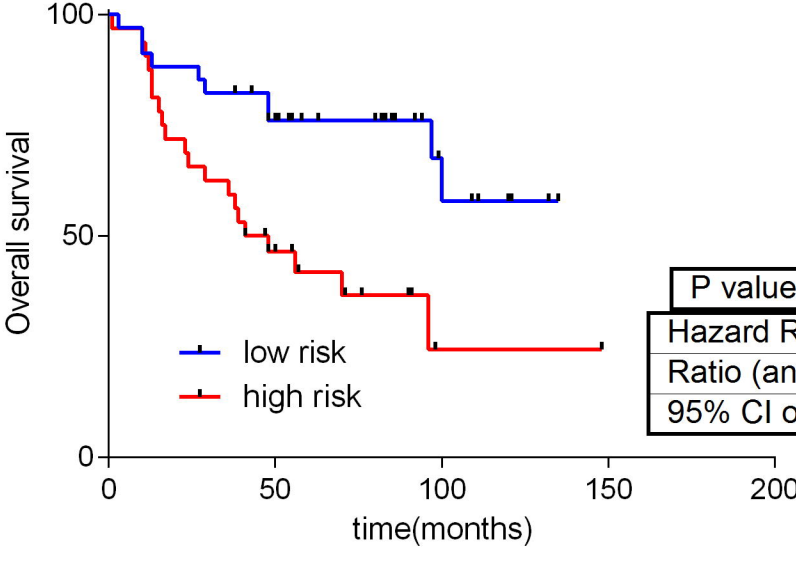
530 stage, N = lymph node status, G = tumor grade.

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

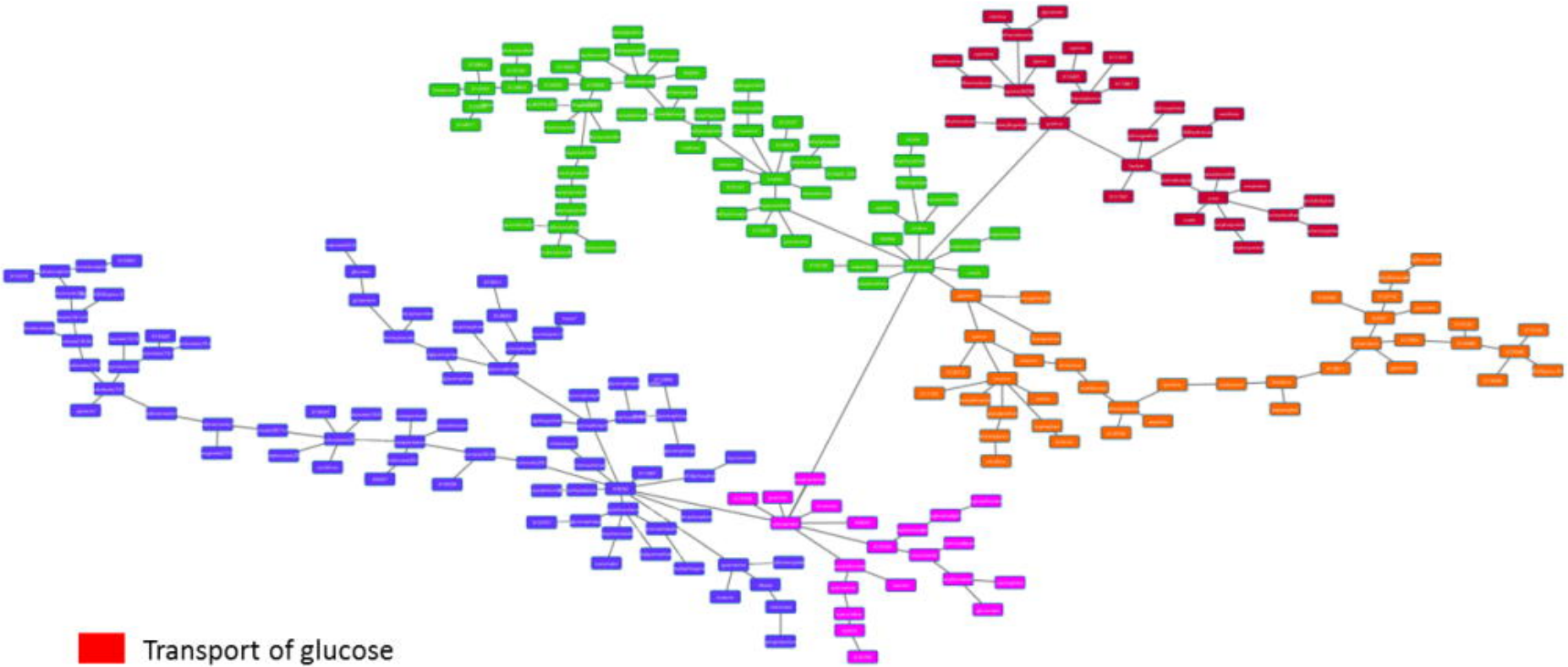
532 **S1 Fig: Node activities from the metabolic network.**

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

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

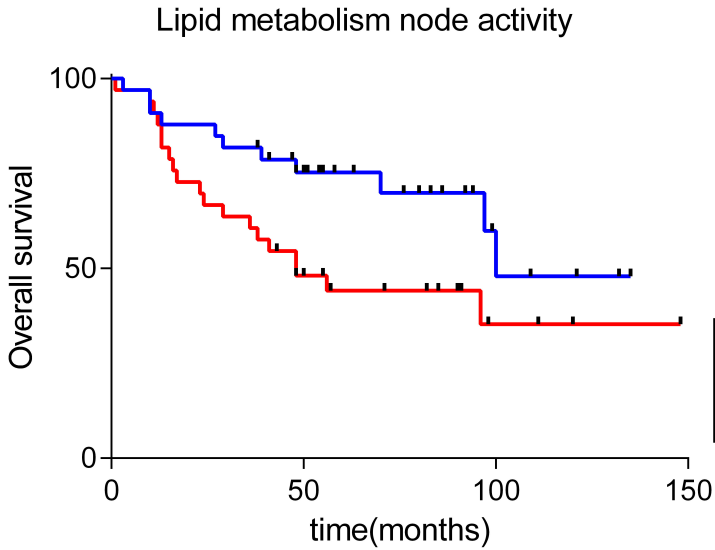


P value	0.0030
Hazard Ratio (logrank)	A/B
Ratio (and its reciprocal)	0.3422
95% CI of ratio	0.1557 to 0.6687



Transport of glucose

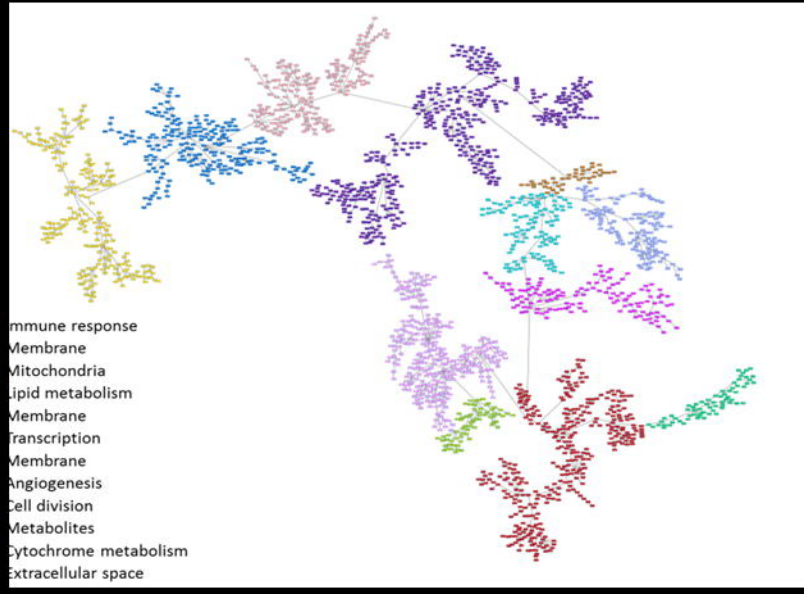
Aminoacid interconversion

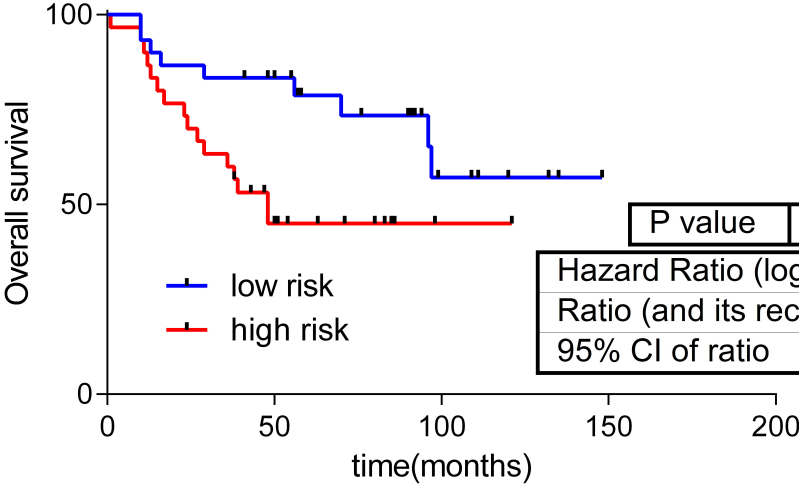


P value	0.0452

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

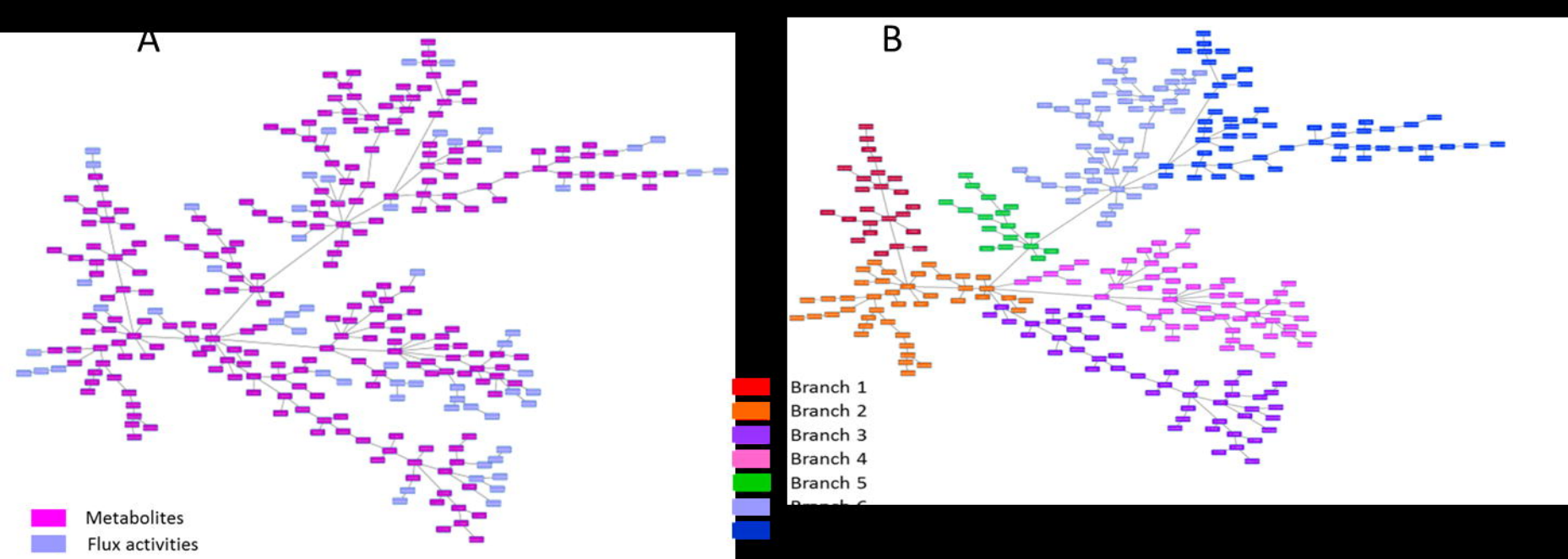
Genes
Metabolites


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

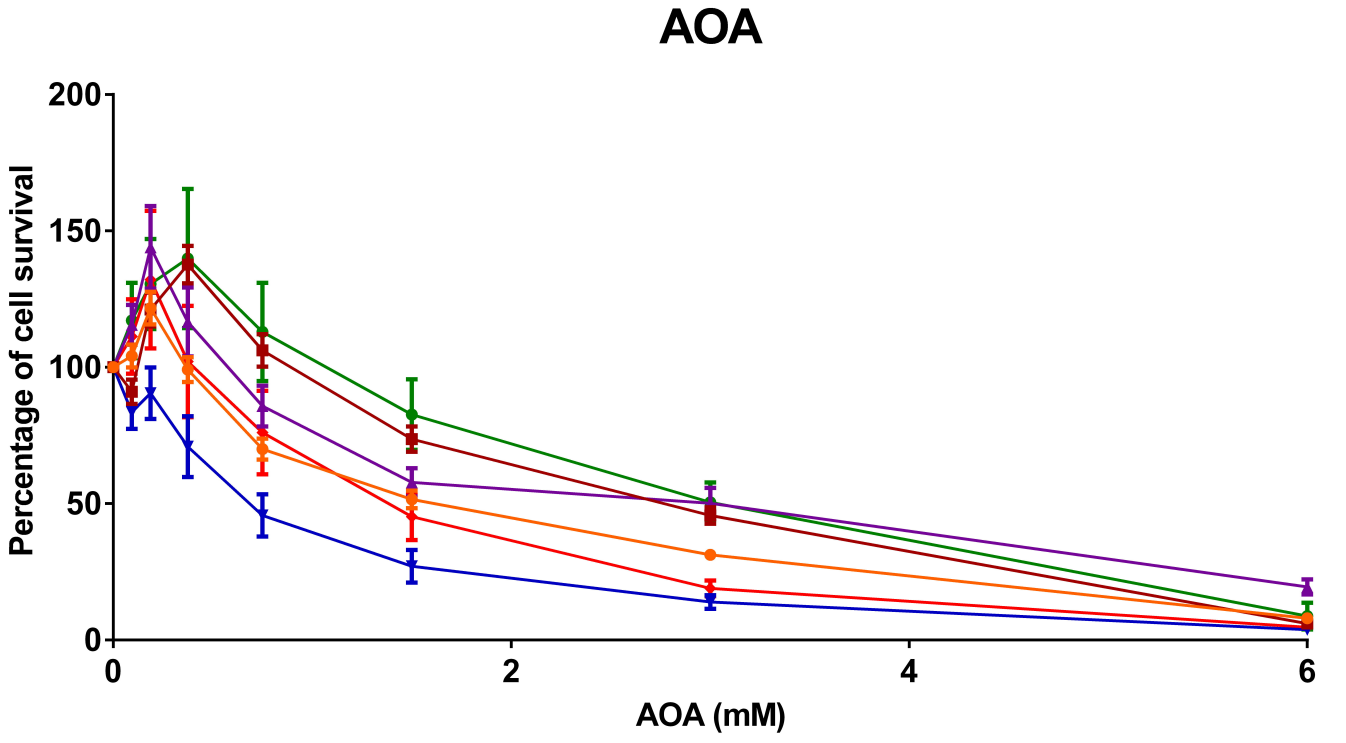


P value	0.0243

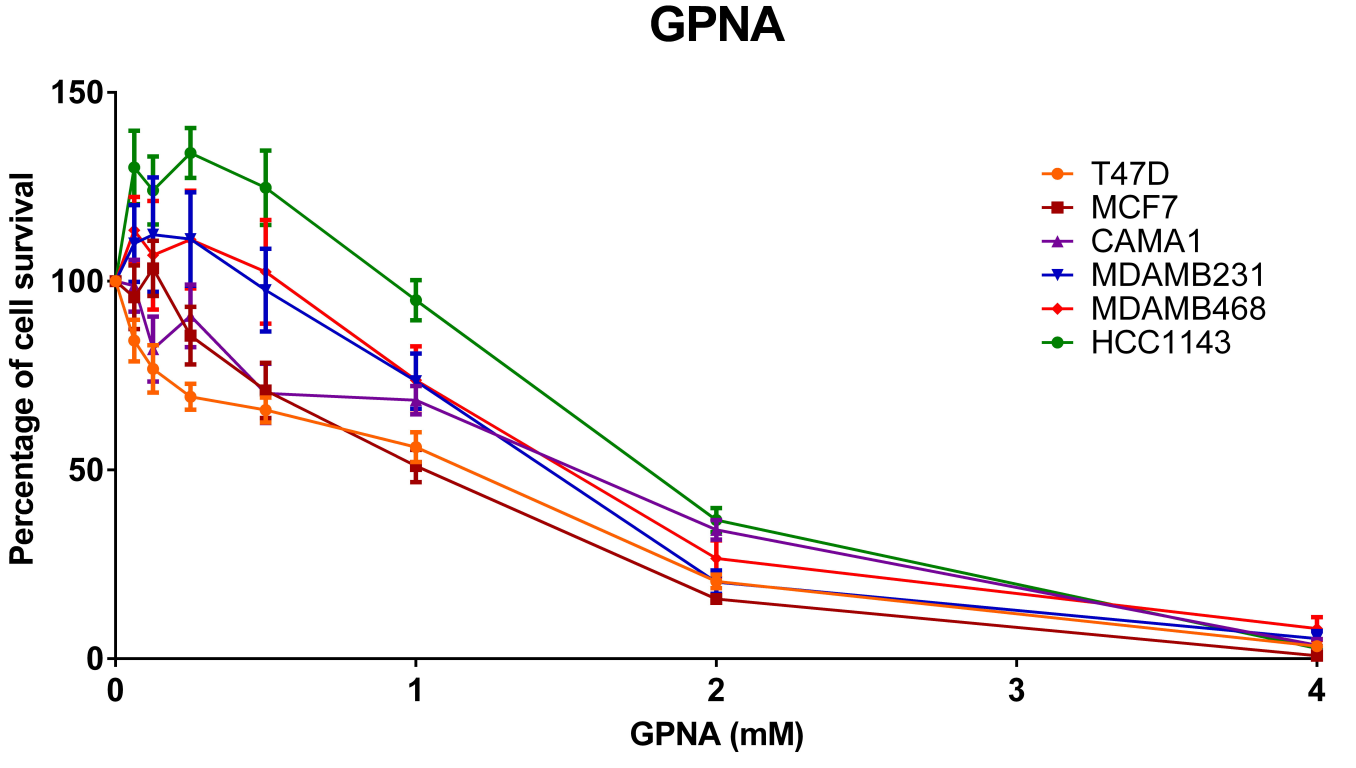
Hazard Ratio (logrank)	A/B
Ratio (and its reciprocal)	0.4110
95% CI of ratio	0.1782 to 0.8713


A
B
Metabolites
Flux activities
Branch 1
Branch 2
Branch 3
Branch 4
Branch 5

A


AOA
Percentage of cell survival
200
150
100
50
0
0
2
4
6
AOA (mM)

B


GPNA
Percentage of cell survival
150
100
50
0
0
1
2
3
4
GPNA (mM)
T47D
MCF7
CAMA1
MDAMB231
MDAMB468
HCC1143