METAtarget – extracting key enzymes of metabolic regulation from high-throughput metabolomics data using KEGG REACTION information

Jan Budczies[1*], Carsten Denkert[1], Berit M Müller[1], Scarlet F Brockmöller[1], Manfred Dietel[1], Jules L Griffin[2], Matej Oresic[3], Oliver Fiehn[4]

[1]Institute of Pathology, Charité – Universitätsmedizin Berlin, 10117 Berlin, Germany
[2]Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom
[3]VTT Technical Research Centre of Finnland, Espoo, Finnland
[4]Genome Center, University of California Davis, Davis, CA, USA
[*]Email: jan.budczies@charite.de

Abstract: METAtarget is a new method for reverse engineering of metabolic networks and the detection of targets enzymes from high-throughput metabolomics data. Using KEGG REACTION, reactant partners are identified and the ratio of product to substrate metabolite concentrations is employed as surrogate for the reaction activity. A test statistics is introduced to assess changes in the activity of reactions between different disease states. In an application of METAtarget to breast cancer, we investigate the dependence of tumor metabolism on hormone receptor status. To this end, we analyze metabolomics data that were generated within the METAcancer project and compare the identified reactions with data on enzyme expression that are obtained from publicly available breast cancer gene expression series. As result, deregulation of key enzymes and reactions of glycolysis, glutaminolysis and other metabolic pathways are detected.

1 Introduction

In recent years, techniques for metabolic profiling based on mass spectrometry (MS) and nuclear magnetic resonance spectrometry (NMR) advanced and now allow the simultaneous monitoring of hundreds of metabolites [Fi01, GS04]. Metabolomics emerged as an additional high-throughput technology complementary to other -omics approaches like genomics, transcriptomics and proteomics. In cancer research, liquid and gas chromatography-based MS have been successfully applied to the analysis of body fluids and tissues [De08, De09, Sr09].

Uncovering of the biochemical pathways that constitute the human metabolism represents one of the major achievements of biochemical research over the past 100 years [GHW05]. This knowledge has been an invaluable information source to improve human health by providing new insights into nutrition, disease mechanisms and the effect of drugs. In the postgenomic era metabolic pathway knowledge has been integrated with information from sequencing of genomes and is publicly available from databases like KEGG [Ka10], Reactome [Ma09] and BioCarta (www.biocarta.com).

Consequently, suitable tools for the integration of metabolomics data together with pathway knowledge are urgently needed and will facilitate and accelerate the interpretation of experimental data. Currently, several tools are available for the visualization of metabolomics data in context of biochemical networks. Metscape [Ga10] and MetaNetter [Jo08] are plugins for cytoscape, a powerful and widely-used software environment for models of biomolecular interaction networks [Sh03]. Web-based metabolic network explorers include the KEGG atlas [Ok08] and iPath [Le08]. Going beyond a network visualization, PROFILE clustering orders metabolites according to their distance in KEGG pathways and visualizes metabolic changes in context of the functional clustering [De08]. TICL is a tool for network generation from metabolite lists that includes a significance assessment for the relevance of the generated networks [An09].

Here we present METAtarget, a method for quantitative analysis of metabolomics data in context of biochemical pathways. METAtarget employs the ratio of product to substrate concentrations as surrogate for the activity of metabolic reactions. A suitable test statistic is defined in order to measure changes of reaction activities between two disease states. METAtarget delivers a list significantly changed reactions and the associated enzymes that are possible targets for a therapeutic intervention.

As an application of METAtarget, we analyze GC-MS data that were generated in the framework of METAcancer, a European collaboration on the metabolism of breast cancer. Comparing metabolic profiles of estrogen receptor positive (ER+) and receptor negative (ER-) breast cancer, METAtarget delivers a list of metabolic reactions that are regulated depending on hormone receptor status. Using an independent gene expression data set on breast cancer, we evaluate the hypothesis that the detected changes in metabolism are associated with transcriptional regulation of enzymes.

2 Material and Methods

2.1 Assessing the activity of metabolic reactions

The simplest design of a metabolomics experiment deals with the comparison of diseased and healthy tissue or of different tissues types in disease states *a* and *b*. Let us denote the concentration of a metabolite Z in two tissue types by Z_a and Z_b. As it is commonpraxis for the analysis of –omics data, we assume that the variables Z are transformed to the log-scale. For the log-scale concentrations of the product X and the substrate Y of a metabolic reaction we define the statistics

$$t = \frac{\mathrm{E}(X_a) - \mathrm{E}(X_b) - \mathrm{E}(Y_a) + \mathrm{E}(Y_b)}{\sqrt{\mathrm{VAR}(X_a - Y_a)/N_a + \mathrm{VAR}(X_b - Y_b)/N_b}},$$

wherein N_a and N_b are the numbers of tissues in disease states *a* and *b*. The numerator of *t* can be interpreted in two ways: (i) as a different effect of the disease state on the product compared to the substrate and (ii) as a measure how the product-substrate ratio changes when comparing the two disease states *a* and *b*. The statistics defined above can be read as Welch's t-statistics in the difference variable $X - Y$. Significance is assessed by Welch's t-test.

METAtarget extracts information about reactions and enzymes from the KEGG database (www.genome.jp/kegg). Only metabolites annotated as "main" reaction partners in KEGG RPAIR are considered as pair of substrate and product. For each pair of substrate and product, the regulation of the reaction is assessed by the statistics *t* and Welch's t-test. P-values < 0.05 after Bonferroni correction for the number of tested reactions are considered statistically significant.

2.2 Breast cancer metabolomics data

METAcancer (www.metacancer-fp7.eu) is an EU-funded project aiming at the analysis of the breast cancer metabolome and the discovery of new molecular markers. A series of more than 200 breast cancers was investigated using three different metabolic platforms, GC-MS, LC-MS and NMC. Here, we analyze the METAcancer GC-MS data that include measurements of 124 KEGG annotated metabolites in 188 ER+ and 58 ER- breast cancers. Estrogene receptor status of the METAcancer samples was determined immunohistologically, tumors with ≥ 10% ER positive cells were considered as ER positive. Prior to analysis, metabolomics data were transformed to the log2-scale.

2.3 Breast cancer transcriptomics data

Three publicly available breast cancer gene expression series GSE2034, GSE7390 and GSE11121 were downloaded from the GEO repository (www.ncbi.nlm.nih.gov/geo). All data sets were generated using the same kind of microarrays (Affymetrix HG-U133A GeneChips). The expression series were merged to a large expression data set of 684 nodal negative breast cancers. Microarray data were preprocessed with the standard mas5 method and transformed to the log2-scale. As immunohistological data for estrogen receptor (ER) status were not available for all samples, ER status was derived from gene expression data. Samples with ESR1 absolute expression ≥ 10 (measured by probe set 205225_at) were considered as ER positive, samples with ESR1 expression < 10 as ER negative. 176 of the 684 breast cancers were ER negative, 508 ER positive. Significance of differential expression between ER+ and ER- tumors was assessed by Welch's t-test. P-values < 0.05 after Bonferroni correction for the number of genes were considered statistically significant.

3 Results

GC-MS profiling of 246 breast cancers within the METAcancer project led to the identification of 468 metabolites. 162 out of these could be mapped to known chemical structures and metabolite names, 124 could be found in the KEGG database. Using KEGG RPAIR we identified 91 substrate-product pairs that were main reactants in metabolic reactions.

Next, we analyzed the substrate-product pairs for differential regulation between ER+ and ER- breast cancers. Using the statistics *t* (cf. material and methods section) we detected 13 differentially regulated pairs of reactants. Bar plots show the differential expression between ER+ and ER- breast cancer for substrates and products (Fig. 1A) and the substrate / product ratio for ER+ and ER- breast cancer (Fig. 1B).

Using KEGG REACTION the 13 reactant pairs could be mapped to 51 metabolic reactions (Tab. 1). In KEGG REACTION, the reactant pairs are stored with the EC numbers of the catalysing reactions. This information, together with the information on the human genome was used to map the reactant pairs to 29 human genes.

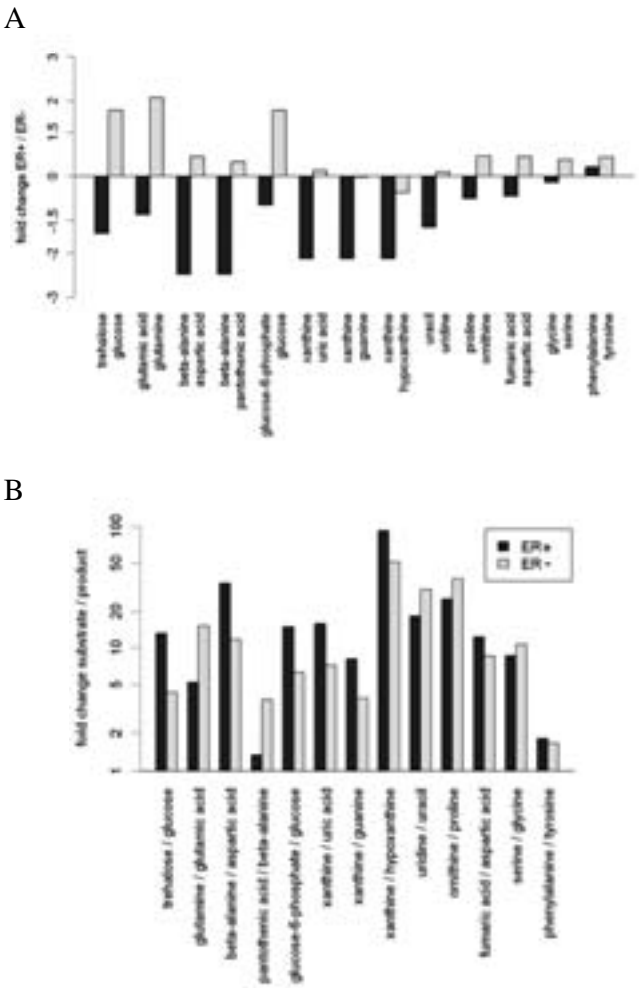

Fig. 1: Substrate-product pairs that are differently regulated in ER+ compared to ER- breast cancer. Significance was assessed by the statistics t (cf. material and methods). **A** Fold change between ER+ and ER- breast cancer in pairs of substrates and products. **B** Fold change between substrates and products in ER+ and ER- breast cancer.

Tab. 1: Differently regulated metabolic reactions between ER+ and ER- breast cancer. After mapping of metabolites to reactions using KEGG RPAIR, differently regulated substrate-product pairs were detected by the analysis of metabolomics data. For each substrate-product pair the numerator of the statistics *t*, a difference of differences (dd), is reported. Catalyzing enzymes were identified using KEGG ENZYME. Enzymes were investigated for differential expression between ER+ and ER- tumors using an independent breast cancer genes expression data set. Meaning of the check marks behind the gene symbols: "+" = significant up-regulation, "-" = significant down-regulation, "~" = no differential regulation, no check mark = not represented by the microarray.

substrate	product	dd	reaction	enzymes
glucose	trehalose	-3.1	alpha,alpha-Trehalose + H2O <=> 2 D-Glucose	TREH~
glutamic acid	glutamine	2.9	ATP + L-Glutamate + NH3 <=> ADP + Orthophosphate + L-Glutamine	GLUL+
			L-Glutamine + H2O <=> L-Glutamate + NH3	ASNS-, CAD-,GLS-, GLS2+
			ATP + Deamino-NAD+ + L-Glutamine + H2O <=> AMP + Diphosphate + NAD+ + L-Glutamate	NADSYN1+
			ATP + UTP + L-Glutamine + H2O <=> ADP + Orthophosphate + CTP + L-Glutamate	CTPS-, CTPS2~
			2 ATP + L-Glutamine + HCO3- + H2O <=> 2 ADP + Orthophosphate + L-Glutamate + Carbamoyl phosphate	CAD-
			ATP + L-Aspartate + L-Glutamine + H2O <=> AMP + Diphosphate + L-Asparagine + L-Glutamate	ASNS-
			L-Glutamine + D-Fructose 6-phosphate <=> L-Glutamate + D-Glucosamine 6-phosphate	GFPT1-, GFPT2~
			5-Phosphoribosylamine + Diphosphate + L-Glutamate <=> L-Glutamine + 5-Phospho-alpha-D-ribose 1-diphosphate + H2O	PPAT-
			ATP + Xanthosine 5'-phosphate + L-Glutamine + H2O <=> AMP + Diphosphate + GMP + L-Glutamate	GMPS-
			ATP + 5'-Phosphoribosyl-N-formylglycinamide + L-Glutamine + H2O <=> ADP + Orthophosphate + 2-(Formamido)-N1-(5'-phosphoribosyl)acetamidine + L-Glutamate	PFAS+
aspartic acid	beta-alanine	-2.9	L-Aspartate <=> beta-Alanine + CO2	GAD1+, GAD2~
beta-alanine	pantothenic acid	2.8	ATP + (R)-Pantoate + beta-Alanine <=> AMP + Diphosphate + Pantothenate	
			Pantothenate + H2O <=> (R)-Pantoate + beta-Alanine	
glucose	glucose-6-phosphate	-2.4	ATP + D-Glucose <=> ADP + D-Glucose 6-phosphate	GCK~, HK1~, HK2~, HK3-
			D-Glucose 6-phosphate + H2O <=> D-Glucose + Orthophosphate	G6PC~, G6PC2~
			ITP + D-Glucose <=> IDP + D-Glucose 6-phosphate	HK1~, HK2~, HK3-
			dATP + D-Glucose <=> dADP + D-Glucose 6-phosphate	HK1~, HK2~, HK3-
uric acid	xanthine	-2.2	Xanthine + NAD+ + H2O <=> Urate + NADH + H+	XDH~
			Xanthine + H2O + Oxygen <=> Urate + H2O2	XDH~
guanine	xanthine	-2.1	Guanine + H2O <=> Xanthine + NH3	GDA
hypoxanthine	xanthine	-1.8	Hypoxanthine + NAD+ + H2O <=> Xanthine + NADH + H+	XDH~
			Hypoxanthine + Oxygen + H2O <=> Xanthine + H2O2	XDH~
uracil	uridine	1.7	Uridine + H2O <=> Uracil + D-Ribose	
			Uridine + Orthophosphate <=> Uracil + alpha-D-Ribose 1-phosphate	UPP1-, UPP2
ornithine	proline	-1.5	L-Ornithine <=> L-Proline + NH3	
aspartic acid	fumaric acid	-1.4	L-Aspartate <=> Fumarate + NH3	
glycine	serine	1.2	5,10-Methylenetetrahydrofolate + Glycine + H2O <=> Tetrahydrofolate + L-Serine	SHMT1~, SHMT2-
phenylalanine	tyrosine	1.1	Tetrahydrobiopterin + L-Phenylalanine + Oxygen <=> Dihydrobiopterin + L-Tyrosine + H2O	PAH+
			L-Phenylalanine + Tetrahydrobiopterin + Oxygen <=> L-Tyrosine + 4a-Hydroxytetrahydrobiopterin	PAH+

To evaluate the hypothesis that metabolic changes are associated with transcriptional regulation of enzymes, we analyzed an independent gene expression data set of 684 breast cancers. 27 of the 29 enzymes identified before were represented on the microarray, 16 were differentially expressed between ER+ and ER- breast cancer. As shown in Tab. 1, six enzymes were up-regulated in ER+ breast cancer, while ten enzymes were down-regulated.

Examples for the detected reactions are conversion of glucose to glucose-6-phosphate being the first step of glycolysis and conversion of glutamine to glutamate being the first step of glutaminolysis. Both reactions belong to catabolic pathways that can be used for the production of energy and have been described as up-regulated in cancer cells [VCT09]. Fig. 2 shows the differential expression of the human enzymes catalysing these reactions.

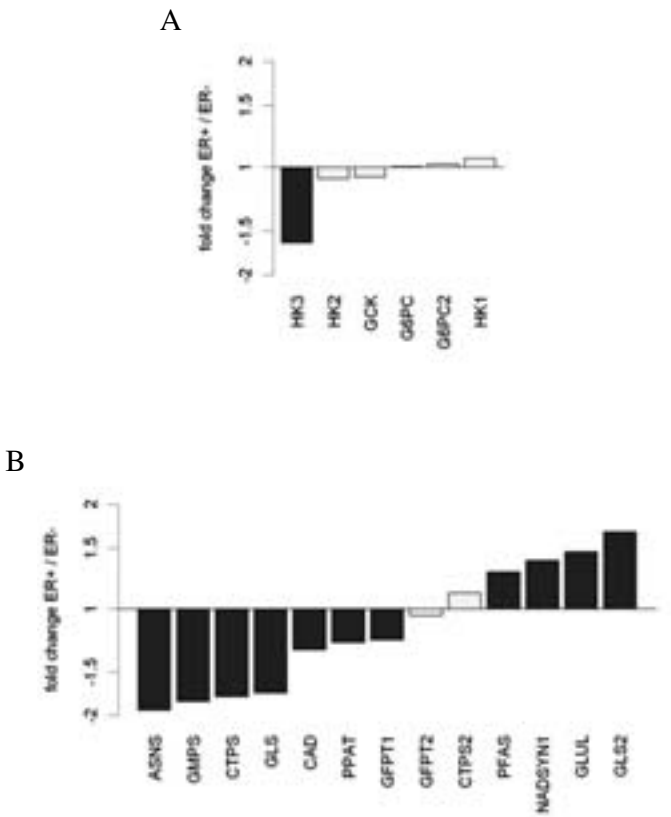

Fig. 2: Differentially expressed enzymes between ER+ and ER- breast cancer. **A** Enzymes catalyzing the first step of glycolysis, glucose -> glucose-6-phosphate. **B** Enzymes catalyzing the first step of glutaminolysis, glutamine -> glutamate. Green bars indicate significant differential expression (after Bonforroni correction for testing 27 genes).

For glycolysis, only hexokinase 3 (HK3) turns out to be differentially expressed between ER+ and ER- breast cancer (down-regulated in ER+ tumors with fold change 1.6). For glutaminolysis, a number of enzymes are differentially expressed between ER+ and ER- tumors (4 up-regulated, 7 down-regulated). Many of these enzymes catalyse several reactions, for example the asparagine synthetase (ASNS) and the tri-functional carbamoyl-phosphate synthetase 2, aspartate tracarbamylase, and dihydroorotase (CAD) that use glutamine as amide-N-donor.

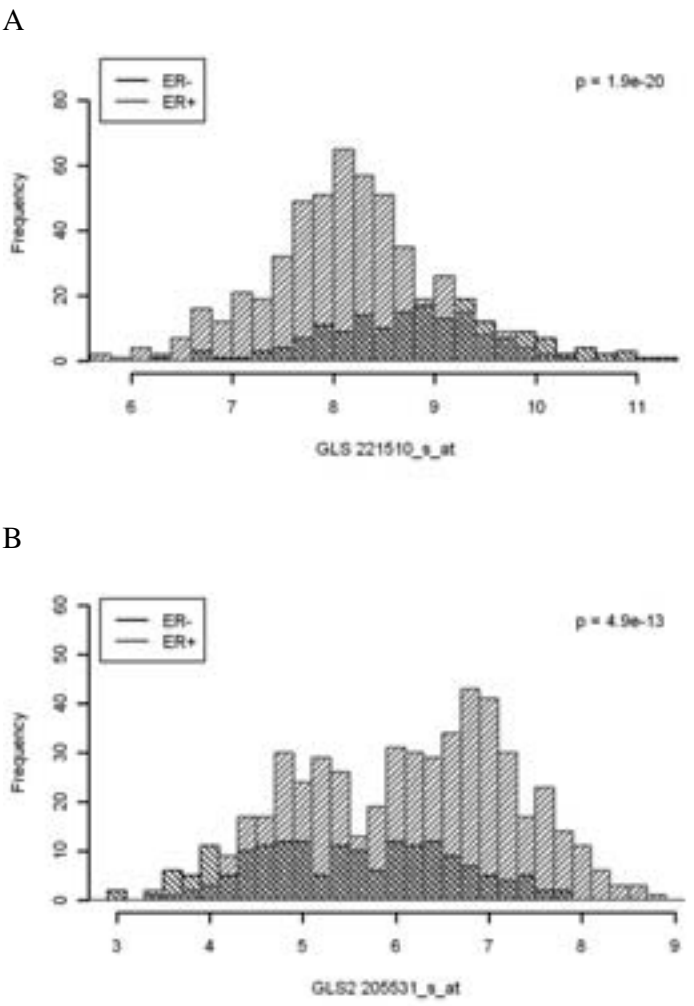

Fig. 3: Differential expression of glutaminase (GLS) and glutaminase 2 (GLS2) between ER+ and ER- breast cancer. Histograms show the distribution of GLS and GLS2 expression as they are measured by the microarrays (log-2 scale).

Among human genes, only the two isoenzymes glutaminase (GLS, also termed kidney type glutaminase) and glutaminase 2 (GLS2, also termed liver type glutaminase) exclusively convert glutamine to glutamate. Fig. 3 shows the expression of GLS (down-regulated in ER+ tumors, fold change 1.7) and the expression of GLS2 (up-regulated in ER+ tumors, fold change 1.7) in ER+ and ER- breast cancer.

4 Discussion

Changes in metabolite concentrations are the final response of a cell to genetic or environmental changes. Metabolite concentrations reflect the outcome of regulation at many molecular levels that takes place in living cells. Indeed, to monitor the final outcome of many regulatory layers is one of the strengths of the metabolomics approach. Regulation of metabolite concentrations can take place at the following levels:

- DNA level: loss of function mutations of enzymes
- RNA level: regulation of enzyme expression (epigenetic, transcriptional or post-transcriptional)
- Protein level: phosphorylation or other post-translational modifications of enzymes
- Interaction level: allosteric regulation of enzymes

A difficulty in the analysis of metabolomics data is connected with backward analysis of the causal chain, in order to understand the mechanism of metabolic regulation and to detect targets for a possible therapeutic intervention. METAtarget implements a reverse engineering step for metabolic networks by using the ratio of product to substrate concentrations as surrogate for the activity of enzymes. METAtarget is based on information on reactant pairs and enzymes that is obtained from the KEGG REACTION database.

In this paper we have assessed the significance metabolic changes by Welch's t-test. Validity of this approach depends on at least approximate normal distribution of the difference variables $X - Y$. As a more conservative alternative, but with costs of losing power, a rank statistics based approach (Mann-Whitney test) can be applied. As unimportant for significance assessment, we did not take into account reversibility or the direction of reactions. Denotation of a reactant as substrate or product is arbitrary.

Using METAtarget, we have analyzed the metabolism of breast cancer cells in dependence of hormone receptor status. Worldwide, immunohistological determination of estrogene receptor status is part of the breast cancer routine diagnostics. Patients with ER+ tumors are known to benefit from hormone therapy (for example treatment with tamoxifen), while ER- breast cancer is known to be a more aggressive breast cancer subtype. Most of the ER- tumors are highly proliferating and have tumor grades 2 or 3 [DRL07]. Furthermore, these tumors include the triple-negative subtype that is difficult to treat and has a poor prognosis.

Analyzing metabolomics data generated within the METAcancer project, we have identified 13 reactant pairs with a shifted equilibrium depending on ER status. The expression pattern of the corresponding 29 enzymes was analyzed in three publicly available gene expression series. 7 enzymes turned out to be significantly up-regulated, 10 enzymes significantly down-regulated in ER+ tumors compared to ER- tumors, while 12 enzymes remained unchanged. All 29 enzymes are interesting as targets, because manipulation of enzyme activity could restore the metabolism towards a less aggressive type. On the other side, differential expression of the 17 regulated enzymes is expected to contribute to the regulation of breast cancer metabolism in dependence of hormone receptor status.

In particular, we detected an up-regulation of glycolysis and glutaminolysis in ER- tumors, compatible with a higher demand on energy of a more aggressive cancer. Targeting these pathways could be an opportunity for treatment. More inside in the regulation processes has been provided by analyzing the gene expression of enzymes that catalyze the entry reaction of glycolysis and of glutaminolysis (Fig. 2).

The expression of glutaminases has been extensively studied before und shown to exhibit tissue-specific expression profiles [SO09]. Co-expression of GLS and GLS2 is a frequent event in human cancer [Pé05]. Furthermore, high GLS expression has been described as being associated with high proliferation rates, whereas repression of GLS and prevalence of GLS2 has been described to be related to quiescent or resting states [LKM69]. This observation is compatible with high expression of GLS in the strong proliferation ER- tumors and higher average expression GLS2 in ER+ tumors, an entity that also contains weak proliferating G1 tumors.

In summary, METAtarget is a new method for reverse engineering of metabolic networks and the detection of targets enzymes from high-throughput metabolomics data. In an application to METAcancer, deregulation of key reactions of glycolysis, glutaminolysis and other pathways could be detected between ER+ and ER- breast cancer.

Acknowledgements

The work was funded by the European Commission, FP7 grant #200327 (METAcancer).

Bibliography

[An09] Antonov, A.V.; Dietmann, S.; Wong, P. et. al.: TICL - a web tool for network-based interpretation of compound lists inferred by high-throughput metabolomics. FEBS J. (2009) 276; pp. 2084-2094.

[De08] Denkert, C.; Budczies, J.; Weichert, W. et. al.: Metabolite profiling of human colon carcinoma - deregulation of tca cycle and amino acid turnover. Mol. Cancer (2008) 7; p. 72.

[De09] Denkert, C.; Budczies, J.; Darb-Esfahani, S. et. al.: A prognostic gene expression index in ovarian cancer - validation across different independent data sets. J. Pathol. (2009) 218; pp. 273-280

[DRL07] Dunnwald, L.K.; Rossing, M.A.; Li, C.I.: Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. Breast Cancer Res. (2007) 9; p. R6.

[Fi01] Fiehn, O.: Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. Comp. Funct. Genomics (2001) 2; pp. 155-168.

[Ga10] Gao, J.; Tarcea, V.G.; Karnovsky, A. et. al.: Metscape: a cytoscape plug-in for visualizing and interpreting metabolomic data in the context of human metabolic networks. Bioinformatics (2010) 26; pp. 971-973.

[GHW05] German, J.B.; Hammock, B.D.; Watkins, S.M.: Metabolomics: building on a century of biochemistry to guide human health. Metabolomics (2005) 1; pp. 3-9.

[GS04] Griffin, J.L.; Shockcor, J.P.: Metabolic profiles of cancer cells. Nat. Rev. Cancer (2004) 4; pp. 551-561.

[Jo08] Jourdan, F.; Breitling, R.; Barrett, M.P. et. al.: MetaNetter: inference and visualization of high-resolution metabolomic networks. Bioinformatics (2008) 24; pp. 143-145.

[Ka10] Kanehisa, M.; Goto, S.; Furumichi, M. et. al.: KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. (2010) 38; p. D355-60.

[Le08] Letunic, I.; Yamada, T.; Kanehisa, M. et. al.: iPath: interactive exploration of biochemical pathways and networks. Trends Biochem. Sci. (2008) 33; pp. 101-103.

[LKM69] Linder-Horowitz, M.; Knox, W.E.; Morris, H.P.: Glutaminase activities and growth rates of rat hepatomas. Cancer Res. (1969) 29; pp. 1195-1199.

[Ma09] Matthews, L.; Gopinath, G.; Gillespie, M. et. al.: Reactome knowledgebase of human biological pathways and processes. Nucleic Acids Res. (2009) 37; p. D619-22.

[Ok08] Okuda, S.; Yamada, T.; Hamajima, M. et. al.: KEGG atlas mapping for global analysis of metabolic pathways. Nucleic Acids Res. (2008) 36; p. W423-6.

[Pé05] Pérez-Gómez, C.; Campos-Sandoval, J.A.; Alonso, F.J. et. al.: Co-expression of glutaminase K and L isoenzymes in human tumour cells. Biochem. J. (2005) 386; pp. 535-542.

[Sh03] Shannon, P.; Markiel, A.; Ozier, O. et. al.: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. (2003) 13; pp. 2498-2504.

[Sr09] Sreekumar, A.; Poisson, L.M.; Rajendiran, T.M. et. al.: Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature (2009) 457; pp. 910-914.

[SO09] Szeliga, M.; Obara-Michlewska, M.: Glutamine in neoplastic cells: focus on the expression and roles of glutaminases. Neurochem. Int. (2009) 55; pp. 71-75.

[VCT09] Vander Heiden, M.G.; Cantley, L.C.; Thompson, C.B.: Understanding the warburg effect: the metabolic requirements of cell proliferation. Science (2009) 324; pp. 1029-1033.