BIOINFORMATICS ORIGINAL PAPER Vol. 26 no. 9 2010, pages 1225–1231
doi:10.1093/bioinformatics/btq113

Systems biology

Advance Access publication March 24, 2010

PathWave: discovering patterns of differentially regulated enzymes in metabolic pathways

Gunnar Schramm[1,2], Stefan Wiesberg[3,4], Nicolle Diessl[1], Anna-Lena Kranz[1,2], Vitalia Sagulenko[5], Marcus Oswald[3,4], Gerhard Reinelt[3], Frank Westermann[5], Roland Eils[1,2,*] and Rainer König[1,2,*]

[1]Department of Bioinformatics and Functional Genomics, Institute of Pharmacy and Molecular Biotechnology, and Bioquant, University of Heidelberg, Im Neuenheimer Feld 267, [2]Department of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, [3]Institute of Computer Science, [4]Interdisciplinary Center for Scientific Computing, University of Heidelberg and [5]Department of Tumor Genetics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

Associate Editor: Alfonso Valencia

ABSTRACT

Motivation: Gene expression profiling by microarrays or transcript sequencing enables observing the pathogenic function of tumors on a mesoscopic level.

Results: We investigated neuroblastoma tumors that clinically exhibit a very heterogeneous course ranging from rapid growth with fatal outcome to spontaneous regression and detected regulatory oncogenetic shifts in their metabolic networks. In contrast to common enrichment tests, we took network topology into account by applying adjusted wavelet transforms on an elaborated and new 2D grid representation of curated pathway maps from the Kyoto Enzyclopedia of Genes and Genomes. The aggressive form of the tumors showed regulatory shifts for purine and pyrimidine biosynthesis as well as folate-mediated metabolism of the one-carbon pool in respect to increased nucleotide production. We spotted an oncogentic regulatory switch in glutamate metabolism for which we provided experimental validation, being the first steps towards new possible drug therapy. The pattern recognition method we used complements normal enrichment tests to detect such functionally related regulation patterns.

Availability and Implementation: PathWave is implemented in a package for R (www.r-project.org) version 2.6.0 or higher. It is freely available from http://www.ichip.de/software/pathwave.html

Contact: r.koenig@dkfz.de; r.eils@dkfz.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on November 24, 2009; revised on February 17, 2010; accepted on March 13, 2010

1 INTRODUCTION

Cancer cells exhibit a dramatically disturbed metabolism to satisfy their high bioenergetic demands for cell proliferation (Jones and Thompson, 2009). Accordingly, a long-standing strategy for cancer treatment is to attack basic tumor metabolism. Mainly, these treatments are rather unspecific and hinder nucleotide biosynthesis (Chen and Pankiewicz, 2007). Besides this, the introduction of experimental high-throughput methods in functional genomics such as gene expression profiling using microarrays and genome-wide sequencing has evoked the challenging task to observe and understand the pathogenic function of tumors on a mesoscopic level. However, the large volume of information generated in these experiments must be funneled into manageable and functionally sensible partitions to select components that well describe the tumor pathologies. Intelligent embedding of the expression data into the underlying topology of the metabolic network may enable the detection of tumor-specific metabolism for directed and specific targeting of tumor cells.

Neuroblastoma is the most common solid, extracranial tumor of early childhood, mainly affecting children at the age of about 1 year. It is derived from primitive cells of the sympathetic nervous system. In many patients, neuroblastoma is metastatic at the time of diagnosis and undergoes rapid progression with fatal outcome. Alternatively, neuroblastomas, especially in infants younger than 1 year at diagnosis can regress spontaneously, and the tumor can differentiate into benign ganglioneuroma in older infants (Schwab *et al*., 2003). Detailed diagnosis and appropriate adjustment of therapy requires the support of investigations at the molecular level. DNA microarray technology improved the prediction of patient outcome in comparison to established risk markers (Oberthuer *et al*., 2006).

We wanted to track how aggressive neuroblastomas have specifically regulated their metabolism to optimize oncogenetic fitness, and to elucidate ways to severely perturb this process. An analysis method was required that discovers pathways in the metabolic network, especially showing significant shifts (similar and contrasting) in regulation. When analyzing data on a metabolic network, enzymes can be represented by their corresponding genes. Transcriptional data, and the topological information derived from the metabolic network, was connected by calculating *Z*-scores of highly correlated sub-networks (Patil and Nielsen, 2005). Chuang *et al*. (2007) improved classification of breast cancers with expression patterns of small subnets of a signal transduction network. Common gene expression levels of neighboring nodes in a metabolic network were calculated by averaging over all neighbors

*To whom correspondence should be addressed.

© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

Downloaded from https://academic.oup.com/bioinformatics/article/26/9/1225/200098 by guest on 20 August 2022

of a gene and revealed several interesting regulated pathways for the human immune system (Nacu *et al*., 2007). However, these approaches were not developed to detect highly contrasting expression of neighboring genes that undergo a switch-like shift of regulation in a tumor cell. Especially, these switches can be highly relevant to identify potential drug targets that specifically attack the tumor at sites of flux-redirections with which the tumor established parasitic advantages.

Wavelet transforms have been commonly applied in information technology and image processing to track congeneric and contrasting signatures (Chang and Kuo, 1993; Mallat, 1998). However, applying this powerful technology to analyze cellular networks is challenging. While the underlying topology of an ordinary image is a simple lattice grid, cellular networks exhibit a rather complex scale-free architecture (Jeong *et al*., 2000). In our initial approach, we mapped the gene expression data onto (lattice grid-like) adjacency matrices of networks and applied Haar wavelet transforms onto dense subsets of the metabolic network. Features of the wavelet transforms that could significantly separate samples from different treatments were used to extract metabolic pathways showing the most significant gene expression patterns (König *et al*., 2006).

In the present study, we substantially improved this technology by (i) using cell-physiologically well-defined and curated pathways from the Kyoto Enzyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al*., 2008) which extensively simplified interpretation of the results, (ii) developing a new and elaborated metric to arrange the order of enzyme representations as a lattice grid-like architecture of single pathways taking network priorities from curators of KEGG into account, (iii) implementing a one-step frameshift concept for wavelet transforms to overcome their rigidity, and (iv) we made the software freely available and easy applicable by a package for R (www.r-project.org). When applying this method to the neuroblastoma tumors, the most significant expression patterns were detected in purine and pyrimidine biosynthesis and folate-mediated one-carbon metabolism. These pathways can account for increased nucleotide production for proliferation. We spotted a significant switch-like regulation pattern in glutamate metabolism that hints towards de-regulated neurotransmitter production and glutamine uptake from the bloodstream. A potential drug target was proposed for this pathway and experimentally validated by drug-treatment of several neuroblastoma cell lines.

2 METHODS

For simplification, we explained the method with a toy example of a small synthetic pathway and simulated expression patterns (Supplementary Material). For all multiple testing corrections in this study, we used the method of Bonferroni, 1935 (Gordi and Khamis, 2004). *P*-values for the pathways of all gene set enrichment tests and our method were corrected for multiple testing by this method.

2.1 Assembling the metabolic pathways

Pathways were defined according to curated pathway maps of the KEGG database (version from February 4, 2009) (Kanehisa *et al*., 2008). Each metabolic pathway was established by defining neighbors of reactions using the information from KEGG (ftp://ftp.genome.jp/pub/kegg/xml/kgml/metabolic/organisms/hsa/). Two reactions were neighbors if a metabolite existed that was the product of one and the substrate of the other. We defined reactions as the nodes and metabolites as the edges between them. Pathways without any connected reaction were discarded. This resulted in 99 pathways with 1826 different reactions. Each pathway was represented by its adjacency-matrix. An entry at row a and column b was set to one if there existed a metabolite that was produced by reaction a and consumed by reaction b or vice versa. The sizes of the symmetric adjacency-matrices were between 2×2 and 92×92 reactions.

2.2 Ordering the 2D pathway representation with the grid arrangement method

To apply our feature extraction method we required a 2D arrangement of the metabolic network. We calculated an embedding of the metabolic networks for every KEGG pathway into a 2D, regular square lattice grid. To preserve neighborhood characteristics of the reactions, we were looking for embeddings in which adjacent nodes of the network were placed onto the grid as close to each other as possible. As a measure of distance in the lattice, we used the Manhattan distance, i.e. for any two grid points $u=(i_1, j_1)$ and $v=(i_2,j_2)$ the distance was given by $d=|i_1-i_2|+|j_1-j_2|$. We wanted to determine an optimal neighborhood in which the total edge length of the graph on the lattice was minimized while conserving the network topology. This resulted in an NP-hard combinatorial optimization problem. We stated this problem as an integral linear program (IP; see Nemhauser and Wolsey, 1999 for an introduction to integer programming). We formulated the IP by introducing 3D binary variables x_{vij} for every node v and every grid point (i, j) stating whether or not node v has to be placed on grid point (i, j). For each pair of nodes (u, v), we calculated their distance d_{uv}. For a given lattice grid g, the undirected network graph $G=(V, E)$ with node set V, edge set E and adjacency matrix M, the most basic IP was given by finding an optimum for

$$\min_{x,d} \sum\nolimits_{a,b\in V, a<b} M(a,b)\cdot d_{ab} \tag{1}$$

with the constraints

$$\sum\nolimits_{(i,j)\in g} x_{vij}=1,\ \forall v\in V \tag{2}$$

$$\sum\nolimits_{v\in V} x_{vij}\le 1,\ \forall (i,j)\in g \tag{3}$$

$$d_{ab}\ge \begin{cases} A+B \\ A-B \\ -A+B \\ -A-B \end{cases},\ \forall (a,b)\in V\times V, a<b \tag{4}$$

where

$$A := \sum\nolimits_{(i,j)\in g} i\cdot x_{aij} - \sum\nolimits_{(i,j)\in g} i\cdot x_{bij}, \tag{5}$$

$$B := \sum\nolimits_{(i,j)\in g} j\cdot x_{aij} - \sum\nolimits_{(i,j)\in g} j\cdot x_{bij}$$

and

$$x_{vij}\ge 0,\quad x_{vij}\in \mathrm{Z},\ \forall v\in V, (i,j)\in g. \tag{6}$$

Constraints (2) and (3) guaranteed that all nodes were placed exactly once and that each grid point could be used at most once. Constraints (4) ensured that the distance of node a and b is given by $|A|+|B|$, where A and B are computed by Equation (5) as $A=i_a-i_b$ and $B=j_a-j_b$. All variables were enforced to values 0 or 1 by constraint (6). The problem was solved by CPLEX 8.1 (ILOG, Gentilly, France) for 99 lattice grids (representing 99 KEGG-maps) with an average optimality of 96% for embeddings on square grids of side length $\sqrt{|V|}+1$, rounded up to the next integer.

2.3 Network motifs constrain the optimization problem

This basic model was enhanced by a number of graph dependent, additional constraints on the distance variables. They provided lower bounds for the distance sums of well-known sub-graph motifs. That is, for an edge induced sub-graph $G'\subset G$ with a least objective function contribution of $lb(G')$, the following inequality can be added or dynamically separated:

$$\sum\nolimits_{(u,v)\in E(G')} d_{uv}\ge lb\left(G'\right) \tag{7}$$

Downloaded from https://academic.oup.com/bioinformatics/article/26/9/1225/200098 by guest on 20 August 2022

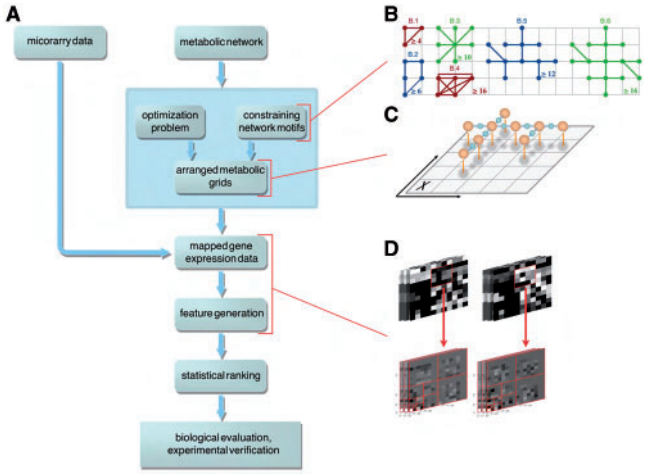

Fig. 1. Workflow of data integration, analysis and experimental validation. (**A**) Workflow of the method: each pathway of the metabolic network was represented on optimally arranged 2D grids, gene expression data were mapped onto these grids for every patient, features were generated of combined expression values of neighboring reactions in the grid, the discriminative power for each feature was statistically estimated, pathways and patterns of significance were given out for functional interpretation and experimental validation. (**B**) Specific network motifs constrained the optimal arrangement of the metabolic network leading to (**C**) an optimal arrangement of the network on a lattice-grid. (**D**) Gene expression data were mapped onto the optimally arranged grids and features were generated allowing the identification of discriminative patterns.

We considered the sub-graph motifs of star graphs, cliques consisting of up to 10 vertices and odd cycles ($2k$+1-cycles). Moreover, a certain class of trees with maximum vertex degree $\Delta(T) \leq 4$ (Fig. 1B) decreased computation time and enhanced separation ability.

These motifs included six representatives of typical sub-graph constraint classes: the $2k$+1-cycles for $k = 1$ (Fig. 1B.1) and $k = 2$ (Fig. 1B.2), a star graph with eight vertices (Fig. 1B.3), a 5-clique (Fig. 1B.4) and neighborhood star graphs with 12 (Fig. 1B.5) and 15 (Fig. 1B.6) vertices. As the graphs were embedded optimally, the numbers gave the total edge lengths of the embeddings as well as the right-hand sides $lb(G')$. Furthermore, calculation time was reduced by symmetry-breaking constraints eliminating all but a few representative embeddings from each equivalence class of symmetrical embeddings. For this, grid symmetries due to translation, rotation and reflection of the embeddings were considered as well as vertex subsets whose inner permutations did not change the value of the objective function.

2.4 Assembling the gene expression data and mapping onto the network

Our metabolic analysis was performed with gene expression data from an earlier study in which we supported clinical diagnoses of neuroblastoma tumors (Oberthuer *et al*., 2006). The gene expression profiling was performed for 251 patients diagnosed between 1989 and 2004, in duplicate as dye-swap experiments on Agilent oligonucleotide microarrays (www.agilent.com) with 10 163 neuroblastoma-specific probes. The age of the patients was between 0 and 296 months (median age: 15 months). For our study, we compared stage 1 patients without MYCN amplification, 65 in total, to 19 stage 4 patients with MYCN amplification. According to the International Neuroblastoma Staging System (INSS), stage 1 tumors are localized and confined to the area of origin. Stage 4 tumors disseminate to distant lymph nodes, bone marrow, bone, liver or other organs and have a very poor prognosis if the oncogene MYCN is genetically amplified, i.e. abundant in high copy number (Schwab *et al*., 2003). We normalized the data with the variance normalization method (Huber *et al*., 2002). The raw and normalized data are deposited at ArrayExpress (http://www.ebi.ac.uk/arrayexpress; experiment accession number E-TABM-38). The expression data of each dataset was mapped onto the corresponding reactions of the transcribed enzymes using the gene–protein information from KEGG (ftp://ftp.genome.jp/pub/kegg/xml/kgml/metabolic/organisms/hsa). Mean values were taken if a reaction was catalyzed by a complex of proteins. The expression values of each reaction were z-transformed to facilitate combinations of the values that were needed for the wavelet transforms. Expression data were available for 1103 enzymatic reactions extracted from KEGG. The expression data of all samples were mapped onto the optimally ordered grid representations of all KEGG pathways, respectively. This resulted in 84 different patterns (of stage 1 and stage 4 patients) for each KEGG pathway.

2.5 Pattern recognition on lattice grids with Haar wavelet transforms

We wanted to explore every possible expression pattern of neighboring genes and groups of genes within a KEGG pathway that showed significant differences between samples of different conditions. For this, we performed a Haar wavelet transform for each optimized grid representation of the pathways. The wavelet transformed expression values were statistically tested to identify pathways with a discriminative pattern between tumors of favorable and unfavorable outcome. Such a Haar wavelet transform can be regarded as systematically applying low pass and high pass filters from fine grain to coarse grain resolutions. It is therefore well suited to not only detect commonly regulated pathways (low-pass filter), but also to elucidate switch-like behaviors within the pathways (high-pass filter), for more details see Mallat (1998). To overcome rigidity of wavelet transforms, we covered any possible combination of neighboring reactions by shifting the frame for applying the transforms: we conducted the Haar wavelet transform on the original grid, on the grid without the first row and column, respectively, and finally without the two, first row and column. If the number of rows (columns) became odd after deletion of the first row (column) then the last row (column) was also removed. By this, we avoided an unnecessary weighting of the last and isolated rows (columns). This procedure was carried out for all KEGG pathways of every sample. The results of the transforms were stored as the corresponding features for every sample.

2.6 Significance tests for the pathways

Pathways were ranked according to the statistical significance of an assigned score. T-tests were applied on all features F_i^π of pathway π returning P-values P_i^π. The pathway-score S_π was derived by

$$S_\pi = \max \left| \log_{10} P_i^\pi \right| \tag{8}$$

To estimate the statistical significance of pathway-score S_π, we randomly sampled the samples (patients, drawing without replacement) n times ($n = 10\,000$ for the neuroblastoma study). Assuming an extreme value distribution (Gumbel distribution), for each pathway a curve was fitted to the distribution of the pathway-scores of the permutated samples. The P-value for each pathway resulted from this fitting-curve and was corrected for multiple testing (Bonferroni, 1935; Gordi and Khamis, 2004). To focus on the most relevant features, only pathways with more than five significantly, differentially regulated reactions and genes were further investigated ($P \leq 0.01$, not corrected for multiple testing). For this, two reactions that consisted of exactly the same genes were counted as one reaction. Pathways with significant features were further functionally characterized by analyzing the literature (Section 3 and Supplementary Material).

Downloaded from https://academic.oup.com/bioinformatics/article/26/9/1225/200098 by guest on 20 August 2022

2.7 Discovering local patterns

All features were statistically tested with *t*-tests to separate between favorable and unfavorable tumors and corrected for multiple testing. Statistically significant features contained those sub-graphs of the metabolic network that showed differentially regulated patterns. Such regions of interest could directly be accessed by reconstructing the regions that were represented by the significant features and were given out by PathWave (expression patterns in Supplementary Figs S4–S13 and Table S2).

2.8 Cell culture, reagents, treatments and data analysis for the experimental D-cycloserine study

We used four human neuroblastoma cell lines (SK-N-SH, SH-EP, IMR32 and Kelly) for our analyses. They were maintained in DMEM medium (Lonza, Verviers, Belgium) supplemented with L-glutamine and 10% fetal calf serum (FCS) at 37°C in 5% CO_2 atmosphere. D-cycloserine was purchased from Sigma (Munich, Germany). For cytotoxicity assays approximately 1E+5 cells were seeded onto 24-well plates in DMEM without L-glutamine and supplemented with 5% FCS. D-cycloserine was added to the cells 18 h after seeding at a final concentration of 5 mM. The concentration of D-cycloserine was one order of magnitude below the reported toxic concentration in rats [see DrugBank (Wishart *et al*., 2006, 2008)]. Total amount of cells in culture was determined with crystal violet (Serva) after 20 min (0 h), 24 h, 48 h and 72 h. The cells were fixed with 3.7% formaldehyde and stained with 0.5% crystal violet. After several washings with DPBS (Lonza), crystal violet was re-solubilized in a buffer containing 0.1 M sodium citrate, pH 4.2 and 50% ethanol. The absorbance of crystal violet was measured at 580 nm and taken as a measure for cell proliferation. Measurements were done in a FLUOstar OPTIMA plate reader (BMG Labtech, Offenburg, Germany). All experiments were done in quadruplicates.

For each time point, raw values were subtracted by the mean background intensity coming from four blank wells treated similar to the test wells. To obtain replication levels, for each cell line and replicate, the intensities were divided by their intensities at time point 0 h, respectively. Mean values and standard deviations were calculated for each time point (after 0 h) using the data from each cell line and replicate. The difference of the distributions was tested with a Student's *t*-test.

Table 1. Identified significant differentially regulated pathways with more than five differentially regulated KEGG reactions

Rank	Pathway	*P*-value	Score size[a]
1	Purine metabolism	< 1E-16	1
2	Glutamate metabolism	< 1E-16	1
3	Glycolysis/Gluconeogenesis	1.1E-14	1
4	Pyrimidine metabolism	5.5E-14	2
5	One carbon pool by folate	3.2E-13	1
6	Phosphatidylinositol signaling system	9.1E-12	2
7	Pyruvate metabolism	1.3E-11	3
8	Valine, leucine and isoleucine degradation	6.9E-11	1
9	Lysine degradation	1.5E-10	3
10	Glycine, serine and threonine metabolism	2.1E-10	1
11	Urea cycle and metabolism of amino groups	2.5E-10	1
12	Inositol phosphate metabolism	1.3E-9	1
13	Fatty acid metabolism	1.7E-9	3
14	Folate biosynthesis	1.8E-9	2
15	Glutathione metabolism	2.3E-9	1
16	*N*-Glycan biosynthesis	6.2E-8	2
17	Starch and sucrose metabolism	1.7E-7	1
18	Glycerophospholipid metabolism	3.9E-7	1
19	Glycosphingolipid biosynthesis—neo-lactoseries	5.3E-7	1
20	Sphingolipid metabolism	1.4E-6	2
21	Tyrosine metabolism	6.6E-6	2
22	Aminoacyl-tRNA biosynthesis	1.3E-5	2
23	Fatty acid biosynthesis	4.1E-5	1

[a]Size of the most significant pattern with which the score was calculated, 1 = 1st wavelet with 2×2 pixel on the grid; 2 = 2nd wavelet with 4×4 pixel, 3 = 3rd wavelet with 8×8 pixel.

3 RESULTS

3.1 Combining regulation patterns of enzymes

We compared gene expression profiles of 19 aggressive neuroblastomas having unfavorable prognosis (stage 4 according to the international neuroblastoma staging system and amplification of the *MYCN* oncogene) with 65 tumors having favorable prognosis (stage 1, no *MYCN* amplification). Our aim was to compare the regulation of metabolism of these two tumor entities. The workflow was as follows (Fig. 1): metabolic pathways were extracted from the KEGG database and a lattice grid-like representation was constructed for each pathway. Enzymes were optimally arranged on the grid to preserve the relevant pathway topology. Gene expression data were mapped onto the corresponding enzymes in the optimally arranged grid. Neighboring enzymes were grouped and their gene expression values combined using wavelet transforms. These transforms yielded combined expression values ('features') by applying low-pass filters to detect similar expression changes and high-pass filters to detect contrasting regulation patterns. We tested the performance of all non-trivial features (10 377) to separate the two tumor entities by a *t*-test. Pathways were ranked according to their significance. Only pathways with more than five significantly, differentially regulated reactions and genes were further investigated to focus on the most relevant features. This procedure yielded significant features from 23 different pathways (Table 1). All reactions of these features are shown in the Supplementary Material (Supplementary Table S2) together with the regulation patterns in the pathways from KEGG for the first 10 pathways (Supplementary Figs S4–S13). The Supplementary Material also includes a summary of information from the literature regarding the oncogenic relevance of pathways not discussed in the main text. We denoted an enzyme as being up- or down-regulated in the aggressive tumors only if its differential regulation was significant ($P \leq 0.01$, not multiple testing corrected).

We included the enzyme commission (EC) numbers for all enzymes to a convenient tracking in the pathway maps. We used the same abbreviations for metabolites as in KEGG.

3.2 A cellular switch in the glutamate metabolism

A significant pattern was identified in glutamate metabolism. It consisted of six reactions ($P < 1E\text{-}16$, Fig. 2 and Supplementary Fig. S5). Glutamine is produced by the host, and is highly abundant in the blood. It is consumed by the parasitic tumor cells to provide an ammonium supply (Medina, 2001). It has previously been shown that tumors depress systemic glutamine

Downloaded from https://academic.oup.com/bioinformatics/article/26/9/1225/200098 by guest on 20 August 2022

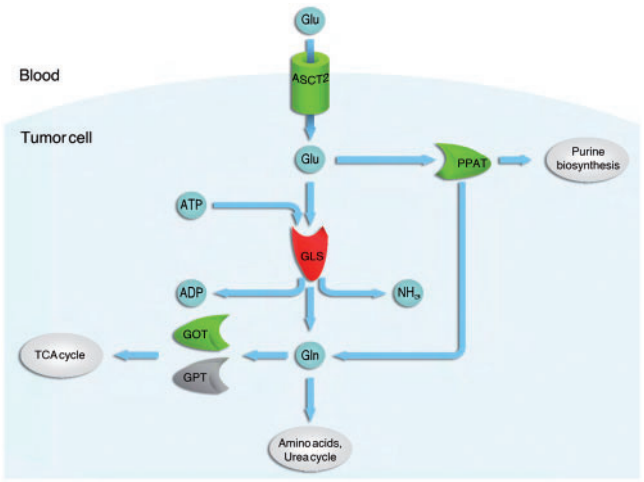

Fig. 2. Regulation pattern in glutamate metabolism. In the aggressive tumors, the ASCT2 glutamine transporter, amido phosphoribosyltransferase (PPAT) and one amino transferase (GOT) were significantly up-regulated to take up glutamine from the bloodstream and metabolize it for purine biosynthesis, amino acids biosynthesis and anapleurosis of the TCA cycle. Ammonium detoxification by glutaminase (GLS) was significantly down-regulated. Probes for the amino transferase GPT were not included in the microarrays.

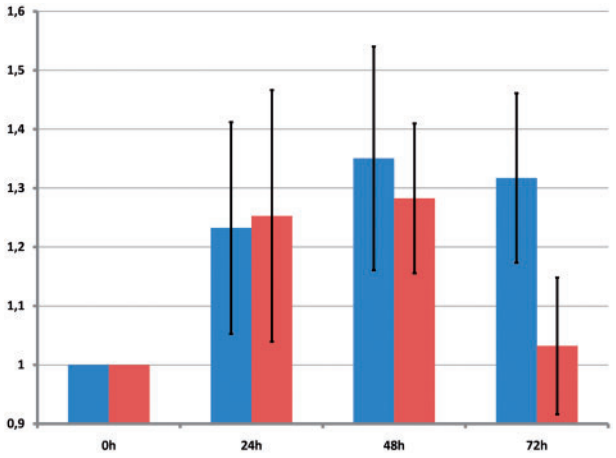

Fig. 3. Proliferation of neuroblastoma cells treated with D-cycloserine. Proliferation is shown for cells treated with 5 mM D-cycloserine (red) and untreated control cultures (blue). Error bars are the standard deviations (1σ) for the corresponding time point and treatment. The *y*-axis denotes replication levels in accordance to absorbance of a dye that stained fixed cells (crystal violet), normalized to 0 h ($1\times$ proliferation). In comparison to non-treated cells, the proliferation of the treated cells decreased significantly at 72 h ($P = 1.3E\text{-}5$). There was no significant difference in proliferation at 24 h and 48 h.

levels in cancer patients (Klimberg and McClellan, 1996). Glutamine is catabolized into ribosylamine-5-phosphate by amidophosphoribosyl transferase to synthesize purines. Amidophosphoribosyl transferase has been proposed as a drug target to treat cancer (Christopherson *et al*., 1995). It may be combined with glutamate antagonists. Glutamate has previously been shown to be important for tumor growth, since inhibition using glutamate antagonists led to reduced proliferation (Rzeski *et al*., 2002). We observed a significant regulation pattern for glutamine metabolism in the investigated neuroblastomas, which is in line with these observations from the literature (Fig. 2). More specifically, ATP-dependent glutaminase (GLS, EC 3.5.1.2) was down-regulated, reducing glutamine catabolism to ammonium and glutamate. Glutamine flux was redirected into purine biosynthesis via down-regulation of ATP-dependent GLS and up-regulation of amidophosphoribosyl transferase (PPAT, EC 2.4.2.14). Amidophosphoribosyl transferase also produces glutamate that is used as a building block for synthesizing further non-essential amino acids. Glutamate was shown to be used for TCA cycle anapleurosis in glioblastomas (DeBerardinis *et al*., 2007). To identify an anapleurotic tendency in neuroblastoma, we experimentally investigated the proliferation of neuroblastoma cell lines after treatment with D-cycloserine. D-cycloserine compromises the action of two aminotransferases in the glutamate metabolism (GOT, EC 2.6.1.1 and GPT, EC 2.6.1.1) (Fischer *et al*., 1997) that are necessary for this TCA anapleurosis. Growth was significantly reduced after 72 h in neuroblastoma cells treated with 5 mM D-cycloserine ($P = 1.3E\text{-}5$, Fig. 3). Interestingly, Wise and co-workers (Wise *et al*., 2008) have recently shown that *c-MYC* regulated the stimulation of just such a glutaminolysis program, in which glutamine is used for TCA anapleurosis, in glioma cells. It is likely that this oncogenetic regulation scheme is transferable to neuroblastoma, for which we have shown that *MYCN* can compensate for *c-MYC* activity and that *MYCN/c-MYC* signaling is more active in more aggressive neuroblastoma subtypes (Westermann *et al*., 2008).

3.3 The aggressive tumors employ increased nucleotide biosynthesis

Highly significant regulation patterns were detected for the pathways of purine and pyrimidine biosynthesis as well as folate-mediated one-carbon metabolism. The pathways ranked at positions 1, 4 and 5 ($P \leq 1E\text{-}16$, 5.5E-14 and 3.2E-13, respectively). These pathways were mainly up-regulated to enable enforced nucleotide biosynthesis for increased cell cycle activity of aggressive tumors. All enzymes involved in the biosynthetic pathway for the purines ATP and GTP were up-regulated (Supplementary Fig. S4). We defined the pathway for ATP and GTP biosynthesis as starting with ribose-5-phosphate originating from the pentose phosphate pathway and glutamate metabolism using, among others, AICAR, IMP, XMP, GMP, GDP, ADP. Phosphatases (EC 3.1.3.5, EC 3.6.1.5, EC 3.6.1.6) that degrade compounds for purine production were down-regulated. We also detected a significant regulation pattern for TTP and CTP biosynthesis in pyrimidine metabolism (Supplementary Fig. S7). Enzymes were up-regulated for RNA synthesis (RNA polymerase, EC 2.7.7.6), the final biosynthetic step of UTP and CTP (nucleotide phosphate kinase, EC 2.7.4.6), dUPD production from dUMP (thymidylate kinase, EC 2.7.4.9) and the conversions of CDP to dCDP and dUMP to dTMP (ribonucleotide reductase, EC 1.17.4.1 and thymidylate synthetase, EC 2.1.1.45). Enzymes reversing pyrimidine anabolism were also down-regulated (EC 3.1.3.5, EC 3.6.1.5, EC 3.6.1.6 and EC 3.5.4.5), similar to the regulatory patterns for purine metabolism.

A significant regulation pattern was found in folate-mediated metabolism of the one-carbon pool. Ten reactions were up-regulated,

Downloaded from https://academic.oup.com/bioinformatics/article/26/9/1225/200098 by guest on 20 August 2022

specifically, EC 1.5.1.3 (2), EC 2.1.2.1, EC 2.1.2.2 (2), EC 2.1.2.3, EC 2.1.1.45, EC 3.5.4.9, EC 6.3.4.3, and EC 1.5.1.15 (Supplementary Fig. S8 and Table S2). This pathway serves to carry and activate single carbons for purine and pyrimidine biosynthesis, utilizing pyrimidine thymidylate synthase (EC 2.1.1.45), purine GAR formyltransferase (EC 2.1.2.2) and AICAR formyltransferase (EC 2.1.2.3). Folate derivatives are also needed for methionine synthesis, which is essential for cancer cell survival (Stankova *et al*., 2005). To synthesize methionine, 5,10-methylene tetrahydrofolate is processed by methylene tetrahydrofolate reductase (EC 1.5.1.20) into 5-methyl tetrahydrofolate from which methionine synthase synthesizes methionine. In fact, inhibition of methylene tetrahydrofolate reductase has been shown to reduce tumor growth by depleting the cellular methionine pool (Stankova *et al*., 2005). Interestingly, methylene tetrahydrofolate reductase itself was not significantly regulated in the pattern we identified. Instead, thymidylate synthase (EC 2.1.1.45), was up-regulated, and may have taken over producing 5,10-methylene tetrahydrofolate in these tumor cells. Based on the regulatory patterns identified by PathWave, we propose that inhibiting thymidylate synthetase in combination with methylene tetrahydrofolate reductase will reduce tumor growth more effectively.

3.4 Comparison to established methods

Although PathWave was not designed as an enrichment test but rather to point to regulatory patterns in pathways and regions therein, we were interested in the results from an established gene set enrichment method. Therefore, we applied the Gene Set Enrichment Analysis (GSEA; Mootha *et al*., 2003) on the expression data of all pathways that we also had analyzed with PathWave. GSEA revealed three significantly enriched pathways, i.e. pyrimidine metabolism, purine metabolism and polyunsaturated fatty acid biosynthesis ($P \leq$1E-16, $\leq$1E-16, 9.9E-4, respectively, corrected for multi-testing, results for all pathways are in the Supplementary Table S4). Furthermore, we used DAVID which is another commonly used gene set enrichment test (Dennis *et al*., 2003; Huang da *et al*., 2009). Although DAVID revealed various enriched KEGG Pathways (see Supplementary Table S5a and S5b) only three *metabolic* pathways were identified (pyrimidine and purine metabolism, and glycolysis/gluconeogenesis). Hence, GSEA and DAVID were capable to identify enriched pathways, but showed less sensitivity in comparison to PathWave

3.5 Applying PathWave to another neuroblastoma dataset

To verify our findings, we analyzed a further independent expression dataset of neuroblastomas (Wang *et al*., 2006) consisting of primary tumors from 27 stage 1 patients without MYCN amplification and 20 stage 4 patients with MYCN amplification. Raw expression data were downloaded from NCBI, normalized and analyzed with PathWave. The significance threshold was again set to $P = 0.01$. As the fraction of significantly regulated reactions ($P \leq 0.01$) was lower in this dataset (29%) compared to our first study (43%), we focused on pathways containing four or more differentially regulated reactions. PathWave revealed 20 pathways (Supplementary Table S3) with significant regulation patterns, 15 out of which were also found in the first dataset we studied. Specifically, all high ranking pathways (ranks 1–6) of the first analysis showed up again in this dataset confirming our initial results.

4 DISCUSSION AND CONCLUSION

We mapped gene expression data from neuroblastomas having two very distinct clinical courses onto the human metabolic network. We revealed interesting insights into tumor cell regulation when applying our novel method PathWave. The aggressive tumors showed significantly up-regulated pathways for purine and pyrimidine synthesis. This was expected, as tumors need these building blocks to maintain quick mitotic cycles. We observed an interesting regulatory switch in glutamate metabolism: the energy consuming ammonium elimination from glutamine was down-regulated, while amidophosphorybosyl transferase was up-regulated. This may have redirected toxic ammonium from degradation into nucleotide anabolism in the aggressive tumors. Additionally, neuroblastoma tumor cells may utilize systemic glutamine for anapleurosis of the TCA cycle, similar to what has been recently reported for glioblastoma cells (DeBerardinis *et al*., 2007). We provided experimental evidence for this regulatory switch in tumor cell glutamate metabolism by targeting the amino transferases at the necessary metabolic junction, which resulted in reduced proliferation of neuroblastoma cell lines. Folate-mediated one-carbon metabolism was also differentially regulated in the aggressive neuroblastomas. We suggest targeting thymidylate synthetase in combination with methylene tetrahydrofolate reductase in neuroblastoma cell lines to assess the relevance of these results for therapeutic intervention. These interesting regulation switches could be found by our pattern recognition method as our implementation of wavelet transforms systematically tracked co-regulated and anti-co-regulated neighboring nodes in the network. Our network representation by lattice grids simplified their topology as it avoided hub-like structures by including only compounds that were selected to be relevant for the corresponding pathway by KEGG curators. The grid arrangement method placed players in pathways on a 2D map while conserving the direct neighborhoods of the players. This method is new and has potential for other applications. It is a generalization of the one-dimensional linear arrangement optimization problem (Bar-Yehuda *et al*., 2001) to two dimensions. The model can be extended to higher dimensions, but becomes more difficult to solve with each additional dimension. The newest version of our branch-and-cut algorithm tests whether a given set of values d_{uv}, $(u,v) \in E$ exhibits a feasible embedding.

PathWave enabled to focus on pathways with distinct regulated patterns in the network and pointed to sections in these pathways at which a switch-like regulation may have occurred. However, manual inspection and interpretation of the regulation of these sub-graphs is still necessary to derive their relevance within the functional context. When using such a global scanning device, lack of specificity and sensitivity must still be tackled. For about one-third of the enzymes in KEGG, we could not assign any probe from our microarray chips. Especially, for small networks this may have led to inappropriate overestimation of single nodes. For this reason, we discarded pathways with to few reactions from our analyses.

In general, we detected cellular switches in the metabolism of the tumor under study. The presented analysis technique is capable of reducing the relevant pathways to those having significant patterns of functionally connected proteins.

Downloaded from https://academic.oup.com/bioinformatics/article/26/9/1225/200098 by guest on 20 August 2022

ACKNOWLEDGEMENTS

We thank Kathy Astrahantseff for stylistic corrections.

Funding: BMBF-FORSYS Consortium Viroquant (#0313923); the Helmholtz Alliance on Systems Biology and the Nationales Genom-Forschungs-Netz (NGFN+) for the neuroblastoma project, ENGINE (#01GS0898).

Conflict of Interest: none declared.

REFERENCES

Bar-Yehuda,R. *et al.* (2001) Computing an optimal orientation of a balanced decomposition tree for linear arrangement problems. *J. Graph. Alg. Appl.*, **5**, 1–27.

Bonferroni,C.E. (1935) Il calcolo delle assicurazioni su gruppi di teste. In, Studi in Onore del Professore Salvatore Ortu Carboni. Rome, 13-60.

Chang,T. and Kuo,C.-C. (1993) Texture analysis and classification with tree-structured wavelet transform. *IEEE Trans. Image Process.*, **2**, 429–441.

Chen,L. and Pankiewicz,K.W. (2007) Recent development of IMP dehydrogenase inhibitors for the treatment of cancer. *Curr. Opin. Drug Discov. Dev.*, **10**, 403–412.

Christopherson,R.I. *et al.* (1995) Inhibitors of dihydro-orotase, amidophosphoribosyltransferase and IMP cyclohydrolase as potential drugs. *Biochem. Soc. Trans.*, **23**, 888–893.

Chuang,H.Y. *et al.* (2007) Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.*, **3**, 140.

DeBerardinis,R.J. *et al.* (2007) Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc. Natl Acad. Sci. USA*, **104**, 19345–19350.

Dennis,G. Jr. *et al.* (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.*, **4**, P3.

Fischer,Y. *et al.* (1997) Glucose transport and glucose transporter GLUT4 are regulated by product(s) of intermediary metabolism in cardiomyocytes. *Biochem. J.*, **321** (Pt 3), 629–638.

Gordi,T. and Khamis,H. (2004) Simple solution to a common statistical problem: interpreting multiple tests. *Clin. Ther.*, **26**, 780–786.

Huang da,W. *et al.* (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protocols*, **4**, 44–57.

Huber,W. *et al.* (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18** (Suppl. 1), S96–S104.

Jeong,H. *et al.* (2000) The large-scale organization of metabolic networks. *Nature*, **407**, 651–654.

Jones,R.G. and Thompson,C.B. (2009) Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev.*, **23**, 537–548.

Kanehisa,M. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.

Klimberg,V.S. and McClellan,J.L. (1996) Claude H. Organ, Jr. Honorary Lectureship. Glutamine, cancer, and its therapy. *Am. J. Surg.*, **172**, 418–424.

König,R. *et al.* (2006) Discovering functional gene expression patterns in the metabolic network of Escherichia coli with wavelets transforms. *BMC Bioinformatics*, **7**, 119.

Mallat,S. (1998) *A Wavelet Tour of Signal Processing*. Academic Press, New York.

Medina,M.A. (2001) Glutamine and cancer. *J. Nutrition*, **131**, 2539S–2542S; discussion 2550S–2531S.

Mootha,V.K. *et al.* (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.*, **34**, 267–273.

Nacu,S. *et al.* (2007) Gene expression network analysis, and applications to immunology. *Bioinformatics*, **23**, 850–858.

Nemhauser,G. and Wolsey,L. (1999) *Integer and Combinatorial Optimization*. John Wiley & Sons, New York.

Oberthuer,A. *et al.* (2006) Customized oligonucleotide microarray gene expression-based classification of neuroblastoma patients outperforms current clinical risk stratification. *J. Clin. Oncol.*, **24**, 5070–5078.

Patil,K.R. and Nielsen,J. (2005) Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc. Natl Acad. Sci. USA*, **102**, 2685–2689.

Rzeski,W. *et al.* (2002) Glutamate antagonists limit tumor growth. *Biochem. Pharmacol.*, **64**, 1195–1200.

Schwab,M. *et al.* (2003) Neuroblastoma: biology and molecular and chromosomal pathology. *Lancet Oncol.*, **4**, 472–480.

Stankova,J. *et al.* (2005) Antisense inhibition of methylenetetrahydrofolate reductase reduces cancer cell survival in vitro and tumor growth in vivo. *Clin. Cancer Res.*, **11**, 2047–2052.

Wang,Q. *et al.* (2006) Integrative genomics identifies distinct molecular classes of neuroblastoma and shows that multiple genes are targeted by regional alterations in DNA copy number. *Cancer Res.*, **66**, 6050–6062.

Westermann,F. *et al.* (2008) Distinct transcriptional MYCN/c-MYC activities are associated with spontaneous regression or malignant progression in neuroblastomas. *Genome Biol.*, **9**, R150.

Wise,D.R. *et al.* (2008) Myc regulates a transcriptional program that stimulates mitochondrial glutaminolysis and leads to glutamine addiction. *Proc. Natl Acad. Sci. USA*, **105**, 18782–18787.

Wishart,D.S. *et al.* (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36**, D901–D906.

Wishart,D.S. *et al.* (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.*, **34**, D668–D672.

Downloaded from https://academic.oup.com/bioinformatics/article/26/9/1225/200098 by guest on 20 August 2022