

RESEARCH

Open Access

# Discovery of metabolite biomarkers: flux analysis and reaction-reaction network approach

Limin Li<sup>1</sup>, Hao Jiang<sup>2</sup>, Yushan Qiu<sup>3</sup>, Wai-Ki Ching<sup>3\*</sup>, Vassilios S Vassiliadis<sup>4</sup>

From The 6th International Conference on Computational Systems Biology (ISB2012) Xi'an, China. 18-20 August 2012

## Abstract

**Background:** Metabolism is a vital cellular process, and its malfunction can be a major contributor to many human diseases. Metabolites can serve as a metabolic disease biomarker. An detection of such biomarkers plays a significant role in the study of biochemical reaction and signaling networks. Early research mainly focused on the analysis of the metabolic networks. The issue of integrating metabolite networks with other available biological data to reveal the mechanics of disease-metabolite associations is an important and interesting challenge.

**Results:** In this article, we propose two new approaches for the identification of metabolic biomarkers with the incorporation of disease specific gene expression data and the genome-scale human metabolic network. The first approach is to compare the flux interval between the normal and disease sample so as to identify reaction biomarkers. The second one is based on the Reaction-Reaction Network (RRN) to reveal the significant reactions. These two approaches utilize reaction flux obtained by a Linear Programming (LP) based method that can contribute to the discovery of potential novel biomarkers.

**Conclusions:** Biomarker identification is an important issue in studying biochemical reactions and signaling networks. Two efficient and effective computational methods are proposed for the identification of biomarkers in this article. Furthermore, the biomarkers found by our proposed methods are shown to be significant determinants for diabetes.

## Background

Deficiency in essential metabolites can directly cause metabolic diseases. Metabolic diseases profiling is promising in uncovering the mechanism of disease-metabolite associations. Existing research mainly emphasized on the analysis of metabolic networks [1-3]. Models in investigation of large-scale metabolic networks outperform other quantitative approaches [4,5]. The widespread appearance of gene expression data gives a clue for the integration of metabolite network data to reveal significant biomarkers.

Flux Balance Analysis (FBA) [6] is a constraint-based and traditional approach for predicting flux distribution. It has been employed in [7] to identify a number of important metabolic reactions. Drug targets are adopted to

reduce abnormal metabolites through formulating an optimal combinatoric problem on metabolic networks [8,9]. In [10], drug target prediction can be formulated as an integer linear programming model. A quantitative method based on two-stage FBA has been proposed in [11] for drug target identification. In [12], by profiling human metabolic reactions, a drug-reaction network was established for predicting enzyme targets. Here we develop a computational approach to identify metabolic biomarkers using human metabolic reactions incorporating disease-specific gene-expression data [13]. Metabolic biomarkers are metabolites demonstrating consistent variation in concentration in disease state; they can be very useful for a diagnostic purpose, see for instance [14]. As an efficient diagnostic tool and a safe evaluator for drug candidates, metabolomics will play an important role.

Furthermore, we also construct a RRN which provides a platform for ranking the significance of reactions by

\* Correspondence: wching@hku.hk

<sup>3</sup>Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong  
Full list of author information is available at the end of the article

using the PageRank algorithm. The reaction network can be constructed in such a way that the nodes represent reactions and an edge is placed whenever the reactions share the common metabolite. Note that one reaction can have thousands of edges with the related reactions which makes the network extremely complicated. This issue can be addressed by identifying densely connected subgraphs. The clustering toolbox is therefore employed to find the subnetworks. The graph clustering is based on the assumption that a group of functionally related nodes are likely to highly interact with each other while being more separate from the rest of the network [15]. In [16], the challenge of the clustering network graphs was presented. In particular, the results of most methods are highly sensitive to their parameters and the predicted clusters can vary from one method to another. Here we focus on analyzing the cluster with the largest number of entities. The selected metabolic reactions provide a platform for us to draw the RRN which depicts the interactions of the reactions. For the ease of interpretation and visualization of the networks, we apply Cytoscape to construct the network. PageRank approach is a promising method to evaluate the importance of a webpage. We then integrate the PageRank algorithm and the FBA method to evaluate the significance of the reactions. We also propose simple statistical criteria to select significant reactions which enable us to identify the corresponding metabolites.

Diabetes Mellitus is a group of metabolic diseases that are amongst the major human malnutrition diseases. Risk assessment is one of the possible ways to prevent the disease. Metabolic profiling, an unbiased technique, can potentially trigger the identification of high-risk candidates and therefore it can reduce the related costs [17].

We then integrate the human metabolic network with disease-specific gene expression data to analyze the flux profiles within the network. In the following section, we introduce two methods for metabolite biomarkers discovery. The validity of the two approaches will be further discussed. The identified metabolite biomarkers may have potential applications for disease diagnosis.

## Materials and methods

### Materials

The genome-scale human metabolic network reconstructed by Duarte et al. [18] consists of 3742 reactions, 2766 metabolites and 1905 genes. Three types of information have been used to describe a metabolic network. One of them is stoichiometry, which is used to depict the quantitative associations among reactants and products in all the involved reactions. Another part consists of enzymes corresponding to each reaction in the network. The last part is the flux capacity of each reaction. We employ Human Recon 1, one of the two independently

developed human metabolic networks [18,19] in our study. The data is available at the BiGG database (<http://bigg.ucsd.edu/>). One can retrieve the reactions and the involved genes using MATLAB. And the RRN can be implemented in the Cytoscape software package which is available at (<http://www.cytoscape.org>).

### Methodology

We introduce two novel methods for integrating gene expression data and the human metabolic network in biomarker discovery.

#### Flux profile comparison (FPC) method [13]

There are several major steps that we have to conduct before the construction of LP model, and eventually allow us for the detection of biomarkers.

1. **Expression levels in reactions**
2. **The LP model**
3. **Flux profiles in disease/normal samples**
4. **Identification of significant reactions**
5. **Significant metabolite discovery**

Comparing to the well known model using the human metabolic network to predict metabolic biomarkers of human inborn errors of metabolism [14], our model takes more realistic constraints into consideration. Firstly, the genome-scale human metabolic network we utilize here consists of 1905 boundary metabolites and 3742 reactions in total. Secondly, we integrate gene expression data in both normal and disease state to mark highly and lowly expressed reactions. Without forcing the reactions to be active in the normal state or inactive in the disease state, we adopt a probability measure for the reaction to be active or inactive instead. We use two pairs of gene expression data in both healthy and disease status and consider the overlap of the discovered metabolic biomarkers. Regarding the solutions of the LP problems, we use a large-scale optimization method which is based on Linear Interior Point SOLver (LIPSOL) [20] in MATLAB on a Windows Vista machine. These characteristics of our approach contribute to the discovery of metabolic biomarkers in a more significant way.

#### RRN construction method

In this section, we propose the second novel approach to identify the metabolite biomarkers based on RRN.

##### • **Clustering toolbox for identifying the subnetworks.**

The availability of the human metabolic network from Duarte et al. [18] enables us to retrieve the reactions and the genes. It includes 3742 reactions with 2766 metabolites and 1905 genes, which suggests a potential way to build a network if the reactions share the common metabolites. The nodes of the network represent the reactions,

while the reactions are linked if they have the same metabolite. Some reactions have thousands of edges based on the foundation of constructing the network. Therefore the RRN can be extremely complex which makes it difficult to draw the network by using Cytoscape. We address this problem by using a clustering toolbox to identify the subnetworks which have similar properties. Clustering analysis [21] aims to classify a set of observations into two or more mutually exclusive unknown groups based on combinations of variables. Thus, cluster analysis is usually adopted in the context of unsupervised classification [22]. It can be applied to a wide range of biological study cases, such as microarray, sequence and phylogenetic analysis [23]. The purpose of clustering is to group different objects together by observing common properties of elements in a system. In a biological network, this can help identify similar biological entities, like proteins that are homologous in different organisms or that belong to the same complex and genes that are co-expressed [24,25]. Among all the clustering algorithms, the  $k$ -means algorithm [26] aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean. The  $k$ -means method and its modifications are widely used for gene expression data analysis [27]. Here we utilize this method to classify the reactions into several groups. Reactions in the same cluster have similar behavior. We focus on analyzing the cluster with the largest number of elements and we choose  $k$  to be 50. With different  $k$  the classification is consequently different. It is interesting to note that the elements in the largest cluster are quite similar. Thus parameter  $k$  does not have much influence on the results of the cluster. We finally choose the largest cluster with 134 reactions for our analysis. And the network is shown in Figure 1.

• **The PageRank algorithm for evaluating the reactions in the RRN.**

With the obtained reaction network, one can evaluate the significance of each network. The page rank of a webpage is a number for representing the relative importance of the webpage based on the number of inbound and outbound links. Inbound links are links from outside pointing to a webpage. Outbound links are links from a webpage to any other webpages [28]. The page rank of a webpage can be obtained from the following formula:

$$P_i = (1 - d) + d \sum_{j \in M(i)} \frac{P_j}{L(j)}$$

where  $P_i$  is the page rank of the webpage  $i$ ,  $M(i)$  is the set of the webpages linked to webpage  $i$ ,  $L(j)$  is the number of outbound links of webpage  $j$ , and  $d$  is a residual probability which is usually set to be 0.85 [28]. Here we remark that the numerical results are similar in our

experiments for other  $d \geq 0.85$ . The values of the PageRank of the webpages are the entries of the dominant eigenvector of the modified adjacency matrix. We denote the eigenvector

$$\mathbf{r} = [P_1, P_2, \dots, P_N]^T$$

where  $N$  is the total number of pages and  $\mathbf{r}$  is the solution of the recursive formula:

$$\mathbf{r} = \begin{bmatrix} 1 - d \\ 1 - d \\ \vdots \\ 1 - d \end{bmatrix} + d \begin{bmatrix} \ell(1, 1) & \ell(1, 2) & \dots & \ell(1, N) \\ \ell(2, 1) & \ddots & & \vdots \\ \vdots & & & \vdots \\ \ell(N, 1) & \ell(N, 2) & \dots & \ell(N, N) \end{bmatrix} \mathbf{r}$$

and the adjacency matrix  $\ell(i, j)$  is 0 if webpage  $i$  does not link to webpage  $j$ , and we have the normalization condition that, for each  $j$

$$\sum_{i=1}^N \ell(i, j) = 1$$

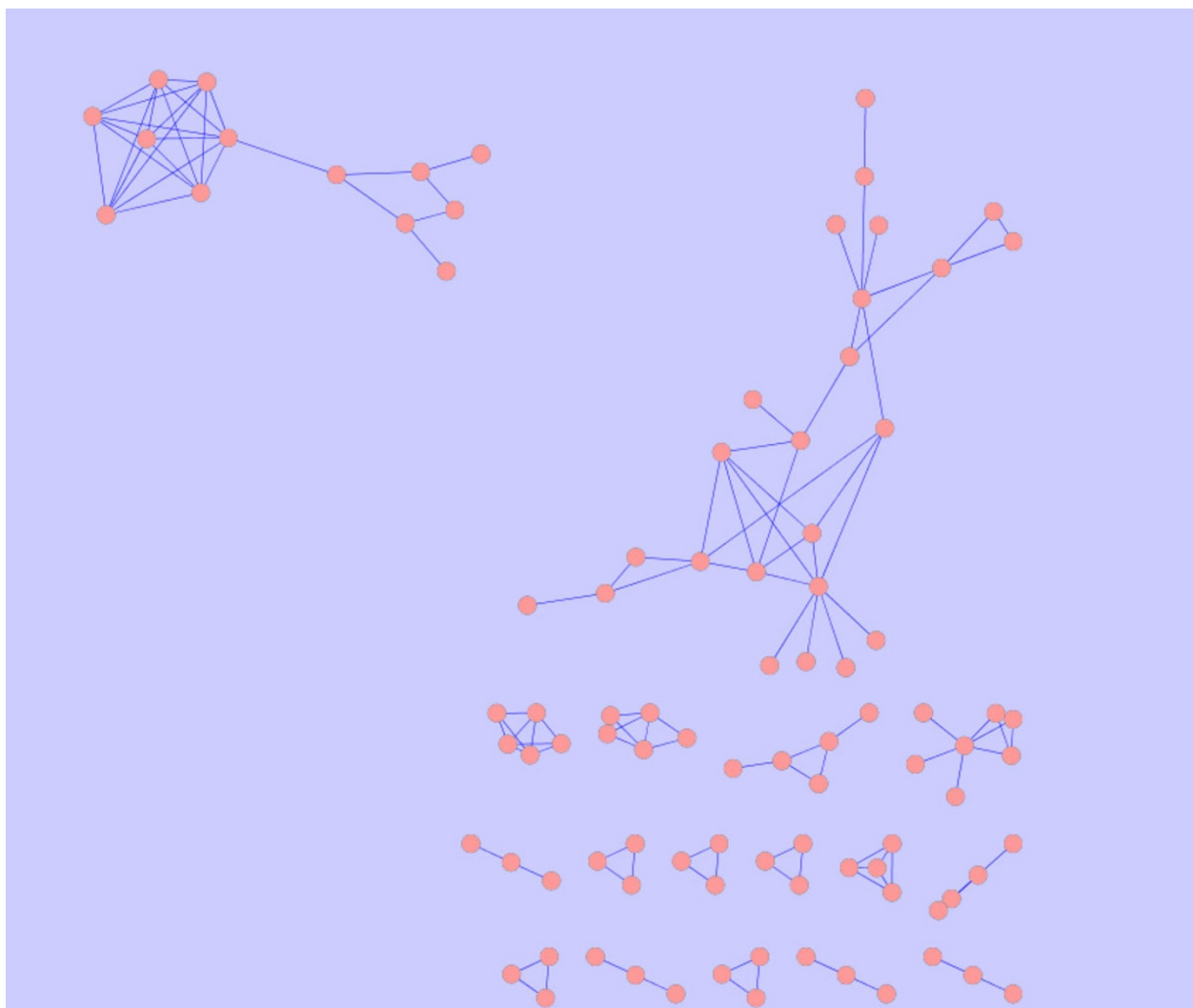
i.e., the sum of each column is 1. The value of the ranking indicates an importance of a particular page. Inspired by [29], one can apply the PageRank algorithm to rank the importance of the reaction in the RRN. Here we apply this approach on two subgraphs which are exactly the subsets of Figure 1. The subgraphs are described in Figures 2 and 3. Figure 2 is the left upper corner of Figure 1 and Figure 3 is the right upper side of Figure 1. Then we can evaluate the nodes (reactions) in these two subnetworks. Furthermore, each reaction also has the flux value which represents the significance. Here we propose a simple approach to integrate flux value and rank value to yield a final significant score for each reaction. Let  $\mathbf{z}$  be the vector containing the final values used for ranking the reactions. We define

$$\mathbf{z}_k = \mathbf{p}_k \times \mathbf{v}_k$$

where  $\mathbf{p}$  is the ranking result obtained from the PageRank algorithm,  $\mathbf{v}$  is the flux vector for all the reactions generating from the flux analysis, and  $k$  represents the  $k$ th entry (reaction) of the vector. We then select several reactions with the comparatively high value from the RRN using this criteria.

**Results and discussions**

In this section, we discuss some of our findings by our proposed two approaches: FPC and RRN. For FPC, we have filtered out 5 reactions and all the participating genes in these reactions [13]. In terms of genes, we have identified 11 significant genes involved in diabetes, "ALDH" and its variants(9 in total), "HSD3B2" and "KHK". In [30],



**Figure 1** Reaction network.

“ALDH” activity has been experimentally shown to be related to the increasing risk of large vessel disease in diabetes. Direct intra-pancreatic delivery of ALDH activity could be a potential feasible strategy for diabetes [31]. Furthermore, researchers have found that, ALDH2.487Lys allele was related to the decreasing prevalence odds in type II diabetes in a clinical study on diabetes [32]. “HSD3B2” is discovered highly expressed with regulation of FXR (farnesoid  $\times$  receptor) while FXR agonists are an appearing therapeutic treatment for diabetes [33]. The value of “KHK” as a pharmacological target needs further verification [34], but it can be a possible biomarker in diabetes treatment.

Considering the metabolites related diabetes, “ac[e]” acting as an inhibitor, is very helpful for patients in clinical trials, see for example [35]. Both “nadph” and “nadp” are valuable metabolites in l-xylulose (l-xylulose is

obtained by “nadph” and “nad” reduction with “d-xylulose” [36]) which is intensively used in diabetes diagnosis. While reaction 1951 is glycolaldehyde dehydrogenase, glycolaldehyde has been shown to play a significant role in diabetic cardiomyopathy [37]. And “pi” is a critical component in the disturbance of diabetes [38].

Furthermore, we analyze the significant reactions selected by integrating the flux analysis and the PageRank algorithm based on the subnetworks of the RRN. We filter out five reactions which are reported in Table 1. All the participating genes in these five reactions are listed in Table 2.

We remark that the symbol “ $\rightleftharpoons$ ” means the reaction is reversible and “ $\rightarrow$ ” means the reaction is irreversible. The number inside the parentheses (.) is the quantity of the metabolite. For example, in the Reaction 3511 of Table 1, we need “coa[m]:tetpent6crn[m] = 1:1” to produce “crn

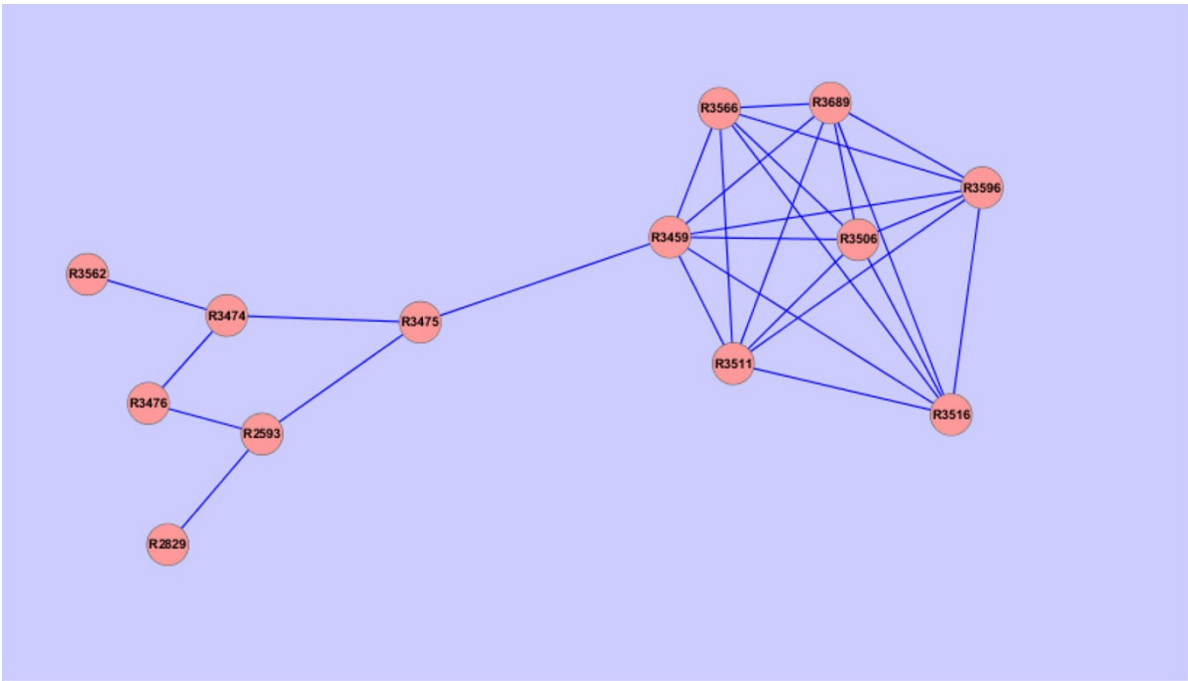


Figure 2 The left upper corner of figure 1.

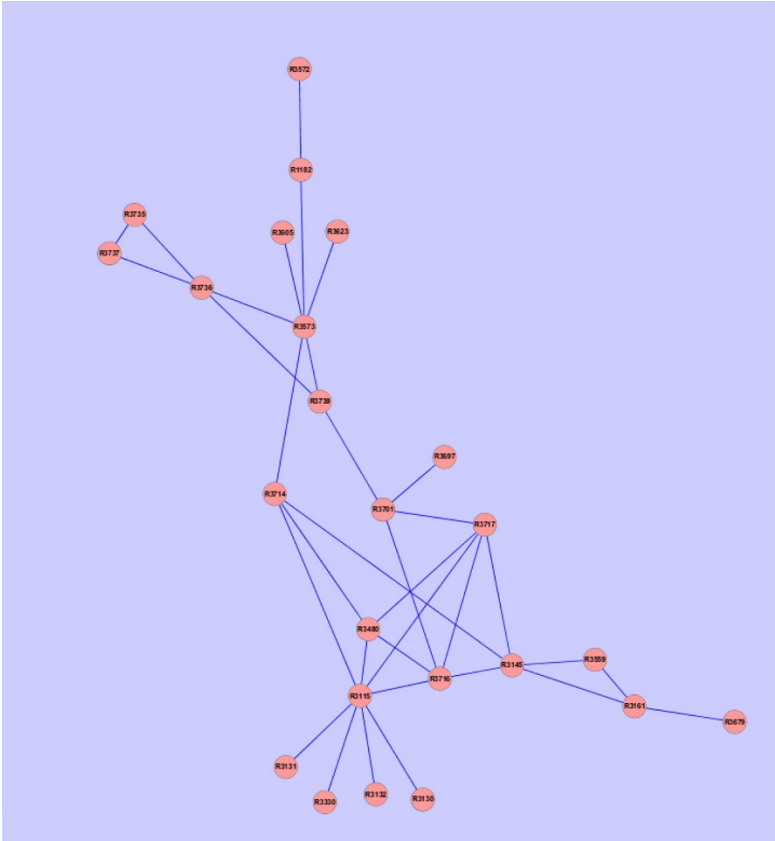


Figure 3 The right upper corner of figure 1.

**Table 1 Significant reactions selected from RRN**

Index	Reaction Equation
3115	"[x]:h2o+prpncoa→3hpcoa"
3161	"[c]:pi+uri ⇌ r1p+ura"
3573	"[c]:h+nadh+q10→nad+q10h2"
3474	"[m]:fad+succ ⇌ fadh2+fum"
3511	"[m]:coa+tetpent6crn→crn+tetpent6coa"

[m]:tetpent6coa[m] = 1:1". In the associated genes of Table 2, "CPT2" and "SDHD" etc are the gene symbols.

Considering the genes involved in the reactions, we have identified 9 important genes participating in Diabetes in Table 2. In [39], experiments have shown that gene "EHHADH" is involved in mitochondrial fatty acid  $\beta$ -oxidation and the variants in "EHHADH" are associated with type 2 Diabetes. In [40], it is demonstrated that obese diabetes impairs the rhythmic expression of various genes, including "UPP2". It has been shown that Diabetes appears to be associated with increased levels of oxidative stress in [41]. And gene "TXNRD1" exhibited increases in oxidative stress. "SDHB", "SDHA", "SDHC" and "SDHD" are four protein subunits forming succinate dehydrogenase. The role of "SDH" needs further investigation, but it is suspected that malfunction of the SDH complex can cause a hypoxic response in the cell that leads to tumor formation. Pharmacological inhibition of the CPT system by the glycidic acid derivative etomoxir, an irreversible and nonisoform-specific active site inhibitor of CPTs, has been demonstrated to reduce fasting blood glucose in an animal model of type 2 diabetes mellitus [42].

From the perspective of the metabolites related to the disease, reaction 3115 is an active reaction involved in fatty acid cholesterol metabolism during prostate cancer progression [43]. Here "pi" is a determining factor in regulation of metabolism in diabetes [38] and "nadh" involves in lactate formation (3-4-hydroxyphenyl lactate formation) [43]. Both "fad" and "fadh2" play an important role in fatty acids metabolism [44]. The role of malonyl CoA as a key glucose-derived metabolite, an allosteric inhibitor of fatty acid oxidation, has attracted many attentions. Recent studies have investigated the effects of manipulating this metabolite in various tissues [45]. For the issue of diabetes, a study in [46]

**Table 2 Significant genes for diabetes in RRN**

Reactions Index	Genes
3115	"EHHADH"
3161	"UPP1" "UPP2"
3573	"TXNRD1"
3474	"SDHD" "SDHC" "SDHB" "SD-HA"
3511	"CPT2"

demonstrated that high levels of malonyl CoA and reduced fat oxidation enhance glucose disposal in primary human skeletal myocytes.

One can see that these two approaches are extremely different, but some of the metabolite biomarkers are the same. It has been shown that "nadh" and "pi" are the important factors for detecting diabetes. We can conclude that both our proposed methods are effective for identifying the metabolite biomarkers for diseases.

## Conclusions

In this paper, we first develop a computational method to identify significant genes and metabolites for metabolic diseases. LP based strategy is then utilized to obtain flux profiles in disease/normal samples. Gene expression data in two pairs of samples at disease/normal states contributes to discovering genes and metabolites that can be potential biomarkers. We then further present a second novel approach to identify the significant metabolites for the metabolic disease. We also employ the constraint-based flux distribution to analyze the metabolic network. The clustering method makes it possible to identify the subgraphs with the common properties, which is a key step to construct the RRN with Cytoscape. To evaluate the reaction in the network, we propose the PageRank algorithm to evaluate the node. We integrate the flux value and the rank result to select the significant reactions, from which the related metabolites biomarkers can be identified. The integration of genome-scale human metabolic network data with gene expression levels offers a new way for systematically identifying potential biomarkers.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

LL and HJ came up with the idea. WKC, LL, HJ and VSV designed the research. HJ and YQ performed the research and analyzed the results. WKC, HJ, LL, YQ and VSV wrote the paper. All authors read and approved the final manuscript.

## Acknowledgements

The authors would like to thank the three anonymous referees for their helpful comments and constructive suggestions. The preliminary version of the paper has been presented in ISB 2012 and published in the corresponding conference proceedings [13].

## Declarations

The publication of this article has been funded by the GRF Grant, HKU CERG Grants, Hung Hing Ying Physical Research Grant and National Natural Science Foundation of China Grant Nos. 10971075 and S201201009985. This article has been published as part of *BMC Systems Biology* Volume 7 Supplement 2, 2013: Selected articles from The 6<sup>th</sup> International Conference of Computational Biology. The full contents of the supplement are available online at <http://www.biomedcentral.com/bmcsystbiol/supplements/7/S2>.

## Authors' details

<sup>1</sup>Institute of Information and System Science, Xi'an Jiaotong University, Xi'an, China. <sup>2</sup>Department of Mathematics, School of Information, Renmin University of

China, Beijing, China. <sup>3</sup>Advanced Modeling and Applied Computing Laboratory, Department of Mathematics, The University of Hong Kong, Pokfulam Road, Hong Kong. <sup>4</sup>Department of Chemical Engineering and Biotechnology, University of Cambridge, Pembroke Street, Cambridge, UK.

Published: 17 December 2013

## References

1. Reich JG, Selkov EE: *Energy Metabolism of the Cell: A Theoretical Treatise* New York: Academic Press; 1981.
2. Fell DA: *Understanding the Control of Metabolism* London: Portland Press; 1996.
3. Varner J, Ramkrishna D: **Metabolic Engineering from a Cybernetic Perspective. 1. Theoretical Preliminaries.** *Biotechnology Progress* 1999, **15**(3):407-425.
4. Bailey JE: **Complex Biology with No Parameters.** *Nat Biotechnol* 2001, **19**:503-504.
5. Palsson B: **The Challenges of in Silico Biology.** *Nat Biotechnol* 2000, **18**:1147-1150.
6. Kauffman KJ, Prakash P, Edwards JS: **Advances in Flux Balance Analysis.** *Current Opinion in Biotechnology* 2003, **14**:491-496.
7. Almaas E, Oltvai ZN, Barabasi AL: **The Activity Reaction Core and Plasticity of Metabolic Networks.** *Plos Computational Biology* 2005, **1**:e68.
8. Sridhar P, Kahveci T, Ranka S: **An Iterative Algorithm for Metabolic Network-based Drug Target Identification.** *Pacific Symposium on Biocomputing* 2007, **12**:88-99.
9. Sridhar P, Song B, Kahveci T, Ranka S: **Mining Metabolic Networks for Optimal Drug Targets.** *Pacific Symposium on Biocomputing* 2008, **13**:291-302.
10. Li Z, Wang RS, Zhang XS, Zhang L: **Detecting drug targets with minimum side effects in metabolic networks.** *IET Syst Biol* 2009, **3**(6):523-33.
11. Li Z, Wang RS, Zhang XS: **Two-stage flux balance analysis of metabolic networks for drug target identification.** *BMC Syst Biol* 2011, **5**(Suppl 1):S11.
12. Li L, Zhou X, Ching W, Wang P: **Predicting Enzyme Targets for Cancer Drugs by profiling Human metabolic Reactions in NCI-60 Cell Lines.** *BMC Bioinformatics* 2010, **11**:501.
13. Li L, Jiang H, Ching W, Vassiliadis V: *Metabolite Biomarker Discovery For Metabolic Diseases By Flux Analysis* Proceedings of the 6th IEEE International Conference on Systems Biology (ISB 2012), Xian China; 2012, 1-5, IEEE.
14. Shlomi T, Cabili MN, Ruppin E: **Predicting Metabolic Biomarkers of Human Inborn Errors of Metabolism.** *Molecular Systems Biology* 2009, **5**:263.
15. Aittokallio Tero, Schwikowski benno: **PGraph-based Methods for Analysing Networks in Cell Biology.** *Bioinformatics* 2006, **7**(3):243-255.
16. Dhaeseleer P: **How Does Gene Expression Clustering Work?** *Nat Biotechnol* 2005, **23**:1499-1501.
17. Harrigan GG, Goodacre R: *Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis* London: Kluwer Academic Publisher; 2003.
18. Duarte NC, Becker SA, Jamshidi N, Thiele I, MO ML, Vo TD, Srivas R, Palsson BO: **Global Reconstruction of the Human Metabolic Network Based on Genomi and Bibliomic data.** *Proc Natl Acad Sci USA* 2007, **104**(6):1777-1782.
19. Ma H, Goryanin I: **Human Metabolic Network Reconstruction and Its Impact on Drug Discovery and Development.** *Drug Discovery Today* 2008, **13**:402-408.
20. Zhang Y: **Solving Large-Scale Linear Programs by Interior-Point Methods Under the MATLAB Environment.** Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD; 1995, Technical Report TR96-01.
21. Jain AK, Murty MN, Flynn PJ: **Data Clustering: A Review.** *ACM Computing Surveys (CSUR)* 1999, **31**(3):264-323.
22. Duda RO, Hart PE, Stork DG: **Pattern Classification. ch.10: Unsupervised Learning and Clustering.** *John Wiley & Sons, New York* 2001.
23. Saitou N, Nei M: **The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees.** *Mol Biol Evol* 1987, **4**(4):406-425.
24. Borate BR, Chesler EJ, Langston MA, Saxton AM, Voy BH: **Comparison of Threshold Selection Methods for Microarray Gene Co-expression Matrices.** *BMC Res Notes* 2009, **2**:240.
25. Perkins AD, Langston MA: **Threshold Selection in Gene Co-expression Networks Using Spectral Graph Theory Techniques.** *BMC Bioinformatics* 2009, **10**:S4.
26. Macqueen B: **Some methods for classification and Analysis of Multivariate Observations.** In *In proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability. Volume 1.* Berkeley, University of California press; 1967:281-297.
27. Lu Y, Lu S, Fotouhi F, Deng Y, Brown SJ: **Incremental Genetic k-means Algorithm and Its Application in Gene Expression Data Analysis.** *BMC Bioinformatics* 2004, **5**:172.
28. Brin S, Page L: **The Anatomy of a Large-Scale Hypertextual Web Search Engine.** *Computer Networks and ISDN Systems* 1998, **30**:107-117.
29. Ding Chris HQ, Zha Hongyuan, He Xiaofeng, Husbands Parry, Simon Horst D: **Link Analysis: Hubs and Authorities on the World WideWeb.** *Society for Industrial and Applied Mathematics* 2004, **46**(2):256-268.
30. Jerntorp P, Ohlin H, Almér LO: **Aldehyde dehydrogenase Activity and Large Vessel Disease in Diabetes Mellitus: A Preliminary Study.** *Diabetes* 1986, **35**:291-294.
31. Bell GI, Putman DM, Hughes-Large JM, Hess DA: **Intrapancreatic Delivery of Human Umbilical Cord Blood Aldehyde Dehydrogenase-producing Cells Promotes Islet Regeneration.** *Diabetologia* 2012, **55**:1755-1760.
32. Guang Y, Ohnaka K, Morita M, Tabata S, Tajima O, Kono S: **Genetic Polymorphisms of Alcohol Dehydrogenase and Aldehyde Dehydrogenase: Alcohol Use and Type 2 Diabetes in Japanese Men.** *Epidemiology Research International* 2011, **2011**(2011):583682.
33. Xing Y, Saner-Amigh K, Nakamura Y, Hinshelwood MM, Carr BR, Mason JI, Rainey WE: **The Farnesoid x Receptor Regulates Transcription of 3beta-hydroxysteroid Dehydrogenase Type II in Human Adrenal Cells.** *Mol Cell Endocrinol* 2009, **2**:153-162.
34. Diggle CP, Shires M, McRae C, Crellin D, Fisher J, Carr IM, Markham AF, Hayward BE, Asipu A, Bonthron DT: **Both Isoforms of Ketohexokinase are Dispensable for Normal Growth and Development.** *Physiol Genomics* 2010, **4**:235-43.
35. Codario RA: *Type 2 Diabetes, Pre-Diabetes, and the Metabolic Syndrome.* 2 edition. NewYork: Humana Press; 2011.
36. Bhagavan NV: **Carbohydrate Metabolism II: Gluconeogenesis, Glycogen Synthesis and Breakdown, and Aternative Pathways.** *Medical Biochemistry.* 4 edition. London: Academic Press; 2001.
37. Lorenzi R, Andrades ME, Bortolin RC, Nagai R, Dal-Pizzolo F, Moreira JC: **Glycolaldehyde Induces Oxidative Stress in the Heart: A Clue to Diabetic Cardiomyopathy?** *Cardiovasc Toxicol* 2010, **4**:244-249.
38. Ditezl J, Rvang HH, Nagai R: **Disturbance of Inorganic Phosphate Metabolism in Diabetes Mellitus: Its Impact on the Development of Diabetic Late Complications.** *Curr Diabetes Rev* 2010, **5**:323-333.
39. Banasik K, Justesen JM, Hornbak M, et al: **Bioinformatics-Driven Identification and Examination of Candidate Genes for Non-Alcoholic Fatty Liver Disease.** *PLoS One* 2011, **6**:e16542.
40. Andoa H, Oshimaa Y, Yanagihara H, Hayashia Y, Takamurab T, Kanekob S, Fujimura A: **Profile of Rhythmic Gene Expression in the Livers of Obese Diabetic KK-A Mice.** *Biochemical and Biophysical Research Communications* 2006, **346**(4):1297-1302.
41. Merrill CL, Ni H, Yoon LW: **Etomoxir-Induced Oxidative Stress in HepG2 Cells Detected by Differential Gene Expression Is Confirmed Biochemically.** *Toxicological Sciences* 2002, **68**:93-101.
42. Barnett M, Collier GR, O'Dea K: **The Longitudinal Effect of Inhibiting Fatty Acid Oxidation in Diabetic Rats Fed a High Fat Diet.** *Horm Metab Res* 1992, **24**(8):360-362.
43. Balbin OA, Shellman ER: **The metabolism of Human Prostate Cancer Progression: A Systems Approach.** *Bioinformatics* 2008, **5**:27.
44. Maassen JA, Romijn JA, Heine RJ: **Fatty acid-induced mitochondrial uncoupling in adipocytes as a key protective factor against insulin resistance and beta cell dysfunction: a new concept in the pathogenesis of obesity-associated type 2 diabetes mellitus.** *Diabetologia* 2007, **50**(10):2036-2041.
45. Muoio DM, Newgard CB: **Fatty Acid Oxidation and Insulin Action When Less Is More.** *Diabetes* 2008, **57**(6):1455-1456.
46. Bouzraki K, Austin R, Rune A, Lassman ME, Garcia-Roves PM, Berger JP, Krook A, Chibalin AV, Zhang BB, Zierath JR: **Malonyl Coenzyme-A Decarboxylase Regulates Lipid and Glucose metabolism in Human Skeletal Muscle.** *Diabetes* 2008, **57**:1508-1516.

doi:10.1186/1752-0509-7-S2-S13

Cite this article as: Li et al.: Discovery of metabolite biomarkers: flux analysis and reaction-reaction network approach. *BMC Systems Biology* 2013 **7**(Suppl 2):S13.