Profiling the human response to physical exercise: a computational strategy for the identification and kinetic analysis of metabolic biomarkers

Michael Netzer1, Klaus M Weinberger1,3,4, Michael Handler1, Michael Seger1, Xiaocong Fang5, Karl G Kugler2, Armin Graber2 and Christian Baumgartner1*

Abstract

Background: In metabolomics, biomarker discovery is a highly data driven process and requires sophisticated computational methods for the search and prioritization of novel and unforeseen biomarkers in data, typically gathered in preclinical or clinical studies. In particular, the discovery of biomarker candidates from longitudinal cohort studies is crucial for kinetic analysis to better understand complex metabolic processes in the organism during physical activity.

Findings: In this work we introduce a novel computational strategy that allows to identify and study kinetic changes of putative biomarkers using targeted MS/MS profiling data from time series cohort studies or other crossover designs. We propose a prioritization model with the objective of classifying biomarker candidates according to their discriminatory ability and couple this discovery step with a novel network-based approach to visualize, review and interpret key metabolites and their dynamic interactions within the network. The application of our method on longitudinal stress test data revealed a panel of metabolic signatures, i.e., lactate, alanine, glycine and the short-chain fatty acids C2 and C3 in trained and physically fit persons during bicycle exercise.

Conclusions: We propose a new computational method for the discovery of new signatures in dynamic metabolic profiling data which revealed known and unexpected candidate biomarkers in physical activity. Many of them could be verified and confirmed by literature. Our computational approach is freely available as R package termed BiomarkerR under LGPL via CRAN http://cran.r-project.org/web/packages/BiomarkerR/.

Introduction

In metabolomics the bioinformatics-driven search for highly-discriminatory biomarker candidates has become a key task in the biomarker discovery process with the objective of introducing novel biomarkers aiding in diagnosis or therapeutic management [1-4].

A wide spectrum of feature selection methods including filter, wrapper or embedded algorithms is available for the identification of significant features in biomedical datasets [5-9]. In particular filter algorithms calculate a measure (score), allowing to rank and prioritize putative biomarker candidates according to their predictive value [8]. However, research is still needed to provide bioinformatics methods for the scientific community that address paired/dependent test hypotheses or time series studies. In addition, the quantitative analysis of networks has increasingly become an important technique for the biological interpretation of changes in disease-associated metabolic pathways, allowing the study of interconnectivity, interaction or correlation among analytes. For this type of analysis, different types of topological graph descriptors (e.g., parametric or partition-based entropy measures) can be used to analyze such complex biological networks [10,11].

In this short report we propose a new computational strategy that identifies metabolic biomarker candidates according to their discriminatory ability from dependent...
samples, and we review and interpret them using a network-based approach. For the biomarker identification and prioritization step we apply a recently published filter algorithm, named **Biomarker Identifier** (BI), which calculates a score measure for every analyte, representing the discriminatory ability in terms of the product of sensitivity and specificity, and in an analogous way for paired samples [12]. After BI prioritization we apply a new method to infer a network from the data by calculating analyte ratios, representing interactions of analyte pairs in the network. This discovery step aims at verifying analyte ratios, representing interactions of analyte pairs in the network. This discovery step steps at verifying analyte pairs in the network the threshold \( \tau \) is set to 1 if 
\[
G_{ij} = \begin{cases} 
1 & |s_{ij}| > \tau \\
0 & \text{else}, 
\end{cases}
\]
for \( i, j \in 1, ..., |M| \). A ratio \( r \in R \) is designated as a putative pathway reaction of the form \( A \rightarrow B \), where a reactant \( A \) is metabolized into a product \( B \) via single or multiple reaction paths. To consider significant predictor pairs in the network the threshold \( \tau \) has been evaluated using controlled simulated data in form of \( D(D \sim N(10, 1)) \), see “Additional file 1”) as proposed by Guo et al. [15]. Next, we inferred the network for different values for \( \tau \) and used vertices (metabolites) with at least one edge (i.e., degree > 0) as input for classification and calculated the mean accuracy of the classifier using 10-fold cross-validation (see “Additional file 1”).

In contrast to a static network, typically constructed from data of independent case/control studies, a kinetic network can be inferred on single analytes or analyte pairs (as done in this work) with changes in levels greater than the fixed threshold \( \tau \) at timepoint \( t_0 \) vs. baseline \( t_0 \), representing the dynamics of circulating metabolites over time.

Coupling step 1 with step 2 of our discovery strategy allows for verifying preselected metabolites of step 1 as highly connected vertices (hubs) in the network in step 2. Note that a high degree of interconnectivity of a hub (i) represents analogously a high discriminatory and predictive value of this hub, and (ii) embedded in a network of pathway reactions a key role of normal or abnormal metabolism.

Our computational approach is freely available as R package termed **Biomarker** under LGPL via CRAN http://cran.r-project.org/web/packages/Biomarker/. We chose to implement our method in R because of the broad abundance of this programming language in the bioinformatics community and its open source nature. Additionally, there is a multitude of packages available for the handling and analysis of network-based data (e.g., igraph[16], QuACN[11] or BioNet[17]).

**Bicycle stress test**

We here present a new computational method for the search for stress biomarkers in physically fit persons using targeted MS/MS profiling technology. Using this approach 60 analytes were identified by applying
characteristic mass transitions in multiple reaction monitoring, precursor and neutral loss scans, and quantified using internal standards added to the samples [18].

Briefly, metabolite profiling of blood samples of a total of 30 active and physically fit individuals (22 males, 8 females) with a mean age of 38.33 ± 7.16 years and mean body mass index 23.88 ± 2.50 kg/m² was carried out. The used data represents analyses of residual material of a standard sport-physical examination where each proband had to bicycle on an exercise bike for increasing steps of Watt (W) levels (each step 25 W) until the individual’s maximum capacity was reached.

Capillary blood samples were obtained from the ear lobe before starting the exercise (at rest, \( t_0 \)), and at all Watt levels up to the individual’s maximum performance (\( t_{\text{max}} \)). A total of 60 metabolites (lactate, amino acids and acyl carnitines) were measured in absolute concentration values (\( \mu \text{mol/L} \)) [19]. All individuals gave written informed consent to the attending physician.

### Results and discussion

Using our computational approach the pBI priority model was applied as first step to preselect key metabolites from the measured pool of 60 analytes by computing scores at timepoint \( t_0 \) (at rest) vs. \( t_{\text{max}} \) (at individual maximum performance). Figure 1 shows the pBI scores, exhibiting five performance (\( t_{\text{max}} \)). A total of 60 metabolites (lactate, alanine, glycine and the two short-chain acyl carnitines C2 and C3) categorized as strong pre-

Figure 1 pBI scores pBI scores of top ranked analytes lactate, alanine, C2, C3 and glycine showing increased levels when comparing subjects at rest vs. individual maximum. Dashed lines indicate the score cut-off for strong predictors (\(|\text{pBI}| > 73\)), as defined in [12], images/pBIScores.png.

Results and discussion

Using our computational approach the pBI priority model was applied as first step to preselect key metabolites from the measured pool of 60 analytes by computing scores at timepoint \( t_0 \) (at rest) vs. \( t_{\text{max}} \) (at individual maximum performance). Figure 1 shows the pBI scores, exhibiting five metabolites (i.e., lactate, alanine, glycine and the two short-chain acyl carnitines C2 and C3) categorized as strong predictors. Figure 2 demonstrates the corresponding dynamic network for \( \tau = 73 \) again with lactate, alanine, C2 and C3 as major hubs, and now in the role as representatives of a panel of reaction pairs, coinciding with the univariate pBI metabolite ranking. Interestingly, using the controlled simulated data a threshold of \( \tau = 73 \), which corresponds to the cut-off score for a strong predictor as defined in [12], led to the maximum mean accuracy using a K-nearest-neighbor classifier [20], outperforming also commonly used correlation-networks [21] (see “Additional file 1”).

To validate our findings a literature review and analysis of related KEGG pathways were performed. Among the total of 60 detected metabolites, our analysis revealed 5 key metabolites (lactate, alanine, glycine, and the two short-chain acyl carnitines C2 and C3) associated with physical exercise. These results are consistent with previous reports [19,22,23]. As is well known, anaerobic glycolysis is the main way for energy supply during exercise workout. Lactate is a major end product of the metabolism of glucose through the glycolytic pathway [23]. The skeletal muscle is the main organ producing large amounts of lactate. Typically, the production of lactate is greatly increased during exercise via the oxidative metabolism [23]. In our work, we detected a pBI score for lactate greater than 400 which represents a strong correlation with physical exercise. The known biochemical or physiologic effects of carnitine suggest that supplementation of carnitine may improve exercise performance [24,25]. First, carnitine is required for mitochondrial fatty acid oxidation, which would permit glucose utilization to decrease, and thus preserve muscle glycogen content and ensure maximal rates of oxidative ATP production [24,26]. Second, generation of acetylcarnitine would potentially decrease acetyl-CoA content, relieving inhibition of pyruvate dehydrogenase and decrease the production of lactate [23,24,27]. All of these can potentially improve physical performance during high-intensity exercise. Besides, recent studies demonstrated that short-term administration of glycine propionyl-L-carnitine (GPLC) significantly elevates levels of nitric oxide metabolites at rest and in response to reactive hyperemia [28-30], and can also enhance exercise performance in healthy, trained individuals [28]. Carnosine is synthesized in skeletal muscle from L-histidine and A-alanine amino acids [22]. One important physiological role of carnosine is the maintenance of acid-base homeostasis [22,31]. Studies have shown that supplementation with A-alanine or exercise can increase muscle carnosine content and therefore total muscle buffer capacity with the potential to cause improvements in physical exercise [19,22].

Overall, more than 20 pathways were revealed. We selected the most related pathways which include at least 2 metabolites identified by our coupled 2-step discovery step (Table 1). It shows that the citrate cycle (TCA cycle), multiple amino acid and fatty acid metabolisms are greatly activated in physical exercise.
Conclusion

We have introduced a powerful tool for the search, prioritization and network analysis of putative biomarker candidates in metabolomic studies. Our 2-step approach has several benefits: 1) BI can be applied to dependent samples, calculating absolute scores for prioritizing biomarker candidates into classes of weak, moderate or strong predictors, and computes positive and negative scores, indicating whether the metabolites concentration is increased or decreased compared to its reference. 2) The proposed approach allows to review and interpret findings and thus aids in biochemical interpretation of (ab)normal metabolism by reviewing pathway reactions within the network.

Using our coupled 2-step discovery strategy, we were able to identify and confirm multiple metabolites, i.e., lactate, glycine, alanine, C2 and C3 that are closely associated with metabolism of physical activity [22,23,32].

Figure 2 Inferred network. The resulting kinetic network ($\tau = 73$) exhibits lactate, alanine, C2 and C3 as major hubs (green color) in the center of the network with the highest degree of connectivity. Note that an edge between metabolite $m_i$ and $m_j$ indicates a pBI value of the absolute logarithmic ratio of $m_i$ and $m_j$ greater than $\tau$ (e.g., for the edge between C2 and C16 $|$pBI$|\log_2 (|C2/C16|)| > 73$). Identified key metabolites during physical activity could also be validated by literature [22,23,32]. The number following the underscore of acyl-carnitines symbols indicates the number of double bonds of the carbon-carbon bonds (e.g., for C10_1 the number of double bounds is 1). The network is visualized using Cytoscape [33]. images/pbiGraph.png.
Table 1 KEGG Pathways

| KEGG ID   | KEGG pathway                                      | Related metabolites                                                                 |
|-----------|---------------------------------------------------|--------------------------------------------------------------------------------------|
| Map00330* | Arginine and proline metabolism                    | Alanine, aspartate, ornithine, proline, citrulline, arginine, glutamic acid, lysine, methionine, glycine |
| Map00020* | Citrate cycle                                      | Glycose, leucine, valine, arginine, proline, tyrosine, alanine, aspartate, glutamic acid |
| Map00260* | Citrate cycle                                      | Serine, glycine, threonine, valine, aspartate, lysine, proline, methionine, leucine, arginine |
| Map00250* | Alanine, aspartate and glutamate metabolism        | Aspartate, arginine, histidine, glutamic acid, alanine, proline, lysine, serine, glycine, glucose |
| Map00620* | Pyruvate metabolism                               | Glycine, lactate, serine, leucine, lysine, valine, threonine                         |
| Map00640* | Propanoate metabolism                             | Alanine, lactate, leucine, methionine, valine                                       |
| Map00270 | Cysteine and methionine                            | Methionine, threonine, alanine, aspartic acid, glycine, serine                       |
| Map00473  | D-alanine metabolism                               | Alanine, aspartate, glutamate, glucose, ornithine                                    |
| Map00410  | α-alanine metabolism                               | Aspartate, alanine, histidine, lysine, arginine                                      |
| Map00460  | Cyano amino acid metabolism                        | Serine, glycine, alanine, glutamate, aspartate                                       |
| Map00300* | Lysine biosynthesis                                | Glycine, serine, threonine, lysine, alanine                                         |
| Map00710* | Carbon fixation in photosynthetic organisms        | Tyrosine, phenylalanine, alanine, aspartate, glycine                                 |
| Map00071* | Fatty acid metabolism                              | Multiple carnitines, alanine, aspartate                                             |

Identified KEGG pathways in which metabolites selected by the inferred network are shown. The asterisks (*) indicate metabolic pathways associated with physical exercise (see also section “Results and discussion”).

Acknowledgements
This work was supported by the Austrian Genome Research Program GEN-AU (Bioinformatics Integration Network, BIN III). The authors also thank Matthias Dehmer and Laurin Müller for helpful discussion.

Author details
1Research Group for Clinical Bioinformatics, Institute of Electronic, Electronic Bioengineering, UMIT, 6060 Hall in Tirol, Austria. 2Institute for Bioinformatics and Translational Research, UMIT, 6060 Hall in Tirol, Austria. 3Boccard J, Veuthey JL, Rudaz S: Knowledge discovery in metabolomics: an overview of MS data handling. J Sep Sci 2010, 33(3):290-304.

Competing interests
The authors declare that they have no competing interests.

Received: 3 August 2011 Accepted: 19 December 2011

References
1. Lewis GD, Asnani A, Gurzen RE: Application of metabolomics to cardiovascular biomarker and pathway discovery. J Am Coll Cardiol 2008, 52(2):117-123.
2. Boccard J, Veuthey JL, Rudaz S: Knowledge discovery in metabolomics: an overview of MS data handling. J Sep Sci 2010, 33(3):290-304.
3. Baumgartner C, Osl M, Netzer M, Baumgartner D: Bioinformatic-driven search for metabolic biomarkers in disease. J Clin Bioinforma 2011, 1:2.
4. Manns M, Dunn WB, Neyes L, Goodacre R: The role of metabolites and metabolomics in clinically applicable biomarkers of disease. Arch Toxicol 2011, 85:5-17.
5. Hall MA, Holmes G: Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. IEEE T Knowl Data Eng 2003, 15:1437-1447.
6. Enot DP, Beckmann M, Overly D, Draper J: Predicting interpretability of metabolome models based on behavior, putative identity, and biological relevance of explanatory signals. Proc Natl Acad Sci USA 2006, 103(40):14865-14870.
7. Saeys Y, Inza I, Larrañaga P: A review of feature selection techniques in bioinformatics. Bioinformatics 2007, 23:2507-2517.
8. Netzer M, Millong G, Osl M, Pfeifer B, Praun S, Viliinger J, Vogel W, Baumgartner C: A new ensemble-based algorithm for identifying breath gas marker candidates in liver disease using ion molecule reaction mass spectrometry. Bioinformatics 2009, 25(7):941-947.
9. Inza I, Calvo B, Armananzas R, Bengoetxea I, Larrañaga P, Lozano JA: Machine learning: an indispensable tool in bioinformatics. Methods Mol Biol 2010, 593:23-46.
10. Dehmer M, Barbarini N, Varmuza K, Graber A: A novel topological descriptors for analyzing biological networks. BMC Struct Biol 2010, 10:18.
11. Mueller LAJ, Kugler KG, Dander A, Graber A, Dehmer M: QuACN: an R package for analyzing complex biological networks quantitatively. Bioinformatics 2011, 27:140-141.
12. Baumgartner C, Lewis GD, Netzer M, Pfeifer B, Gurzen RE: A new data mining approach for profiling and categorizing kinetic patterns of metabolic biomarkers after myocardial injury. Bioinformatics 2010, 26(14):1745-1751.
13. Kanehisa M, Goto S: KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000, 28:27-30.
14. Tang Y, Dogru AH, Kurless FJ, Tank MM: Computing Cyclomatic Complexity with Cubic Flowgraphs. Journal of Systems Integration 2001, 10:395-409.
15. Guo Y, Graber A, McBourny RN, Balasubramanian R: Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. BMC Bioinformatics 2010, 11:447.
18. Baumgartner C, Graber A. In Successes and new directions in data mining. Volume 7. Idea Group Inc; 2007:141-166.

19. Weinberger KM. Metabolomics in diagnosing metabolic diseases. Ther Umsch 2008, 65(9):487-491.

20. Parry RW, Jones W, Stokes TH, Phan JH, Moffitt RA, Fang H, Shi L, Oberthuer A, Fischer M, Tong W, Wang MD. k-Nearest neighbor models for microarray gene expression analysis and clinical outcome prediction. Pharmacogenomics J 2010, 10(4):292-309.

21. Mutwil M, Usadel B, Schutte M, Loraine A, Ebenhöh O, Persson S. Assembly of an interactive correlation network for the Arabidopsis genome using a novel heuristic clustering algorithm. Plant Physiol 2010, 152:29-43.

22. Artioli GG, Gualano B, Smith A, Stout J, Lancha AH. Role of beta-alanine supplementation on muscle carnosine and exercise performance. Med Sci Sports Exerc 2010, 42(6):1162-1173.

23. van Hall G. Lactate kinetics in human tissues at rest and during exercise. Acta Physiol (Oxf) 2010, 199(4):499-508.

24. Brass EP, Hiatt WR. The role of carnitine and carnitine supplementation during exercise in man and in individuals with special needs. J Am Coll Nutr 1998, 17(3):207-215.

25. Ahmadi S, Granier P, Taoutaou Z, Mercier J, Dubouchaud H, Prefaut C. Effects of active recovery on plasma lactate and anaerobic power following repeated intensive exercise. Med Sci Sports Exerc 1996, 28(4):450-456.

26. Costill DL, Coyle E, Dalsky G, Evans W, Fink W, Hoopes D. Effects of elevated plasma FFA and insulin on muscle glycogen usage during exercise. J Appl Physiol 1977, 43(4):695-699.

27. Bremer J. Pynuvate dehydrogenase, substrate specificity and product inhibition. Eur J Biochem 1969, 8(4):535-540.

28. Jacobs PL, Goldstein ER, Blackburn W, Orem I, Hughes JR. Glycine propionyl-L-carnitine produces enhanced anaerobic work capacity with reduced lactate accumulation in resistance trained males. J Int Soc Sports Nutr 2009, 6:9.

29. Smith WA, Fry AC, Tschumie LC, Bloomer RJ. Effect of glycine propionyl-L-carnitine on aerobic and anaerobic exercise performance. Int J Sport Nutr Exerc Metab 2008, 18:19-36.

30. Bloomer RJ, Tschumie LC, Smith WA. Glycine propionyl-L-carnitine modulates lipid peroxidation and nitric oxide in human subjects. Int J Vitam Nutr Res 2009, 79(3):131-141.

31. Harris RC, Tallon MJ, Dunnett M, Boobis L, Orem I, Hughes JR. The absorption of orally supplied beta-alanine and its effect on muscle carnosine synthesis in human vastus lateralis. Amino Acids 2006, 30(3):279-289.

32. Negrao CE, Ji LL, Schauer JE, Nagle FJ, Lardy HA. Carnitine supplementation and depletion: tissue carnosines and enzymes in fatty acid oxidation. J Appl Physiol 1987, 63:315-321.

33. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo L, Creech M, Gross B, Hamsperk K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schubevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2007, 2(10):2366-2382.

doi:10.1186/2043-9113-1-34
Cite this article as: Netzer et al.: Profiling the human response to physical exercise: a computational strategy for the identification and kinetic analysis of metabolic biomarkers. Journal of Clinical Bioinformatics 2011:34.