PeerJ Preprints NOT PEER-REVIEWED

Exploiting transcriptomic data in genome-scale metabolic networks: new insights into obesity

Systems Biology is a holistic approach, based on the integration of multiscale models and different kinds of data, aimed at studying the underlying mechanisms of complex biological systems. A *GEnome-scale metabolic Model* (GEM) is the representation of the metabolic structure of a cell in terms of chemical reactions, involved metabolites, and associated genes. GEMs provide a functional scaffold for constraint-based mathematical methods aimed at simulating and predicting metabolic fluxes in living organisms. The most widely used constraint-based method is the Flux Balance Analysis (FBA), that exploits the *stoichiometric matrix*, a mathematical representation of the relations between substrates and products of all the reactions in the GEM. Recently, the increasing availability of large amounts of high-throughput sequencing data has fostered the research of new approaches in which the structural information described by GEMs is integrated with the knowledge coming from omics data, with the aim to build more accurate descriptions of metabolic states. Here we propose to use a recently published method, in which transcriptomic data are integrated into genome-scale metabolic models through the maximization of the correlation between the steady-state pattern of the predicted fluxes and the corresponding absolute gene expression data generated under the condition of interest. This approach has the interesting property that no cell growth function must be minimized to execute the model. We used this methodology to simulate a novel GEM of the human adipocyte (*iAdipocytes1809*), with the aim of getting new insights into the metabolic mechanisms underlying obesity and its relationships with cancer. Obesity is a complex disorder associated with an increased risk of developing several comorbid chronic diseases, ranging from cardiovascular alterations to diabetes, hypertension and cancer. In particular, weight increase and obesity have been identified as the most important risk and prognostic factors for breast cancer, especially in postmenopausal women. We discuss some

PeerJ Preprints NOT PEER-REVIEWED

preliminary results obtained with this approach, hilighting the importance of data integration, and the need for developing new methods that could help in improving our interpretation of biological phenomena.

PeerJ Preprints NOT PEER-REVIEWED

Exploiting transcriptomic data in genome-scale metabolic networks: new insights into obesity

Ilaria Granata[1*#], Enrico Troiano[1*], Mara Sangiovanni[1], and Mario R. Guarracino[1]

[1]ICAR-CNR, Via Pietro Castellino 111, 80131, Napoli, Italy
[*]Authors contributed equally to this work
[#]Corresponding author: ilaria.granata@icar.cnr.it

Introduction

Obesity is a complex disorder associated with an increased risk of developing several comorbid chronic diseases, ranging from cardiovascular alterations to diabetes, hypertension and cancer [1]. In particular, weight increase and obesity have been identified as the most important risk and prognostic factors for breast cancer, especially in postmenopausal women [2].
Systems Biology is a holistic approach, based on the integration of multiscale models and different kinds of data, aimed at studying the underlying mechanisms of complex biological systems. More specifically, Systems Biology approaches permit to reproduce and describe, through computational and mathematical models, the biochemical transformations occurring into cells and living organisms. A *GEnome-scale metabolic Model* (GEM) is the representation of the metabolic structure of a cell in terms of chemical reactions, involved metabolites, and associated genes and provides a functional scaffold for constraint-based mathematical methods aimed at simulating and predicting metabolic fluxes in living organisms. The most widely used constraint-based method is the Flux Balance Analysis (FBA) [3], that exploits the *stoichiometric matrix*, a mathematical representation of the relations between substrates and products of all the reactions in the GEM. Steady-state solutions are searched while maximizing a given objective function usually corresponding to the biomass production [3].
Recently, the increasing availability of large amounts of high-throughput sequencing data has fostered the research of new approaches in which the structural information described by GEMs is integrated with the knowledge coming from omics data, with the aim to build more accurate descriptions of metabolic states. Omics data could be quantitatively integrated in FBA as constraints on the metabolic fluxes, used to reduce to search space of steady-state solutions.

Here we propose to use a recently published integrative method, proposed by Lee and colleagues [4], to simulate a novel GEM of the human adipocyte, *iAdipocytes1809* reconstructed by Adil Mardinoglu et al. [5], to get new insights into the metabolic mechanisms underlying obesity and its relationships with cancer. Lee's algorithm integrates transcriptomic data into genome-scale metabolic models through the maximization of the correlation between the steady-state pattern of the predicted fluxes and the corresponding absolute gene expression data generated under the condition of interest.
This algorithm, that has been used so far only on smaller models, such as the one of *S. cerevisiae*, has the interesting property that no cell growth function must be minimized to execute the model. Indeed, the definition of a single objective function is a big approximation when dealing with complex organisms where many cellular functions may be performed at the same time. Lee's method allowed us to overcome this limitation and execute the *iAdipocyte1809* model in different experimental conditions, driven only by the knowledge of the expression data. Here we present the approach we used and the results we obtained exploiting a dataset of normal and obese women affected by different subtypes of breast cancer.

Methods

The human adipocyte specific model used in this work is *iAdipocytes1809*, a GEM presented by Adil Mardinoglu et al. in [5]. Correlation between gene expression data and metabolic fluxes was maximised by using the Lee algorithm [4] in the MATLAB environment.
Since constraint-based methods rely on the formulation of Linear Programming (LP) problems, we used Gurobi (version 6) optimized solver (http://www.gurobi.com/) to lower the processing time.
In order to test the prediction power of our approach we used the same dataset as [5] and

compared our results with those obtained by the authors with standard FBA on the formation of lipid droplets (LDs). We set oxygen, glucose and triacylglycerols uptake fluxes, since they are the essential metabolites needed for LDs formation. Both the lower and upper bound of those metabolites were set to the experimental flux values used in [5], to constrain the searched solution space, and LDs production was predicted for several time points (pre- and post-prandial).
Then we used the same approach on a different dataset. Gene expression data of 405 postmenopausal breast cancer samples were downloaded from Gene Expression Omnibus (GEO) portal (accession number: GSE78958). From the whole dataset we selected the samples based on Body Mass Index (BMI) and retained 131 with a BMI<25 (normal weight) and 142 with a BMI>30 (obese). The samples were further grouped based on breast cancer subtype (Basal-like, Her2-enriched, luminal B and luminal A). Mean gene expression values of each group, obtained from the processing of raw microarray data by “GEOquery” and “Affy” R packages, were used as input to the algorithm. Predicted fluxes were obtained by executing the model for each cancer subtype and for both normal and obese conditions, using pre- and post-prandial fluxomic data. The log2FoldChange of the predicted rates in lean and obese samples were calculated for each group and subtype, and reactions showing a log2FC > |1| were selected and grouped by subsystem annotation. Genes associated with these reactions were extracted and their behaviour analyzed through differential expression analysis. The latter was performed with the “Limma” R package.

Results

The rates predicted by our approach were comparable to the ones obtained with the standard FBA approach in [5], for the output rates of LDs and other fluxes (such as Not Esterified Fatty Acids, NEFAs), encouraging us to further use the proposed methodology.
Flux predictions were obtained for all the cancer subtypes of GSE78958, in both the normal and obese subjects.
From the intersection of the different subgroups was possible to extract the dysregulated reactions specific to a group or in common. Among all, it is worth noticing the different rate of lipid transfer reactions associated to the *STARD3* gene (StAR Related Lipid Transfer Domain Containing 3) between lean and obese patients of Luminal A and Her2 subgroups. Vassilev et al. [6] demonstrated that highly StARD3-expressing breast cancer cells and membrane-altering properties of the protein product might enhance the oncogenic signaling and thereby contribute to a bad prognosis.
These preliminary findings, obtained through analyses of metabolic fluxes modulated by gene expression data, highlight the importance of data integration, and the need for developing new methods that could help in improving our interpretation of biological phenomena

References

1. Egger, Garry, and John Dixon. "Beyond obesity and lifestyle: a review of 21st century chronic disease determinants." *BioMed research international* 2014 (2014).
2. Cheraghi, Zahra, et al. "Effect of body mass index on breast cancer during premenopausal and postmenopausal periods: a meta-analysis." *PloS one* 7.12 (2012): e51446.
3. Orth, Jeffrey D., Ines Thiele, and Bernhard Ø. Palsson. "What is flux balance analysis?." *Nature biotechnology* 28.3 (2010): 245-248.
4. Lee, Dave, et al. "Improving metabolic flux predictions using absolute gene expression data." *BMC systems biology* 6.1 (2012): 73.
5. Mardinoglu, Adil, et al. "Integration of clinical data with a genome-scale metabolic model of the human adipocyte." *Molecular systems biology* 9.1 (2013): 649.
6. Vassilev, Boris, et al. "Elevated levels of StAR-related lipid transfer protein 3 alter cholesterol balance and adhesiveness of breast cancer cells: potential mechanisms contributing to progression of HER2-positive breast cancers." *The American journal of pathology* 185.4 (2015): 987-1000.