Integrated computational extraction of cross-cancer poly-omic signatures

Extended Abstract*

Guido Zampieri
Department of Computer Science and Information Systems, Teesside University
Middlesbrough, United Kingdom
g.zampieri@tees.ac.uk

Claudio Angione
Department of Computer Science and Information Systems, Teesside University
Middlesbrough, United Kingdom
c.angione@tees.ac.uk

ABSTRACT

Understanding the interplay between metabolism and genetic regulation is considered key to shed light on the mechanisms underlying cancer onset and progression. In this work, we reconstruct a number of tumor-specific genome-scale metabolic models and inspect estimated flux profiles via statistical analysis, characterizing the detailed metabolic response associated to altered regulation in various tissues. We thus demonstrate that combining complementary computational techniques it is possible to identify poly-omic differences and commonalities across cancer types.

KEYWORDS

Genome-scale modeling, flux balance analysis, statistical data analysis, cancer metabolism.

ACM Reference Format:
Guido Zampieri and Claudio Angione. 2018. Integrated computational extraction of cross-cancer poly-omic signatures: Extended Abstract. In *Proceedings of 10th International Workshop on Bio-Design Automation (10th IWBDA).* ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

Several recent studies have shown how cancer cells present distinct metabolic hallmarks, such as deregulated uptake of glucose and amino acids. Even the gene theory of cancer has been recently object of revision in light of old and new observations [1]. It is therefore clear that alterations on a genomic and a metabolic level do not work in isolation, but rather co-participate in malignant transformation. However, the precise rewiring in the metabolism of transformed cells requires more extensive elucidation. Here, we address this problem through the investigation of the entire metabolic states associated to altered genetic regulation in the NCI60 cancer cell line panel, which covers nine different tissues [2]. By combining genome-scale metabolic models (GSMMs) and statistical analysis we characterize the corresponding cross-cancer poly-omic landscape.

2 METHODS

Experimental data sets here employed are transcriptomic profiles, nutrient uptake rates and proliferation rates for 56 NCI60 cell lines, obtained from previous studies [3, 4]. We used this data to build and evaluate an array of cell line-specific GSMMs, starting from the human cell model Recon 2.2 [5]. In this process, a novel version of METRADE [6] was adopted to (i) transform normalized gene expression profiles by gene set rules (ii) obtain tumor-specific flux bounds taking into account both genetic and metabolic uptake constraints. The estimation of associated flux configurations is conducted by a regularized flux balance analysis (FBA) optimization task, as follows:

$$\begin{aligned} &\max_{\mathbf{v}} \ \mathbf{w}^\top \mathbf{v} - \frac{\sigma}{2}\mathbf{v}^\top \mathbf{v} \\ &\text{subject to } \ \mathbf{S}\,\mathbf{v} = 0\,, \\ &\mathbf{v}_{lb}\,\varphi(\boldsymbol{\Theta}) \le \mathbf{v} \le \mathbf{v}_{ub}\,\varphi(\boldsymbol{\Theta})\,. \end{aligned} \tag{1}$$

Here $\mathbf{w}$ is a real vector expressing the contribution of each reaction to the objective and $\sigma = 10^{-6}$ is a regularization parameter. Vectors $\mathbf{v}_{lb}$ and $\mathbf{v}_{ub}$ represent native flux bounds in Recon, while vector $\varphi(\boldsymbol{\Theta})$ models the reaction-level gene regulation state in any cell line based on the following map:

$$\varphi(\boldsymbol{\Theta}) = \delta\,(1 + \gamma\,|log(\boldsymbol{\Theta})|)^{sgn(\boldsymbol{\Theta}-1)}\,. \tag{2}$$

In this equation, $\boldsymbol{\Theta}$ is obtained from transcript abundances by converting logical gene-protein-reaction rules into max/min operations, as originally implemented in METRADE [6]. Moreover, γ is a parameter representing the magnitude with which gene expression affects reaction rates, while δ is a scaling factor introduced to adjust native flux bounds to experimental uptake rates.

*Oral presentation

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

10th IWBDA, August 2018, Berkeley, California USA

© 2018 Copyright held by the owner/author(s).

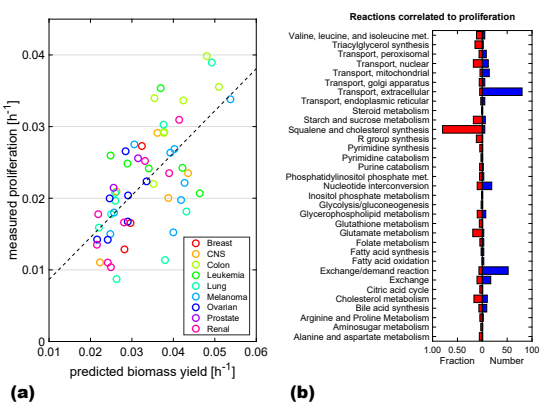

Figure 1: (a) Comparison between biomass yield predicted by each cell line-specific GSMM and the corresponding experimentally measured proliferation rates at the optimal γ and δ values. (b) Overview of metabolic reactions whose predicted fluxes significantly correlate with measured cellular proliferation (1% threshold). For each pathway, number and fraction of significantly correlated reactions are visualized in blue and red, respectively.

We performed a sensitivity analysis on parameters γ and δ in Eq. (2) to evaluate the obtained flux profiles in terms of the Pearson correlation coefficient (PCC) r between predicted cellular growth and experimentally measured proliferation rate. The predicted growth was computed through Eq. (1) assuming biomass accumulation as a proxy for cell proliferation and thus as a meaningful FBA objective to model cancerous metabolism. Repeated PCC estimation allowed identifying optimal γ and δ values across several orders of magnitude. We carried out regularized FBA using the COBRA toolbox in Matlab and the quadratic solver Gurobi [7]. Finally, using the FactoMineR package in R [8] we performed principal component analysis (PCA) to characterize the cross-tumor variation at a genome-scale metabolic flux level.

3 RESULTS

As a result of the sensitivity analysis on parameters γ and δ in Eq. (2), we obtained a PCC peak where $r \simeq 0.66$, $p\text{-}value \simeq 1.5 \cdot 10^{-8}$ (Fig. 1a). We thus inspected the whole flux profiles of tumor cells by studying their PCC with respect to cellular proliferation rates. We observed a significant PCC (threshold 1%) for reactions in a number of cancer-associated pathways, supporting the reliability of our GSMMs, as well as in less obvious pathways (Fig. 1b). These may suggest or corroborate unknown mechanisms for tumor development. In particular, the majority of cholesterol synthesis pathway emerges as correlated to proliferation, supporting its debated involvement in cancer. As another example, the exchange

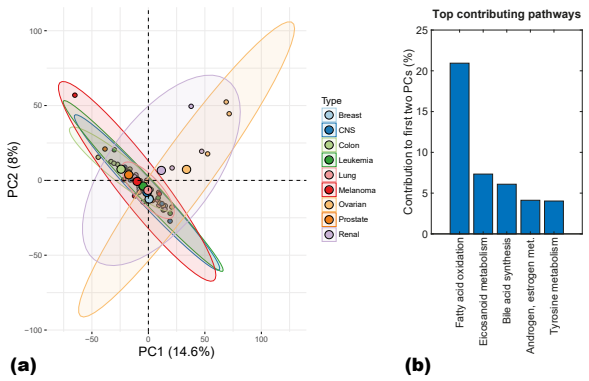

Figure 2: (a) Variability across flux profiles relative to different tumor types in the space of the first two principal components. (b) Contribution to the first two principal components of the most highly contributing pathways, obtained by summing the contributions of their associated reactions.

of dietary compounds such as maltodextrins also results associated to proliferation.

Next, PCA of the flux profiles allowed detecting poly-omic heterogeneities across the cell lines. As Fig. 2a shows, the ovarian and renal cell tumors present a markedly distinct metabolic behavior, almost orthogonal to all other tissues. A closer look at the composition of first principal components allowed identifying key pathways underlying such variation, like fatty acid oxidation or eicosanoid metabolism (Fig. 2b). This analysis thus highlights potential links in the metabolic reprogramming of the two cancer types, suggesting also precise reactions to focus experimental verification on.

4 CONCLUSIONS

In this work, we analyzed the poly-omic configurations of multiple cancer types through an integrated computational pipeline and within a comprehensive cross-tumor framework. Our analysis led to the identification of both variation and common patterns across the tumors, providing novel insights in the general cancer molecular landscape. We thus showed that the joint application of GSMMs and statistical analysis techniques can help elucidate the mechanisms underlying cancer development and progression.

REFERENCES

[1] Thomas N. Seyfried et al. 2014. *Carcinogenesis* 35, 3 (2014), 515–527.
[2] Uwe Scherf et al. 2000. *Nat Genet* 24, 3 (2000), 236–244.
[3] Christian Diener and Osbaldo Resendis-Antonio. 2016. *Front Physiol* 7 (2016), 644.
[4] Daniel C. Zielinski et al. 2017. *Sci Rep* 7 (2017), 41241.
[5] Neil Swainston et al. 2016. *Metabolomics* 12, 7 (2016), 109.
[6] Claudio Angione and Pietro Lió. 2015. *Sci Rep* 5 (2015), 15147.
[7] Jan Schellenberger et al. 2011. *Nat Protoc* 6 (2011), 1290.
[8] Sébastien Lê et al. 2008. *J Stat Softw* 25, 1 (2008), 1–18.