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. Moreover, vectors $\mathbf{v}_{lb}$ and $\mathbf{v}_{ub}$ represent native flux bounds in Recon. These are altered by a factor determined by the effective gene expression $\boldsymbol{\Theta}$ for each reaction, according to the following map:

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

In this equation γ is a parameter representing the magnitude with which expression affects reaction rates, while δ is a scaling factor adjusting native flux bounds relatively to uptake data. The vector $\boldsymbol{\Theta}$ is computed starting from gene expression values and transforming logical gene-protein-reaction rules into max/min operations [6].

*Both poster and oral presentation

10th IWBDA, August 2018, Berkeley, California USA

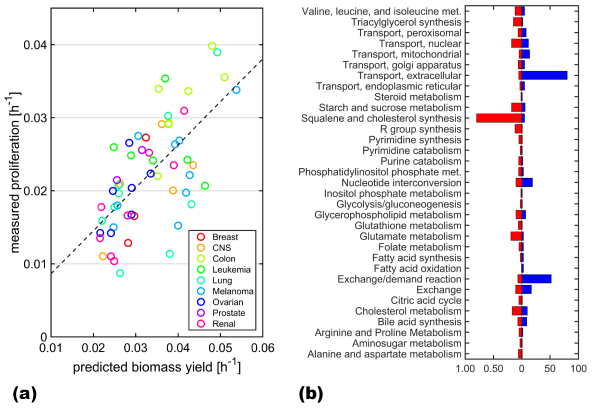

Figure 1: (a) Comparison between biomass yield predicted by each cell line-specific GSMM and the corresponding experimentally measured proliferation rates at the optimal γ and δ values. (b) Overview of metabolic reactions whose predicted fluxes significantly correlate with measured cellular proliferation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

We performed a sensitivity analysis on parameters γ and δ in Eq. (2) to evaluate the obtained flux profiles. The Pearson correlation coefficient (PCC) r between predicted cellular growth and experimentally measured proliferation rate was computed through Eq. (1) assuming biomass accumulation as objective. Repeated PCC estimation allowed identifying optimal γ and δ values across several orders of magnitude. We carried out FBA using the COBRA toolbox in Matlab [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 correlation (threshold 1%) in a number of cancer-associated pathways, supporting the reliability of our GSMMs and suggesting other potential mechanisms of tumor growth (Fig. 1b). In particular, the majority of the cholesterol synthesis pathway emerges as correlated to proliferation, supporting previous data that suggest a dependence of patient survival on alterations in cholesterol homeostasis.

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

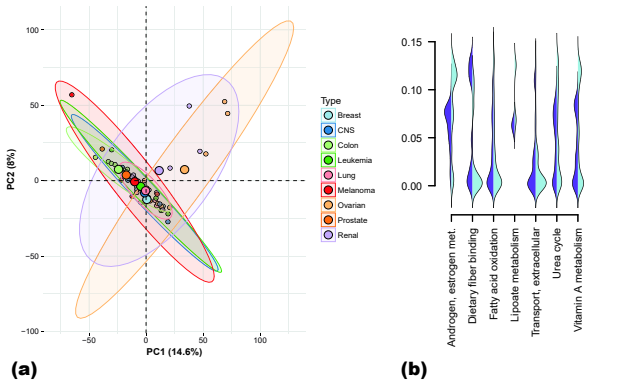

Figure 2: (a) Variability across flux profiles relative to different tumor types in the first two principal components space. (b) Distribution of single reaction contributions to first (purple) and second (light blue) principal components for the most highly contributing pathways.

closer look at the composition of first principal components allowed identifying key associated pathways and metabolic reactions, such as androgen and estrogen metabolism or dietary fiber binding (Fig. 2b).

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 lea to the identification of both variation and common patterns across the tumors, providing novel insights in the general cancer molecular landscape. We thus also 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. Cancer as a metabolic disease: implications for novel therapeutics. *Carcinogenesis* 35, 3 (2014), 515–527.
[2] Uwe Scherf et al. 2000. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 24, 3 (2000), 236–244.
[3] Christian Diener and Osbaldo Resendis-Antonio. 2016. Personalized Prediction of Proliferation Rates and Metabolic Liabilities in Cancer Biopsies. *Front Physiol* 7 (2016), 644.
[4] Daniel C. Zielinski et al. 2017. Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism. *Sci Rep* 7 (2017), 41241.
[5] Neil Swainston et al. 2016. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* 12, 7 (2016), 109.
[6] Claudio Angione and Pietro Lió. 2015. Predictive analytics of environmental adaptability in multi-omic network models. *Sci Rep* 5 (2015), 15147.
[7] Jan Schellenberger et al. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 6 (2011), 1290.
[8] Sébastien Lê et al. 2008. FactoMineR: An R Package for Multivariate Analysis. *J Stat Softw* 25, 1 (2008), 1–18.