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:

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:

\[
\begin{align*}
\max \mathbf{w}^T \mathbf{v} - \frac{\sigma}{2} \mathbf{v}^T \mathbf{v} \\
\text{subject to } \mathbf{S} \mathbf{v} = \mathbf{0}, \quad (1) \\
\mathbf{v}_{lb} \phi(\Theta) \leq \mathbf{v} \leq \mathbf{v}_{ub} \phi(\Theta).
\end{align*}
\]

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 \( \phi(\Theta) \) models the reaction-level gene regulation state in any cell line based on the following map:

\[
\phi(\Theta) = \delta (1 + \gamma |\log(\Theta)|)^{sgn(\Theta-1)}.
\]

In this equation, \( \Theta \) is obtained from transcript abundances by converting logical gene-protein-reaction rules into max/min operations, as originally implemented in METRADE [6]. Moreover, \( \gamma \) is a parameter representing the magnitude with which gene expression affects reaction rates, while \( \delta \) is a scaling factor introduced to adjust native flux bounds to experimental uptake rates.
We carried out regularized FBA using the COBRA toolbox in Eq. (2), where we obtained a PCC peak where $r = 0.66, p-value \approx 1.5 \cdot 10^{-6}$ (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 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