Bioinformatics, YYYY, 0–0
doi: 10.1093/bioinformatics/xxxxx
Advance Access Publication Date: DD Month YYYY
Original Paper

Systems Biology

rMTA: Robust Metabolic Transformation Analysis

Luis V. Valcárcel[1,2,4], Verónica Torrano[3,4], Luis Tobalina[5], Arkaitz Carracedo[3,4,6,7] and Francisco J. Planes[1,*]

[1]Tecnun, University of Navarra, Manuel de Lardizábal 15, 20018 San Sebastián, Spain;

[2]Area de Hemato-Oncología, IDISNA, Centro de Investigación Médica Aplicada (CIMA), University of Navarra, Pío XII 55, 31008 Pamplona, Spain;

[3]CIC bioGUNE, Bizkaia Technology Park, 801 Building, 48160 Derio, Spain;

[4]CIBERONC;

[5]RWTH Aachen University, Faculty of Medicine, Joint Research Centre for Computational Biomedicine, MTZ Pauwelsstrasse 19 D-52074 Aachen;

[6]Ikerbasque, Basque foundation for science, 48011 Bilbao, Spain;

[7]Biochemistry and Molecular Biology Department, University of the Basque Country (UPV/EHU), P. O. Box 644, E-48080 Bilbao, Spain.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: The development of computational tools exploiting -omics data and high-quality genome-scale metabolic networks for the identification of novel drug targets is a relevant topic in Systems Medicine. Metabolic Transformation Algorithm (MTA) is one of these tools, which aims to identify targets that transform a disease metabolic state back into a healthy state, with potential application in any disease where a clear metabolic alteration is observed.

Results: Here, we present a robust extension to MTA (rMTA), which additionally incorporates a worst-case scenario analysis and minimization of metabolic adjustment (MOMA) to evaluate the beneficial effect of gene knockouts. We show that rMTA complements MTA in the different datasets analyzed (gene knockout perturbations in different organisms, Alzheimer's disease and prostate cancer), bringing a more accurate tool for predicting therapeutic targets.

Availability: rMTA is freely available on The Cobra Toolbox: https://opencobra.github.io/cobra-toolbox/latest/.

Contact: fplanes@tecnun.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

With the release of high-quality human genome-scale metabolic networks, together with the increasing –omics data availability, the use of constraint-based models in the drug discovery process is expanding, particularly for the elucidation of novel and better targets (Oberhardt et al., 2013). Early approaches have focused on cancer and bacterial infections, namely by directly targeting the growth of target cells (Apaolaza et al., 2017; Folger et al., 2011; Plata et al., 2010). A more general approach, termed Metabolic Transformation Algorithm (MTA), was presented in Yizhak et al. (2013), aiming to identify targets that transform a disease metabolic state into a healthy one, with potential application in any disease where a clear metabolic alteration is observed. In that work, Ruppin and co-workers applied MTA to ageing, and was later used in Alzheimer's disease (Stempler et al., 2014). Importantly, this work opened new avenues to systematically target cellular metabolism in human disease.

In the present study, we show that the MTA scoring procedure constitutes a best-case scenario, which may over-estimate the beneficial effect of a particular perturbation, typically gene knockouts or drugs, to transform a disease phenotype into a healthy one. We propose a more robust approach, which takes into account a worst-case scenario and results from the minimization of metabolic adjustment (MOMA) (Segre et al., 2002). We term our approach Robust Metabolic Transformation Algorithm (rMTA).

To illustrate our approach, we first applied rMTA to gene knockout experiments in different organisms and Alzheimer's disease (AD) data, previously analyzed with MTA (Stempler et al., 2014; Yizhak et al., 2013), showing that rMTA effectively enhances and improves MTA. In addition, we applied rMTA to a recently published gene expression dataset in prostate cancer, where the role of the metabolic co-regulator PGC1α in tumor progression and metastasis was analyzed (Torrano et al., 2016)**.**

2 Methods

2.1 Metabolic Transformation Algorithm (MTA)

MTA only requires two inputs: gene expression levels of the target (healthy) and source (disease) states, as well as a genome-scale metabolic reconstruction. Full details of the algorithm can be found in Yizhak et al. (2013). Here, we summarize the most relevant steps of MTA.

1. Using a three-level classification of reactions based on absolute expression data, network contextualization of the source state is carried out using iMAT (Shlomi *et al.*, 2008). Flux sampling techniques are then applied to generate a mean flux distribution at the source state, denoted as v^{ref}.

2. Determination of changed and unchanged reactions based on v^{ref}, differential gene expression analysis and gene-to-reaction mappings. The aim here is to define the subset of reactions whose flux should change in the forward (R_F) and backward (R_B) direction to transform the source state into the target state. To that end, we first categorize reactions as elevated or reduced based on differential gene expression data and gene-to-reactions mappings. Then, reaction i is in R_F if $v_i^{ref} > 0$ and it is elevated or if $v_i^{ref} < 0$ and it is reduced. Similarly, reaction i is in R_B if $v_i^{ref} < 0$ and it is elevated or if $v_i^{ref} > 0$ and it is reduced. The rest of reactions (R_S) are unchanged and their flux should be similar in both states.

3. Selection of a threshold for each reaction, ε_i, from which a significant flux deviation from v^{ref} in the direction of the target (healthy) state is achieved.

4. Upon gene knockout, determination of the maximum number of significant flux alterations in the desired direction while respecting the flux of unchanged reactions. This is done via mixed-integer quadratic programming (MIQP) using the following objective function and mathematical constraints:

$$\min\ (1-\alpha)\sum_{i\in R_S}(v_i^{ref}-v_i)^2 + \frac{\alpha}{2}\sum_{i\in R_F} y_i + \frac{\alpha}{2}\sum_{i\in R_B} y_i \qquad (1)$$

s.t.

$$v_i^{min} \le v_i \le v_i^{max}, \qquad i \in R \qquad (2)$$

$$S \cdot v = 0 \qquad (3)$$

$$v_i - y_i^F \cdot \left(v_i^{ref} + \varepsilon_i\right) - y_i \cdot v_i^{min} \ge 0, \qquad i \in R_F \qquad (4)$$

$$y_i^F + y_i = 1, \qquad i \in R_F \qquad (5)$$

$$v_i - y_i^B \cdot \left(v_i^{ref} - \varepsilon_i\right) - y_i \cdot v_i^{max} \le 0, \qquad i \in R_B \qquad (6)$$

$$y_i^B + y_i = 1, \qquad i \in R_B \qquad (7)$$

Eq. (2) integrates thermodynamic, gene knockout and growth medium constraints, where v are the reaction fluxes, while v^{min} and v^{max} their associated lower and upper bounds, respectively. The mass balance constraint is enforced in Eq. (3), where S represents the stoichiometric matrix. In Eqs. (4)-(5), y and y^F are binary variables associated with changed reactions in the forward direction R_F. With these constraints, we guarantee for a particular reaction i in R_F that, if y_i=0 (or y_i^F=1), its associated flux is significantly altered in the desired forward direction, *i.e.* $v_i \ge v_i^{ref} + \varepsilon_i$. Similarly, Eqs. (6)-(7) guarantee for reaction i in R_B that, if y_i=0 (or y_i^B=1), its associated flux is significantly altered in the desired backward direction, *i.e.* $v_i \le v_i^{ref} - \varepsilon_i$. α in Eq. (1) is a trade-off parameter. The resulting flux distribution is termed v^{res}.

5. Calculation of the Transformation Score (TS), as in Eq. (8), which captures the transformation upon gene knockout from the source to the target state:

$$\frac{\sum_{i\in R_{success}} abs\left[\left(v_i^{ref}-v_i^{res}\right)\right] - \sum_{i\in R_{unsuccess}} abs\left[\left(v_i^{ref}-v_i^{res}\right)\right]}{\sum_{i\in R_S} abs\left[\left(v_i^{ref}-v_i^{res}\right)\right]} \qquad (8)$$

where $R_{success}$ and $R_{unsuccess}$ denote fluxes in v^{res} involving an alteration in the desired and undesired direction, respectively. MTA algorithm ranks gene knockouts according to their TS (Eq. (8)) by using the transformed flux distribution (v^{res}) obtained after solving Eqs. (1)-(7). The more positive the TS is, the greater the capacity of the related intervention to move in the target (healthy) direction.

2.2 Robust Metabolic Transformation Algorithm (rMTA)

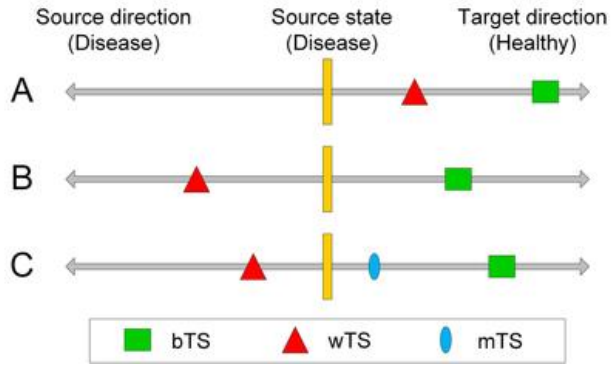
As explained above, the MTA algorithm ranks gene knockouts according to their TS (Eq. (8)) by using the transformed flux distribution (v^{res}) obtained after solving Eqs. (1)-(7). In our view, this scoring procedure constitutes a **best-case scenario**, since it minimizes the number of unsuccessful changes, providing knockouts with maximum capacity to move in the target direction. **Best-case TS (bTS)** can be either positive or negative, namely the more positive, the more capacity to move in the desired (target) direction.

However, one may question whether a particular gene knockout could also reinforce the source state, particularly by evaluating the maximum capacity to further move in the source direction. To that end, for each gene knockout, we propose to swap R_F and R_B subsets and recalculate Eqs. (1)-(7) and Eq. (8), resulting in what we term **worst-case TS (wTS)**. For a given gene knockout, similar to the best-case TS, the more positive the wTS is, the more capacity to further move the flux distribution in the source (disease) direction. In terms of drug targets, we are certainly interested in those knockouts having a positive bTS and a negative wTS (see Figure 1A). In cases where we have both a positive bTS and wTS, the interpretation is unclear and the algorithm cannot clearly distinguish the beneficial effect of the knockout. An illustrative example of an underdetermined scenario can be found in Figure 1B, where bTS=wTS.

As shown in Figure 1C, where bTS>wTS (both being positive), we can have less extreme situations, with values of bTS and wTS differing significantly. In these cases, the use of minimization of metabolic adjustment (MOMA) can be informative and constitutes the most realistic approach to decide whether a particular gene knockout pushes metabolism into the healthy or disease direction. This is equivalent to fixing α=0, including the whole set of reactions (not only R_S) in the first term of Eq. (1) and resolving Eqs. (1)-(3) and recalculating (8), resulting in what we term **MOMA TS (mTS).** Note here that we also evaluated ROOM (regulatory on/off

minimization of metabolic flux changes) (Shlomi *et al.*, 2005). However, we found ROOM significantly more computationally demanding than MOMA with no increase in accuracy (see Supplementary Table 1 for details).

Figure 1: Illustration of the problem addressed by rMTA. A) Transformation Scores (TS) after gene knockout are skewed to the healthy direction in both the best-case (bTS) and worst-case scenario (wTS); B) TSs are similar in value and skewed to the opposite direction in the best-case and worst-case scenario and, therefore, under-determination arises; C) The same as B, but TS is higher in the best-case scenario and skewed to the healthy direction when MOMA is applied (mTS>0). Under-determination is resolved here using mTS.



Gene expression datasets and analysis As shown in Figure 1C, where bTS>wTS (both being positive), we can have less extreme situations, with values of bTS and wTS differing significantly. In these cases, the use of minimization of metabolic adjustment (MOMA) constitutes the most realistic approach to decide whether a particular gene knockout pushes metabolism into the healthy or disease direction. This is equivalent to fixing α=0, including the whole set of reactions (not only R_S) in the first term of Eq. (1) and recalculating Eq. (1) and (2), resulting in what we term **MOMA TS (mTS).** Note here that we also evaluated ROOM (regulatory on/off minimization of metabolic flux changes) (Shlomi *et al.*, 2005). However, we found ROOM significantly more computationally demanding than MOMA with no increase in accuracy (see Supplementary Table 1 for details).

In order to account for these different factors, we propose in Eq. (9) a unified score (rTS), which integrates bTS, wTS and mTS, providing a more robust approach:

$$rTS = mTS \cdot (k \cdot (bTS - wTS))^{a} \qquad (9)$$

where a=1 if bTS>0, mTS>0 and wTS<0, a=0 otherwise; k is a big positive number, here k=100. As reflected in Figure 1, our aim is to rank higher perturbations skewed to the target direction, *i.e.* bTS>0, mTS>0 and wTS<0, and, in these situations, $rTS=mTS*k*(bTS-wTS)$. We included k in Eq. (9) to guarantee that $rTS>mTS$, *i.e.* $k*(bTS-wTS)>1$, when a=1. In the rest of the cases (a=0), we rely on MOMA and fixed $rTS=mTS$ (see Supplementary Figure 1). We call this extension **robust Metabolic Transformation Analysis** (rMTA).

The value of our rMTA approach is analyzed in three different cases: gene knockout predictions in *E. coli*, mouse and human samples, Alzheimer's disease and prostate cancer (see Results section). Below, we provide the technical details of our implementation of MTA and rMTA.

2.3 Implementation of MTA and rMTA

For studies using human cells, we used the genome-scale metabolic network reconstruction Recon1 (Duarte *et al.*, 2007), while for those using mouse models, we used the available reconstruction that is based on Recon 1 (Sigurdsson *et al.*, 2010). This was done to compare our approach with the results presented in Stempler et al. (2014) and Yizhak et al. (2013).

n addition, for each of the dataset analyzed here, we simulated the same growth medium that was used in the experiment. In human brain data from AD patients, in which the growth medium is unknown, we assumed the most general unconstrained case where all nutrients are available. In the prostate cancer study, based on human cell lines, DMEM medium was simulated. In the case of gene knockout experiments with mouse and human cells, RPMI, DMEM or general unconstrained media were simulated depending on the case considered (Supplementary Table 2). For flux sampling, we used *optGpSampler* (Megchelenbrink *et al.*, 2014), an extension of *GpSampler* (Schellenberger and Palsson, 2009) that converges up to 500 faster in large networks. 2000 different flux distributions were determined and averaged to obtain v^{ref}.

As noted above, a flux is significantly altered in the desired direction ($y_i=0$) if $v_i^{res} \geq v_i^{ref} + \varepsilon_i$ if $i \in R_F$, or if $v_i^{res} \leq v_i^{ref} - \varepsilon_i$ if $i \in R_B$. Following the methodology described in Yizhak et al. (2013), for studies using mouse and human cells, we fixed a different parameter ε_i for each reaction. However, we set a minimum ε_i of 0.001 to avoid numerical issues. In addition, we set the trade-off parameter α in Eq. 1 to 0.66, as done in Yizhak et al. (2013). We also conducted a sensitivity analysis on the α and ε parameters, finding that the main results shown in the Results section are stable for different values (Supplementary Figure S2-S7).

2.4 Gene expression datasets and analysis

Alzheimer's Disease study. Illumina microarrays data was obtained from the Gene Expression Omnibus (GEO) database (Barrett *et al.*, 2012), particularly GSE15222, which involves data from 363 cortical samples of Alzheimer's and control patients' post-mortem brains (Webster *et al.*, 2009). Data was processed with the R statistical framework. Data was log2 transformed and quantile normalized using the *lumi* package (Du *et al.*, 2008). Summarized gene-level intensities were obtained using the *limma* package (Ritchie *et al.*, 2015).

For the absolute classification of genes into highly, moderately and lowly expressed required for iMAT analysis, for each different AD sample, we selected as highly expressed those having a larger expression value than the mean expression + δ SD and as lowly expressed those having a lower expression than the mean expression - δ SD. We fixed δ=0.3, as done in Stempler et al. (2014). The rest of the genes were considered as moderately expressed**.** A consensus gene classification in AD was then determined, particularly genes were classified as highly/lowly expressed if this outcome was obtained in at least β% of AD samples, with β=100. We used gene-protein-reaction rules available in Recon1 to convert highly, moderately and lowly expressed genes into high, medium and low activity reactions, as required for the iMAT analysis. We conducted a sensitivity analysis on δ and β, finding that v^{ref} was not significantly altered (see Supplementary Figure S8).

In order to determine differentially expressed genes between control patients and AD, as required for the calculation of TS scores, we again used the *limma* package. We selected those genes having a probability of being differentially expressed greater than 0.95 (B statistic greater than 4.25) and being up-regulated or down-regulated in healthy subjects vs. AD patients by at least a fold change of ±20%. Genes not satisfying these restrictions were classified as unchanged. Up-regulated, unchanged and down-regulated genes were converted into elevated, unchanged and reduced activity reactions using gene-protein-reaction rules. Note that the threshold for differential expression defined above was selected to obtain 100-200

changed reactions, as suggested in Yizhak et al. (2013) in order to guarantee a tractable running time. As noted above, in conjunction with v^{ref}, this information was used to determine R_F and R_B.

Prostate cancer study. Illumina microarray data was obtained from Torrano et al. (2016), which includes 3 samples of wild-type prostate cancer cell line PC3, where tumor suppressor PGC1α is repressed (source state, metastatic phenotype), and 3 samples of engineered PGC1α-expressing PC3 (target state, non-metastatic phenotype). This data is also accessible in GEO under reference number GSE75193.

For absolute gene expression classification, we used the same procedure as in the AD study. However, given the limited number of samples, we fixed a stricter threshold to obtain highly and lowly expressed genes in the source state, namely δ=0.7 (see Supplementary Figure S9). For the differential expression analysis between the source and target states, we again used the *limma* package and selected those genes with an adjusted (FDR) p-value < 0.05, which again resulted in a number around 100-200 of altered reactions.

Gene knockout experiments. In the case of human cells, we considered the same 2 experiments reported in (Yizhak *et al.*, 2013), namely the knockout of SDHB in the hepatocellular carcinoma cell line Hep3B (Cervera *et al.*, 2008) and the mutation of SDHA in acute quadriplegic myopathy patients (Bakay *et al.*, 2006). Gene expression data for these two cases is accessible in GEO under reference numbers GSE10289 and GSE3307. In addition, we added 2 more recently published cases that knocked out RRM1 and RRM2 in the Multiple Myeloma cell line H929 (GEO accession number GSE93425) (Sagawa *et al.*, 2017). With respect to mouse cells, we only considered 1 out of 2 experiments reported in Yizhak et al, 2013, (GKD knockout, GEO accession number GSE12748) (MacLennan *et al.*, 2009). Microarray data of the other experiment (FH1 knockout) was not available. However, we used a very similar experiment with the same gene knockout (GEO accession number GSE10989).

All these experiments have been carried out using Affymetrix microarrays. Data was processed with RMA, available in *aroma-affymetrix* (Bengtsson *et al.*, 2008). Log2 normalized signals were summarized to the gene-level. For the absolute gene expression classification, we used the same procedure as in the AD study, with β=100%, but with δ=0.7-1, depending on the distribution of intensities of metabolic genes and the number of samples for each scenario.

For the differential expression analysis, we used the *limma* package. As noted above, we varied the filter of p-value, FDR and B-statistic in order to obtain around 100-200 altered reactions, as suggested in Yizhak et al. (2013).

3 Results

3.1 Comparison of rMTA and MTA in different organisms

To examine the performance of rMTA, we first applied it to predict the effects of gene knockouts in *E. Coli*, as previously done in Yizhak et al. (2013). We used gene expression data measured before and after a specific metabolic gene knockout, which allows us to evaluate whether rMTA is able to rank high the gene knockout causing the metabolic transformation from the wild-type into the knockout state. We fixed the same parameters used in Yizhak et al. (2013) (α=2/3, ε_i=0.01) and applied rMTA in the same conditions (v^{ref}, R_F, R_B, R_S) (data provided by the authors). Our approach was compared with the results obtained with MTA. Results are shown in Table 1 and Supplementary Data 1.

Table 1. Comparison of rMTA and MTA in E. coli

Experiment number	Gene name	Sign of:			Gene Ranking	
		bTS	mTS	wTS	rMTA	MTA
1	pgi	+	+	-	4	4
2	pgi	+	+	-	8	8
3	ppc	-	+	+	13	50
4	ppc	+	+	+	14	10
5	tpiA	+	+	-	4	3
6	tpiA	+	+	-	9	10

bTS represents a best-case transformation score (TS); mTS is the TS score obtained using MOMA; wTS represents the worst-case TS; rMTA is Robust MTA and ranks gene knockout according to robust TS. The column entitled 'MTA' includes the results reported in Yizhak et al. (2013). Experiments 1 and 2 represent two different gene expression studies that knocked out pgi. The same for ppc in experiments 3 and 4 and tpiA in experiments 5 and 6.

We can observe in Table 1 that the ranking of the gene knockout causing the perturbation is similar in both rMTA and MTA: it is better ranked in rMTA in experiments 3 and 6, while the opposite happens in experiments 4 and 5; in the other 2 cases we have the same position with both approaches. In 4 out of 6 cases (experiments 1, 2, 5 and 6), the true gene knockout does have a positive bTS and mTS, with a negative wTS. As shown in Figure 1, gene knockouts with this pattern constitute the most natural candidates to revert to the original scenario. However, in these 4 cases, rMTA could not improve the performance of MTA and both provided very good results. In contrast, this pattern of signs is not followed in experiments 3 and 4: wTS>0 in both cases and bTS<0 in experiment 3. In these 2 cases, rMTA (now based on MOMA, see Eq. (9)) provided very good results, particularly in the case of experiment 3, where *ppc* is better ranked in rMTA than in MTA (see Supplementary Data 1 for full details).

The effect of rMTA was more clearly observed when we replicated the study of Alzheimer's Disease reported in Stempler et al. (2014), where MTA was used to predict drug targets (see Methods section for details). Figure 2A displays bTS and wTS for the different gene knockouts simulated in the AD study. The gene knockout ranked first in rMTA, SLC5A8 (Solute Carrier Family 5, Member 8), is far from the best-ranked positions in MTA (position 313). However, it is the gene knockout with the most negative wTS, which implies that this gene knockout, even in the worst-case scenario, leads metabolic fluxes into the healthy direction. In addition, mTS is also the highest for this gene knockout (Figure 2B). This result shows that rMTA may significantly alter the ranking obtained with MTA. Of course, we found cases where rMTA and MTA obtained similar results, for example SLC5A3 and SLC5A11, which were ranked first and second in MTA and in the top 20 in rMTA. However, a remarkable number of gene knockouts were ranked at very distant positions in rMTA and MTA, *e.g.* SLC6A5.

In summary, Figure 2 shows that some genes ranked high in rMTA are not found in the top positions of the MTA ranking, showing that integrating a worst-case scenario and MOMA with MTA (as done in rMTA) do have an impact in the ranking of the most relevant gene knockouts to transform a metabolic state into a healthy state. Full details of our AD study can be found in Supplementary Data 2.

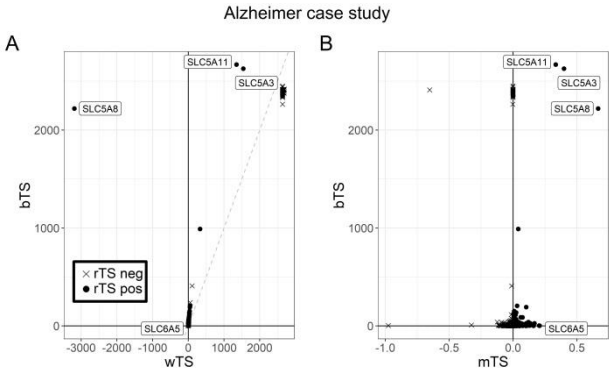


Figure 2: Comparison of rMTA and MTA in the Alzheimer's Disease study. A) Best-case Transformation Score (bTS) vs worst-case transformation score (wTS) for different gene knockouts analyzed. Source and target states are AD patients and healthy controls, respectively; B) bTS vs MOMA TS (mTS) for different gene knockout analyzed.

Given the results in AD, we conducted a similar analysis to the one in Table 1 (known gene knockout perturbations) with human and mouse cells (see Methods section for details), as done in Yizhak et al. (2013). The benefit of using rMTA can now be clearly observed, as we always rank the gene knockouts that cause the perturbation significantly higher (one-sided Wilcoxon signed-rank test, p-value=0.016). The pattern of signs in experiments 1 and 2 follows the one discussed in *E. coli*. The rest of cases presents under-determination, but it is importantly smoothed by rMTA. Full details can be found in Supplementary Data 3.

Table 2. Comparison of rMTA and MTA in human and mouse cells

Exp. number	Gene name	Organism	Sign of:			Ranking	
			bTS	mTS	wTS	rMTA	MTA
1	RRM1	Homo Sapiens	+	+	-	28	353
2	RRM2	Homo Sapiens	+	+	-	36	372
3	SDHA	Homo Sapiens	+	+	+	119	546
4	SDHB	Homo Sapiens	+	+	+	88	402
5	FH1	Mus Musculus	+	+	+	11	575
6	GKD	Mus Musculus	+	+	+	92	453

bTS represents a best-case transformation score (TS); mTS is the TS score obtained using MOMA; wTS represents the worst-case TS; rMTA is Robust MTA and ranks gene knockout according to robust TS. The column entitled 'MTA' includes the ranking provided when bTS was only considered. The metabolic model of Homo Sapiens (Recon1) and Mus Musculus (iMM1415) contain 1496 and 1375 genes, respectively.

3.2 rMTA in prostate cancer

In a recent work (Torrano *et al.*, 2016), it was shown that the master metabolic co-regulator PGC1α suppresses prostate cancer progression and metastasis. They proved that reconstituting the expression of PGC1α in different prostate cancer cell lines leads to a decrease in growth, proliferation and metastasis. In order to evaluate the transcriptomic changes mediated by the induction of PGC1α, they compared the gene expression data between wild-type and engineered PGC1α-expressing PC3 prostate cancer cell line (see Methods section).

Here, we apply rMTA to identify gene knockouts that transform PC3 cells (source state) into a less aggressive state, which is defined by those expressing PGC1α (target state). Results are shown in Figure 3A. It can be observed that the best-ranked gene knockouts in MTA show a high level of under-determination, since they have an even larger value for wTS (*e.g.* SLC12A3), which once again emphasizes the limitations of MTA. Instead, the best-ranked genes in rMTA have a reduced transformation score (Figure 3B), but they are biased towards the healthy direction, *e.g.* MTHFD1L. Full details can be found in Supplementary Data 4.

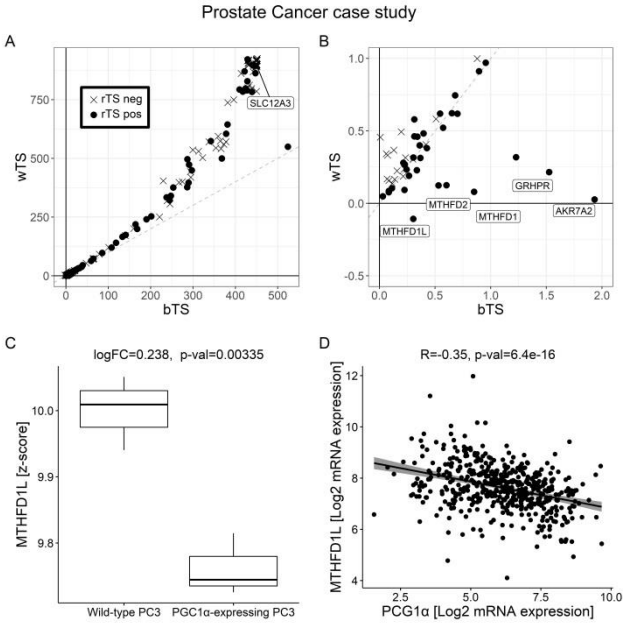


Figure 3: rMTA application to prostate cancer. A) Best-case TS (bTS) vs worst-case TS (wTS) to transform wild-type into engineered PGC1α-expressing PC3 cells for different gene knockouts analyzed; B) Zoomed-in Figure3a to highlight top-ranked genes in rMTA; C) Expression of MTHFD1L in wild-type into engineered PGC1α-expressing PC3 cells; D) Expression plot of PGC1α (horizontal axis) and MTHFD1L (vertical axis) in prostate cancer samples from TCGA.

Whether top-ranked gene knockouts in rMTA are sufficient to recapitulate the effect of PGC1α is debatable and requires further experimental validation. Among them, we have several genes involved in folate metabolism: MTHFD2, MTHFD1 and MTHFD1L. This is particularly interesting, since a recent work provided evidence that the PGC1α/ERRα axis regu-

lates folate metabolism in breast cancer (Audet-Walsh *et al.*, 2016). Particularly, they showed that PGC1α and MTHFD1L have an inverse and significant correlation at the expression level, which is also observed in our prostate cancer study (Figure 3C) and in primary prostate cancer samples from CANCERTOOL (Cortazar *et al.*, 2018). Figure 3D shows the expression of PGC1a and MTHFD1L in primary prostate cancer samples from TCGA, but the same result is found in three additional prostate cancer datasets available in CANCERTOOL (Supplementary Figure S10). On the other hand, the knockdown of MTHFD1L in hepatocellular carcinoma significantly decreased proliferation and growth (Lee *et al.*, 2017). Based on our data, we propose the inhibition of MTHFD1L as a promising target to recapitulate (at least part of) the effect of PGC1α as a tumor suppressor in prostate cancer.

4 Discussion

A key challenge in Systems Medicine is the development of computational methods, driven by –omics data, capable of predicting novel and more effective therapeutic strategies in different diseases. Existing in-silico approaches to target cellular metabolism, based on the constrained-based modeling paradigm, have mainly concentrated on cancer and parasitology (Apaolaza et al., 2017; Plata et al., 2010). Certainly, cellular growth and proliferation are clear phenotypes relevant in these diseases, which have been typically analyzed by constraint-based models (through the biomass reaction). These methods could hardly be applied, if at all, for diseases where a clear targetable metabolic phenotype is unknown.

Ruppin and co-workers overcame this issue with the release of MTA. The logic behind MTA is the construction of data-driven metabolic objectives that can be systematically analyzed through constraint-based models. This was done based on differential expression analysis, namely by defining a list of reactions that need to be up/down regulated in order to transform a disease phenotype into a healthy one. The first applications of MTA to different diseases, reported in Yizhak et al. (2013) and Stempler et al. (2014), were encouraging and promising for the field of Systems Medicine.

In order to complement the path opened by MTA, we here present a more robust formulation, termed rMTA, aimed to more accurately quantify the capacity of selected perturbations to transform a disease phenotype into a healthy one. To that end, we pondered three different scores: best-case TS (bTS), the one given by MTA; worst-case TS (wTS), which results from swapping RF and RB in MTA and recalculating transformation scores, and MOMA TS (mTS), a more realistic scenario that minimizes the distance to reference fluxes upon perturbation. The rMTA score ranks high perturbations satisfying that 1) mTS>0 and 2) bTS>0 and wTS<0, as we showed in the different cases analyzed in the Results section.

A crucial insight that our rMTA approach uncovered was that most of the best-ranked gene knockouts in MTA suffer from under-determination, i.e. they have a similar capacity to move in the desired and undesired direction. Thus, users should be cautious with the results from MTA (bTS) and apply more restrictive filters, as we proposed here using bTS, wTS and mTS. In the controlled gene knockout experiments with human and mouse samples (Table 2), the benefit of using the score proposed by rMTA over MTA is clearly observed.

We believe that this intrinsic under-determination can only be broken by adding more experimental data. The integration of metabolomics data and isotope labeling data will allow us to more accurately define the reference fluxes in the source state. A better definition of references fluxes is a key point in both MTA and rMTA, but it is not the only question to be addressed. A second challenge, perhaps more relevant, is to disentangle internal rules and regulatory constraints as to how cells accomplish their metabolic adaptation after perturbations. Overall, the rational study of these questions will allow us to identify better therapeutic strategies to target aberrant metabolism in human disease.

Acknowledgements

We thank Eytan Ruppin and Keren Yizhak for providing us with the E.coli data for conducting the comparison of MTA and rMTA and two anonymous reviewers for their insightful comments and suggestions.

Funding

The work of L.V.Valcarcel is supported by Instituto de Salud Carlos III (ISCIII) [FI17/00297]; F.J. Planes is supported by the Ministry of Economy and Competitiveness of Spain [BIO2013-48933, BIO2016-77998-R]; V.Torrano is funded by Fundación Vasca de Innovación e Investigación Sanitarias, BIOEF (BIO15/CA/052), the AECC J.P. Bizkaia and the Basque Department of Health (2016111109); A. Carracedo is supported by the Basque Department of Industry, Tourism and Trade (Etortek) and the department of education (IKERTALDE IT1106-16), the BBVA foundation, the MINECO (SAF2016-79381-R (FEDER/EU); Severo Ochoa Excellence Accreditation SEV-2016-0644; Excellence Networks SAF2016-81975-REDT), European Training Networks Project (H2020-MSCA-ITN-308 2016 721532), the AECC (IDEAS175CARR; GCTRA18006CARR) and the European Research Council (Starting Grant 336343, PoC 754627). CIBERONC was co-funded with FEDER funds.

Conflict of Interest: none declared.

References

Apaolaza,I. et al. (2017) An in-silico approach to predict and exploit synthetic lethality in cancer metabolism. Nature communications, 8, 459.

Audet-Walsh,E. et al. (2016) The PGC-1a/ERRa Axis Represses One-Carbon Metabolism and Promotes Sensitivity to Anti-folate Therapy in Breast Cancer.

Bakay,M. et al. (2006) Nuclear envelope dystrophies show a transcriptional fingerprint suggesting disruption of Rb–MyoD pathways in muscle regeneration. Brain, 129, 996–1013.

Barrett,T. et al. (2012) NCBI GEO: archive for functional genomics data sets—update. Nucleic acids research, 41, D991–D995.

Bengtsson,H. et al. (2008) aroma. affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory tech report.

Cervera,A.M. et al. (2008) Cells silenced for SDHB expression display characteristic features of the tumor phenotype. Cancer Research, 68, 4058–4067.

Cortazar,A.R. et al. (2018) CANCERTOOL: a visualization and representation interface to exploit cancer datasets. Cancer research, 78, 6320–6328.

Du,P. et al. (2008) lumi: a pipeline for processing Illumina microarray. Bioinformatics, 24, 1547–1548.

Duarte,N.C. et al. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proceedings of the National Academy of Sciences, 104, 1777–1782.

Folger,O. et al. (2011) Predicting selective drug targets in cancer through metabolic networks. Molecular Systems Biology, 7.

Lee,D. et al. (2017) Folate cycle enzyme MTHFD1L confers metabolic advantages in hepatocellular carcinoma. The Journal of clinical investigation, 127, 1856–1872.

MacLennan,N.K. et al. (2009) Weighted gene co-expression network analysis identifies biomarkers in glycerol kinase deficient mice. Molecular genetics and metabolism, 98, 203–214.

Megchelenbrink,W. et al. (2014) optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. PloS one, 9, e86587.

Oberhardt,M.A. et al. (2013) Metabolically re-modeling the drug pipeline. Current opinion in pharmacology, 13, 778–785.

Plata,G. et al. (2010) Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. Molecular Systems Biology, 6.

Ritchie,M.E. et al. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research, 43, e47–e47.

Sagawa,M. et al. (2017) Ribonucleotide Reductase Catalytic Subunit M1 (RRM1) as a Novel Therapeutic Target in Multiple Myeloma. Clinical Cancer Research.

Schellenberger,J. and Palsson,B.Ø. (2009) Use of randomized sampling for analysis of metabolic networks. Journal of Biological Chemistry, 284, 5457–5461.

Segre,D. et al. (2002) Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences, 99, 15112–15117.

Shlomi,T. et al. (2008) Network-based prediction of human tissue-specific metabolism. Nat Biotech, 26, 1003–1010.

Shlomi,T. et al. (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proceedings of the National Academy of Sciences, 102, 7695–7700.

Sigurdsson,M.I. et al. (2010) A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. BMC systems biology, 4, 140.

Stempler,S. et al. (2014) Integrating transcriptomics with metabolic modeling predicts biomarkers and drug targets for Alzheimer's disease. PLoS One, 9, e105383.

Torrano,V. et al. (2016) The metabolic co-regulator PGC1α suppresses prostate cancer metastasis. Nature cell biology, 18, 645.

Webster,J.A. et al. (2009) Genetic control of human brain transcript expression in Alzheimer disease. The American Journal of Human Genetics, 84, 445–458.

Yizhak,K. et al. (2013) Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. Nature communications, 4, 2632.