## SUPPLEMENTARY DATA

**SUPPLEMENTARY NOTES**

**1 Mouse dataset**

Mouse RNA-seq data were retrieved from Hill *et al.* (2018) (NCBI: PRJNA453187). Two conditions of Ly6c adipose tissue macrophage transfer and PBS-injected control in epididymal white adipose tissue with four replicates each were considered. The BiGG iMM1415 model (http://bigg.ucsd.edu/models, Sigurdsson *et al.*, 2010) was used as the comprehensive mouse model with 2484 reactions, 1137 genes, and 1704 metabolites.

**2 Genome-scale metabolic network analysis**

***2.1 FBA, FVA, and blocked reactions***

These classic algorithms are summarized here as used in this paper. Let be a GSMN with *n* reactions, *m* metabolites. **S** denotes the stoichiometric matrix of dimension , which contains the stoichiometric coefficients of metabolites in each reaction, with positive coefficients for products and negative coefficients for substrates. Let be a vector representing the fluxes on the *n* reactions, which are constrained by lower and upper bounds, given by the vectors **lb** and **ub**, respectively. Assuming that the flux distribution exists in a quasi-steady state, *i.e.* the rate of production of each metabolite is equal to its rate of consumption, the balance in the system is expressed as

.

Flux Balance Analysis (FBA) uses linear programming (LP) to optimize an objective function where **c** is a vector of weights. Maximizing the biomass production is usually of interest and can be achieved by formulating the particular vector with a weight equal to 1 for the growth reaction and weights equal to 0 for all the other reactions. In this context, the optimization problem addressed by FBA can be stated as

find subject to

and the model is said to be *viable* if .

Flux Variability Analysis (FVA) can be used to detect the reactions that are blocked in the model, *i.e.* that cannot carry a flux and could be removed from the model without affecting its predictive properties. It consists in solving the following LP problem:

find and for *i* = 1..*n*

subject to

The *i*th reaction is considered to be blocked if (Henry *et al.*, 2007).

***2.2 Software used***

The analyses were done using R (v3.4.2). FBA, FVA, dead-end detection, reaction and gene knock-out analyses were performed with the *sybil* R package (v2.0.4) (Gelius-Dietrich *et al.*, 2013). The *clp* solver (v1.16.10) with *inibarrier* algorithm via the *clpAPI* R package (v1.2.7) (Gelius-Dietrich, 2016) was used for the simulation of *Y. lipolytica*. The *glpk* solver (v4.59) with simplex algorithm via the *glpkAPI* R package (v1.3.0) (Gelius-Dietrich, 2015) was used for the simulation of mouse. The *sybil* tolerance was set to 1e-8.

The *createTissueSpecificModel* function implemented in the COBRA Toolbox v3.0 (Heirendt *et al.*, 2017) was employed to perform the analyses with the GIMME (Becker and Palsson, 2008) and iMAT (Zur *et al.*, 2010) algorithms in Matlab (R2018a).

**3 RNA-seq data analysis**

RNA-seq reads were aligned on the *Y. lipolytica* genome GCA\_000002525.2\_ASM252v1 or the mouse genome GCA\_000001635.8 using *tophat2* (v2.0.13) (Kim *et al.*, 2013). The number of reads mapped onto each gene locus was obtained with *HTSeq-count* (v0.6.1) (Anders *et al.*, 2015). The RNA-seq read counts were normalized using the TMM method from the *edgeR* R package (v3.12.1) (Robinson and Oshlack, 2010), and subsequently transformed to log2-counts with *voom* transformation (Law *et al.*, 2014). Then, the average log2-counts in each condition was computed and referred to as **expr** in our analysis. RPKM expression pkmExpr denoted normalized expression by gene length and library size and was computed as number of reads per kilobase per million mapped reads.

To determine differentially expressed genes, we built a linear model with one factor per condition on the transformed data, and performed differential analysis via moderated *t*-statistics on 21 pair-wise contrasts between the seven conditions studied in Maguire *et al.* (2014), as well as one contrast between the two conditions in Hill *et al*. (2018) with the *limma* R package (v3.30.13) (Smyth, 2005),. Cut-offs of fold-change = 2 and FDR = 0.05 were used to determine differentially expressed genes.

**4 GO annotation**

The GO annotations of *Y. lipolytica* consisting of 4629 GO terms and 4747 annotated genes (out of 6453) were retrieved from UniProt (UniProt Consortium, 2015). An initial enrichment was performed via topGO with *weight01* algorithm and *fisher* statistic on the set of genes in the model iMK735 *versus* the complete genome (gene universe). This yielded 135 GO terms (denoted *GO135* and listed in Table S1) in the biological process category, which were associated with at least 3 and at most 50 genes in the iMK735 model and enriched with *p*-value < 0.1. This subset of GO terms helped to reduce the computing time while preserving most relevant information documented by iMK735. The comparison of our algorithm, topGO, GSEA, and Maguire *et al*. (2014) was performed on GO135. topGO (v2.24.0) was performed with *weight01* algorithm and *fisher* statistic (Alexa and Rahnenführer, 2016) using the GO annotation and gene universe of the whole genome. Twenty-one sets of differentially expressed genes in 21 pair-wise contrasts in Maguire *et al*. (2014) as mentioned above were studied. Gene set enrichment analysis was performed with the GSEA software (v2.2.3) (Subramanian *et al.*, 2005) using the same GO annotation and gene universe. The gene set of iMK735 with its RNA-seq transformed count data was provided as input, and all the 21 pair-wise contrasts were studied. *P*-values obtained with metaboGSE, topGO, and GSEA were adjusted using Benjamini-Hochberg correction (FDR) (Benjamini and Hochberg, 1995).

The mouse GO annotation consisting of 16429 GO terms and 23905 annotated genes was retrieved via the *org.Mm.eg.db* R package (v3.6.0) (Carlson, 2018). An initial enrichment was performed via topGO on the set of genes from the model iMM1415 *versus* the complete genome (gene universe) to obtain all iMM1415-related GO terms. The latter was further filtered to those related to inflammatory response, cholesterol and lipid biosynthesis, as suggested by Hill *et al.* (2018) and resulted in 24 GO terms (denoted *GO24* and listed in Figure S4). The comparisons of metaboGSE, topGO with both *classic* and *weight01* algorithms, and GSEA were performed on these GO terms for the unique contrast between PBS and Ly6c RNA-seq conditions from Hill *et al*. (2018).

**5 Randomization test for an individual condition**

In addition to the test for significance of discrepancy of a given gen set between conditions, metaboGSE also permits investigating the significance of a gene set in a single experimental condition via a test for randomness of the gene set. For each condition, we generate depletion curves for random gene sets and compare their area under the curve (AUC) with that of *g*. The single *p*-value of *g* for each condition indicates the probability that AUC of *g* is higher than AUC of a random gene set. This test is available in the *metaboGSE* R package, yet not discussed in the paper. It can be used to filter out the gene sets with significant discrepancy between conditions that only occurs in sub-models of very low fitness.

**SUPPLEMENTARY FIGURES**



**Figure S1: Comparison of genes in *Y. lipolytica* sub-models constructed with metaboGSE, GIMME, and iMAT for RNA-seq data from Maguire *et al*. (2014).** Intersection: number of common genes across conditions, Union: number of all genes across conditions, k: number of genes in the comprehensive model.

[FigS2.pdf]

**Figure S2.** **Enrichment** **of *GO135* via metaboGSE with *Y. lipolytica* sub-models for RNA-seq data from Maguire *et al*. (2014).**

****

**Figure S3.** **Venn diagram of the** **top 50 GO terms enriched via each of the three methods metaboGSE, topGO *weight01*, and GSEA for RNA-seq data from Maguire *et al.* (2014).**



**Figure S4.** **Enrichment of GO:0006696 ergosterol biosynthetic process with *Y. lipolytica* sub-models for RNA-seq data from Maguire *et al*. (2014).**

(A) Depletion fraction in function of number of removed genes

(B) Depletion curve: depletion fraction in function of fitness times fraction of remaining genes in the model

(C) Expression of associated genes. \* Genes causing sudden drops in different depletion curves, yet not belonging to the GO:0006696 gene set.



**Figure S5. Enrichment –log10 *p*-values of GO terms with different levels of specificity** concerning inflammatory response (orange), cholesterol (green) and lipid biosynthesis (magenta) by metaboGSE (in red), topGO *weight01* (in blue), topGO *classic* (in cyan), and GSEA (in yellow) for the mouse contrast PBS versus Ly6c from *Hill et al*. (2018). NA values indicate that the GO terms were not identified by the corresponding method.

**SUPPLEMENTARY TABLES**

[TableS1.xlsx]

**TableS1.** **Performance of different rankings measured by performance index in all RNA-seq samples from Maguire *et al***. **(2014) and Hill *et al.* (2018)**. **random**: random draw, **expr**: voom-normalized expression, **pkmExpr**: voom-normalized expression in RPKM, **relExpr1**: relative expression **expr/<expr>**, **revExpr**: reverse expression, **zExpr**: *z*-score. Lowest in **bold**, highest in *italic*.

[TableS2.xlsx]

**Table S2. Enrichment of GO135 based on integration of RNA-seq data and metabolic networks compared to those obtained in Maguire *et al***. **(2014)**. FDR from permutation test (metaboGSE), Fisher test (topGO), phenotype-based permutation test (GSEA) and Maguire *et al*. (2014) are reported. NA values indicate that the GO terms were not found by the corresponding methods. The top 50 GO terms enriched by metaboGSE are marked in red, topGO weight01 in blue, and GSEA in yellow.

**References**

Alexa,A. and Rahnenführer,J. (2016) topGO: Enrichment Analysis for Gene Ontology. R package version 2.24.0.

Anders,S. *et al.* (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.

Becker,S.A. and Palsson,B.O. (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.*, **4**, e1000082.

Benjamini,Y. and Hochberg,Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.*, **57**, 289–300.

Carlson,M. (2018) org.Mm.eg.db: Genome wide annotation for Mouse. R package version 3.6.0.

Gelius-Dietrich,G. (2016) clpAPI: R Interface to C API of COIN-OR Clp. R package version 1.2.7.

Gelius-Dietrich,G. (2015) glpkAPI: R Interface to C API of GLPK. R package version 1.3.0.

Gelius-Dietrich,G. *et al.* (2013) sybil – Efficient constraint-based modelling in R. *BMC Syst. Biol.*, **7**, 125.

Heirendt,L. *et al.* (2017) Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. *ArXiv171004038 Q-Bio*.

Henry,C.S. *et al.* (2007) Thermodynamics-Based Metabolic Flux Analysis. *Biophys. J.*, **92**, 1792–1805.

Hill,D.A. *et al.* (2018) Distinct macrophage populations direct inflammatory versus physiological changes in adipose tissue. *Proc. Natl. Acad. Sci. U. S. A.*, **115**, E5096–E5105.

Kim,D. *et al.* (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.*, **14**, R36.

Law,C.W. *et al.* (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.*, **15**, R29.

Maguire,S.L. *et al.* (2014) Zinc finger transcription factors displaced SREBP proteins as the major Sterol regulators during Saccharomycotina evolution. *PLoS Genet.*, **10**, e1004076.

Robinson,M.D. and Oshlack,A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**, R25.

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

Smyth,G.K. (2005) limma: Linear Models for Microarray Data. In, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Statistics for Biology and Health. Springer, New York, NY, pp. 397–420.

Subramanian,A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 15545–15550.

UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204-212.

Zur,H. *et al.* (2010) iMAT: an integrative metabolic analysis tool. *Bioinforma. Oxf. Engl.*, **26**, 3140–3142.