



HAL
open science

Large-Scale Modeling Approach Reveals Functional Metabolic Shifts during Hepatic Differentiation

Nathalie Poupin, Anne Corlu, Nicolas J. Cabaton, H el ene Dubois-Pot-Schneider, C ecile Canlet, Elodie Person, Sandrine S. Bruel, Cl ement Frainay, Florence Vinson, Florence Maurier, et al.

► **To cite this version:**

Nathalie Poupin, Anne Corlu, Nicolas J. Cabaton, H el ene Dubois-Pot-Schneider, C ecile Canlet, et al.. Large-Scale Modeling Approach Reveals Functional Metabolic Shifts during Hepatic Differentiation. *Journal of Proteome Research*, 2019, 18 (1), pp.204-216. 10.1021/acs.jproteome.8b00524 . hal-01937261

HAL Id: hal-01937261

<https://univ-rennes.hal.science/hal-01937261v1>

Submitted on 3 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

1
2
3
4
5
6
7 1 A large-scale modelling approach reveals functional
8
9
10
11 2 metabolic shifts during hepatic differentiation
12
13
14
15

16 3 *Nathalie Poupin*^{*1}, *Anne Corlu*², *Nicolas J. Cabaton*¹, *Hélène Dubois-Pot-Schneider*^{2‡}, *Cécile*
17
18 4 *Canlet*¹, *Elodie Person*¹, *Sandrine Bruel*¹, *Clément Frainay*¹, *Florence Vinson*¹, *Florence*
19
20 5 *Maurier*^{1‡}, *Fabrice Morel*², *Marie-Anne Robin*², *Bernard Fromenty*², *Daniel Zalko*¹ and *Fabien*
21
22
23 6 *Jourdan*¹
24
25

26 7
27
28 8 ¹UMR1331 Toxalim (Research Centre in Food Toxicology), Université de Toulouse, INRA,
29
30 9 ENVY, INP-Purpan, UPS, Toulouse, France
31
32

33
34 10 ²Univ Rennes, INSERM, INRA, Institut NUMECAN (Nutrition Metabolisms and Cancer)
35
36 11 UMR_A 1341, UMR_S 1241, F-35000 Rennes, France
37
38

39 12
40
41
42 13
43
44
45 14
46
47 15 **Corresponding Author**
48

49
50 16 *E-mail: nathalie.poupin@inra.fr
51
52
53
54 17
55
56
57
58
59
60

18 ABSTRACT

19 Being able to explore the metabolism of broad metabolizing cells is of critical importance in many
20 research fields. This article presents an original modelling solution combining metabolic network
21 and omics data to identify modulated metabolic pathways and changes in metabolic functions
22 occurring during differentiation of a human hepatic cell line (HepaRG). Our results confirm the
23 activation of hepato-specific functionalities and newly evidence modulation of other metabolic
24 pathways, which could not be evidenced from transcriptomic data alone. Our method takes
25 advantage of the network structure to detect changes in metabolic pathways that do not have gene
26 annotations, and exploits flux analyses techniques to identify activated metabolic functions.
27 Compared to usual cell-specific metabolic network reconstruction approaches, it limits false
28 predictions by considering several possible network configurations to represent one phenotype,
29 rather than one arbitrarily selected network. Our approach significantly enhances the
30 comprehensive and functional assessment of cell metabolism, opening further perspectives to
31 investigate metabolic shifts occurring within various biological contexts.

32 KEYWORDS: Genome scale metabolic modelling / global metabolic shifts / HepaRG cell line /
33 hepatic differentiation / transcriptomics and metabolomics

34 INTRODUCTION

35 Getting a global picture of how cell metabolism is reprogrammed or shifted under different
36 conditions is a crucial step towards the understanding of how cellular metabolic functionalities can
37 be altered or modified under the induction of internal processes (*e.g.*, cell differentiation) or
38 external factors (*e.g.*, toxics or drugs). Technological advances in high-throughput DNA and RNA
39 sequencing methods now enable the analysis of the complete genome or transcriptome of a given
40 cell or organism, providing a first large-scale picture of its biological status under a given condition.
41 However, this information is often not sufficient to describe nor predict the metabolic phenotype
42 of the studied system, as metabolic processes involve many interconnected reactions, whose final
43 resulting activities are not always directly and strictly related to the observed gene expression
44 levels. Understanding the metabolic behavior of cells and tissues in given conditions requires
45 apprehending this complex network of reactions at the whole genome scale.

46 In this perspective, genome-scale metabolic network reconstructions (GSMNR) open promising
47 possibilities for shaping a comprehensive picture of the metabolism and its modulation within
48 various contexts, such as development, pathological processes, or under drug exposure. These
49 reconstructions aim at assembling all the biochemical transformations known to occur in a given
50 organism, based on its genome annotation and data available from the literature ¹. They include
51 information about the gene-protein-reactions (GPR) associations, which describe the relationships
52 between the metabolic reactions, the enzymes that catalyze them, and the genes these enzymes are
53 encoded by. Hence, GSMNR provide a relevant scaffold for the mechanistic analysis and
54 interpretation of gene expression profiles in terms of metabolic phenotype. To date, several generic
55 reconstructions have been published for the human metabolic network (Recon 1 ²; EHMN ³; HMR2
56 ⁴; Recon 2 ⁵; Recon2.1 ⁶; Recon 2.2 ⁷; iHsa ⁸; Recon 3D ⁹). These reconstructions represent the
57 overall theoretical metabolic capacity of a generic human cell or tissue, regardless of external

1
2
3 58 conditions or cell or tissue specificities, and encompass up to 13000 reactions and 3700 genes for
4
5 59 the latest reconstruction. Because of their genericity, these reconstructions cannot directly be used
6
7
8 60 to predict accurately cell- or tissue- specific metabolic behaviors. They need to be tailored to more
9
10 61 specific subnetworks, by identifying the reactions from the original generic GSMNR that are
11
12 62 specifically active in the studied cell, tissue or conditions. Several methods have been proposed to
13
14 63 tackle this challenge and have already succeeded in capturing the specific metabolic features of
15
16 64 tissues ¹⁰⁻¹² or cancer cells ¹³⁻¹⁷. These methods use experimental data (*e.g.*, transcriptomic or
17
18 65 proteomic data) to predict a consistent subnetwork of reactions specifically active in the studied
19
20
21 66 condition that best matches with the experimental data. The algorithms used for this purpose rely
22
23 67 on constraint-based modeling (CBM) approaches, which aim at calculating the steady-state fluxes
24
25
26 68 of metabolites, *i.e.*, the rates of metabolic conversion or transfer, through each network reaction ¹⁸.

27
28 69 In this study, we propose a generic approach, based on the human GSMNR Recon 2, and on
29
30
31 70 transcriptomic and metabolomic data, to identify global changes in metabolic functions between
32
33 71 two cell populations. We based our new method on a previously published algorithm for tissue-
34
35 72 specific network reconstruction, the iMat algorithm ^{10,19}, with the originality that we depicted the
36
37 73 metabolism of the each cell population by a set of subnetworks rather than one unique subnetwork.
38
39
40 74 We hypothesize that this approach enables to better characterize the whole metabolic capability of
41
42 75 a cellular system independently of specific conditions and limits the risk of false predictions. We
43
44 76 applied this approach for comparing the metabolism of the human hepatic cells HepaRG in two
45
46
47 77 distinct differentiation stages: undifferentiated progenitor cells and fully differentiated hepatocyte-
48
49 78 like cells. The HepaRG cell line, which is derived from a human hepatocellular carcinoma, is a
50
51 79 relevant human model to study hepatic cell differentiation, as it has the particularity to exhibit
52
53
54 80 features of bipotent progenitor cells but is also able to differentiate into mature hepatocyte-like
55
56 81 cells. These fully differentiated hepatocyte-HepaRG cells have been demonstrated to display

1
2
3 82 metabolic capacities close to primary human hepatocytes, and in particular, to express numerous
4
5 83 xenobiotic metabolizing enzymes ^{20,21} and to present functional mitochondria ²². Hepatic cell
6
7
8 84 differentiation is a key process in liver maintenance and regeneration and many liver diseases may
9
10 85 imply, to some extent, hepatic regenerative processes. Therefore, being able to compare the
11
12 86 metabolic capabilities of cells at different developmental stages in *in vitro* models is a critical step
13
14 87 when aiming to generate fully functional hepatocytes for liver transplantation or to study disease
15
16
17 88 or toxicological mechanisms ²³. Metabolic differences between pluripotent stem cells and
18
19 89 differentiated cells have been reported ^{24–26}, but studies providing a more global overview of the
20
21 90 metabolic shifts occurring during re-programming and differentiation are still lacking. Besides,
22
23 91 most studies investigating the differences between progenitors and differentiated HepaRG cells
24
25
26 92 have so far focused on a set of selected transcripts or proteins associated with the expression of
27
28 93 enzymes involved in xenobiotics metabolism or with liver-specific functions (such as albumin
29
30 94 production) ^{20,27,28}. A few global methylomic or transcriptomic analyses have enabled to further
31
32
33 95 explore the modulations occurring at the regulation level upon differentiation and have reported
34
35 96 the modulation of global cellular processes (such as cell cycle, cell death, apoptosis, cell
36
37 97 morphology ...) ^{29,30}. However, information is still missing about the modulations occurring at the
38
39 98 metabolic level and involving sustainable changes in metabolic capacities and functionality of these
40
41
42 99 cells. The global approach we propose in this study intends to help generating a global picture of
43
44 100 the metabolism of HepaRG cells. In comparison to usual transcriptomic analyses, the method we
45
46
47 101 developed takes advantage of the network structure to go beyond the descriptive analysis based on
48
49 102 metabolic pathways and to gain information about the functional capacity of the cells, without
50
51 103 focusing on one *a priori* defined restricted set of metabolic features or genes.
52
53
54 104

105 **MATERIALS & METHODS**

106 **HepaRG cell culture**

107 The HepaRG cells, kindly given by Dr C. Guguen-Guillouzo, were cultured according to the
108 standard protocol described by Aninat et al.²⁰. Briefly, HepaRG cells were seeded at a density of
109 2.4×10^4 cells/cm² in a growth medium composed of William's E medium (Gibco, Illkirch, France)
110 supplemented with 10% fetal bovine serum (HyClone, Thermo Scientific, Illkirch, France), 100
111 U/ml penicillin and 100 µg/mL streptomycin, 5 µg/mL insulin, 2mM glutamine (Sigma, Saint-
112 Quentin Fallavier, France) and 50 µM hydrocortisone (Serb, Paris, France). Undifferentiated
113 HepaRG cells were collected after 3 days of culture on this medium ("3-day cells"). To obtain
114 differentiated HepaRG cells, cells were cultured in the initial medium for 2 weeks, then switched
115 to a medium supplemented with 2% DMSO (Sigma) and cultured on this supplemented medium
116 for 2 more weeks. Addition of DMSO induces of HepaRG cells to two different cell types:
117 hepatocyte-like cells and biliary like-cells. HepaRG cells corresponding to fully differentiated
118 hepatocytes were called "30-day cells".

119 **Microarray analyses**

120 Total messenger RNA (mRNA) from four biological replicates was extracted from 3-day and 30-
121 day HepaRG cells. mRNA was checked for purity and integrity, using an Agilent Bioanalyser
122 (Agilent Technologies, Palo Alto, CA). Genome-wide expression profiling was performed using
123 the low-input QuickAmp labeling kit and human SurePrint G3 8x60K pangenomic microarrays
124 (Agilent Technologies, Santa Clara, CA, USA). Gene expression data were processed using Feature
125 Extraction and GeneSpring software (Agilent Technologies).

126 Genes were classified in 2 categories according to their expression level. Genes whose expression
127 level was below the defined threshold of 150 were considered as not detected or not expressed
128 (NE), whereas genes whose expression level was above 150 for all replicates were considered as

1
2
3 129 significantly expressed (HE). In addition, genes that were found to be significantly up-regulated
4
5 130 between d3 and d30 (p-value <0.01 with an unpaired t-test adjusted with Bonferoni correction and
6
7 131 with a fold change higher than 2) were considered as significantly expressed (HE) at d30 and,
8
9
10 132 similarly, down-regulated genes were considered as significantly expressed at d3. The threshold
11
12 133 value of 150 corresponds as the 25 percentile of the gene expression level among all genes and all
13
14 134 replicates, which is generally considered as the background noise in Agilent microarray
15
16
17 135 experiments.

19 136 **Metabolomic analyses**

20
21 137 Metabolites were extracted from 1 million of cells with an acetonitrile/water (1:9) solvent
22
23 138 mixture. Samples were centrifuged 10 minutes at 5000 g. Supernatants were evaporated using a
24
25
26 139 SpeedVac and resuspended in deuterated water (D20), samples were vortexed and transferred into
27
28 140 5 mm NMR tubes.

29
30
31 141 ¹H NMR spectra of cell extracts were acquired on a Bruker Avance spectrometer (Bruker,
32
33 142 Karlsruhe, Germany) operating at 600.13 MHz, and equipped with an inverse detection 5 mm TXI
34
35 143 ¹H-¹³C-¹⁵N cryoprobe connected to a cryoplatfrom. Spectra were acquired using a Carr-Purcell-
36
37 144 Meiboom-Gill (CPMG) spin echo pulse sequence with a 2 seconds relaxation delay to attenuate
38
39 145 macromolecules signals. A water suppression signal was achieved by presaturation during the
40
41
42 146 relaxation delay. The spectral width was set to 20 ppm for each spectrum and 512 scans were
43
44 147 collected with 32K points. Free induction decays were multiplied by an exponential window
45
46
47 148 function before Fourier Transformation. The spectra were manually phased and the baseline was
48
49 149 corrected using TopSpin 3.2 software (Bruker, Karlsruhe, Germany).

50
51 150 Metabolites were identified using the literature, home-made and freeware databases such as the
52
53
54 151 Human Metabolome Database (³¹; www.hmdb.ca/).

152 **Implementation and adaptation of the iMat algorithm to generate stage-specific metabolic** 153 **models of HepaRG cells**

154 We used the generic human metabolic network reconstruction Recon 2 (⁵; version 2.04,
155 downloaded from <http://vmh.uni.lu/#downloadview>) as a framework for the prediction of active
156 metabolic reactions in the HepaRG cells. This genome-scale metabolic network reconstruction
157 encompasses 7440 reactions, 2140 genes and 2626 unique metabolites and is supposed to represent
158 the comprehensive metabolic capacity of any human cell or tissue. The network reconstruction is
159 converted into a mathematical model, where the list of metabolic reactions is described as a
160 stoichiometric matrix (S). This matrix defines which metabolites (enumerated as rows) participate
161 in each of the network reactions (enumerated as columns), with numerical entries in the matrix
162 representing the stoichiometric coefficients of the reactions.

163 First, gene expression data were mapped to the network reactions by using the Gene-Protein-
164 Reaction (GPR) association rules defined in the network reconstruction. We used these GPR
165 associations to determine a set of a priori highly expressed reactions (HEr) or not expressed
166 reactions (NEr) according to the expression level of their associated gene(s). In the case of "AND"
167 GPR associations, we classified the reaction as HEr (or NEr respectively) if all genes were HE (or
168 NE respectively), whereas in the case of "OR" associations, the reaction was considered as HEr if
169 any of the associated gene was HE. In the initial Recon 2 reconstruction, 4821 reactions had defined
170 GPR associations.

171 We then implemented the iMat algorithm ¹⁰, that we adapted, in order to predict which reactions
172 from the generic metabolic network Recon 2 were specifically active in the HepaRG cells (Figure
173 S1). As initially proposed by Shlomi et al., the optimization problem was formulated as a Mixed
174 Integer Linear Programming (MILP) problem, to find a steady-state flux distribution that
175 maximizes the number of reactions whose flux is consistent with the measured expression level of

176 their associated gene(s) and that complies with stoichiometric and thermodynamic constraints. In
 177 our modified version of this algorithm, we added some constraints so that the predicted flux
 178 distribution is also in agreement with the metabolomics data, meaning that it allows all detected
 179 metabolites to be produced. More precisely, for each identified metabolite, at least one reaction
 180 that is able to produce it in the model should have a non-zero flux.

$$181 \max \quad \sum_{i \in HE_r} (y_i^+ + y_i^-) + \sum_{i \in NE_r} y_i$$

$$182 \quad v, y^+, y^-, y$$

183 subject to

$$184 \quad S \cdot v = 0 \quad (1)$$

$$185 \quad v_{\min} \leq v \leq v_{\max} \quad (2)$$

$$186 \quad v_i + y_i^+ (v_{\min,i} - \varepsilon) \geq v_{\min,i}, \quad i \in HE_r \quad (3)$$

$$187 \quad v_i + y_i^- (v_{\max,i} + \varepsilon) \leq v_{\max,i}, \quad i \in HE_r \quad (4)$$

$$188 \quad v_i + y_i \cdot v_{\min,i} \geq v_{\min,i}, \quad i \in NE_r \quad (5)$$

$$189 \quad v_i + y_i \cdot v_{\max,i} \leq v_{\max,i}, \quad i \in NE_r \quad (6)$$

$$190 \quad v_{i(j)} + x_{i(j)}^+ (v_{\min,i(j)} - \varepsilon) \geq v_{\min,i(j)}, \quad \text{for } (j \in \text{obs. mets}) \& (i \in R\text{Prod}_j) \quad (7)$$

$$191 \quad v_{i(j)} + x_{i(j)}^- (v_{\max,i(j)} + \varepsilon) \leq v_{\max,i(j)}, \quad \text{for } (j \in \text{obs. mets}) \& (i \in R\text{Prod}_j) \quad (8)$$

$$192 \quad \sum_i x_{i(j)} \geq 1, \quad \text{for } [j \in \text{ExpMets}] \quad (9)$$

$$193 \quad y_i^+, y_i^-, y_i, x_i^+, x_i^- \in [0;1]$$

194 v is the flux vector and S is the stoichiometric matrix.

195 Equation (1) ensures that flux values comply with the mass balance constraints. Thermodynamic
 196 and capacity constraints are imposed by equation (2): restricted direction and values for reaction
 197 flux, according to these constraints, are defined in v_{\min} and v_{\max} vectors (minimal and maximal flux

1
2
3 198 values respectively), which are set as lower and upper bound for reaction flux values. Equations
4
5 199 (3) to (6) set constraints corresponding to gene expression data. y_i^+ , y_i^- and y_i are boolean variables
6
7 200 representing the adequacy between the predicted flux v_i through reaction i and its expression level.
8
9
10 201 For HEr, y_i^+ (y_i^-) represents whether the reaction is active in the forward (or backward, respectively)
11
12 202 direction: equations (3) & (4) enforce that, if $y_i^+=1$ or $y_i^-=1$ (*i.e.*, the reaction is active), the value
13
14 203 of the flux through reaction i is larger than a threshold ϵ , whereas if $y_i^+=0$ or $y_i^-=0$ (*i.e.*, the reaction
15
16 204 is inactive), the flux through reaction i must be 0. For NEr, y_i represents whether the reaction is
17
18 205 inactive: equations (5) & (6) enforce that, if $y_i=1$ (*i.e.*, the reaction is inactive), the flux through
19
20 206 reaction i is 0. A threshold of 0.1 was used to predict the activity of reactions. Different threshold
21
22 207 values were tested and provided qualitatively similar results. Equations (7) to (9) set constraints
23
24 208 corresponding to metabolomic data. For each identified metabolite j ("ExpMets"), "RProd $_j$ "
25
26 209 encompass all reactions i that can produce this metabolite according to the stoichiometric matrix
27
28 210 S . x_i^+ (or x_i^- , respectively) are Boolean variable representing whether the reaction i is active or not
29
30 211 in the forward (or backward, respectively) direction. When $x_i^+=1$, the flux through reaction i must
31
32 212 be larger (or smaller, respectively) than ϵ (equation (7)), forcing the reaction to be forward active.
33
34 213 For reversible reactions, when $x_i^- = 1$, the flux through reaction i must be negative and lower than
35
36 214 ϵ (equation (8)), forcing the reaction to be backward active. If $x_i^+=0$ and $x_i^-=0$, the flux through
37
38 215 reaction i is forced to be 0, forcing the reaction to be inactive. The constraints set in Eq. 9 ensure
39
40 216 that at least one reaction is able to produce the metabolite j is active.

41
42 217 The optimization problem consists in finding the flux distribution v , which maximizes the
43
44 218 number of HEr which are active ($v_{\text{HEr}} \geq \epsilon$) and the number of NEr which are inactive ($v_{\text{NEr}} = 0$).
45
46 219 This sum represents the "adequacy score" of the flux distribution. The percentage of adequacy is
47
48 220 calculated as the adequacy score divided by the theoretical maximal adequacy score that could be
49
50
51
52
53
54
55
56
57
58
59
60

221 obtained (*i.e.*, the sum of all reactions associated with HE genes and the sum of all reactions
222 associated to NE genes).

223 The algorithm was implemented in Matlab 2014a (The Mathworks, Natick, MA, USA) and the
224 resolution of the MILP problem was performed using the ilog CPLEX solver version 12.6.0 (ILOG,
225 Sunnyvale, CA, USA).

226 **Identification of alternative adequate stage-specific subnetworks equally fitting the** 227 **experimental data**

228 The flux distribution obtained by solving the MILP problem is optimal in term of adequacy with
229 transcriptomic and metabolomic data, but is not unique. Alternative solutions exist that have the
230 same adequacy with experimental data. To explore the space of alternative solutions, as proposed
231 by Shlomi *et al.* and in a further study³², we searched whether, for each reaction that was predicted
232 to be active (or inactive, respectively) in the first optimal solution, a solution with a similar
233 adequacy score could be found with this reaction being inactive (or active, respectively) (Figure
234 S1C). From the initial computed flux distribution, each reaction was successively forced to be
235 either active (*i.e.*, to carry a non-zero flux, with $v \geq \epsilon$) or inactive ($v = 0$) and a new MILP
236 optimization was performed. Reversible reactions were successively set to be active in each
237 direction ($v \geq \epsilon$, then $v \leq -\epsilon$). Only flux distributions having a maximal adequacy score were kept,
238 forming the final set of optimal solutions. Each of these solutions represents a subnetwork of
239 predicted active reactions, with optimal adequacy with experimental data. The same computations
240 were performed independently using the data obtained at d3 and d30, providing one set of equally
241 adequate subnetworks with regard to gene expression data for each d3 and d30 stage.

242 For each stage, reactions that were found to be active in all possible subnetworks were considered
243 as "required" (R) reactions, meaning that no maximal adequacy score could be achieved when these
244 reactions were forced to be inactive. Conversely, reactions that display a zero-flux in all the optimal

1
2
3 245 solutions were considered as "Inactive" (I) reactions, since no maximal adequacy score could be
4
5 246 achieved when these reactions were forced to be active. All other reactions that are either active or
6
7
8 247 inactive in the optimal solutions were considered as "Potentially active" (PA) reactions. As a result,
9
10 248 sets of R, PA and I reactions were defined for each stage (Figure S1D).

12 249 **Assessment of the liver-specificity of genes in generated subnetworks**

14 250 To evaluate the liver-specificity of our generated models, we assessed, for each subnetwork, the
15
16
17 251 number of predicted "expressed" genes that can be considered as liver-specific. We used
18
19 252 information about the tissue location of genes from the UNIPROT³³ and the Human Protein Atlas
20
21 253 (HPA³⁴) databases.

23 254 First, for each generated subnetwork, genes predicted as "expressed" were inferred from the
24
25
26 255 predicted active reactions using the GPR rules. UNIPROT identifiers for Recon 2 genes were
27
28 256 retrieved from their EntrezGene identifiers (as available in the initial reconstruction) using the
29
30
31 257 DAVID database³⁵ and predicted "expressed" genes were compared to genes with evidence for
32
33 258 presence in the liver according to the UNIPROT database. For each subnetwork, we calculated the
34
35 259 "recall" of liver-specific genes, as the proportion of Recon 2 genes with evidence of presence in
36
37 260 liver that were predicted to be "expressed" in the subnetwork, and the "precision", as the proportion
38
39 261 of predicted "active" genes in the subnetwork that were liver located.

41
42 262 In a second analysis, we mapped the confidence level for protein expression in liver hepatocytes
43
44 263 assigned in the HPA database (High, Medium, Low or Not detected) to the generated subnetworks
45
46
47 264 and evaluated, in our HepaRG 3-day and 30-day subnetworks, the proportion of predicted active
48
49 265 reactions associated with high, medium or low confidence level for protein expression in liver
50
51 266 according to the HPA database. The recall of reactions associated with high confidence for liver
52
53
54 267 expression was calculated as the proportion of reactions with high confidence level that were

1
2
3 268 predicted to be active in each subnetwork. The precision was calculated as the proportion of
4
5 269 predicted active reactions that had a high confidence level.

8 270 **Simulation of achievable metabolic functions in generated subnetworks**

9
10 271 We performed flux balance analysis to test whether our generated subnetworks were able to
11
12 272 achieve some defined metabolic functions. A list of 155 metabolic functions (SI Table S1),
13
14 273 including 111 generic functions (*i.e.*, functions assumed to be fulfilled by any tissue or cell
15
16 274 independently of its type) and 44 hepato-specific functions, was gathered from previous
17
18 275 publications ^{4,36,37}. For each of these tasks, we defined a corresponding objective function and
19
20 276 maximized it to check whether it could carry a non-zero flux. The sets of metabolites that could be
21
22 277 taken up and released was restrained to mimic a minimal standard medium and adjusted depending
23
24 278 on the specific metabolic objective tested. For instance, we tested the ability of our models to
25
26 279 perform gluconeogenesis (the formation of glucose from various gluconeogenic substrates) by
27
28 280 restraining the set of metabolites that could be taken up to only non-carbon sources (O₂) except for
29
30 281 lactate and glucogenic amino acids (alanine, glutamine ...) and maximize the excretion of glucose.
31
32 282 This functional testing was performed for each of the generated subnetworks at both d3 and d30.
33
34 283 We verified that all the 155 defined metabolic functions could be achieved by the initial Recon 2
35
36 284 model. The complete list of tested metabolic functions and corresponding applied constraints for
37
38 285 uptake and secretion is provided in supplemental Table 1.

44 286 **Analysis of the variability among generated subnetworks**

45
46 287 PCA was performed on all generated subnetworks, independently of the stage. Reactions were
47
48 288 used as variables and each subnetwork was represented by a vector of binary values corresponding
49
50 289 to the predicted activity state of each reaction (0 for predicted inactive reactions and 1 for predicted
51
52 290 active reactions).

56 291 **Pathway enrichment analyses for identification of activated and inactivated pathways**

1
2
3 292 Reactions were considered as inactivated during the differentiation if they were predicted to be
4
5 293 active in at least one subnetwork (R or PA) at d3 and inactive at d30, and conversely activated if
6
7 294 they were predicted to be inactive at d3 and R or PA at d30. Pathway enrichment analyses were
8
9
10 295 performed over activated and inactivated reactions to assess whether the given reactions were
11
12 296 significantly over-represented in a metabolic pathway. Pathway enrichment statistics were
13
14 297 performed using one-tailed exact Fisher test, with a Bonferroni correction for multiple tests³⁸,
15
16
17 298 using the metabolic pathways defined in Recon 2.04. Blocked reactions, identified using flux
18
19 299 variability analysis, were excluded from the background set of Recon 2 reactions. Pathway
20
21 300 enrichment analyses were also performed for the sets of highly or not expressed genes.

22
23
24 301 **Assessing the benefits of the developed approach: comparison with predictions made from**
25
26 302 **transcriptomic data only**

27
28 303 To assess the interest of using the network topology and stoichiometry to predict the metabolic
29
30 304 modulations occurring during the differentiation process, we performed pathway enrichment
31
32 305 analyses on genes identified as up or downregulated between the two stages. Genes were
33
34 306 considered as upregulated if they were classified as NE at d3 and HE at d30 and inversely,
35
36 307 downregulated, if they were classified as HE at d3 and NE at d30.

37
38
39
40 308 **Assessing the benefits of the developed approach: comparison with predictions made from**
41
42 309 **single solutions**

43
44 310 We compared our predictions with predictions that would be obtained when considering only
45
46 311 one individual optimal subnetwork instead of a set of optimally adequate subnetworks for each
47
48 312 stage. For each stage, we selected two distinct individual solutions “iMat-A” and “iMat-B” among
49
50 313 all the equally adequate generated subnetworks. “iMat-A” is the first initial optimal solution,
51
52 314 corresponding to the one returned by the iMat algorithm such as implemented in the CobraToolbox.
53
54 315 “iMat-B” is one other randomly selected solution among all optimal solutions. We also compared
55
56
57
58
59
60

1
2
3 316 our results with results obtained by using the FASTCORE algorithm ³⁹, which provides with one
4
5 317 particular solution minimizing the number of active reactions ("FASTCORE solution"). For each
6
7 318 stage, we used the set of reactions associated with highly expressed genes as input for the "core
8
9 319 reactions set" in the FASTCORE algorithm, which then returns a minimal flux consistent
10
11 320 subnetwork containing all the reactions from this core set and a minimal set of additional reactions.
12
13
14 321 To compare individual solutions to the whole set of subnetworks, we considered the union of
15
16 322 subnetworks. For liver-specific genes, we considered all the genes that were predicted to be
17
18 323 expressed in any of the cell-specific subnetworks. For reactions predicted to be activated in the
19
20 324 union of equally adequate subnetworks, we considered all the reactions that were predicted to be
21
22 325 active in any of the 30-day cell-specific subnetworks but inactive in all 3-day cell-specific
23
24 326 subnetworks. Inversely for reactions predicted to be inactivated in the union of subnetworks, we
25
26 327 considered all the reactions that were predicted to be inactive in any of the 30-day subnetworks but
27
28 328 active in all 3-day subnetworks.
29
30
31

32 **Data availability**

33
34
35 330 Gene expression data have been deposited in the Gene Expression Omnibus (GEO) database
36
37 331 under the accession number GSE112123. NMR metabolomic data are provided as Supporting
38
39 332 Information (Table S2).
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

334 RESULTS

335 In this study, we aimed at comparing the differences in metabolic functions expressed in 3-day
336 (non-differentiated) progenitor HepaRG cells ("3-day cells") vs. 30-day fully differentiated
337 hepatocyte-like HepaRG cells ("30-day cells"), using a large-scale modelling approach. To that
338 purpose, we defined consistent metabolic models that distinctively represent the functional
339 metabolism of 3-day cells and 30-day cells. We used the generic genome scale metabolic
340 reconstruction Recon 2⁵ as a scaffold to integrate transcriptomic and metabolomic data and
341 predicted which of the Recon 2 reactions would be specifically active or inactive at each
342 differentiation stage in HepaRG cells. Prediction of the reaction activities was computed using the
343 iMat algorithm proposed by Shlomi *et al.*¹⁰ that we adapted to fit the specific objectives of our
344 study. Figure S2 illustrates the pipeline followed in this study.

345 **Experimental transcriptomic and metabolomic data provide only partial coverage of the** 346 **metabolic network**

347 For each differentiation stage, we identified expressed (HE) or not expressed (NE) genes from
348 the experimental data set and transferred this information to reactions using the GPR associations.
349 As expected, only a small proportion (9%) of the genes classified as HE or NE from the
350 experimental data set could be mapped into Recon 2, these mapped genes representing 32% of the
351 metabolic genes currently annotated in Recon 2 (Figure S3A&B). According to GPR associations,
352 29-31% of Recon 2 reactions could be linked to gene expression data (Figure S3C). More precisely,
353 16% and 21% of Recon 2 reactions were associated with HE genes at day 3 (d3) and day 30 (d30),
354 respectively, whereas 13% (d3) and 10% (d30) were associated with NE genes. Of note, the relation
355 between reactions and genes is not bijective: one reaction can be linked to several distinct genes,
356 for instance when several isoenzymes are involved, and conversely one gene product can control
357 more than one reaction. For instance, for the carnitine palmitoyltransferase 1A gene (CPT1A,

1
2
3 358 hsa:1374), the encoded protein can catalyze the transfer of the acyl group of different long-chain
4
5 359 fatty acids (FAs) onto carnitine. In parallel, NMR analyses allowed the detection and identification
6
7 360 of 36 and 34 metabolites in 3-day and 30-day HepaRG cells, respectively, out of which 34 (d3) and
8
9 361 32 (d30) metabolites could be mapped into Recon 2 (Table S2).

10
11
12 362
13 363 **Reconstruction of stage-specific cell models from experimental data: multiple possible**
14
15
16 364 **subnetworks for 3-day and 30-day cells**

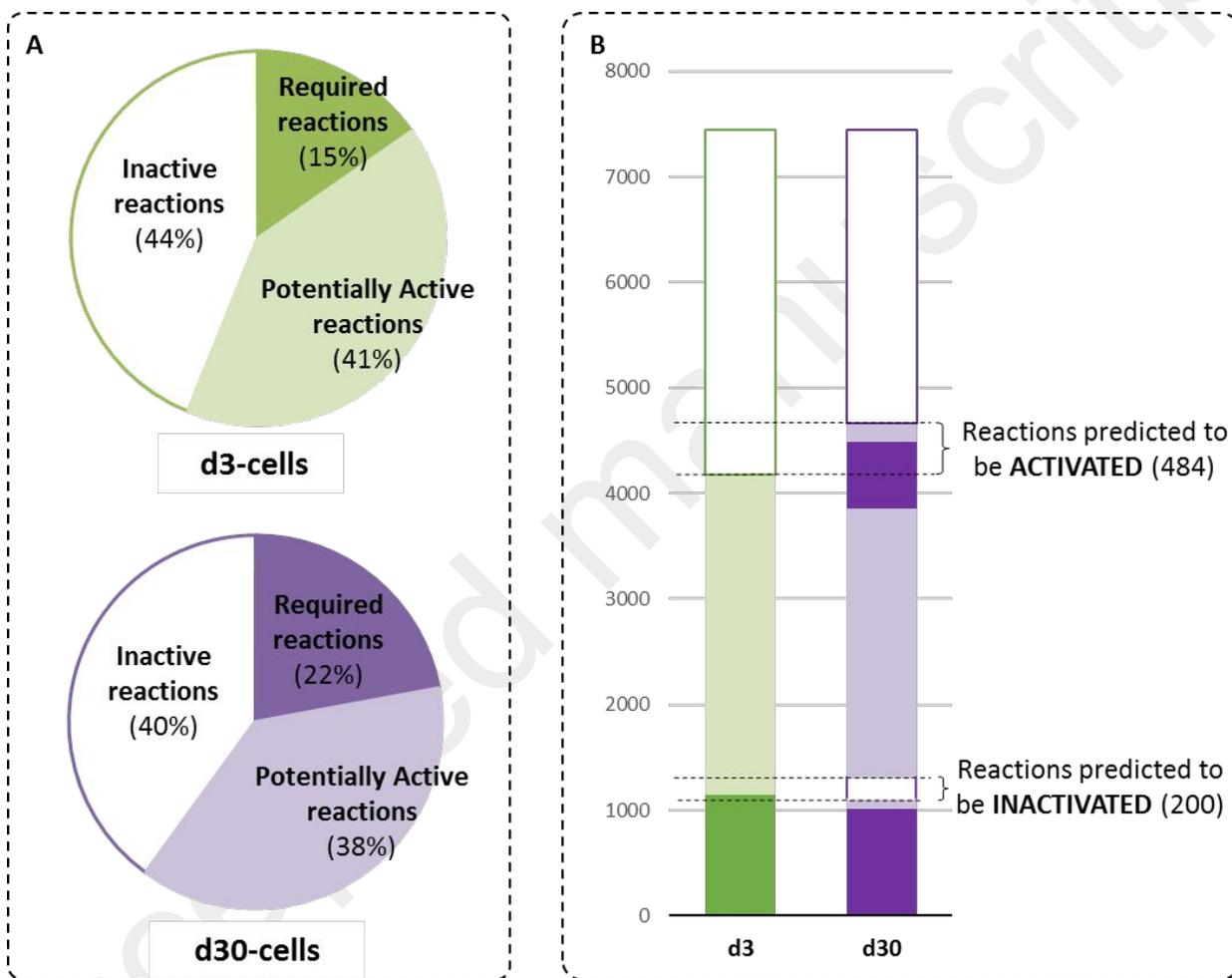
17
18 365 The cell-specific metabolic network reconstruction algorithm uses gene expression levels as
19
20 366 clues to predict the activity state of the associated reactions. It relies on CBM methods to find
21
22 367 metabolic flux distributions, which (1) maximize the number of reactions whose predicted flux is
23
24 368 consistent with associated gene expression level and (2) comply with the stoichiometric constraints
25
26 369 imposed by the network topology. We adapted the iMat algorithm so that the predicted flux
27
28 370 distributions would also take into account metabolomics data (see Material and Methods).

29
30
31 371 We identified several flux distributions having a similar adequacy with the experimental gene
32
33 372 expression data, as assessed by the "adequacy score" (*i.e.*, the proportion of reactions whose
34
35 373 predicted activity state is in adequacy with the transcriptomic data). One predicted flux distribution
36
37 374 corresponds to a set of active reactions (*i.e.*, reactions carrying a non-zero flux) representing a
38
39 375 consistent and fully connected subnetwork. We identified 3534 distinct subnetworks having a
40
41 376 maximal adequacy score of 75% for the 3-day cells, and 3313 distinct subnetworks with an
42
43 377 adequacy score of 72% for the 30-day cells (Table S3). These two sets of equally optimal
44
45 378 subnetworks will be subsequently referred to as "cell-specific models". As imposed by the
46
47 379 modeling constraints, all distributions enable the production of the set of NMR identified
48
49 380 metabolites. We tested that no better adequacy with transcriptomic data could be obtained when
50
51 381 releasing the additional constraints imposed from metabolomic data (see Material & Methods for
52
53
54
55
56
57
58
59
60

1
2
3 382 details), which means that there was no specific disagreement between transcriptomic and
4
5 383 metabolomic data. The number of predicted active reactions was significantly higher in 30-day
6
7 384 cell-specific models than in 3-day cell-specific models, ranging from 2842 to 3036 at d30 (median:
8
9 385 2931) and from 2445 to 2673 (median: 2567) at d3 (Figure S4). The predictions of reactions activity
10
11 386 for all subnetworks are provided in Table S4.
12
13

14 387 Rather than selecting only one arbitrary subnetwork, we chose to take into account all the
15
16 388 alternative solutions we found to be equally adequate with the experimental data for each
17
18 389 differentiation stage. Reactions that were consistently predicted to be active in all subnetworks for
19
20 390 one differentiation stage were considered as "required reactions" (R) whereas reactions whose
21
22 391 predicted activity varied across subnetworks were considered as "potentially active" (PA) (Figure
23
24 392 S2B). 56% and 60% of Recon 2 reactions were globally predicted to be active, either PA or R (*i.e.*,
25
26 393 active in at least one subnetwork), at d3 and at d30 respectively. At d30, 37% of these active
27
28 394 reactions were predicted to be required, whereas a slightly lower proportion of R reactions was
29
30 395 found at d3 (27%) (Figure 1A). We identified two main causes contributing to the high number of
31
32 396 alternative possible subnetworks. The first one is the lack of gene expression information for some
33
34 397 reactions. Globally, we observed that reactions with a variable predicted activity state (PA
35
36 398 reactions) are associated with gene expression data in a higher proportion (2%) than reactions with
37
38 399 consistent predicted activity among stage-specific models (R reactions; 58%). The second one is
39
40 400 the existence of several alternative reactions performing the same metabolic transformation. For
41
42 401 instance, some biochemical conversions happen to be described by two alternative paths that use a
43
44 402 different number of reaction steps or the same reactions are sometimes duplicated with more than
45
46 403 one tissue-specific annotation. As a striking example, in the N-glycan synthesis pathway of the
47
48 404 Recon 2 network, two alternative paths of 22 reactions enable the synthesis of the glycan precursor
49
50 405 before it is bound to a protein in the endoplasmic reticulum (ER) and further processed in the ER
51
52
53
54
55
56
57
58
59
60

406 and Golgi apparatus. The two paths only differ by the fact that the reactions are annotated as liver
 407 reactions ("_L") in one and as uterine reactions ("_U") in the other. These 44 reactions have been
 408 identified through Principal Component Analysis (PCA) as the reactions that contribute the most
 409 to the discrimination between subnetworks within each set of stage-specific subnetworks (Figure
 410 S5) demonstrating that path redundancy can contribute to the variability among subnetworks.



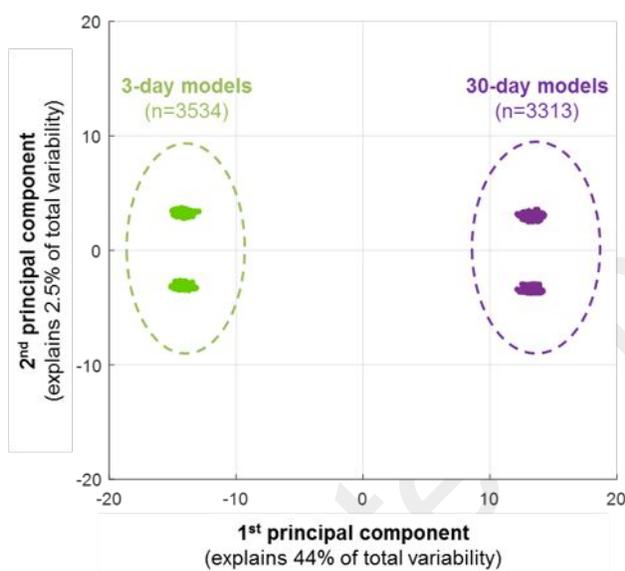
411

412 **Figure 1.** Predicted activity of Recon 2 reactions in stage-specific subnetworks.

413

414 **Comparison of stage-specific subnetworks: evidences for global metabolic changes during**
 415 **the differentiation process**

1
2
3 416 To investigate which significant metabolic changes occur between 3-day and 30-day HepaRG
4
5 417 cells, we compared the cell-specific models generated for each stage. We first performed an
6
7 418 unsupervised discriminant analysis (PCA) of all the cell-specific models together. The score plot
8
9 419 of the PCA analysis showed a clear separation between the two stages, with more than 40% of the
10
11 420 variability explained by the differentiation stage represented along the first axis (Figure 2). This
12
13 421 demonstrates not only that there are metabolic differences between these two stages, but also that
14
15 422 these differences exceed the differences between the subnetworks within each stage, which is
16
17 423 represented by the second axis and accounts for only 2.5% of the total variability.
18
19
20
21



424
425 **Figure 2.** Principal component analysis of 3-day and 30-day subnetworks

426 We identified reactions that are activated during the differentiation process, by considering the
427 reactions predicted to be inactive in all d3 cell-specific networks but active in at least one of the
428 30-day cell-specific model (*i.e.*, PA or R at d30). Conversely, reactions predicted to be active at d3
429 but inactive at d30 were identified as inactivated during differentiation (Figure 1B). According to
430 our predictions, 484 reactions were activated and 200 reactions inactivated during HepaRG cell
431 differentiation. These reactions were further analyzed through pathway enrichment. Activated

432 reactions were significantly over-represented in known hepato-specific pathways (bile acid
 433 synthesis and cytochrome metabolism) but also in the following pathways: biotin metabolism, FA
 434 oxidation and tryptophan metabolism. The biotin metabolism pathway acts as a cycle, where the
 435 biotin binds to carboxylase enzymes (biotinylation), enabling their activation. Biotin is then
 436 released by hydrolysis, to be recycled and re-used for coenzyme activity. Carboxylase enzymes
 437 activated by biotinylation are involved in many important cellular functions, related in particular
 438 to the production and breakdown of proteins, fats, and carbohydrates and to urea metabolism ⁴⁰,
 439 which are indeed essential in fully functional hepatic cells. The essential amino acid tryptophan is
 440 mainly metabolized in the liver through the kynurenine pathway, which accounts for about 95% of
 441 its degradation in normal physiological conditions ⁴¹. The prediction of the activation of this
 442 pathway in mature hepatocytes is consistent with the observation of the increased activity of
 443 enzymes involved in this pathway during the development in rat liver ⁴². Conversely, the activity
 444 of the extracellular transport pathway and of the FA synthesis pathway was predicted to be lower
 445 in 30-day HepaRG cells (Table 1). Our results therefore suggest a shift in the metabolism of FAs,
 446 with a decrease in the FA synthesis and a concomitant increase in FA oxidation in differentiated
 447 cells.

448 **Table 1.** Predictions of metabolic pathways significantly modulated between d3 and d30

ACTIVATED pathways			INACTIVATED pathways		
Pathways	% pathway reactions	corrected p-value	Pathways	% pathway reactions	corrected p-value
Biotin metabolism	100	1.31e ⁻¹¹	Transport, extracellular	9.6	5.60e ⁻³¹
Fatty acid oxidation	15.9	6.80e ⁻⁰⁷	Fatty Acid Synthesis	45.2	1.30e ⁻¹⁵
Bile acid synthesis	27.9	5.13e ⁻⁰⁵			

Tryptophan metabolism	40.0	2.44e ⁻⁰⁴
Cytochrome metabolism	57.1	3.76e ⁻⁰⁴
Blood group synthesis	30.4	1.59e ⁻⁰³
Lysine metabolism	50.0	1.44e ⁻⁰³
Limonene & pinene degradation	100	3.45e ⁻⁰²

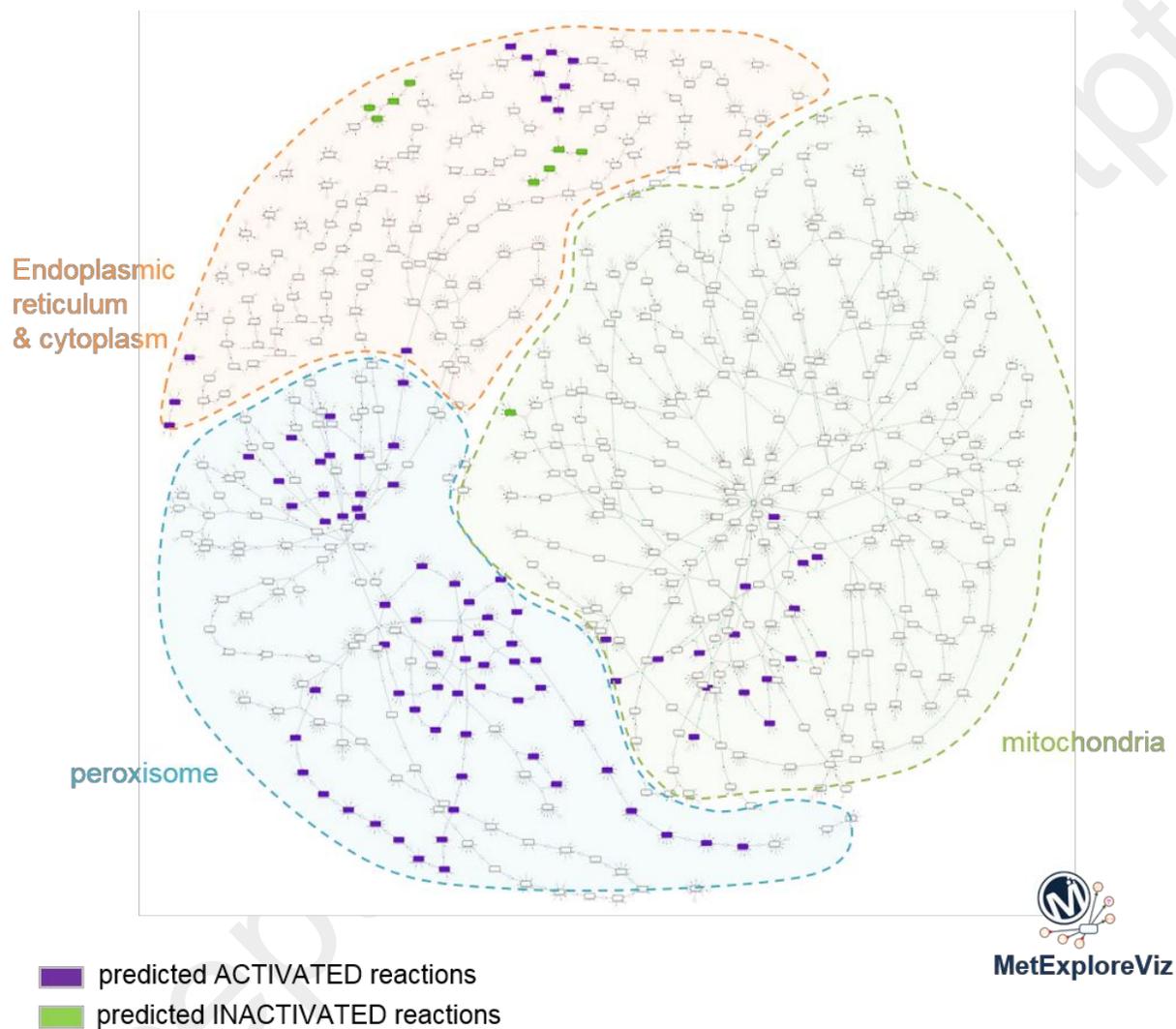
449

450 **Visualization-based mining strengthens metabolic activity comparisons**

451 Although pathway enrichment analysis offers a first global view of the pathways modulated
452 during the differentiation process, some of these pathways, such as the FA synthesis and oxidation
453 pathways, include a very large number of reactions (126 and 868 reactions, respectively). A deeper
454 examination of the reactions modulated within these pathways was carried out using the
455 MetExplore web server for visualization^{38,43}. We observed that FA oxidation reactions predicted
456 to be activated between d3 and d30 were specifically located in the peroxisome, whereas most
457 mitochondrial FA oxidation reactions were predicted to be active at both differentiation stages
458 (Figure 3). Notably, peroxisomal FA oxidation allows the specific degradation of very long-chain
459 FAs (>C₂₀) whereas the shortened FAs are further oxidized in the mitochondria⁴⁴. Our results
460 therefore imply that increased peroxisomal FA oxidation induced by drugs or FA overload⁴⁵ cannot
461 occur in progenitor cells. Although limited peroxisomal FA oxidation can reduce energy
462 production, it primarily induces the accumulation of toxic very long-chain FA metabolites⁴⁶.
463 Regarding the FA synthesis pathway, the visualization highlighted that reactions predicted to be
464 inactivated between d3 and d30 were more specifically involved in the elongation of the carbon
465 chain, suggesting that the synthesis of long-chain FAs cannot be achieved anymore in fully

1
2
3 466 differentiated hepatocytes. However, transformation of FAs or acyl-CoA, through desaturation or
4
5 467 hydrolysis reactions respectively, can be performed at both stages of development (Figure S6).

6
7
8 468



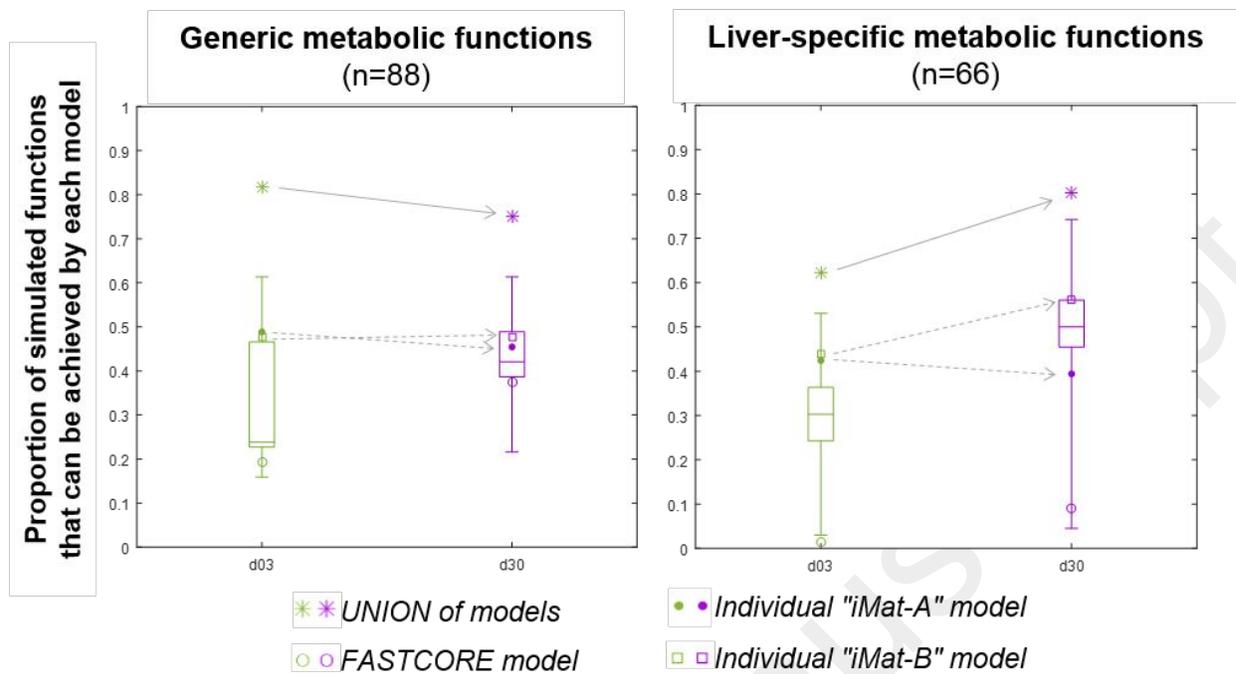
469

470 **Figure 3.** Visualization of predicted modulated reactions in the fatty acid oxidation pathway

471

472 **Analysis of flux-consistent cell-specific models reveals functional metabolic changes during**
473 **the differentiation process**

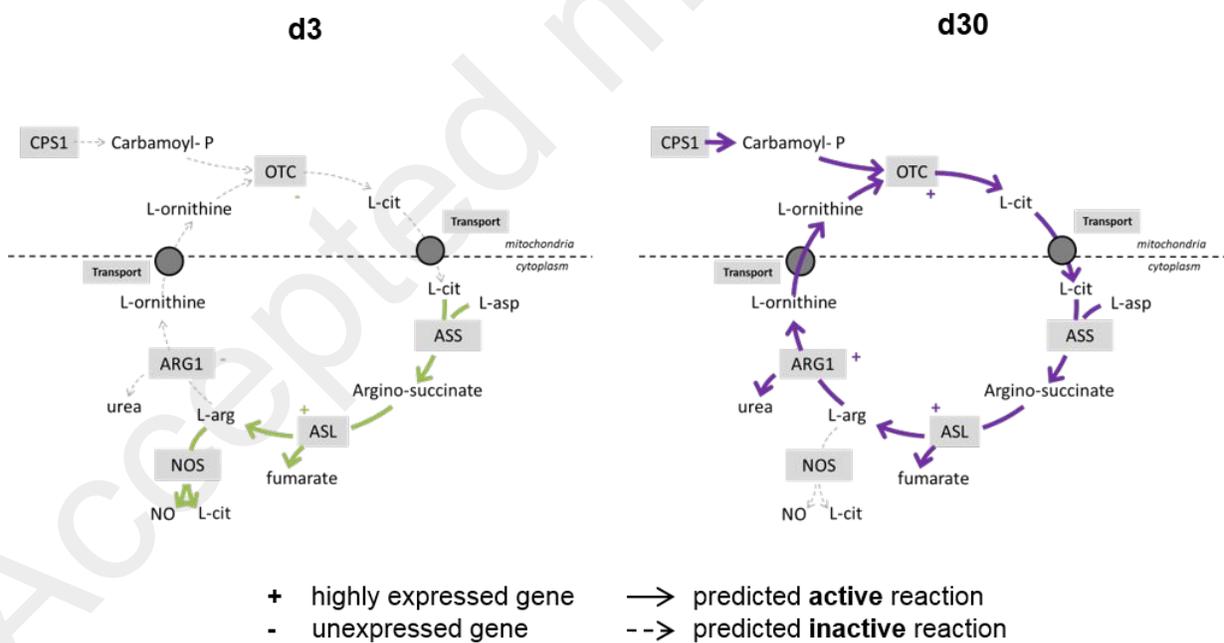
1
2
3 474 To assess whether the generated 3-day and 30-day cell-specific models actually accounted for
4
5 475 distinct liver-specific metabolic functionalities, we challenged the capacity of each cell-specific
6
7 476 subnetwork to perform 154 defined metabolic functions ^{4,36,37} (Table S1). Out of this list, 88
8
9 477 functions are generic functions, which can theoretically be achieved by any cell type, and 66 are
10
11 478 hepato-specific metabolic functions known to specifically take place in liver, such as ammonia
12
13 479 detoxification (through ureagenesis), ketogenesis, biliary acid formation, or gluconeogenesis. On
14
15 480 average, a higher number of tested metabolic functions, and especially of liver-specific functions,
16
17 481 could be achieved by 30-day cell-specific models compared to 3-day models: 50% vs. 31% for
18
19 482 hepatic functions and 42% vs. 33% for generic functions (Figure 4). Similarly, when comparing
20
21 483 the set of all metabolic functions that can be fulfilled by the whole set of models at each stage (*i.e.*,
22
23 484 "UNION" of models), we observed that a larger range of hepatic functions can be fulfilled at 30
24
25 485 days than 3 days. More specifically, our models predict that 30-day cells are able to degrade a
26
27 486 wider range of amino acids and to produce the ketone body β -hydroxybutyrate compared to 3-day
28
29 487 cells (Table 2). Interestingly, this end product of mitochondrial FA oxidation presents different
30
31 488 cellular signaling functions in addition to its role in energy production ⁴⁷. Regarding the
32
33 489 gluconeogenesis capacities, which were assessed by testing the ability of the models to account for
34
35 490 the production of glucose in the absence of carbohydrate sources (*e.g.*, from non-carbohydrates
36
37 491 metabolites including lactate, pyruvate, glycerol and glucogenic amino acids), results indicate that
38
39 492 gluconeogenesis can theoretically be carried out by HepaRG cells at both stages, although by a
40
41 493 higher proportion of cell-specific models at d30 than d3. In agreement with the acquisition of liver-
42
43 494 specificity over the differentiation process, our models predict that only 30-day cells can produce
44
45 495 urea from the degradation of AA (Table 2 & Figure 5) and can synthesize bile acids such as
46
47 496 glycocholate, glycochenodeoxycholate and taurocholate.
48
49
50
51
52
53
54
55
56
57
58
59
60



497

498 **Figure 4.** Simulations of metabolic functions in d3 and d30 models.

499



500

501 **Figure 5.** Predicted activities for reactions in the urea cycle502 **Table 2.** Comparison of liver-specific metabolic functionalities between 3-day and 30-day models

	3-day	30-day
Bile acid formation	1 ^a / 6 ^b [0 – 2] ^c	2 / 6 [0 – 6]
Ureogenesis	0 / 2 [0 – 0]	2 / 2 [0 – 2]
gluconeogenesis	4 / 9 [0 – 9]	6 / 9 [0 – 9]
ketogenesis	0 / 4 [0 – 2]	2 / 4 [0 – 4]
Amino acid degradation	12 / 23 [2 – 17]	17 / 23 [2 – 23]

503

504 We further addressed the ability of progenitor and fully differentiated HepaRG models to account

505 for liver-specificity acquisition, by assessing the proportion (1) of liver-specific genes that were

506 included in each cell-specific subnetwork (based on UNIPROT database annotations, see Material

507 & Methods) and (2) of reactions associated with high confidence level for protein expression in

508 liver (based on the Human Proteome Atlas database information on tissue protein expression, see

509 Material & Methods). We identified that 46% of the Recon 2 genes had evidence for expression in

510 liver. On average, a higher proportion of these liver-expressed genes was recalled in the generated

511 subnetworks representing the 30-day cells (72 to 84% depending on the subnetworks) compared to

512 3-day cells (64 to 76%), consistent with our expectations (Figure S7). We also obtained a higher

513 recall (and proportion) of reactions expressed in hepatocytes with high confidence level, according

514 to the HPA database, in 30-day HepaRG networks (49 to 53%) compared to 3-day HepaRG

515 networks (36 to 41%) (Figure S8). Globally, these results tend to validate the ability of our

516 developed models to account for the hepatic specialization occurring during cell-differentiation,

517 when considering conjointly all the cell-specific subnetworks identified for each stage.

518

519 DISCUSSION

1
2
3 520 The metabolic status of many *in vitro* models still requires to be duly characterized and this is
4
5 521 especially true for models used to study hepatic metabolic diseases, most of which imply, to a
6
7 522 certain degree, hepatic regeneration mechanisms. In this study, we propose a methodology, based
8
9
10 523 on large-scale modelling approaches, which enables to point out functional shifts in the global
11
12 524 metabolism of cells and apply it to compare the metabolic capacity of hepatic cells at two distinct
13
14 525 differentiation stages (progenitor *vs.* fully differentiated cells). Our developed methodology allows
15
16 526 gaining a better understanding of metabolic shifts as it can unveil critical information not directly
17
18
19 527 available from experimental data such as gene expression or metabolomics. Transcriptomic data
20
21 528 often provide a restricted coverage of the generic metabolic network genes. In our study, gene
22
23 529 expression data covered only 31% and 32% of Recon 2 reactions at d3 and d30 respectively when
24
25
26 530 considering the genes present in the initial transcriptomic dataset (Figure S3). The problem related
27
28 531 to the partial representation of metabolic network genes with commonly used microarrays is an
29
30
31 532 acknowledged fact, with a coverage that varies depending on the type of microarray used ⁴⁸. It
32
33 533 should also be noted that the partial coverage of metabolic reactions is partly due to a lack of gene
34
35 534 related information for some reactions in the network (35% in Recon 2): these reactions would not
36
37
38 535 be covered whatever the transcriptomic dataset used.. Having such scattered transcriptomic
39
40 536 information does not allow to get an overall picture of the metabolic changes occurring during
41
42 537 differentiation. By integrating transcriptomic data within the context of a genome-scale metabolic
43
44 538 network and computing flux distributions, one takes into account the fact that metabolites used in
45
46
47 539 a reaction need to be produced and consumed, in stoichiometric proportions, by other reactions in
48
49 540 the network. Therefore, the activity of a reaction is not only predicted from the expression level of
50
51 541 its associated gene but can also be inferred from the activity of downstream and upstream reactions
52
53
54 542 in the network. In our study, 79.2% and 76.5% of the reactions predicted to be active at d3 and d30
55
56 543 (42% of the reactions if solely considering “required” reactions), respectively, were not associated
57
58
59
60

1
2
3 544 with any gene expression data and therefore their activity could be newly predicted thanks to the
4
5 545 computational method (Figure S9). A large proportion of these reactions belongs to the transport
6
7 546 and exchange pathways, which are less well annotated in term of associated genes, but are,
8
9
10 547 nevertheless, key players in cellular and tissue biology and absolutely need to be taken into account
11
12 548 to obtain fully flux-consistent networks. Interestingly, the biotin and tryptophan metabolism
13
14 549 pathways, for which gene expression data was available for only a small proportion of the reactions
15
16 550 (33-53%), could not be predicted as activated on the sole basis of transcriptomic data but their
17
18 551 predicted activation was inferred by taking into account the network structure and the connections
19
20 552 between reactions (Table S5). On the contrary, a few pathways such as "xenobiotics metabolism",
21
22 553 "androgen and estrogen synthesis and metabolism" and "arachidonic acid metabolism" were not
23
24 554 predicted as activated, although they would be evidenced as activated if basing solely on gene
25
26 555 expression data. This is because most of the reactions in these pathways are blocked reactions
27
28 556 (Table S5), meaning that their products cannot be further consumed or their substrates cannot be
29
30 557 produced. The Recon 2 network contains almost 30% of "blocked" reactions. Such reactions are
31
32 558 always predicted to be inactive when using constraint-based modeling approaches, and can
33
34 559 generate false negatives and "artificial" divergences between gene expression data and model
35
36 560 predictions that do not have any actual biological meaning. For instance, we observed that 7% (at
37
38 561 d3) and 8% (at d30) of the reactions are predicted as inactive although they are associated with HE
39
40 562 genes (Figure S9), but 95 to 99% of these reactions are actually blocked reactions, meaning that
41
42 563 the observed divergence cannot, in this case, be interpreted as a post-regulation process. More
43
44 564 interestingly, when not taking into account blocked reactions, about 1% of the reactions have a
45
46 565 predicted activity state that is not in accordance with their associated gene expression data. The
47
48 566 discrepancy between data and predictions might also be linked to the presence in the network of
49
50 567 promiscuous genes, associated with many different reactions (*e.g.*, the gene EHHADH or hsa:1962
51
52
53
54
55
56
57
58
59
60

1
2
3 568 controls 194 reactions). Enforcing a strict translation of the expression or non-expression of such
4
5 569 genes to the reaction level would amount to consider that all proteins encoded by a single gene are
6
7
8 570 active (or inactive, respectively) when this gene is expressed (or not/lowly expressed, respectively),
9
10 571 which is not biologically relevant. Using the iMat algorithm prevents from doing this strict
11
12 572 association, but might consequently generate some discrepancies between predictions and data. In
13
14 573 our study, most reactions whose predicted activity is in opposition with their gene expression level
15
16 574 belong to extracellular transport and FA oxidation pathways, in which a few genes (respectively
17
18 575 47 and 93) are associated with a large number of reactions (361 and 709 respectively).
19
20

21 576 Conventional transcriptomic analyses result in lists of identified expressed genes, which are
22
23 577 commonly interpreted using pathway enrichment analyses, providing information about which
24
25 578 parts of the network are more likely to be differentially expressed in specific experimental
26
27 579 conditions ^{49,50}. However, the definition of metabolic pathways is very dependent on the database
28
29 580 or network used and is not always relevant regarding functional metabolism. Indeed, some
30
31 581 metabolic functions span over several metabolic pathways, while some metabolic pathways are
32
33 582 very generic and encompass several distinct metabolic functions ⁵¹. Thus, one may miss a specific
34
35 583 or biologically relevant functionality when relying on these global pathways for interpretation.
36
37 584 Generating flux-consistent models of the functional metabolic network of cells offers the clear
38
39 585 advantage of going beyond the analysis of metabolic pathways by enabling a more holistically
40
41 586 identification of modulated metabolic functions and therefore providing a finer assessment of the
42
43 587 metabolic capacities of a system. For instance, in our study, we predicted that ureagenesis is
44
45 588 activated during the differentiation process, although the "urea cycle" pathway was not evidenced
46
47 589 as activated if relying only on the pathway enrichment analysis (p-value = 1). Indeed, in Recon 2,
48
49 590 the "urea cycle pathway" encompasses 69 reactions, but only a few of them are truly involved in
50
51 591 urea synthesis (other reactions contribute to creatine or spermidine metabolism) so that the
52
53
54
55
56
57
58
59
60

1
2
3 592 predicted activation of these reactions was masked by the lack of changes of many other reactions
4
5 593 in this pathway.

6
7 594 The originality of the approach we propose in this study mainly relies on the fact that we chose
8
9
10 595 not to arbitrarily select one specific model among all equally optimal cell-specific models
11
12 596 identified, but rather to consider them conjointly for each differentiation stage. We observed that
13
14 597 the existence of a high number of alternative possible subnetworks is partly due to the existence of
15
16 598 redundant reactions in the initial generic reconstruction that often reflect artifactitious redundancies
17
18 599 rather than a true biological alternative. Manual curation of these redundancies would contribute
19
20 600 to decrease the number of computed equally adequate stage-specific models and therefore reduce
21
22 601 the uncertainty in the prediction of reaction activity. Despite this well identified issue, we showed
23
24 602 that the comparison of sets of several similarly adequate subnetworks for two distinct conditions
25
26 603 still enables evidencing stage-dependent differences, as stage-dependent differences contributed
27
28 604 more to the discrimination among all subnetworks than the variability between the subnetworks at
29
30 605 each stage (Figure 2). Other methods used to generate cell-specific metabolic models have adopted
31
32 606 a different strategy, which consists in selecting one specific solution which optimizes an additional
33
34 607 criteria, such as minimizing the total sum of fluxes in the network (Euclidean FBA), maximizing
35
36 608 the biomass production or cell growth ⁵², minimizing the number of active reactions ³⁹ or being
37
38 609 able to achieve a set of predefined metabolic tasks ¹⁴. This entails to make some additional and *a*
39
40 610 *priori* assumptions about the "optimal" metabolic state of the cells or some target metabolic
41
42 611 functions. This can be relevant in the case of rapidly growing cells, such as cancer cells. However,
43
44 612 it is much less relevant for cells that are not in a permanent proliferative status and are able to
45
46 613 achieve a wide span of metabolic activities, such as liver cells. We assessed the benefit of our
47
48 614 approach by comparing our results with results obtained when considering only one subnetwork
49
50 615 for each stage, either arbitrarily selected among all the subnetworks identified by the iMat
51
52
53
54
55
56
57
58
59
60

1
2
3 616 algorithm (iMat-A and iMat-B, in Figure 4 & Figures S7 & S8) or generated using the FASTCORE
4
5 617 algorithm ³⁹ (see Material and Methods). When considering single solutions, we observed that
6
7 618 some metabolic pathways, with little gene expression information, appeared as significantly
8
9 619 enriched in activated (*e.g.*, ubiquinone synthesis) or inactivated reactions (*e.g.*, keratan sulfate
10
11 620 degradation, N-glycan synthesis) solely as a result of arbitrariness in the choice of the subnetwork
12
13 621 at each stage (Table S6), whereas the method developed in this work prevents from making such
14
15 622 biased predictions. For instance, the ubiquinone synthesis pathway (Figure S10) was predicted as
16
17 623 activated when using the FASTCORE algorithm or considering only one iMat solution. Because
18
19 624 no gene-expression information is available at d3 for any of the reactions in this pathway, all these
20
21 625 reactions can be predicted to be either active or inactive at d3. As a result and because reactions
22
23 626 are necessarily active at d30, whatever the selected subnetwork, it may then be concluded that the
24
25 627 pathway is activated if reactions are predicted to be inactive in the retained d3-subnetwork, which
26
27 628 is what happens when using the FASTCORE algorithm that aims to minimize the number of
28
29 629 reactions (Figure S10C&D). Conversely, with the strategy we implemented in our method, which
30
31 630 takes into consideration all the possibilities of reactions activity at d3, we do not predict that this
32
33 631 pathway is activated (Figure S10B), retaining a more cautious conclusion. The same interpretations
34
35 632 can be made for the N-glycan synthesis and keratin sulfate degradation pathways (Figures S11 &
36
37 633 S12 and Supplementary Notes). The FASTCORE algorithm provides one unique minimal
38
39 634 subnetwork, which contains all the reactions associated with highly expressed genes and a minimal
40
41 635 set of additional reactions required in order to get a flux-consistent network. Although it might
42
43 636 reduce the bias in the comparison of the conditions by providing a deterministic way to define one
44
45 637 unique model for each condition, we suspect that these minimal networks are more dependent on
46
47 638 the transcriptomic data and are likely biased toward assuming that reactions are not active when
48
49 639 they have no associated expression data. In addition, it is questionable whether the metabolic state
50
51
52
53
54
55
56
57
58
59
60

1
2
3 640 involving a minimal number of biochemical reactions as computed by the FASTCORE algorithm
4
5 641 is truly a good picture of the actual metabolic capacity of a mammalian cell. We reached the
6
7 642 conclusion that, in the case of insufficient experimental data, arbitrarily selecting one possible
8
9 643 subnetwork could lead to erroneous conclusions and we argue that our method, which takes into
10
11 644 account and compares sets of equally optimal subnetworks, prevents from making potentially false
12
13 645 positive, and therefore erroneous, predictions regarding the activation or inactivation of pathways.
14
15

16
17 646 Getting different results when using different algorithms for identifying cell-specific models is a
18
19 647 well acknowledged issue, which has already been reported in several studies^{53,54}. No method
20
21 648 provides more accurate or exact results than others, and the choice of the method should be made
22
23 649 depending on the available experimental data, as well as the context and the aim of the study.
24
25
26 650 However, the approach that we propose here has the advantage not to require any other information
27
28 651 than transcriptomic data and does not make any biological assumption about the metabolic state or
29
30 652 objective of the cell system. For this reason, it can interestingly be applied to poorly characterized
31
32 653 cell systems. Such approaches are nowadays absolutely relevant and necessary for a sound and
33
34 654 untargeted assessment of metabolic modulations because they allow for an accurate modeling of
35
36 655 cellular metabolic networks out of omics data with a minimal number of *a priori* assumptions and
37
38 656 therefore limit the risk of false-predictions. Also, considering all adequately possible subnetworks
39
40 657 seem relevant under the hypothesis that the variability observed between the distinct possible
41
42 658 subnetworks partly reflects the actual metabolic heterogeneity in a population of cells. Therefore,
43
44 659 taking into consideration this variability should provide a better picture of the whole metabolic
45
46 660 capability of heterogeneous cellular systems.
47
48
49
50

51
52 661
53 662 **CONCLUSION**
54
55
56
57
58
59
60

1
2
3 663 In conclusion, by implementing a new strategy integrating transcriptomic and metabolomic data
4
5 664 within the context of the global human genome-scale metabolic model, we were able to better
6
7 665 characterize the metabolic state of HepaRG cells at two stages of differentiation and to identify
8
9
10 666 some metabolic functions that are set up during the differentiation process of this human hepatic
11
12 667 model. Our predictions are consistent with the known activation of pathways corresponding to
13
14 668 hepato-specific functionalities (bile acid synthesis, cytochrome metabolism) during the
15
16 669 differentiation process, but also newly evidenced the modulation of other metabolic pathways,
17
18 670 involving the metabolism of biotin, tryptophan and FAs, which could not have been evidenced
19
20
21 671 solely from transcriptomic data. The approach we applied in this study offers several advantages
22
23 672 compared to both analyses based on transcriptomic data and common computational approaches
24
25 673 used to identify cell- or tissue-specific metabolic models: it especially allows to point out metabolic
26
27 674 pathways which do not have gene annotations, to go further the analysis of traditionally defined
28
29 675 metabolic pathways by identifying modulated metabolic functions, and to reduce the risk of false
30
31 676 predictions by retaining a set of networks instead of a particular (sometime arbitrarily defined) one.
32
33 677 The result is that it allows a comprehensive and functional assessment of the metabolic capacity of
34
35 678 cells with a minimal number of a priori assumptions. It opens interesting potential for the
36
37 679 comparison of global metabolic shifts occurring during differentiation process of various cell types,
38
39 680 but also in the broader context of long-term metabolic changes, such as induced during the onset
40
41
42 681 of chronic metabolic diseases or chronic exposure to toxics.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

682 FIGURES

683 **Figure 1. Predicted activity of Recon 2 reactions in stage-specific subnetworks.** (A) The
684 activity of Recon 2 reactions was predicted using the iMat algorithm. Several solutions were found
685 and the reactions were classified as "required" (R) if they were found to be active in all solutions
686 and conversely "Inactive" (I) if they display a zero-flux in all the optimal solutions. All other
687 reactions, found to be either active or inactive in the optimal solutions, were classified as
688 "Potentially active" (PA). This analysis was made independently for the d3 and d30 differentiation
689 stages. (B) Reactions identified as I at d3 but R or PA at d30 were considered as activated, whereas
690 reactions identified as R or PA at d3 but I at d30 were considered as activated.

691 **Figure 2. Principal component analysis of 3-day and 30-day subnetworks.** PCA analysis was
692 performed on the whole set of equally adequate subnetworks obtained for both stages (n=6847).
693 Reactions were used as variables. Each subnetwork was represented as a binary vector with 0
694 values for inactive reactions and 1 values for active reactions.

695 **Figure 3. Visualization of predicted modulated reactions in the fatty acid oxidation pathway.**
696 Reactions belonging to the fatty acid oxidation pathway in Recon 2 were extracted and visualized
697 using the MetExplore webserver (Chazalviel *et al.*, 2018). Reactions predicted to be activated or
698 inactivated between the 2 differentiation stages were mapped onto the pathway.

699 **Figure 4. Simulations of metabolic functions in 3-day and 30-day models.** Metabolic functions
700 were simulated for each of the cell-specific models identified at each stage and for the model
701 obtained using the FASTCORE algorithm. Boxplots represent the distribution of the proportion of
702 simulated functions that can be achieved by 3-day and 30-day cell specific models for all generic
703 (left panel) or all liver-specific (right panel) metabolic functions. UNION of cell-specific models,

704 represented by a star, corresponds to the proportion of functions that can be achieved by at least
 705 one of the cell-specific model. The iMat-A and iMat-B points correspond to the proportion of
 706 functions that can be achieved by these 2 randomly selected subnetworks individually. Arrows
 707 represent the sense of variation between d3 and d30, which can be different depending on which
 708 subnetworks we compare.

709 **Figure 5. Predicted activities for reactions in the urea cycle.** Reactions predicted to be active
 710 are represented with green plain arrows for d3 and purple plain arrows for d30. "+" and "-" signs
 711 indicate reactions associated with highly or unexpressed genes. The following reactions belonging
 712 to the urea cycle are represented (names in italic correspond to the reaction identifiers in Recon2):
 713 CPS1, carbamoyl-phosphate synthase (*CBPSam* & *r0034*); OTC, ornithine carbamoyltransferase
 714 (*OCBTm*); ASS, argininosuccinate synthase (*ARGSS*); ASL, argininosuccinate lyase (*ARGSL*);
 715 ARG1, arginase (*ARGN*); NOS, L-Arginine,NADPH:oxygen oxidoreductase (NO-forming)
 716 (*r0145*); ornithine mitochondrial transport exchange with citruline (ORNt4m) (grey circle).

717 TABLES

720 **Table 1.** Predictions of metabolic pathways significantly modulated between d3 and d30

ACTIVATED pathways		INACTIVATED pathways	
Pathways	corrected p-value	Pathways	corrected p-value
Biotin metabolism	$1.31e^{-11}$	Transport, extracellular	$5.60e^{-31}$
Fatty acid oxidation	$6.80e^{-07}$	Fatty Acid Synthesis	$1.30e^{-15}$
Bile acid synthesis	$5.13e^{-05}$		
Tryptophan metabolism	$2.44e^{-04}$		
Cytochrome metabolism	$3.76e^{-04}$		

Blood group synthesis 1.59e⁻⁰³

Lysine metabolism 1.44e⁻⁰³

Limonene & pinene degradation 3.45e⁻⁰²

721 Pathway enrichment analysis was performed on set of reactions identified as inactivated (n=200)
722 or activated (n=484) between 3-day and 30-day cell models, after removing blocked reactions. P-
723 values were obtained by performing a hypergeometric test followed by a Bonferroni correction.

724
725 **Table 2.** Comparison of liver-specific metabolic functionalities between 3-day and 30-day models

	d3	d30
Bile acid formation	1 ^a / 6 ^b [0 – 2] ^c	2 / 6 [0 – 6]
Ureogenesis	0 / 2 [0 – 0]	2 / 2 [0 – 2]
gluconeogenesis	4 / 9 [0 – 9]	6 / 9 [0 – 9]
ketogenesis	0 / 4 [0 – 2]	2 / 4 [0 – 4]
Amino acid degradation	12 / 23 [2 – 17]	17 / 23 [2 – 23]

726 Results from liver-specific metabolic functions simulated for each of the cell-specific models
727 identified at each stage: for each type of function, values in bold (a) indicate the average number
728 of simulated functions that can be achieved out of the total number of tested functions (b). Values
729 in square brackets (c) correspond to the minimal and maximal numbers of simulated functions that
730 can be achieved.

731
732 ASSOCIATED CONTENT

733 **Supporting information.**

734 Principle of the iMat algorithm illustrated on a Toy model (Figure S1); Workflow of analyses
735 implemented for comparison of stage-specific subnetworks (Figure S2); Distribution of genes
736 expression levels in experimental data set and Recon 2 network (Figure S3); Comparison of
737 predicted active reactions in subnetworks between the 2 stages (Figure S4); Analysis of
738 heterogeneity in 30-day subnetworks (Figure S5); Visualization of predicted modulated reactions

1
2
3 739 in the fatty acid synthesis pathway (Figure S6); Assessment of the liver-specificity of generated
4
5 740 subnetworks: comparison with UNIPROT data (Figure S7); Assessment of the liver-specificity of
6
7 741 generated subnetworks: comparison with Human Proteome Atlas data (Figure S8); Comparison of
8
9
10 742 predicted reaction activity with gene expression data (Figure S9); Comparison between predictions
11
12 743 made from conjoint or individual analysis of cell-specific models for the ubiquinone synthesis
13
14 744 pathway (Figure S10); Comparison between predictions made from conjoint or individual analysis
15
16
17 745 of cell-specific models for the N-glycan synthesis pathway (Figure S11); Comparison between
18
19 746 predictions made from conjoint or individual analysis of cell-specific models for the keratin sulfate
20
21 747 degradation pathway (Figure S12); Adequacy of model predictions with gene expression data
22
23
24 748 (Table S3); Modulated pathways identified using pathway enrichment: comparison between
25
26 749 predictions made from generated models & transcriptomic data (Table S5); Modulated pathways
27
28 750 identified using pathway enrichment: comparison between predictions made from conjoint or
29
30
31 751 individual analysis of cell-specific models (Table S6). Interpretations of activated or inactivated
32
33 752 reactions depending on the selected optimization solution for the Ubiquinone, N-glycan synthesis
34
35 753 and keratin sulfate degradation pathways (Supplementary Notes) (PDF)

36
37
38 754 Metabolic functions used for simulations (Table S1, XLSX)

39
40
41 755 NMR identified metabolites for 3-day and 30-day HepaRG cells (Table S2, XLSX)

42
43
44 756 Computed subnetworks for 3-day and 30-day HepaRG cells from our method and the
45
46
47 757 FASTCORE algorithm (Table S4, XLSX)

48
49
50 758

51
52
53 759 This material is available free of charge via the Internet at <http://pubs.acs.org>.

1
2
3 760
4
5 761 AUTHOR INFORMATION
6
7 762 **Present adress**
8
9
10 763 † SRSMC, UMR 7585, Université de Lorraine, France. ‡INSB Institut des sciences biologiques,
11
12 764 UMR8204 Centre d'infection et d'immunité de Lille, France.

13
14
15 765 **Author contributions**

16
17
18 766 FJ, NP, AC, FMo and MAR designed the research; FMo, HDS, CC, SB, EP and NC performed the
19
20 767 experiments; NP developed the method and performed the computational analysis; FMa, CF and
21
22 768 FV participated in the computational analysis; NP wrote the manuscript; FJ, DZ, BF, NC, AC and
23
24 769 NP contributed in data interpretation, discussion on results and writing and revising the manuscript.
25
26
27

28 770 **Funding source**

29
30 771 This work was supported by the French Ministry of Research and National Research Agency as
31
32 772 part of the French MetaboHUB, the national metabolomics and fluxomics infrastructure (Grant
33
34 773 ANR-INBS-0010) and the NISTEC project (Grant ANR-09-CESA-003).
35
36
37

38 774
39
40
41 775 **ACKNOWLEDGMENT**

42
43
44 776 This work was supported in part by the French Ministry of Research and the National Research
45
46 777 Agency, as part of the French MetaboHUB infrastructure (the national metabolomics and
47
48 778 fluxomics infrastructure, Grant ANR-INBS-0010) and the NISTEC project (Grant ANR-09-
49
50 779 CESA-003). The authors thank Maxime Chazalviel for its work on metabolic network visualization
51
52 780 and Dr Marc Audebert for providing valuable help on experimental procedures.
53
54
55
56
57
58
59
60

- 1
2
3 781
4
5 782 REFERENCES
6
7
8 783 (1) Thiele, I.; Palsson, B. Ø. Reconstruction Annotation Jamborees: A Community Approach to
9
10 784 Systems Biology. *Mol. Syst. Biol.* **2010**, *6*, 361.
11
12
13 785 (2) Duarte, N. C.; Becker, S. A.; Jamshidi, N.; Thiele, I.; Mo, M. L.; Vo, T. D.; Srivas, R.;
14
15 786 Palsson, B. Ø. Global Reconstruction of the Human Metabolic Network Based on Genomic
16
17 787 and Bibliomic Data. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (6), 1777–1782.
18
19
20
21 788 (3) Ma, H.; Sorokin, A.; Mazein, A.; Selkov, A.; Selkov, E.; Demin, O.; Goryanin, I. The
22
23 789 Edinburgh Human Metabolic Network Reconstruction and Its Functional Analysis. *Mol Syst*
24
25 790 *Biol* **2007**, *3*, 135.
26
27
28
29 791 (4) Mardinoglu, A.; Agren, R.; Kampf, C.; Asplund, A.; Uhlen, M.; Nielsen, J. Genome-Scale
30
31 792 Metabolic Modelling of Hepatocytes Reveals Serine Deficiency in Patients with Non-
32
33 793 Alcoholic Fatty Liver Disease. *Nat Commun* **2014**, *5* (May 2013), 3083.
34
35
36
37 794 (5) Thiele, I.; Swainston, N.; Fleming, R. M. T.; Hoppe, A.; Sahoo, S.; Aurich, M. K.;
38
39 795 Haraldsdottir, H.; Mo, M. L.; Rolfsson, O.; Stobbe, M. D.; et al. A Community-Driven
40
41 796 Global Reconstruction of Human Metabolism. *Nat Biotechnol* **2013**, *31* (5), 419–425.
42
43
44
45 797 (6) Smallbone, K. Striking a Balance with Recon 2.1. **2013**.
46
47
48 798 (7) Swainston, N.; Smallbone, K.; Hefzi, H.; Dobson, P. D.; Brewer, J.; Hanscho, M.; Zielinski,
49
50 799 D. C.; Ang, K. S.; Gardiner, N. J.; Gutierrez, J. M.; et al. Recon 2.2: From Reconstruction
51
52 800 to Model of Human Metabolism. *Metabolomics* **2016**, *12* (7), 109.
53
54
55 801 (8) Blais, E. M.; Rawls, K. D.; Dougherty, B. V.; Li, Z. I.; Kolling, G. L.; Ye, P.; Wallqvist, A.;

- 1
2
3 802 Papin, J. A. Reconciled Rat and Human Metabolic Networks for Comparative
4
5 803 Toxicogenomics and Biomarker Predictions. *Nat. Commun.* **2017**, *8*, 14250.
6
7
8 804 (9) Brunk, E.; Sahoo, S.; Zielinski, D. C.; Altunkaya, A.; Dräger, A.; Mih, N.; Gatto, F.; Nilsson,
9
10 805 A.; Preciat Gonzalez, G. A.; Aurich, M. K.; et al. Recon3D Enables a Three-Dimensional
11
12 806 View of Gene Variation in Human Metabolism. *Nat. Biotechnol.* **2018**, *36* (3), 272–281.
13
14
15
16 807 (10) Shlomi, T.; Cabili, M. N.; Herrgard, M. J.; Palsson, B. Ø. O.; Ruppín, E.; Herrgård, M. J.;
17
18 808 Palsson, B. Ø. O.; Ruppín, E. Network-Based Prediction of Human Tissue-Specific
19
20 809 Metabolism. *Nat Biotech* **2008**, *26* (9), 1003–1010.
21
22
23
24 810 (11) Jerby, L.; Shlomi, T.; Ruppín, E. Computational Reconstruction of Tissue-Specific
25
26 811 Metabolic Models: Application to Human Liver Metabolism. *Mol Syst Biol* **2010**, *6*, 401.
27
28
29 812 (12) Bordbar, A.; Jamshidi, N.; Palsson, B. O. IAB-RBC-283: A Proteomically Derived
30
31 813 Knowledge-Base of Erythrocyte Metabolism That Can Be Used to Simulate Its
32
33 814 Physiological and Patho-Physiological States. *BMC Syst Biol* **2011**, *5*, 110.
34
35
36
37 815 (13) Agren, R.; Bordel, S.; Mardinoglu, A.; Pornputtapong, N.; Nookaew, I.; Nielsen, J.
38
39 816 Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and
40
41 817 16 Cancer Types Using INIT. *PLoS Comput Biol* **2012**, *8* (5), e1002518.
42
43
44
45 818 (14) Agren, R.; Mardinoglu, A.; Asplund, A.; Kampf, C.; Uhlen, M.; Nielsen, J. Identification of
46
47 819 Anticancer Drugs for Hepatocellular Carcinoma through Personalized Genome-Scale
48
49 820 Metabolic Modeling. *Mol. Syst. Biol.* **2014**, *10*, 1–13.
50
51
52
53 821 (15) Frezza, C.; Zheng, L.; Folger, O.; Rajagopalan, K. N.; MacKenzie, E. D.; Jerby, L.;
54
55 822 Micaroni, M.; Chaneton, B.; Adam, J.; Hedley, A.; et al. Haem Oxygenase Is Synthetically
56
57
58
59
60

- 1
2
3 823 Lethal with the Tumour Suppressor Fumarate Hydratase. *Nature* **2011**, 477 (7363), 225–
4
5 824 228.
6
7
8
9 825 (16) Goldstein, I.; Yizhak, K.; Madar, S.; Goldfinger, N.; Ruppin, E.; Rotter, V. P53 Promotes
10
11 826 the Expression of Gluconeogenesis-Related Genes and Enhances Hepatic Glucose
12
13 827 Production. *Cancer Metab.* **2013**, 1 (1), 9.
14
15
16 828 (17) Yizhak, K.; Le Dévédec, S. E.; Rogkoti, V. M.; Baenke, F.; de Boer, V. C.; Frezza, C.;
17
18 829 Schulze, A.; van de Water, B.; Ruppin, E. A Computational Study of the Warburg Effect
19
20 830 Identifies Metabolic Targets Inhibiting Cancer Migration. *Mol. Syst. Biol.* **2014**, 10, 744.
21
22
23
24 831 (18) Orth, J. D.; Thiele, I.; Palsson, B. Ø. O. What Is Flux Balance Analysis? *Nat Biotechnol*
25
26 832 **2010**, 28 (3), 245–248.
27
28
29 833 (19) Zur, H.; Ruppin, E.; Shlomi, T. IMAT: An Integrative Metabolic Analysis Tool.
30
31 834 *Bioinformatics* **2010**, 26 (24), 3140–3142.
32
33
34
35 835 (20) Aninat, C.; Piton, A.; Glaise, D.; Le Charpentier, T.; Langouët, S.; Morel, F.; Guguen-
36
37 836 Guillouzo, C.; Guillouzo, A. Expression of Cytochrome P450 Enzymes and Nuclear
38
39 837 Receptors in Human Hepatoma HepaRG Cells. *Drug Metab. Dispos.* **2006**, 34 (1), 75–83.
40
41
42
43 838 (21) Rodrigues, R. M.; Heymans, A.; De Boe, V.; Sachinidis, A.; Chaudhari, U.; Govaere, O.;
44
45 839 Roskams, T.; Vanhaecke, T.; Rogiers, V.; De Kock, J. Toxicogenomics-Based Prediction of
46
47 840 Acetaminophen-Induced Liver Injury Using Human Hepatic Cell Systems. *Toxicol. Lett.*
48
49 841 **2016**, 240 (1), 50–59.
50
51
52
53 842 (22) Peyta, L.; Jarnouen, K.; Pinault, M.; Guimaraes, C.; Pais de Barros, J.-P.; Chevalier, S.;
54
55 843 Dumas, J.-F.; Maillot, F.; Hatch, G. M.; Loyer, P.; et al. Reduced Cardiolipin Content
56
57
58
59
60

- 1
2
3 844 Decreases Respiratory Chain Capacities and Increases ATP Synthesis Yield in the Human
4
5 845 HepaRG Cells. *Biochim. Biophys. Acta - Bioenerg.* **2016**, *1857* (4), 443–453.
6
7
8 846 (23) Ma, X.; Duan, Y.; Tschudy-Seney, B.; Roll, G.; Behbahan, I. S.; Ahuja, T. P.; Tolstikov, V.;
9
10 847 Wang, C.; McGee, J.; Khoobyari, S.; et al. Highly Efficient Differentiation of Functional
11
12 848 Hepatocytes From Human Induced Pluripotent Stem Cells. *Stem Cells Transl. Med.* **2013**, *2*
13
14 849 (6), 409–419.
15
16
17
18 850 (24) Zhang, J.; Nuebel, E.; Daley, G. Q.; Koehler, C. M.; Teitell, M. A. Metabolic Regulation in
19
20 851 Pluripotent Stem Cells during Reprogramming and Self-Renewal. *Cell Stem Cell* **2012**, *11*
21
22 852 (5), 589–595.
23
24
25
26 853 (25) Xu, X.; Duan, S.; Yi, F.; Ocampo, A.; Liu, G.-H.; Izpisua Belmonte, J. C. Mitochondrial
27
28 854 Regulation in Pluripotent Stem Cells. *Cell Metab.* **2013**, *18* (3), 325–332.
29
30
31
32 855 (26) Teslaa, T.; Teitell, M. A. Pluripotent Stem Cell Energy Metabolism: An Update. *EMBO J.*
33
34 856 **2015**, *34* (2), 138–153.
35
36
37 857 (27) Bucher, S.; Jalili, P.; Le Guillou, D.; Begriche, K.; Rondel, K.; Martinais, S.; Zalko, D.;
38
39 858 Corlu, A.; Robin, M.-A.; Fromenty, B. Bisphenol a Induces Steatosis in HepaRG Cells
40
41 859 Using a Model of Perinatal Exposure. *Environ. Toxicol.* **2017**, *32* (3), 1024–1036.
42
43
44
45 860 (28) Cerec, V.; Glaise, D.; Garnier, D.; Morosan, S.; Turlin, B.; Drenou, B.; Gripon, P.;
46
47 861 Kremsdorf, D.; Guguen-Guillouzo, C.; Corlu, A. Transdifferentiation of Hepatocyte-like
48
49 862 Cells from the Human Hepatoma HepaRG Cell Line through Bipotent Progenitor.
50
51 863 *Hepatology* **2007**, *45*, 957–967.
52
53
54
55 864 (29) Hart, S. N.; Li, Y.; Nakamoto, K.; Subileau, E. a.; Steen, D.; Zhong, X. b. A Comparison of
56
57
58
59
60

- 1
2
3 865 Whole Genome Gene Expression Profiles of HepaRG Cells and HepG2 Cells to Primary
4
5 866 Human Hepatocytes and Human Liver Tissues. *Drug Metab. Dispos.* **2010**, *38* (6), 988–994.
6
7
8 867 (30) Parent, R.; Beretta, L. Translational Control Plays a Prominent Role in the Hepatocytic
9
10 868 Differentiation of HepaRG Liver Progenitor Cells. *Genome Biol.* **2008**, *9*, R19.
11
12
13
14 869 (31) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.;
15
16 870 Mandal, R.; Aziat, F.; Dong, E.; et al. HMDB 3.0—The Human Metabolome Database in
17
18 871 2013. *Nucleic Acids Res.* **2012**, *41* (D1), D801–D807.
19
20
21
22 872 (32) Rossell, S.; Huynen, M. A.; Notebaart, R. A. Inferring Metabolic States in Uncharacterized
23
24 873 Environments Using Gene-Expression Measurements. *PLoS Comput Biol* **2013**, *9* (3),
25
26 874 e1002988.
27
28
29 875 (33) UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2017**, *45* (D1), D158–
30
31 876 D169.
32
33
34
35 877 (34) Uhlen, M.; Fagerberg, L.; Hallstrom, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.;
36
37 878 Sivertsson, A.; Kampf, C.; Sjostedt, E.; Asplund, A.; et al. Tissue-Based Map of the Human
38
39 879 Proteome. *Science (80-.)*. **2015**, *347* (6220), 1260419–1260419.
40
41
42
43 880 (35) Dennis, G.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H.; Lempicki, R. A.;
44
45 881 Quackenbush, J.; Wheeler, D.; Chappey, C.; et al. DAVID: Database for Annotation,
46
47 882 Visualization, and Integrated Discovery. *Genome Biol.* **2003**, *4* (9), R60.
48
49
50 883 (36) Bordbar, A.; Feist, A. M.; Usaite-Black, R.; Woodcock, J.; Palsson, B. O.; Famili, I. A
51
52 884 Multi-Tissue Type Genome-Scale Metabolic Network for Analysis of Whole-Body Systems
53
54
55 885 Physiology. *BMC Syst Biol* **2011**, *5*, 180.
56
57
58
59
60

- 1
2
3 886 (37) Gille, C.; Bolling, C.; Hoppe, A.; Bulik, S.; Hoffmann, S.; Hubner, K.; Karlstadt, A.;
4
5 887 Ganeshan, R.; Konig, M.; Rother, K.; et al. HepatoNet1: A Comprehensive Metabolic
6
7 888 Reconstruction of the Human Hepatocyte for the Analysis of Liver Physiology. *Mol Syst*
8
9 *Biol* **2010**, *6* (411), 411.
10
11
12
13 890 (38) Cottret, L.; Wildridge, D.; Vinson, F.; Barrett, M. P.; Charles, H.; Sagot, M. F.; Jourdan, F.
14
15 891 MetExplore: A Web Server to Link Metabolomic Experiments and Genome-Scale
16
17 892 Metabolic Networks. *Nucleic Acids Res* **2010**, *38* (Web Server issue), W132-7.
18
19
20
21 893 (39) Vlassis, N.; Pacheco, M. P.; Sauter, T. Fast Reconstruction of Compact Context-Specific
22
23 894 Metabolic Network Models. *arXiv:1304.7992 [q-bio.MN]* **2013**, *10*, 1–15.
24
25
26 895 (40) Tong, L. Structure and Function of Biotin-Dependent Carboxylases. *Cell. Mol. Life Sci.*
27
28 896 **2013**, *70* (5), 863–891.
29
30
31 897 (41) Badawy, A. A.-B. Tryptophan Metabolism, Disposition and Utilization in Pregnancy.
32
33 898 *Biosci. Rep.* **2015**, *35* (5), e00261–e00261.
34
35
36
37 899 (42) Suda, T.; Robinson, J. C.; Fjellstedt, T. A. Developmental Changes in the Enzymatic
38
39 900 Capacity for Reduction and Oxidation of Alpha-Ketoadipate in Rat Liver, Heart, Kidney,
40
41 901 and Brain. *Pediatr. Res.* **1978**, *12* (4 Pt 1), 297–300.
42
43
44
45 902 (43) Chazalviel, M.; Frainay, C.; Poupin, N.; Vinson, F.; Merlet, B.; Gloaguen, Y.; Cottret, L.;
46
47 903 Jourdan, F. MetExploreViz: Web Component for Interactive Metabolic Network
48
49 904 Visualization. *Bioinformatics* **2018**, *34* (2), 312–313.
50
51
52
53 905 (44) Wanders, R. J. A.; Waterham, H. R.; Ferdinandusse, S. Metabolic Interplay between
54
55 906 Peroxisomes and Other Subcellular Organelles Including Mitochondria and the
56
57
58
59
60

- 1
2
3 907 Endoplasmic Reticulum. *Front. Cell Dev. Biol.* **2016**, *3*, 83.
4
5
6 908 (45) Kersten, S. Integrated Physiology and Systems Biology of PPAR α . *Mol. Metab.* **2014**, *3* (4),
7
8 909 354–371.
9
10
11 910 (46) Fan, C. Y.; Pan, J.; Chu, R.; Lee, D.; Kluckman, K. D.; Usuda, N.; Singh, I.; Yeldandi, A.
12
13 V; Rao, M. S.; Maeda, N.; et al. Hepatocellular and Hepatic Peroxisomal Alterations in Mice
14 911 with a Disrupted Peroxisomal Fatty Acyl-Coenzyme A Oxidase Gene. *J. Biol. Chem.* **1996**,
15
16 912 *271* (40), 24698–24710.
17
18 913
19
20
21 914 (47) Newman, J. C.; Verdin, E. β -Hydroxybutyrate: A Signaling Metabolite. *Annu. Rev. Nutr.*
22
23 **2017**, *37* (1), 51–76.
24 915
25
26
27 916 (48) Zelezniak, A.; Pers, T. H.; Soares, S.; Patti, M. E.; Patil, K. R. Metabolic Network Topology
28
29 917 Reveals Transcriptional Regulatory Signatures of Type 2 Diabetes. *PLoS Comput. Biol.*
30
31 918 **2010**, *6* (4), e1000729.
32
33
34
35 919 (49) Chagoyen, M.; Pazos, F. MBRole: Enrichment Analysis of Metabolomic Data.
36
37 920 *Bioinformatics* **2011**, *27* (5), 730–731.
38
39
40 921 (50) Persicke, M.; Rückert, C.; Plassmeier, J.; Stutz, L. J.; Kessler, N.; Kalinowski, J.; Goesmann,
41
42 922 A.; Neuweger, H. MSEA: Metabolite Set Enrichment Analysis in the MeltDB Metabolomics
43
44 923 Software Platform: Metabolic Profiling of *Corynebacterium Glutamicum* as an Example.
45
46 924 *Metabolomics* **2012**, *8* (2), 310–322.
47
48
49
50 925 (51) Frainay, C.; Jourdan, F. Computational Methods to Identify Metabolic Sub-Networks Based
51
52 926 on Metabolomic Profiles. *Brief. Bioinform.* **2017**, *18* (1), 43–56.
53
54
55
56 927 (52) Becker, S. A.; Palsson, B. O. Context-Specific Metabolic Networks Are Consistent with
57
58
59
60

1
2
3 928 Experiments. *PLoS Comput. Biol.* **2008**, *4* (5), e1000082.

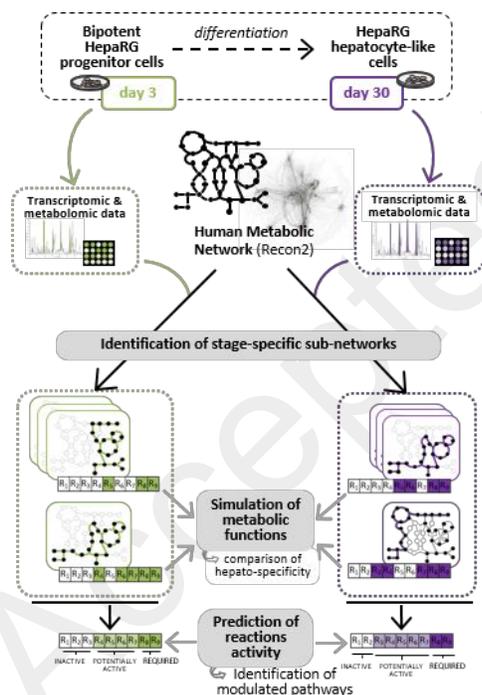
4
5
6 929 (53) Machado, D.; Herrgård, M. Systematic Evaluation of Methods for Integration of
7
8 930 Transcriptomic Data into Constraint-Based Models of Metabolism. *PLoS Comput. Biol.*
9
10 931 **2014**, *10* (4), e1003580.

11
12
13
14 932 (54) Pacheco, M. P.; Pfau, T.; Sauter, T. Benchmarking Procedures for High-Throughput Context
15
16 933 Specific Reconstruction Algorithms. *Front. Physiol.* **2016**, *6*, 410.

17
18
19 934

20
21
22 935

23
24
25
26 936 For Table of Content only



937