Essentiality of local topology and regulation in kinetic metabolic modeling

Gaoyang Li[1,2], Wei Du[1*] and Huansheng Cao[3*]

[1] Key Laboratory of Symbol Computation and Knowledge Engineering of the Ministry of Education, College of Computer Science and Technology, Jilin University, Changchun, 130012, China

[2] Tongji University Cancer Center, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, 200092, China

[3] Division of Natural and Applied Sciences, Duke Kunshan University, Kunshan, 215316, China

* Corresponding authors

Gaoyang Li: lgyzngc@tongji.edu.cn;

Wei Du: weidu@jlu.edu.cn; Huansheng Cao: hc284@duke.edu

Abstract

Genome-scale metabolic networks (GSMs) are mathematic representation of a set of stoichiometrically balanced reactions. However, such static GSMs do not reflect or incorporate functional organization of genes and their dynamic regulation (e.g., operons and regulons). Specifically, there are numerous topologically coupled local reactions through which fluxes are coordinated; and downstream metabolites often dynamically regulate the gene expression of their reactions via feedback. Here, we present a method which reconstructs GSMs with locally coupled reactions and transcriptional regulation of metabolism by key metabolites. The proposed method has outstanding performance in phenotype prediction of wild-type and mutants in *Escherichia coli* (*E. coli*), *Saccharomyces cerevisiae* (*S. cerevisiae*) and *Bacillus subtilis* (*B. subtilis*) growing in various conditions, outperforming existing methods. The predicted growth rate and metabolic fluxes are highly correlated with those experimentally measured. More importantly, our method can also explain the observed growth rates by capturing the 'real' (experimentally measured) changes in flux between the wild-types and mutants. Overall, by identifying and incorporating locally organized and regulated functional modules into GSMs, Decrem achieves accurate predictions of phenotypes and has broad applications in bioengineering, synthetic biology and microbial pathology.

42 Cellular life maintains itself and replicates through the entire set of biochemical reactions in
43 genome-scale metabolic networks (GSMs) operating in a well-coordinated manner [1-3]. But,
44 how these reactions operate and are orchestrated and regulated *in vivo* in a GSM remains a
45 major challenge in systems biology. Some insight into this question has been gained: local
46 metabolite levels are stable in the face of environmental or genetic perturbations [4-10] due to
47 quick compensation from nearby reactions. This suggests that rerouting of fluxes in GSMs
48 operates efficiently, which plays a key role in metabolite homeostasis in maintaining global
49 optima under one condition or across conditions [5]. Furthermore, during the evolution of
50 metabolic networks, reactions and metabolites prefer to actively attach to high-efficiency
51 biochemical reaction chains in an organism, according to the principle of preferential
52 attachment [11-13]. This mode of network organization leads to topological coupling for
53 metabolic flux rerouting. For example, in the TCA (tricarboxylic acid) cycle, the reaction chain
54 D-isocitrate --> a-ketoglutarate --> ... --> malate (K_m=0.029μM and K_{cat}=106.4 for IDH3
55 (Isocitrate dehydrogenase)) is preferred as the primary branch for oxidation of acetyl-CoA over
56 the low-efficiency reaction chain D-isocitrate --> succinate (K_m =8μM and K_{cat} =28.5 for AceA
57 (Isocitrate lyase)). Recent studies have revealed topology-derived module-specific enzyme
58 parameters, e.g., K_m and K_{cat}, network properties, e.g., path length and robustness, and gene
59 expression profiles in GSMs [14]. It has been demonstrated that some reactions are more closely
60 interconnected with high kinetic capabilities than others and form small worlds in GSMs, which
61 is the fundamental organizing principle of pathways [15-17], particularly in central metabolism,
62 for essential precursor conversion in diverse environments [18]. Therefore, coupled reactions
63 constitute the topological foundation of local metabolic flux coordination in GSMs. The

64 importance of these local coupling preferences is manifested in three aspects. First, the
65 changes due to local internal perturbations (gene deletions) are quickly compensated by
66 neighboring reactions [10]. Second, functionally related genes are organized as operons that are
67 coregulated in microorganisms [19-23]. Third, failure to recognize the coupled reactions leads to
68 poor performance in flux estimation by steady-state linear flux optimization, e.g., flux balance
69 analysis (FBA) [24].

70 In response to internal (genetics) and external perturbations, metabolic systems are
71 rather dynamic in adjusting metabolic fluxes [5,25], which results from regulations at the
72 substrate and transcriptional levels [26]. The mechanism of fast-acting and local substrate-level
73 regulation of enzyme activity is well established, including allosteric regulation, feedback
74 inhibition, and covalent modification [27-29]. However, regulator-regulatee pairs of metabolites
75 and enzymes are not available on the genome scale. Similarly, transcriptional regulation of
76 metabolism on a genome scale has just begun [8,30-33]. The expression profiles of metabolic
77 genes are primarily regulated by the growth state via the sequestration or release of
78 transcriptional factors (TFs) by key metabolites with concentrations that vary with growth [32].
79 The activities of over 200 TFs show strong correlations with cognate metabolites following the
80 transition from starvation to growth in *E. coli* [8]. Compared to the global transcriptional
81 regulation of metabolism, regulatory metabolites have not been identified on a genome scale,
82 particularly in the functionally organized local modules of GSMs[8]. The regulation at the
83 metabolic pathway level is particularly important because genes in microbial genomes are
84 organized into operons (coregulated/coexpressed, functionally related and adjacently placed
85 genes), and the expression of operons is regulated mainly by local TFs, which are in turn

86 controlled by the end products of the pathways involving the genes of the operons [8]. These
87 facts underscore the potential of key metabolites directly controlling the pathway fluxes rather
88 than simply being indicators of cell state with precise modeling of the associated enzyme
89 (reaction) kinetics. Therefore, key metabolites may provide an alternative approach for
90 reducing the complexity of dynamic flux prediction without loss in modeling accuracy, given
91 the extreme difficulty of obtaining complete enzyme kinetic parameters (i.e., K_m and K_{cat}) from
92 paired metabonomics and proteomics data in building a reaction kinetics-constrained
93 metabolic flux prediction model in GSMs.

94 In this study, we presented a computational method to integrate flux coordination
95 derived from local coupled reaction topology and key regulator metabolite-mediated dynamic
96 enzyme activity into GSMs to approximate biologically feasible fluxes (Fig. 1). We first
97 incorporated the identified cooperated topological profiles of GSMs into the canonical FBA by
98 representing the synchronously coordinated (coregulated) and closely coupled reactions with
99 a group of independent sparse bases (reactions) according to the stoichiometric matrix
100 decomposition. Then, we reconstructed a decoupled independent basis (reaction)-
101 constrained metabolic model: Decrem, which was carried out on three model organisms: *E.*
102 *coli*, *S. cerevisiae* and *B. subtilis*. The flux distributions predicted by Decrem were highly
103 consistent with experimentally measured fluxes in various cellular behaviors across multiple
104 strains: growth rate and the flux distribution on different carbon sources and in response to
105 genetic and environmental perturbations. Next, a cell state-associated kinetic model (k-
106 Decrem) was utilized to enhance Decrem by incorporating metabolite-derived TF regulation in
107 response to environmental perturbation. A unique advantage of k-Decrem model is that

pathway-associated key metabolite concentrations will suffice to achieve metabolic regulations and thus reduce the requirements for necessary kinetic parameters and paired multi-omics data. In addition, the accurate growth rates predicted by k-Decrem in *E*. coli genome-scale knockout strains revealed that intracellular perturbations are mainly 'buffered' by highly coupled reactions, which reveals the coordination between key precursors of central metabolism and cell growth. To the best of our knowledge, this is the first computational platform to integrate metabolic network topology and regulation by metabolites into GSMs.

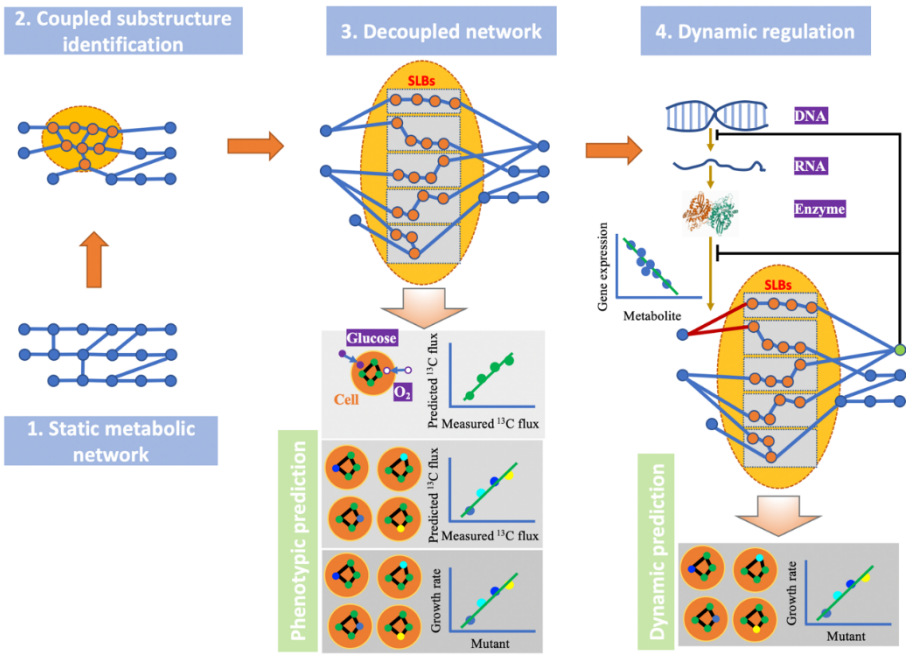


Fig. 1. An overview of Decrem. (1) Schematic representation of a metabolic network. (2) Identification of a closely coupled reaction substructure (orange circles) of a metabolic network. (3) Closely coupled substructure is represented as a group of decoupled sparse linear basis (SLB) reactions (in gray rectangles), and the reconstructed GSMs incorporating these SLB reactions were tested in three scenarios: response in flux and growth rate to genetic perturbation and environmental perturbations. (4) Building a kinetic model integrating regulations at substrate and transcriptional levels, which is tested in growth rate prediction and is interpretation in genome-scale mutant strains.

Results

Reconstruction of GSMs with topologically decoupled reactions

We developed a topologically decoupled linear representation of metabolic network to

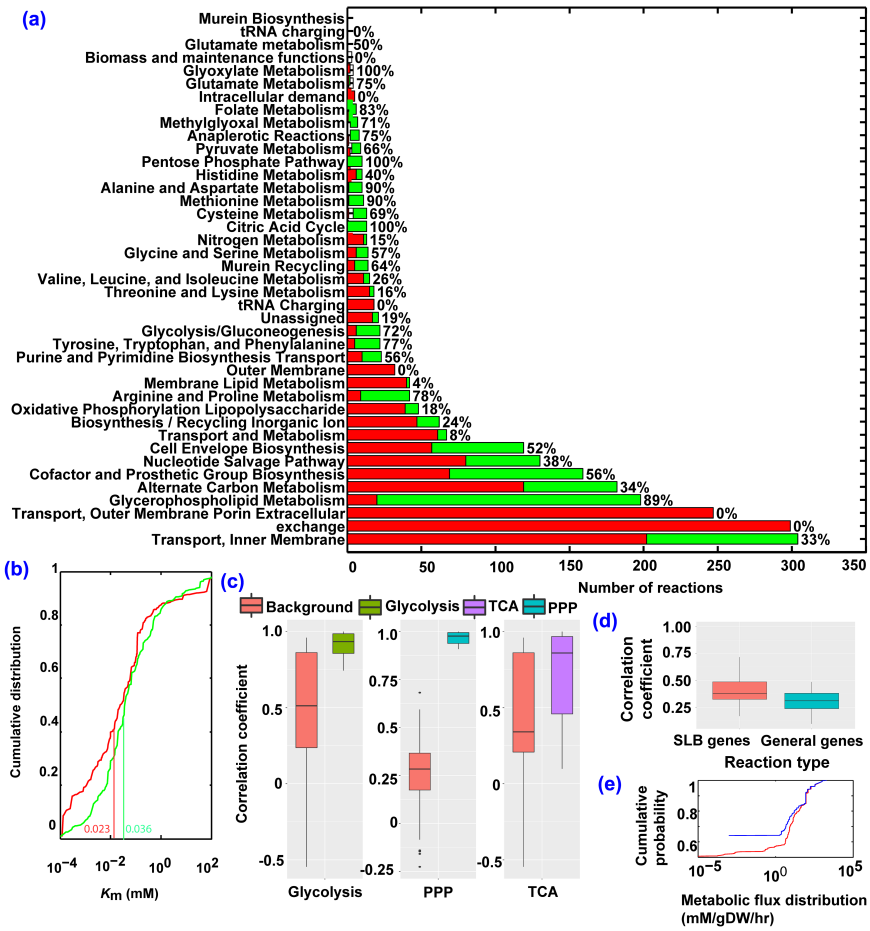
127 characterize the coactivated regulation of highly coupled reactions with three steps. First, to
128 explore the topological organization of highly coupled reactions, substructures composed of
129 tightly connected coupled reactions in the metabolic network were identified from its bipartite
130 graph representation with a graph clustering algorithm (see Methods; Supplementary Fig. 1)
131 [34]. Specifically, a similarity metric for coupled reaction substructure identification was defined
132 as the number of simple cycles between any two reaction nodes in the bipartite graph. Then,
133 reactions involved in multiple simple cycles were treated as candidate highly coupled elements
134 and were classified into one of several coupled reaction-derived substructures using the
135 minimum cut algorithm to further explore their linear decoupled representation (see
136 Methods).

137 Specifically, identified coupled reaction substructures totally included 927 of the 2382
138 reactions in *i*AF1260 model of *E. coli* (Supplementary Table 1). Central metabolism, such as the
139 TCA cycle, pentose phosphate pathway (PPP), glycolysis, and amino acid and
140 glycerophospholipid pathways, primarily consists of coupled reactions and includes 70% (141
141 of 201) reversible reactions. In contrast, tRNA, membrane lipid biosynthesis, membrane
142 transport pathways and the biomass reaction primarily consist of uncoupled linear reaction
143 chains (Fig. 2a). Additionally, the K_m values (0.023 mM) of the coupled reactions in the
144 substructures are smaller (by 56.5%) than those (0.036 mM) in the uncoupled reaction chains
145 (Wilcoxon test, p = 5.74E-4; Fig. 2b). Together, the physical proximity and similar kinetic
146 enzyme parameters suggest that these highly coupled reactions prefer to interact with each
147 other first (as top priority), especially in central metabolism, before reaching out to reactions
148 that are more distant. We then validated the coordination of the identified coupled reactions

149 with publicly available fluxes experimentally measured in 24 knockout strains of *E. coli* [35]. As
150 expected, the correlations among the fluxes of the coupled reactions from the same pathways
151 were higher than those for uncoupled reactions. For example, the average correlation
152 coefficient r is 0.913, 0.975, and 0.794 for the reactions in glycolysis, PPP and TCA cycle,
153 respectively, in contrast to 0.505, 0.267, and 0.421 for each uncoupled reaction set (t test, p =
154 2.33E-9, 3.06E-43 and 2.25E-4, respectively; Fig. 2c), which suggests that the topologically
155 coupled reactions indeed also tend to coordinate their fluxes.

156 Second, we decomposed the highly coupled reaction substructures into representations
157 with minimal independent components. Here, we represented the coupled reaction-derived
158 substructures using sparse linear basis (SLB) vectors of the corresponding null space of the
159 stoichiometry matrix and treated them as the decoupled components of GSMs. That is, each
160 SLB consists of a least number of coupled reactions corresponding to an indivisible
161 independent flux (see Methods). Like elementary flux modes [36], metabolic fluxes under real
162 growth conditions could be decomposed as the weighted linear combination of the identified
163 SLBs [9]. However, unlike the almost infinite number of groups of ordinary linear bases, there is
164 a unique group of SLBs in the null space of the stoichiometry matrix, which is an elegant way
165 to define the mutually independent components for the densely interacting reactions [35]
166 (Supplementary Information: Method S1). We validated the coordinated activation of
167 reactions within the SLBs through gene expression in the 24 knockout strains of *E. coli* [35].
168 Indeed, the correlations (mean r = 0.447) of gene levels among the element reactions from
169 the same SLBs are higher than those (mean r = 0.28) from different SLBs from glycolysis, PPP
170 and the TCA cycle (t test, p = 2.05E-33; Fig. 2d). This suggests that the reactions from the same

171 SLBs tended to be coactivated, but the reactions in different SLBs are more independent.



172

173 **Fig. 2. The statistical analysis of decomposed topological components of *E*. coli *i*AF1260**
174 **model.** (a) The distribution of proportion of coupled reactions (green bar) among the 41
175 metabolic pathways. (b) The comparison of accumulated distribution of kinetic constant, K_m,
176 between the coupled reactions (red line) and non-coupled reactions (green line), the vertical
177 bars are the mean K_m values of the two groups. (c) The correlation distribution of 13C MFA
178 fluxes within the each of three pathways of central metabolism, compared to the
179 corresponding distantly-coupled reactions (examined reactions from different pathways) in 24
180 strains. (d) The correlation distribution of gene expression among the element reactions of
181 sparse linear basis (SLB) vectors against the gene expression among the non-coupled reactions;
182 (e) The cumulative flux distribution of reconstructed decomposition metabolic model vs the
183 original metabolic model according to flux variability analysis.

184 In the last step, we reconstructed a model with decoupled linear representation of a GSM,
185 which was named Decrem, by merging the element reactions of each SLB into a "linear basis
186 reaction (LBR)" with reallocated stoichiometric coefficients (see Methods and Supplementary
187 Fig. 1). To ensure that this decoupled representation has the same flux space as the original

188 GSMs, flux variability analysis (FVA) was conducted in *E. coli* and *S. cerevisiae*. The results
189 confirmed the preservation of solution space (Fig. 2e and Supplementary Fig. 2) [37]. Our model
190 reassigns the trivial fluxes while effectively preserving the primary fluxes of the original GSMs
191 and serves as our working model for subsequent functional analyses.

192 **Prediction of metabolic fluxes in response to environmental changes**

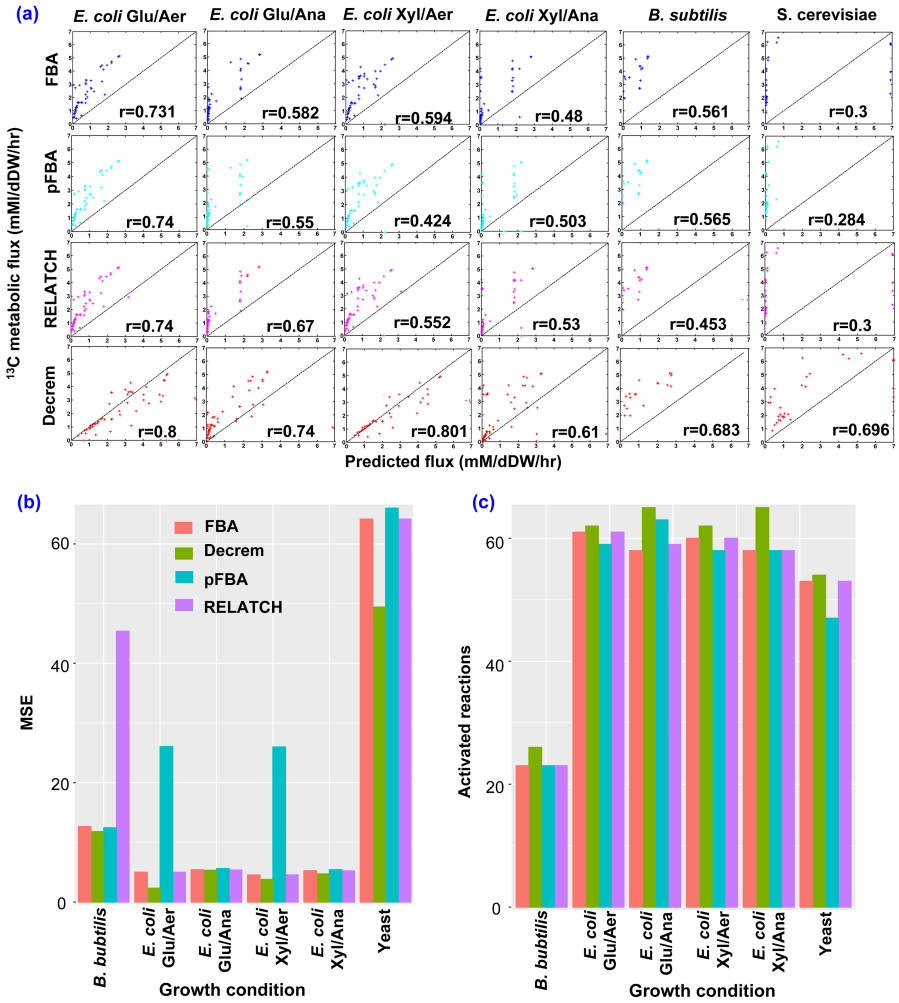
193 We first applied Decrem for metabolic flux prediction, in comparison with three other methods:
194 FBA, pFBA (parsimonious FBA) and RELATCH [38-40] in three model organisms, namely, *E. coli*
195 (*i*AF1260), *S. cerevisiae* (*i*MM904) and *B. subtilis* (*i*YO844), which are composed of 766, 558
196 and 332 LBRs in the corresponding decoupled GSMs, respectively (Supplementary Table 1,
197 Data Availability).

198 We calculated the metabolic flux distribution in *E*.coli K12 MG1655 under four conditions
199 in MOPS medium supplemented with glucose or xylose and under aerobic or anaerobic
200 respiration [41] (Supplementary Information: Material S1), with nutrients as the sole constraints
201 on GSMs (Supplementary Table 2). Using three evaluation metrics—Pearson correlation
202 coefficient *r*, mean squared error (MSE) and the number of activated reactions, the predictions
203 by Decrem have higher correlations with the experimentally measured ^{13}C-MFA (metabolic
204 flux analysis) fluxes (Supplementary Table 3) than the predictions by three other methods
205 under all four conditions (Fig. 3). Decrem also produces much fewer discrepancies (and higher
206 accuracy thereof) between predicted and empirical fluxes than the other methods (Fig. 3a).
207 Furthermore, the fluxes predicted by Decrem have the smallest MSE (maximum upper bound
208 of flux being 1000 mM/gDW/hr), and most of the activated reactions (with nonzero flux)
209 agreed well with the experimental ^{13}C-MFA fluxes (Fig. 3b and Fig.3c). Here, we provided an

210 example of the superior performance of Decrem in the TCA cycle. The Decrem predictions are

211 consistent with the experimentally measured ^{13}C-MFA fluxes (Supplementary Fig. 3), whereas

212 four activated reactions are predicted as inactive (zero fluxes) by FBA.



213

214 **Fig. 3. Comparison of the predictions in three model microorganisms by FBA, pFBA, RELATCH**
215 **and Decrem.** (a) Metabolic flux distribution between the experimental measurements and
216 predictions by the four methods in *E*. coli *i*AF1260 with four conditions: glucose (Gly) or xylose
217 (Oxy) under aerobic (Aer) or anaerobic (Ana) respiration, as well as in the *S*. cerevisiae *i*MM904
218 model and *B*. subtilis *i*YO844 model with glucose under aerobic respiration. r is the Pearson
219 correlation coefficient between the measured ^{13}C MFA fluxes and the predicted fluxes. (b)
220 Comparison of MSE among the four methods across the six growth conditions. (c) Comparison
221 of activated reactions among the four methods across the six growth conditions.

222 Decrem also outperforms the other three methods in flux prediction in two other

223 organisms: *S. cerevisiae* and *B. subtilis* [42,43] (Fig 3; Supplementary Information: Material S1,

224 Supplementary Table 1 and Supplementary Table 2). In particular, Decrem performs well with
225 the complex eukaryote *S*. cerevisiae: *r* is 0.696, 0.3, 0.284 and 0.3 for Decrem, FBA, pFBA and
226 RELATCH, respectively, in which reactions are often strongly coupled based on cellular
227 compartments.

228 **Prediction of metabolic fluxes and growth rates in yeast knockout strains**

229 In addition to fluxes in response to environmental changes, we further evaluated Decrem in
230 predicting fluxes in response to genetic perturbation (single-gene deletions) in two GSMs of *S.*
231 *cerevisiae*, *i*DN750 (1059 metabolites and 1266 reactions) and *i*MM904 (1226 metabolites and
232 1577 reactions). Thirty-eight single-gene-deletion mutants with experimental ^{13}C-MFA fluxes,
233 growth rates, nutrient uptake properties and several extracellular exchange fluxes [44] were used
234 (Supplementary Information: Material S2; Supplementary Table 4). Decrem and four other
235 methods: pFBA, FBA, RELATCH and REPPS [45] were used for flux prediction in these mutants.
236 The predictions by Decrem showed the highest correlations with the experimentally measured
237 fluxes in almost all mutant strains for both GSMs (Fig. 4 and Supplementary Table 5).
238 Specifically, the average Spearman and Pearson correlation coefficients (*r*) determined by
239 Decrem for *i*DN750 are 0.72 and 0.763 in the 38 mutant strains, compared to 0.615 and 0.733,
240 0.607 and 0.681, 0.556 and 0.707, 0.462 and 0.595 for pFBA, FBA, RELATCH and REPPS,
241 respectively (one-way ANOVA, p = 5.52E-20 and 4.21E-07). For the more complete yeast
242 *i*MM904, the mean Spearman and Pearson's correlation coefficients are 0.782 and 0.936,
243 0.765 and 0.747, 0.656 and 0.915, 0.75 and 0.925, 0.645 and 0.841 for Decrem, pFBA, FBA,
244 RELATCH and REPPS, respectively, which are significantly different (one-way ANOVA; p = 4.35E-
245 06 and 8.26E-06, respectively). The differential accuracy of flux predictions between the

246 *i*DN750 and *i*MM904 models shows that Decrem performed better than other methods,

247 particularly on metabolic models that are less complete (e.g., *i*DN750 vs *i*MM904) (Fig 4).

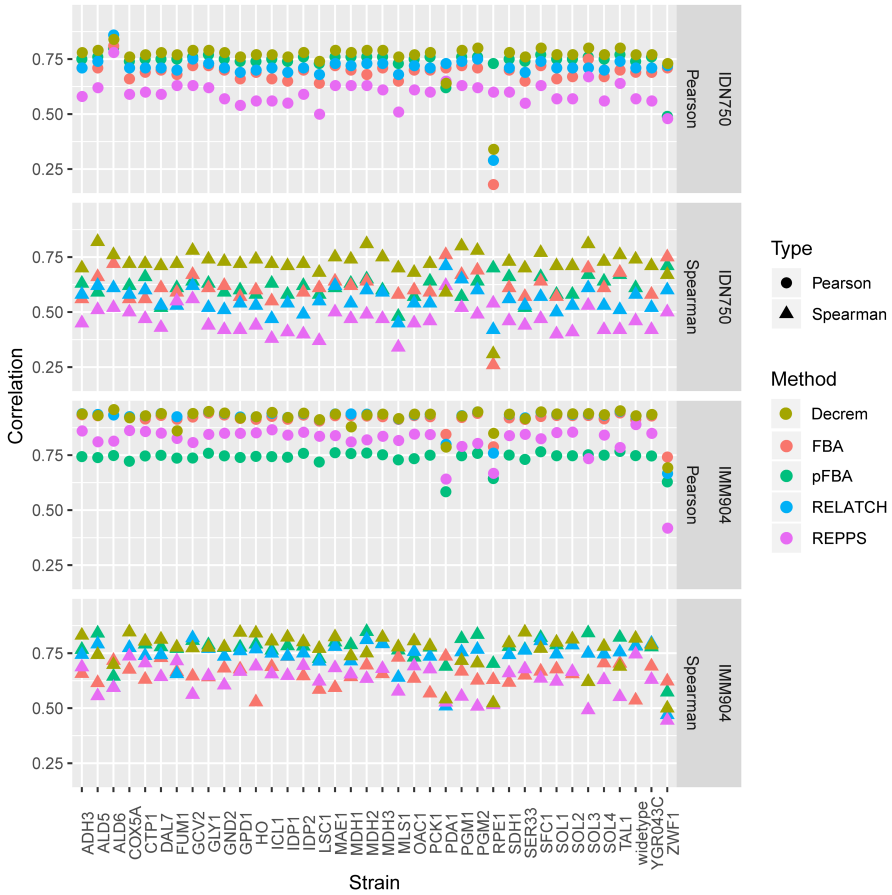
248 Further analysis revealed that such differences exist because a large proportion, 47% (146 of

249 311), of new reactions in *i*MM904 (vs *i*DN750) are highly coupled reactions, compared to an

250 average 35% (558 of 1577) for *i*MM904 (Fisher's exact test, p = 5.427e-04; Data Availability),

251 which clearly optimizes the solution space of the metabolic model and changes the optimal

252 reaction paths in *i*MM904.



253

254 **Fig. 4. Comparison of the flux predictions in two *S. cerevisiae* metabolic models.** The Pearson
255 and Spearman correlation between the ^{13}C MFA measured and predicted metabolic fluxes
256 across each of the five test methods (FBA, pFBA, RELATCH, REPPS and Decrem) among 38
257 strains in the *i*MM904 model and *i*DN750 model of *S. cerevisiae*.

258 We then applied Decrem and pDecrem (parsimonious Decrem, see Supplementary

259 Information: Method S2) to estimate the growth rate in the 38 mutant strains of *S. cerevisiae*

260 using the *i*MM904 model. PCA of the predicted fluxes from all six checked methods shows that
261 the top two principal components (PCs) can explain more than 99% of the variance of all
262 predicted fluxes. The two PCs of Decrem predictions are highly correlated (r > 0.7) with
263 experimentally measured growth rates, while the other methods have, at most, one PC that
264 merely shows moderate correlation (r ~ 0.5) (Supplementary Table 5). Furthermore, a PCA
265 regression reveals that Decrem produces the best flux variance to explain the observed growth
266 rates, as shown by the coefficient of determination (R^2), which was 0.74, 0.731, 0.731, 0.841,
267 0.9 and 0.9 for the six methods (Fig 5a). Most importantly, Decrem successfully identified all
268 the mutations with large effects leading to growth rate < 0.5 h^{-1}: ALD6, FUM1, PDA1, RPE1 and
269 ZWF1 (Fig 5b). Moreover, Decrem-predicted flux distribution through the specific pathways
270 could explain the measured low growth rates. For example, the ZWF1 mutant had a different
271 flux distribution from those of ALD6 and RPE1 (Fig 5c). Interestingly, the exceptionally high
272 fluxes of the mitochondrial transport pathway and citric acid cycle pathway in the ZWF1 strain
273 agree well with the experimental observation that NADPH- and $NADP^+$-dependent
274 mitochondrial malic enzyme flux is significantly increased [44]. The other methods only identified
275 some large-effect mutations on growth rates, leading to a high false-positive rate. In
276 comparison, Decrem demonstrated high accuracy and low false positive rates in assessing the
277 growth rate and well approximated the real metabolic fluxes in the mutants well.

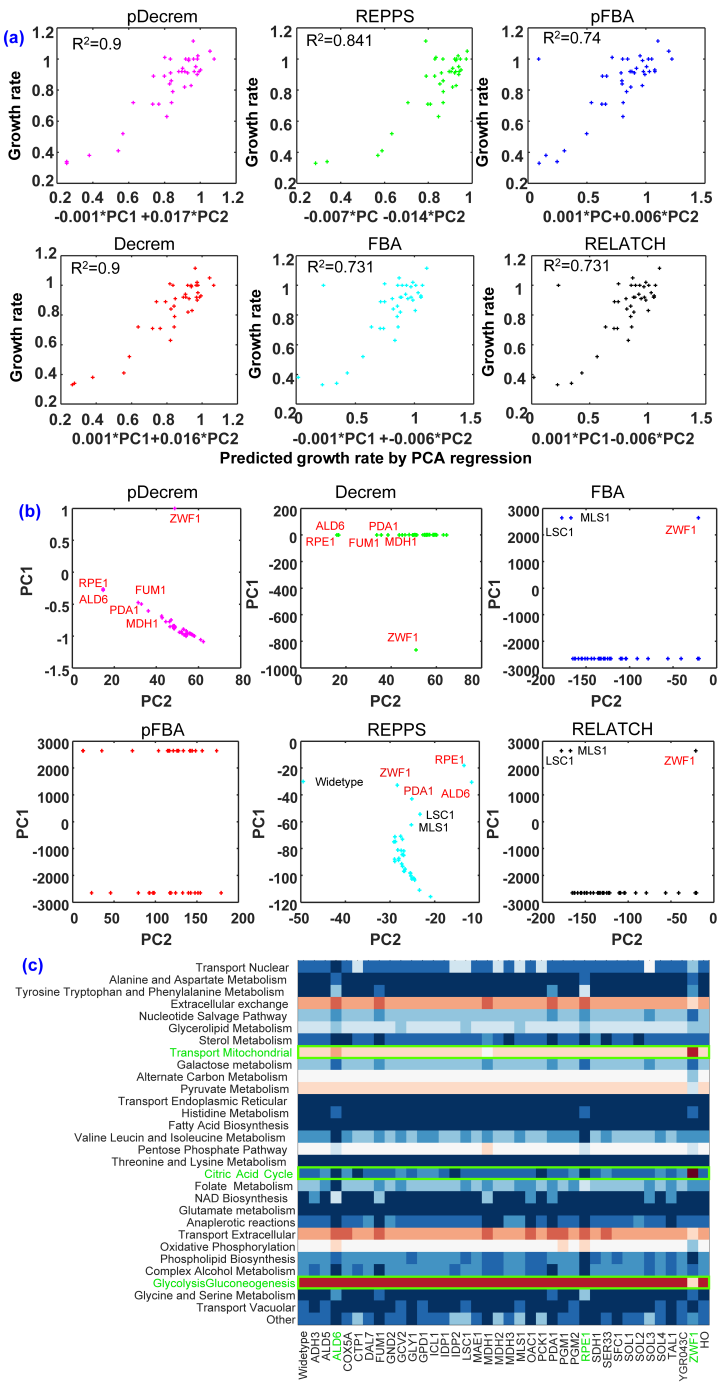


Fig. 5. Cell state predictions for S. cerevisiae by FBA, pFBA, RELATCH, REPPS, Decrem and pDecrem. (a) representation of the experimentally measured (y-axis) and predicted (x-axis) growth rates with PCA regression of the top two PCs from the six test methods. (b) The top two principal components (PCs) of predicted flux variances among the 38 mutant strains, according to the six test methods. The deleted genes with high negative effects on growth are colored in red and the genes with no significant effects on growth are colored in black. (c) The heatmap of predicted fluxes by Decrem through 30 pathways revealed different responding mechanisms among the ZWF1 and the ALD6 mutant, as well as the RPE1 (highlighted pathways in green box indicates the ZWF1 deletion strain specific flux distribution).

289 **Building a metabolic kinetic model incorporating metabolite-TF regulation**

290 Cellular metabolism is not static but rather dynamic because of regulation by substrates and
291 transcription in response to internal and external perturbations. However, for transcriptional
292 regulation of metabolism on a global scale, few regulator-regulatee pairs of metabolites and
293 genes are known. Here, we identified the regulatory roles of key metabolites in central
294 metabolism by incorporating highly coupled coactivated reactions into a metabolic kinetic
295 GSM [46-48], which is verified with empirical multi-omics data in *E. coli* (see Supplementary Fig.
296 4).

297 First, we applied a quantitative approach to identifying pairs of regulator metabolites and
298 regulated genes. As the TFs of metabolic genes in central carbon metabolism are regulated by
299 global growth rate and local pathway-specific essential metabolites, e.g., AMP, fructose-1,6p,
300 and fructose-1p in *E. coli* [48], we hypothesized that there are quantitative relationships between
301 the expression levels of metabolic genes regulated by these TFs and the biomass constituents
302 (growth rate being equivalent to change in biomass), as well as essential metabolites. We then
303 tested this hypothesis with a multi-omics dataset of 24 single-gene knockout strains of *E. coli*
304 [35], which includes the expression of 85 metabolic genes in central metabolism, concentrations
305 of over 100 metabolites and 51 ^{13}C-MFA fluxes for each strain [49]. Through an extensive
306 literature survey [50,51], 45 essential metabolites (and biomass constituents) were selected as
307 potential regulators. They were classified into two groups according to their concentrations:
308 the biomass-constituent group (BG) and the precursor or essential metabolite group (PG) [2]
309 (Supplementary Fig. 5 and Supplementary Table 6). This classification displayed different
310 concentration patterns between the global (biomass constituents) and local (pathway-specific)

311 regulators (Supplementary Information: Material S3).

312 Next, we conducted a partial least squares regression (PLSR) to fit the observed
313 expression profiles of the 85 genes to the concentrations of the potential regulatory
314 metabolites of either group (see Methods). By taking stringent combined thresholds (total
315 regression correlation $r > 0.84$ and the correlation of first PC > 0.38 according to PLSR;
316 Supplementary Information: Material S3), 32 of the 85 genes were identified as the regulatory
317 targets of the 23 BG metabolites, whereas no genes were identified as the regulatory targets
318 of the PG metabolites (Fig. 6a and Supplementary Fig. 6). These identified metabolite-gene
319 regulatory pairs are highly consistent with quantitative proteome-derived metabolite
320 concentrations and the target genes of well-known metabolite-regulated TFs (Supplementary
321 Table 6) [48,52]. The identified metabolite-gene regulatory pairs are further verified with
322 correlations between the measured and predicted gene levels using a wide range of
323 metabolites selected by 10,000 random samplings from the 45 potential metabolic regulators
324 (see Supplementary Information: Material S3). A *p* value of 3.1E-3 for the 23 identified growth-
325 associated metabolites was observed against the randomly selected metabolites (Fig. 6b),
326 while the *p* value was 0.48 for the PG metabolites.

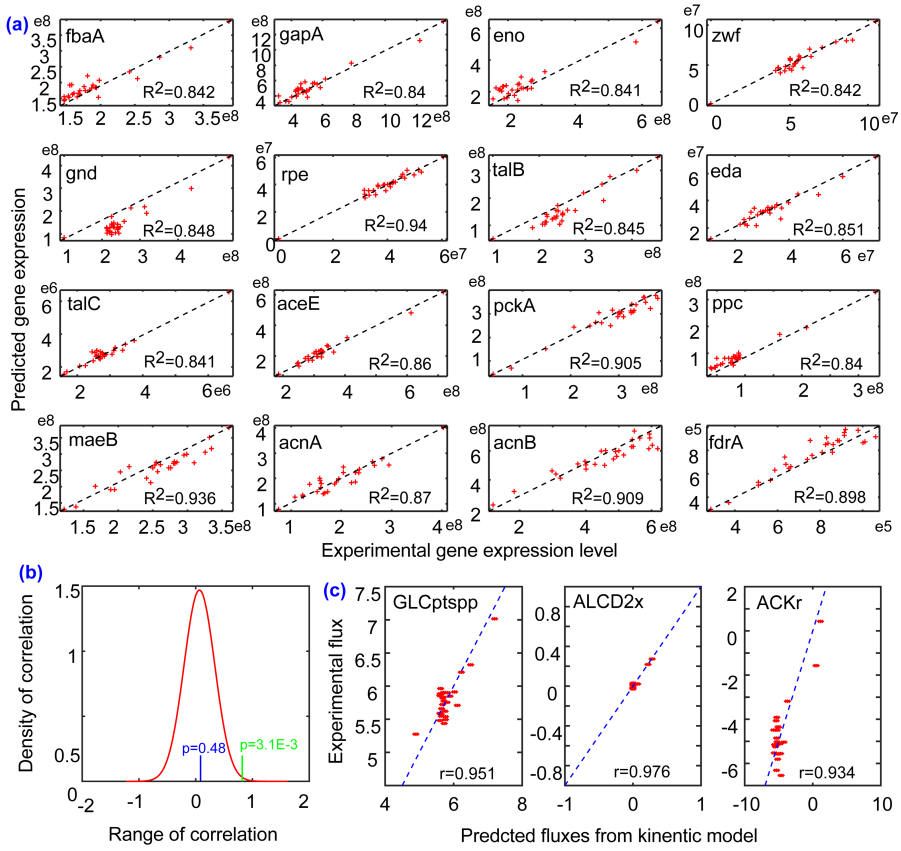


Fig. 6. The analysis of metabolites-based kinetic regulation model of *E*.coli in 24 mutant strains. (a) The predictions of 16 genes expression using the 23 biomass-associated metabolites by the partial least squares regression. (b) The distribution of correlation between the measured and predicted gene expression levels using the metabolites from 10,000 random sampling (red curve), with the statistical significance levels according to the biomass metabolites group (green line) or the precursor metabolites group (blue line). (c) Predicted kinetic fluxes for three key regulatory reactions according to metabolites-determined Michaelis-Menten kinetics.

The identified 32 genes are primarily located in the PPP and pyruvate metabolism in KEGG pathways (Fig 7), which are associated with cell growth for biosynthesis: generating NADPH and pentoses toward nucleotide and amino acid biosynthesis [53], instead of being in energy-producing pathways (TCA and glycolysis). These results suggest that the expression of the genes in growth-associated pathways could be represented as a linear combination of the concentrations of biomass composition.

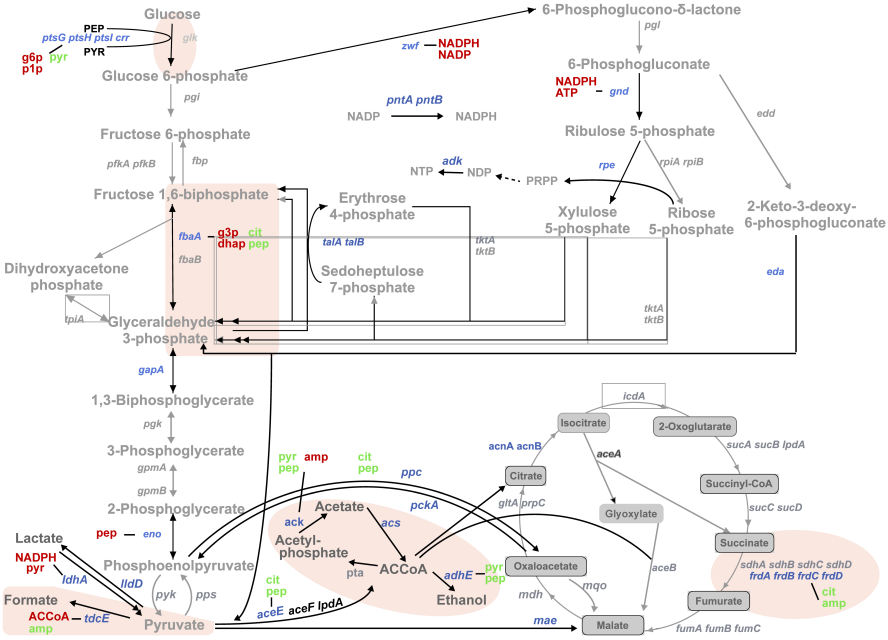


Fig. 7. A map showing of regulator metabolites and the genes regulated by the biomass-constituent metabolites in the central metabolism. The meaning of different backgrounds and colors: red and green metabolites are gene inhibitors and activators; the blue genes present the identified 32 global growth state associated genes. The brown background indicates the identified key regulated reactions which are used to build the kinetic Decrem.

We then constructed a regulation-enabled kinetic model, k-Decrem, for the identified growth-associated metabolic reactions depend only on the concentration of corresponding metabolites above (see Methods; Supplementary Information: Material S3) and evaluated it by optimizing the growth-associated key regulated fluxes in *E. coli* central metabolism. The predicted optimal kinetic fluxes from the kinetic model display high consistency with experimental measurements through ^{13}C isotope tracing (Fig 6c), with *r* values of 0.95, 0.97 and 0.93 for the reactions of glucose transport, alcohol dehydrogenase and acetate kinase, respectively, which are experimentally measured. These fluxes are also utilized to assign key kinetic fluxes in genome-scale metabolic flux prediction below (see Methods).

359 **Growth rate estimation for *E. coli* genome-scale gene deletion strains with k-**

360 **Decrem**

361 We applied k-Decrem constructed above to predict the growth rates of *E. coli* genome-scale

362 single-gene deletion mutants, the growth rates and the corresponding concentrations of over

363 7,000 metabolites of which have been experimentally measured [54]. A total of 1,030 mutants

364 with genes involved in metabolism were selected for growth analysis (see Supplementary

365 Information: Material S4). We first examined the growth rates predicted by the kinetics-free

366 methods pFBA, MOMA and Decrem without any external flux constraints. As expected, poor

367 results are produced, with low correlations with experimentally measured growth rates and *r*

368 values of 0.127, 0.103 and 0.281 for pFBA, MOMA and Decrem, respectively (Supplementary

369 Fig. 7).

370 Next, we constructed the kinetic fluxes of five key regulated reactions in central

371 metabolism with predictions (based on metabolite concentrations) for each of 1,030 mutants

372 to approximate the gene mutant-specific metabolic state, e.g., the branch points of glycolysis

373 and PPP, the flux allocation downstream of pyruvate metabolism, and the growth-associated

374 secretion (see Methods; Fig. 7 and Supplementary Table 7). On the GSMs integrated with

375 kinetic fluxes, the growth rates of the 1,030 mutants were estimated with six methods:

376 Decrem, pDecrem, FBA, pFBA, RELATCH and REPPS. The results showed that all six kinetic

377 methods have significantly improved predictions compared to the kinetic-free methods (Fig

378 8a). Among them, k-Decrem and k-pDecrem produced the highest correlations with the

379 empirical growth rates (r = 0.731 and 0.743 vs 0.421, 0.685, 0.474, and 0.509) (Supplementary

380 Table 7).

381 We then demonstrated the explanatory power of k-Decrem in interpreting the observed
382 growth rates of mutants with corresponding (altered) flux distributions. For that, we calculated the
383 distribution of pathway-specific fluxes across mutant strains, defined as the accumulated flux (AF)
384 [9] of each pathway (the accumulated sum of all nonzero fluxes in a pathway for each mutant strain)
385 (Supplementary Information: Material S4). The correlations between the AFs and growth rate for
386 all strains showed that k-Decrem quantified the greatest number of growth-related pathways that
387 we curated from the literature compared to the other methods. For instance, many well-known
388 important pathways for cell growth—glutamate, nucleic acid and most amino acid metabolism
389 pathways—are 'detected' only by k-Decrem with large correlation coefficients (Fig 8b). Interestingly,
390 although the kinetic pFBA (also k-pDecrem) method can predict the growth rates with relatively
391 high accuracy, the corresponding AFs covered only a few of the curated growth-associated
392 pathways. Moreover, the analyses of the pathway-specific accumulated growth rates (AG; the
393 accumulated sum of growth rates of strains in which the mutated gene is located in the same
394 pathway) suggest that k-Decrem predicted distributions of AGs through all metabolic pathways are
395 highly consistent with the experimentally measured AG distributions (Fig 8c). FBA reaches similar
396 levels of accuracy, but the predictions by pFBA only cover AGs, which were weakly influenced by
397 gene knockouts and shrank the AGs to zero for the pathways containing knockout genes with strong
398 effects (Fig 8c). Such biases are an intrinsic property of the L1 norm-based pFBA (and pDecrem)
399 method [54,55], despite the relatively good fit of correlations.

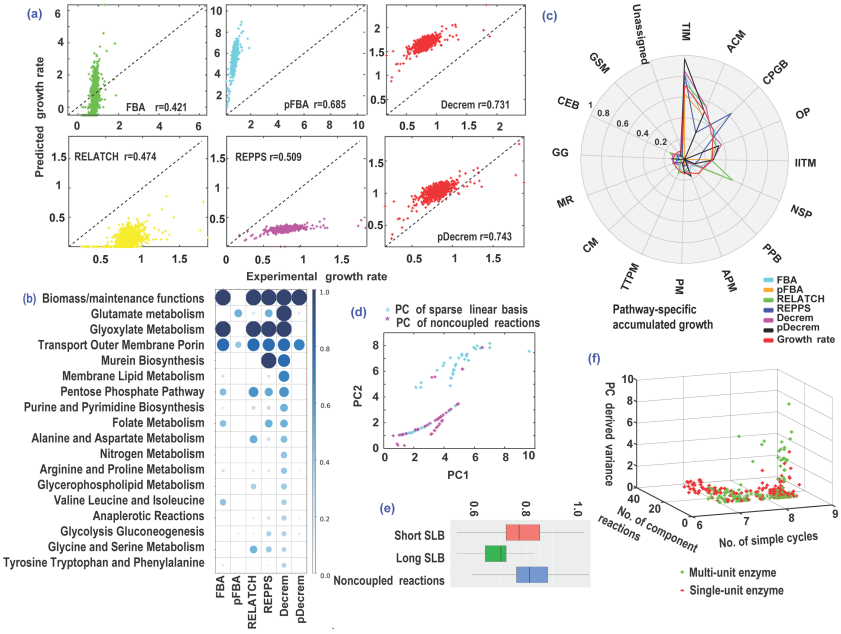


Fig. 8. The growth rate analysis on the genome-scale gene deletion strains. (a) The comparison of measured and predicted growth rates by the k- pDecrem, Decrem, FBA, pFBA, RELATCH and REPPS, on the 1,030 mutants. (b) The correlation between the measured growth rates and the predicted accumulated fluxes though the identified growth-associated pathways. (c) The comparison of the predicted accumulated growth rates through the pathways where the mutants belong, according to the six test methods. (d) The top two PCs revealed difference in flux variances between the coupled reactions (cyan dots) and the non-coupled reactions (purple dots) which are predicted by the Decrem according to the 1,030 strains. (e) The comparison of growth rates distribution among topological components where the mutants belong: the complex LBR (the number of element reactions of corresponding sparse linear basis vector > 1), the simple LBR (the number of element reactions of corresponding sparse linear basis vector = 1) and the non-coupled reactions. (f) The flux variance of LBR based on the multimeric classification: the multi-subunit enzymes (green dots) and single-unit enzymes (red dots), the number of simple cycles and the length (the number of element reactions) of LBR (Fisher's Exact Test, p values = 2.34E-21 and 0.0051).

We further examined the flux variance distribution in each mutant for their changed growth rates. PCA of the fluxes predicted by k-Decrem across the 1,030 mutants is shown in Fig 8d (Supplementary Information: Material S4). The distribution of the top two PCs indicates that the primary flux variances come from the decoupled LBRs, compared to the uncoupled reactions: 1.461 vs 1.10 on average for PC1 (*t* test, p = 3.04E-20); 1.56 vs 1.07 on average for PC2 (*t* test, p = 2.86E-25). This is consistent with the high robustness of the central metabolism [14,56] (primarily consisting of LBRs). Furthermore, we suspected that the effects of deleted genes encoding

422 enzymes for the reactions within the SLBs would be more pronounced than the effect of genes
423 encoding enzymes for uncoupled reactions. Indeed, this is confirmed by the analysis of the reaction
424 type-based growth rates—the complex LBR (the number of element reactions of associated linear
425 basis vector > 1), the simple LBR (the number of element reactions of associated linear basis vector
426 = 1) and the uncoupled reactions (see Supplementary Information: Material S4)—and the average
427 growth rates are 0.666, 0.773 and 0.813 h^{-1} (one-way ANOVA; p = 7.71E-28) (Fig 8e). Finally, we
428 examined the cause for the observed differences among the flux variances, the number of simple
429 cycles and the enzyme properties of LBRs. The flux variances are primarily explained by the
430 multimeric enzymes and the topologically highly connected LBRs: SLBs are involved in a large
431 number of simple cycles and few element reactions (Fisher's exact test, p = 2.34E-21 and 0.0051,
432 respectively) (Fig 8f). Therefore, the topological vulnerability of these reactions will result in
433 functional variability.

434 **Discussion**

435 We proposed Decrem by building the topologically decoupled metabolic network using SLB
436 decomposition incorporating metabolic regulation by metabolites into GSMs, which
437 approximate the kinetic fluxes of cell state-regulated reactions to constrain the feasible region
438 of optimal flux distribution. Decrem effectively reduces the requirements for multi-omics data
439 for genome-scale metabolic kinetic models. Compared to existing methods, Decrem
440 demonstrates superior performance in predicting metabolic fluxes in three model organisms
441 and growth rate in genome-scale knockout strains of *E. coli*. Therefore, it is an effective method
442 for accurately depicting metabolic responses and exploring the self-adapting regulation
443 mechanism of cellular perturbation.

444 By applying SLB decomposition, the (coupled) element reactions within identified SLBs
445 display high coexpression among multiple growth conditions, indicating coordinated
446 activation of topologically highly coupled reactions. Interestingly, similar approaches have
447 been applied in identifying the non-redundant local functional units of metabolism, i.e., the
448 minimal metabolic pathway or flux tope [35,57]. A topological orthogonality principle was
449 successfully used to design bioengineering strains with minimal interaction between desired
450 product-associated pathways and metabolic components related to biomass synthesis [58]. In
451 addition, several specific topological constraints, such as removing the thermodynamically
452 infeasible loops and decoupling two desired phenotypes, have been applied to GSMs to
453 improve their metabolic production in recent studies [59,60]. But Decrem is the first genome-
454 scale topologically decoupled metabolic model for general applications, which clearly
455 demonstrates how the topological preference of a metabolic network can guide the metabolic
456 flux distribution.

457 To explore effective metabolic dynamic robustness or adaption to internal and external
458 perturbations, metabolic kinetic models have attracted great attention by combining fluxomics,
459 metabolomics and transcriptomics into a unified framework [61-63]. However, construction of the
460 GSM kinetic model is obstructed by the limited knowledge about kinetic parameters, e.g., K_m,
461 K_{cat}, scarcity of metabolic regulators and paired genome-scale multi-omics data [46,61]. To date,
462 the largest metabolic kinetic model, k-ecoli457 of *E. coli*, contains only 457 reactions, 337
463 metabolites and 295 substrate-level regulatory interactions according to the computationally
464 predicted kinetic parameters [61]. Alternatively, by integrating the metabolite-TF regulatory
465 regime, kinetic Decrem can predict growth rates and corresponding fluxes. A unique

466 advantage of our kinetic Decrem is that it relies only on the experimental concentrations of
467 identified essential metabolites, which serve as the key indicators of metabolic states and
468 directly regulate enzyme activities or gene expression. Although some metabolites (e.g., those
469 in glycolysis and the TCA cycle) have long been known to regulate enzyme activities in
470 biochemistry textbooks, not much is known on the genome scale, even in model organisms.
471 On the one hand, the regulatory metabolites of specific pathways in central metabolism prefer
472 to frequently interact with the catalytic enzymes by activating or inactivating the functional
473 domains to synchronously adjust the fluxes [46,64]. On the other hand, several studies revealed
474 that approximately 70% of the total variance in the promoter activity of metabolic genes in
475 the central metabolism of *E. coli* can be explained by growth rate-derived global transcriptional
476 regulation across multiple mutant strains [48,65]. These findings suggest a potential relationship
477 between the concentration of biomass-constituent metabolites and the expression of
478 metabolic genes in the regulons of TFs. This relationship was verified by a recent study in which
479 the identified metabolite concentrations were predicted by quantitative proteomics data [66].
480 Overall, both the topology of metabolic networks and regulatory metabolites are utilized to
481 identify coactivated or key regulated reactions, which produces a minimal set of regulatory
482 constraints to develop the genome-scale kinetic models.

483 Compared to other methods for three model organisms, Decrem not only is the best in
484 recovering the real flux distributions with high accuracy in a wide range of strains but also
485 presents excellent predictions and explanations of the observed growth rate. This
486 demonstrated the strong capability of Decrem to approximate the real intracellular state and
487 to be used for designing high-yield mutant strains in bioengineering and synthetic biology.

488 Decrem shows that there is a strong influence of metabolic network topology in the prediction
489 of flux distributions; this phenomenon is found both in the reconstructed LBRs of Decrem and
490 in the two versions of yeast metabolic models. A possible explanation for this observation is
491 that the optimal flux distribution of a metabolic network is strongly determined by its topology,
492 and the rewired or perturbed network structure will have different feasible regions. Moreover,
493 Decrem could be applied to elucidate important regulatory branch points and the self-
494 adaption of regulation mechanisms for the knockout strains, which is helpful for accurately
495 predicting the potential target genes/reactions for designing bioengineering strains.

496 The main limitation of Decrem is that it requires a medium-scale set matched ^{13}C-MFA
497 flux paired with metabolite concentration data to construct reasonable kinetic models.
498 Alternatively, the experimentally measured K_m and K_{cat} can also be used instead of the ^{13}C-MFA
499 flux. However, the well-constructed kinetic models are convenient for transfer to any other
500 applications. Overall, Decrem, a topology and regulatory network-reinforced metabolic
501 analysis model, can accurately predict phenotypes and uncover the complex regulation of cell
502 metabolism.

503 # Methods

504 ## Topology-decoupled reconstruction of the metabolic model

505 The original metabolic network consists of both coupled coactivated reaction cycles (e.g., TCA
506 cycle) and simple linear chain reactions (e.g., biosynthetic reaction chains). We developed the
507 model, Decrem, to capture the contribution of the coactivated and coupled reactions while
508 preserving the consistency of the linear components. The detailed framework for our method
509 is as follows(Supplementary Fig. 1):

Step 1: Identifying reaction cycle-based coupled substructures

Defining a bipartite graph derived from the metabolic network as $G(V_{m+n}, E)$, where the node set V_{m+n} includes both the m metabolite nodes and n reaction nodes, and the edge set E is composed of all the interactions between the metabolite and reaction nodes. We first built the similarity matrix $A_{n\times n}$ for n metabolic reactions based on the number of topological simple cycles of G. Specifically, the element a_{ij} of $A_{n\times n}$, which indicates the similarity between reactions i and j in the metabolic network, is measured by the number of simple directed cycles between the paired reaction nodes (v_i, v_j) of G. According to the similarity matrix $A_{n\times n}$ above, the BestWCut clustering algorithm [34] is then used to identify the dense substructures (also known as network communities) consisting of highly coupled reaction cycles. Here, the substructures are denoted as $C = \{C_k | k = 1 \dots K\}$, where C_k is a subset of reaction index set $[n] = 1, \dots, \mathrm{n}$, k is the index of substructures C_k, and K is the total number of substructures. If we define $D_i = \sum_{j\in[n]} a_{ij}$ as the weighted out-degree of reaction node v_i , $i \in [n]$, then the cluster degree D_k for subnetwork C_k could be defined as follows:

$$D_k = \sum_{i\in C_k} D_i \tag{1}$$

The generalized weighted cut (WCut) associated with C is obtained by minimizing $WCut(C)$:

$$WCut(C) = \sum_{k=1}^{K} \sum_{k'\neq k} \frac{Cut(C_k, C_{k'})}{D_k} \tag{2}$$

where

$$Cut(C_k, C_{k'}) = \sum_{i\in k} \sum_{j\in k'} a_{i,j} \tag{3}$$

and $C_{k'}$ is the complement set of the cluster C_k

Step 2: Reconstructing the decoupled representation of the identified substructure using a sparse linear basis

Inspired by the minimal metabolic pathways [35], we represented the highly coupled substructures with the SLBs of the null space of the corresponding stoichiometric matrix, which is biologically explained as the minimal and indecomposable coupled components, and satisfies the constraint of thermodynamic and mass balance of element reactions. Unlike infinite ordinary linear basis vectors of the null space of the stoichiometric matrix, there is a unique and globally optimal sparsest basis group of the null space [35,67]. Briefly, the orthonormal null space N_{C_k} is initially defined by singular value decomposition (SVD) for the stoichiometric matrix S_{C_k} of the subnetwork C_k. Here, additional artificial exchange reactions are introduced in S_{C_k} to maintain the mass balance of reactions in the subnetwork C_k (more details can be found in Supplementary Information: Method S1). Then, the column vectors of the orthonormal null space N_{C_k} are iteratively replaced by the minimal element reactions that span the removed subspace of vectors. This process is repeated until all the nonzero entries in N_{C_k} converged on a minimum [35]. Here, we utilized the advantage of sparse regularization of the null space N_{C_k} to solve the minimum L1-norm of the null space N_{C_k} of S_{C_k} [67]. The detailed process is showcased in Supplementary Information: Method S1, in which $N_{C_k}^S$ is a minimal sparse basis representation of N_{C_k} in at most $2r_k$ linear programming optimization runs (where $r_k = l_k - rank(S_{C_k})$, l_k is the number of columns (reactions) in S_{C_k}). If we assume $x \in \mathrm{R}^n$ and then each linear programming problem can be formulated as follows:

$$min_{(x,v_{C_k}^m)} \sum_i x_i \tag{4}$$

$$\text{s.t. } S_{C_k}.v_{C_k}^m = 0 \tag{5}$$

$$v_{C_k}^m \leq x \tag{6}$$

$$-v_{C_k}^m \leq x \tag{7}$$

$$lb \leq v_{C_k}^m \leq ub \tag{8}$$

$$w^T.P_{C_k}^{N_m}.v_{C_k}^m > \zeta \ \vee \ w^T.P_{C_k}^{N_m}.v_{C_k}^m < -\zeta \tag{9}$$

where $v_{C_k}^m$ is the SLB of the null space of S_{C_k}at the mth running, and lb and ub are the lower and upper bounds of $v_{C_k}^m$, respectively. The constraint of formula (9) ensures that $v_{C_k}^m$ is linearly independent of the previous $m-1$ SLBs, and w represents a vector of random weights. Here, we employed uniform random weights, and ζ is a small positive constant, e.g., 1.0^{-3}. $P_{C_k}^{N_m}$ is a projection matrix onto the sparse null space $N_{C_k}^{S_m}$, and $N_{C_k}^{S_m} = [v_{C_k}^1, v_{C_k}^2, \dots, v_{C_k}^{m-1}]$. More details of this process are provided in Supplementary Information: Method S1.

The SLBs of S_{C_k}, assumed as $N_{C_k}^S = [v_{C_k}^1, v_{C_k}^2, \dots, v_{C_k}^{r_k}]$, can be solved by the above algorithm. Then, the highly coupled subnetwork C_k can be substituted to construct the mutually independent "linear basis reactions" with the following formula:

$$S_{C_k}^{IBR} = S_{C_k}^*.N_{C_k}^{S^*} \tag{10}$$

where $S_{C_k}^*$ and $N_{C_k}^{S^*}$ are derived from S_{C_k} and $N_{C_k}^S$ after removing the artificial exchange reactions, respectively, and IBR indicates the reconstructed independent "linear basis reactions".

Step 3: Establishing prediction for the decoupled metabolic model

In this step, we reformulated FBA to adapt the reconstructed decoupled metabolic network S^{IR}. The key problem is to determine the flux bounds for each "linear basis reaction":

$$\max_{v^{IR}} cv^{IR} \tag{11}$$

$$\text{s.t.} \quad S^{IR}.v^{IR} = 0 \tag{12}$$

$$S^{IR} = \left[S^{NC}, S_{C_1}^{IBR}, \dots, S_{C_K}^{IBR}\right] = \left[S^{NC}, S_{C_1}^{*} N_{C_1}^{S^*}, \dots, S_{C_K}^{*} N_{C_K}^{S^*}\right] \tag{13}$$

$$v^{IR} = \left[v^{NC}, v_{C_1}^{IBR}, \dots, v_{C_K}^{IBR}\right]^T \tag{14}$$

$$v_{C_k}^{IBR} = \left[v_{C_k}^{1}, \dots, v_{C_k}^{r_k}\right]^T \tag{15}$$

$$lb^{NC} \le v^{NC} \le ub^{NC} \tag{16}$$

$$\max\left(f\left(lb_{C_k}^{i}, NZ\left(N_{C_k}^{i}\right)\right)./f_N\right) \le v_{C_k}^{i} \le \min\left(f\left(ub_{C_k}^{i}, NZ\left(N_{C_k}^{i}\right)\right)./f_N\right) \tag{17}$$

$$f_N = f\left(N_{C_k}^{i}, NZ\left(N_{C_k}^{i}\right)\right) \tag{18}$$

where c and v^{IR} represent the objective function and optimal metabolic flux of S^{IR}, respectively. The superscript IR represents the linearly independent reaction-derived metabolic network, and NC represents the uncoupled reactions (which are composed of linear reaction chains) of the original metabolic network. r_k is the number of columns of $N_{C_k}^{S^*}$, and K is the total number of highly coupled reaction subnetworks identified in *step* 1. $N_{C_k}^{i}$ is the ith column of $N_{C_k}^{S^*}$, and $lb_{C_k}^{i}$ and $ub_{C_k}^{i}$ are the lower and upper bounds of the reaction indicated by $N_{C_k}^{i}$, respectively. $v_{C_k}^{IBR}$ is the flux vector of all the "linear basis reactions" of subnetwork C_k, and $v_{C_k}^{i}$ is the ith flux of $v_{C_k}^{IBR}$. Among them, *i* ranges from 1 to r_k, and k ranges from 1 to K. Then, the fluxes of reactions in the original metabolic network (element reactions) of linear basis vectors will be recovered by the formula $N_{C_k}^{S^*}.v_{C_1}^{IBR}$ according the optimal solution of Decrem outlined above.

The function $NZ(.)$ takes the index of nonzero elements of the input vector, and the function $f(v, I)$ takes elements indexed by the input indicator I from the input vector v. Therefore, $f(N_{C_k}^{i}, NZ(N_{C_k}^{i}))$ represents the nonzero partition coefficient of each element

598 reaction composed of the ith SLB of subnetwork C_k, which is indicated by the nonzero terms

599 of $N_{C_k}^i$. The formula $f(lb_{C_k}^i, NZ(N_{C_k}^i)$ represents the lower bounds of the element reactions

600 composed of the ith SLB of the subnetwork C_k, and $f(ub_{C_k}^i, NZ(N_{C_k}^i)$ represents the upper

601 bounds. In summary, Decrem forces the metabolic fluxes of highly coupled reactions to be

602 incorporated in optimization by representing them as independent linear basis vectors.

603 **Gene expression estimation**

604 Forty-five candidate global and local regulatory metabolites of *E. coli* are collected through

605 KEGG pathway analysis and a literature survey [50,51,68]. These candidates are then categorized

606 into two clusters by hierarchical clustering analysis over the expression profile of their

607 concentrations across 24 mutant strains. Then, the possible regulatory relationship between

608 two identified metabolite groups and the expression of 85 genes are inferred by PLSR [69], which

609 selects the nonredundant and independent factors to maximize the correlation of response

610 variables, using stepwise principal component regression. Furthermore, PLS is used to discover

611 the fundamental quantitative relations between two (expression) observation variable sets,

612 and the general underlying model of multivariate PLS is as follows:

613 $$X = TP^T + E \tag{19}$$

614 $$Y = UQ^T + F \tag{20}$$

615 where X is an $n \times m$ matrix of predictors (each group metabolites concentrations), and Y is

616 an $n \times p$ matrix of responses (genes expression); T and U are $n \times l$ matrices and

617 projections of X (the X score, component or factor matrix) and projections of Y (the Y

618 scores), respectively. P and Q are $m \times l$ and $p \times l$ orthogonal loading matrices,

619 respectively. Matrices E and F are the error terms, assumed to be independent and

identically distributed random normal variables. The decomposition of X and Y was performed to maximize the covariance between T and U.

Finally, the significant metabolite profile and corresponding explicable gene-metabolite regulatory relationships are filtered by setting the proper correlation threshold and the statistical test is built based on random sampling (see Supplementary Information: Material S3).

The metabolite concentration-derived linearized kinetic model

Step 1: The metabolic kinetic model

In this section, we derived a complete reversible rate law for arbitrary reactant stoichiometries. When considering the constraint of thermodynamics and metabolite regulation [46,70], we can rewrite the Michaelis–Menten kinetics for a reversible reaction S <=> P as follows:

$$\mathrm{v}([S],[P],[A],[I],[E]) = [E]\frac{k^{+}[S]/k_m^s - k^{-}[P]/k_m^p}{1+[S]/k_m^s+[P]/k_m^p}\prod_u \frac{[A_u]/k_A^u}{1+[A_u]/k_A^u}\Bigg/\prod_v \frac{1}{1+[I_v]/k_I^v} \quad (21)$$

where $[E]$ is the concentration of enzyme active sites, $[S]$ and $[P]$ are the concentrations of substrates and products, k_m^s and k_m^p are the affinities of the reactants for this enzyme, k^{+} and k^{-} are the maximal forward and reverse catalytic rate constants, $[A]$ and $[I]$ are the concentrations of activators and inhibitors, and $k_a^{A_u}$ and $k_i^{I_v}$ are their corresponding affinities. The positive and negative terms in the numerator are associated with the forward and backward rates, respectively.

We next applied the metabolic thermodynamics constraints given by the Haldane relationship to simplify the term for the backward rate:

$$k^{eq} = \frac{k^{+}k_m^p}{k^{-}k_m^s} = \frac{[P_0]}{[S_0]} = e^{-\Delta_r G'^{\circ}/RT} \quad (22)$$

where $\Delta_r G'^{\circ}$ is the standard Gibbs energy of the reaction (and does not depend on the

enzyme parameters). Using this equality with the above rate law, we can obtain the following:

$$\mathrm{v}([S],[P],[A],[I],[E])$$

$$= ([E]k^{+})\frac{[S]/k_m^s\,(1-[P]/[S]\,e^{-\Delta_r G'^{\circ}/RT})}{1+[S]/k_m^s+[P]/k_m^p}\prod_u \frac{[A_u]/k_A^u}{1+[A_u]/k_A^u}\Bigg/\prod_v \frac{1}{1+[I_v]/k_I^v} \qquad (23)$$

Step 2: The optimal strategy of the linearized kinetic model

In this section, we sought the simplified representation of equation (23) based only on the associated metabolite concentrations. First, we took the negative logarithmic operation of both sides of equation (23) and reorganized the right-hand terms:

$$-\log\big(\mathrm{v}([S],[P],[A],[I],[E])\big) = \underbrace{-\log([E])}_{\text{Enzyme term}} \underbrace{-\log\big(1-[P]/[S]\,e^{-\Delta_r G'^{\circ}/RT}\big)+\sum_v \log(1+[I_v])}_{\text{Metabolite associated terms}}$$

$$\underbrace{-\log(k^{+})-\sum_v k_I^v}_{\text{Kinetic constants}} \underbrace{+\sum_u \log\left(1+\frac{k_A^u}{[A_u]}\right)+\log\left(1+\frac{k_m^s}{[S]}\left(1+\frac{[P]}{k_m^p}\right)\right)}_{\text{Nonlinear terms}} \qquad (24)$$

This model can be solved by collecting the corresponding kinetic parameters, enzyme expression and metabolite concentrations; however, those matched data are often not available in practice, and the metabolic regulators are often unknown. An alternative is to approximate the optimal parameters according to the machine learning method. Specifically, the cell growth state-regulated enzyme expression can be represented as the linear combination of the concentrations of biomass composition (see Supplementary Information: Material S3), which can be marked as $[E]=\sum_{i=1}^{m} w_i[S_i]+b$, where S_i is the concentration of the ith indicator metabolite of the biomass identified in the "Gene expression estimation" section. In addition, we reexamined the nonlinear terms of equation (24) based on the knowledge that systemic experimental analysis revealed that $[S]$ was $\geq k_m^s$ for almost all

of the metabolites in the central metabolism of three model organisms [47]; hence, we have $\frac{k_m^s}{[S]} \leq 1$.

Situation 1: For the growth state-regulated irreversible reaction, if we have $\frac{[P]}{k_m^p} \to 0$, and $\frac{k_A^u}{[A_u]} \leq 1$, then the Taylor expansion of equation (24) is as follows:

$$-\log\big(\mathrm{v}([S],[P],[A],[I],[E])\big)$$

$$= -\log\left(\sum_{i=1}^{m} w_i[S_i] + b\right) - \log\big(1 - [P]/[S]\, e^{-\Delta_r G'^{\circ}/RT}\big) + \sum_{v} \log(1 + [I_v])$$

$$-\log(\boldsymbol{k}^{+}) - \sum_{v} \boldsymbol{k}_{\boldsymbol{I}}^{\boldsymbol{v}} + \frac{\boldsymbol{k}_{\boldsymbol{m}}^{\boldsymbol{s}}}{[S]} + \sum_{u}\left(\frac{\boldsymbol{k}_{\boldsymbol{A}}^{\boldsymbol{u}}}{[A_u]}\right) + O\left(\frac{k_m^s}{[S]}, \frac{k_A^u}{[A_u]}\right) \tag{25}$$

Situation 2: For the generalized reversible reaction in which the enzyme is substrate or product saturated (or both): $\frac{k_m^s}{[S]} \leq 1$ and $\frac{k_m^p}{[P]} \leq 1$. Assuming $v > 0$, we have the following:

$$-\log(k^{+}) + \log\left(1 + \frac{k_m^s}{[S]}\left(1 + \frac{[P]}{k_m^p}\right)\right) \approx \log\left(\frac{1}{k^{+}} + \frac{1}{k^{+}}\frac{[P]}{[S]}\frac{k_m^s}{k_m^p}\right) = \log\left(\frac{1}{k^{+}} + \frac{[P]/[S]}{k^{eq}}\frac{1}{k^{-}}\right) \tag{26}$$

Here, $[P]/[S] < k^{eq}$ according to the reaction Gibbs energy, $\frac{1}{k^{+}} \leq 1$ and $\frac{1}{k^{-}} \leq 1$ for the significantly regulated reactions, and $\frac{1}{k^{+}} + \frac{[P]/[S]}{k^{eq}}\frac{1}{k^{-}} - 1 < 1$; so, the Taylor expansion of (26) is as follows:

$$\log\left(\frac{1}{k^{+}} + \frac{[P]/[S]}{k^{eq}}\frac{1}{k^{-}}\right) = \log\left(1 + \frac{1}{k^{+}} + \frac{[P]/[S]}{k^{eq}}\frac{1}{k^{-}} - 1\right)$$

$$= \frac{1}{k^{+}} + \frac{[P]/[S]}{k^{eq}}\frac{1}{k^{-}} - 1 + O\left(\frac{1}{k^{+}} + \frac{[P]/[S]}{k^{eq}}\frac{1}{k^{-}}\right) \tag{27}$$

$$-\log\big(\mathrm{v}([S],[P],[A],[I],[E])\big)$$

$$\approx -\log\left(\sum_{i=1}^{m} w_i[S_i] + b\right) - \log\big(1 - [P]/[S]\, e^{-\Delta_r G'^{\circ}/RT}\big) + \sum_{v} \log(1 + [I_v])$$

$$-\sum_{v} \boldsymbol{k}_{\boldsymbol{I}}^{\boldsymbol{v}} + \sum_{u}\left(\frac{\boldsymbol{k}_{\boldsymbol{A}}^{\boldsymbol{u}}}{[A_u]}\right) + \frac{1}{\boldsymbol{k}^{+}} + \frac{[P]/[S]}{k^{eq}}\frac{1}{\boldsymbol{k}^{-}} - 1 \tag{28}$$

where all the bold variables in equations (25) and (28) are kinetic parameters. Similarly, when

682 $v > 0$, we have $-\log(k^{-}) + \log\left(1 + \frac{k_m^p}{[P]}\left(1 + \frac{[S]}{k_m^S}\right)\right) \approx \log\left(\frac{1}{k^{-}} + \frac{[S]/[P]}{k^{eq}}\frac{1}{k^{+}}\right)$. This result can

683 be expanded to multisubstrate/multiproduct reactions.

684 Overall, when we neglected the infinitesimal of higher order, the identified regulators and

685 optimal kinetic parameters in models (25) and (28) can be solved with a Bayesian linear

686 regression-derived model selection method, i.e., the SIMMER package [46], which first builds the

687 regulator-free kinetic model, and then uses the BIC criteria to identify the real regulators from

688 the candidate metabolite set (PG metabolites) by evaluating the additional ability to account

689 for the empirical fluxes when integrating a new candidate regulator into the kinetic model in

690 an incremental way (see Supplementary Information: Methods S3). Subsequently, the optimal

691 model is expanded to any other application.

692 ***Step 3: The kinetic Decrem***

693 Finally, the parameterized kinetic model is utilized to describe the growth-associated key

694 regulated reactions in central metabolism. The genome-scale flux distribution is predicted by

695 the kinetic regulated flux-constrained Decrem method, i.e.,

696 $$\max cv_{obj} \tag{29}$$

697 $$\text{s.t.}\ \ S.v = 0 \tag{30}$$

698 $$lb^{NC} \leq v^{NC} \leq ub^{NC} \tag{31}$$

699 $$\max\left(f\left(lb_{C_k}^i, NZ\left(N_{C_k}^i\right)\right)./ff_N\right) \leq v_{C_k}^i \leq \min\left(f\left(ub_{C_k}^i, NZ\left(N_{C_k}^i\right)\right)./f_N\right) \tag{32}$$

700 $$f_N = f\left(N_{C_k}^i, NZ\left(N_{C_k}^i\right)\right) \tag{33}$$

701 $$v_j^{KF} - \delta \leq v_j \leq v_j^{KF} + \delta \tag{34}$$

702 $$i = 1, \dots, r_k; j = 1, \dots, \mathrm{m}\,; k = 1, \dots, K$$

703 where v_j^{KF} is the j-th kinetic flux of the m key regulation reactions, and δ is the tolerance

of kinetic fluxes. The i , k , NC, $lb_{C_k}^i$, $NZ\left(N_{C_k}^i\right)$ values can be found in *step* 3 of the section "Topology-decoupled reconstruction of the metabolic model". Among them, *i* ranges from 1 to r_k, *j* ranges from 1 to m, and k ranges from 1 to K.

Data Availability

All data used are publicly available. The original and reconstructed metabolic models are available online: original metabolic models are available at BIGG models http://bigg.ucsd.edu/models , and reconstructed metabolic models of the three reconstructed models, iAF1260, iMM904, iDN750, are available at https://github.com/lgyzngc/Decrem-1.0.git. All used exchange reactions, nutrient uptake, experimental growth rates, 13C fluxes and gene expression for Decrem modeling and metabolic simulation are found in supplementary tables. And the LS-MS data is sourced from https://www.ebi.ac.uk/biostudies for genome-scale mutant strains of E. coli.

Code Availability

Decrem is implemented as a Matlab package, the source code, user tutorial and demo are available at: https://github.com/lgyzngc/Decrem-1.0.git.

References

1. Thiele, I. & Palsson, B.O. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* **5**, 93-121 (2010).
2. Feist, A.M.*, et al.* A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* **3**, 121 (2007).
3. Reed, J.L., Vo, T.D., Schilling, C.H. & Palsson, B.O. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol* **4**, R54 (2003).
4. Ishii, N.*, et al.* Multiple high-throughput analyses monitor the response of E. coli to perturbations. *Science* **316**, 593-597 (2007).
5. Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. & Sauer, U. Multidimensional optimality of microbial metabolism. *Science* **336**, 601-604 (2012).
6. Deutscher, D., Meilijson, I., Kupiec, M. & Ruppin, E. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nat Genet* **38**, 993-998 (2006).

7. Monk, J.M., *et al.* Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A* **110**, 20338-20343 (2013).
8. Lempp, M., *et al.* Systematic identification of metabolites controlling gene expression in E. coli. *Nat Commun* **10**, 4463 (2019).
9. Kim, P.J., *et al.* Metabolite essentiality elucidates robustness of Escherichia coli metabolism. *Proc Natl Acad Sci U S A* **104**, 13638-13642 (2007).
10. Fuhrer, T., Zampieri, M., Sevin, D.C., Sauer, U. & Zamboni, N. Genomewide landscape of gene-metabolome associations in Escherichia coli. *Mol Syst Biol* **13**, 907 (2017).
11. Barabási, A.-L. & Oltvai, Z.N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101-113 (2004).
12. Oltvai, Z.N. & Barabási, A.-L. Life's Complexity Pyramid. *Science* **298**, 763-764 (2002).
13. Albert, R., Jeong, H. & Barabasi, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378-382 (2000).
14. Li, G., Cao, H. & Xu, Y. Structural and functional analyses of microbial metabolic networks reveal novel insights into genome-scale metabolic fluxes. *Brief Bioinform* **20**, 1590-1603 (2019).
15. Buescher, J.M., *et al.* Global network reorganization during dynamic adaptations of *Bacillus subtilis* metabolism. *Science* **335**, 1099-1103 (2012).
16. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. & Barabasi, A.L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551-1555 (2002).
17. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. & Barabasi, A.L. The large-scale organization of metabolic networks. *Nature* **407**, 651-654 (2000).
18. Wagner, A. & Fell, D.A. The small world inside large metabolic networks. *Proc. R. Soc. Lond. B. Biol. Sci.* **268**, 1803-1810 (2001).
19. Pertea, M., Ayanbule, K., Smedinghoff, M. & Salzberg, S.L. OperonDB: a comprehensive database of predicted operons in microbial genomes. *Nucleic Acids Res.* **37**, D479-D482 (2009).
20. Jacob, F. The Birth of the Operon. *Science* **332**, 767-767 (2011).
21. Che, D., Li, G., Mao, F., Wu, H. & Xu, Y. Detecting uber-operons in prokaryotic genomes. *Nucleic Acids Res.* **34**, 2418-2427 (2006).
22. Dam, P., Olman, V., Harris, K., Su, Z. & Xu, Y. Operon prediction using both genome-specific and general genomic information. *Nucleic Acids Res.* **35**, 288-298 (2007).
23. Jacob, F., Perrin, D., Sánchez, C., Monod, J. & Edelstein, S. The operon: a group of genes with expression coordinated by an operator. *C.R.Acad. Sci. Paris* **250**, 1727-1729 (1960).
24. Kim, J., *et al.* Flux balance analysis of primary metabolism in the diatom Phaeodactylum tricornutum. *Plant J* **85**, 161-176 (2016).
25. O'Brien, E.J., Lerman, J.A., Chang, R.L., Hyduke, D.R. & Palsson, B.Ø. Genome‐scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology* **9**, 1-13 (2013).
26. Lim, H.N., Lee, Y. & Hussein, R. Fundamental relationship between operon organization and gene expression. *Proceedings of the National Academy of Sciences* **108**, 10626 (2011).
27. Aragon, J.J. & Sols, A. Regulation of enzyme activity in the cell: effect of enzyme concentration. *FASEB J* **5**, 2945-2950 (1991).
28. Zhao, X., *et al.* Substrate-driven chemotactic assembly in an enzyme cascade. *Nat Chem* **10**, 311-317 (2018).

29. Sweetlove, L.J. & Fernie, A.R. The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation. *Nat Commun* **9**, 2136 (2018).
30. Pandey, V., Hadadi, N. & Hatzimanikatis, V. Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLoS Comput Biol* **15**, e1007036 (2019).
31. van der Knaap, J.A. & Verrijzer, C.P. Undercover: gene control by metabolites and metabolic enzymes. *Genes Dev* **30**, 2345-2369 (2016).
32. Kochanowski, K.*, et al.* Few regulatory metabolites coordinate expression of central metabolic genes in Escherichia coli. *Mol Syst Biol* **13**, 903 (2017).
33. Dourado, H. & Lercher, M.J. An analytical theory of balanced cellular growth. *Nat Commun* **11**, 1226 (2020).
34. Meilă, M. & Pentney, W. Clustering by weighted cuts in directed graphs. in *Proceedings of the 2007 SIAM International Conference on Data Mining* 135-144 (2007).
35. Bordbar, A.*, et al.* Minimal metabolic pathway structure is consistent with associated biomolecular interactions. *Mol Syst Biol* **10**, 737 (2014).
36. Schuster, S., Fell, D.A. & Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* **18**, 326-332 (2000).
37. Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC bioinformatics* **11**, 489-489 (2010).
38. Orth, J.D., Thiele, I. & Palsson, B.Ø. What is flux balance analysis? *Nature Biotechnology* **28**, 245 (2010).
39. Lewis, N.E.*, et al.* Omic data from evolved E. coli are consistent with computed optimal growth from genome‐scale models. *Molecular Systems Biology* **6**, 390 (2010).
40. Kim, J. & Reed, J.L. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome Biology* **13**, R78 (2012).
41. Gonzalez, J.E., Long, C.P. & Antoniewicz, M.R. Comprehensive analysis of glucose and xylose metabolism in Escherichia coli under aerobic and anaerobic conditions by 13C metabolic flux analysis. *Metabolic Engineering* **39**, 9-18 (2017).
42. Rühl, M., Le Coq, D., Aymerich, S. & Sauer, U. 13C-flux Analysis Reveals NADPH-balancing Transhydrogenation Cycles in Stationary Phase of Nitrogen-starving Bacillus subtilis. *Journal of Biological Chemistry* **287**, 27959-27970 (2012).
43. Moxley, J.F.*, et al.* Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proceedings of the National Academy of Sciences* **106**, 6477-6482 (2009).
44. Blank, L.M., Kuepfer, L. & Sauer, U. Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biology* **6**, R49 (2005).
45. Long, M.R. & Reed, J.L. Improving flux predictions by integrating data from multiple strains. *Bioinformatics* **33**, 893-900 (2016).
46. Hackett, S.R.*, et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, aaf2786 (2016).
47. Park, J.O.*, et al.* Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nature Chemical Biology* **12**, 482 (2016).
48. Kochanowski, K.*, et al.* Few regulatory metabolites coordinate expression of central metabolic

genes in Escherichia coli. *Molecular Systems Biology* **13**, 903 (2017).

49. Ishii, N.*, et al.* Multiple High-Throughput Analyses Monitor the Response of E. coli to Perturbations. *Science* **316**, 593-597 (2007).
50. Chang, A.*, et al.* BRENDA in 2017: new perspectives and new tools in BRENDA. *Nucleic Acids Research* **45**, D380-D388 (2016).
51. Kothari, A.*, et al.* The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Research* **45**, D543-D550 (2016).
52. Zelezniak, A.*, et al.* Machine Learning Predicts the Yeast Metabolome from the Quantitative Proteome of Kinase Knockouts. *Cell Systems* **7**, 269-283.e266 (2018).
53. Voet, D. & Voet, J.G. *Biochemistry, 4th Edition*, (John Wiley & Sons, 2010).
54. Tibshirani, R. Regression shrinkage and selection via the lasso. *J Journal of the Royal Statistical Society: Series B* **58**, 267-288 (1996).
55. Guyon, I. & Elisseeff, A.J.J.o.m.l.r. An introduction to variable and feature selection. **3**, 1157-1182 (2003).
56. Braakman, R. & Smith, E. The compositional and evolutionary logic of metabolism. *Phys Biol* **10**, 011001 (2013).
57. Gerstl, M.P., Muller, S., Regensburger, G. & Zanghellini, J. Flux tope analysis: studying the coordination of reaction directions in metabolic networks. *Bioinformatics* **35**, 266-273 (2019).
58. Pandit, A.V., Srinivasan, S. & Mahadevan, R. Redesigning metabolism based on orthogonality principles. *Nature Communications* **8**, 15188 (2017).
59. Wang, L., Dash, S., Chan, S.H.J. & Maranas, C.D. Accelerating flux balance calculations in genome-scale metabolic models by localizing the application of loopless constraints. *Bioinformatics* **34**, 4248-4255 (2018).
60. Venayak, N., von Kamp, A., Klamt, S. & Mahadevan, R. MoVE identifies metabolic valves to switch between phenotypic states. *Nature Communications* **9**, 5332 (2018).
61. Khodayari, A. & Maranas, C.D. A genome-scale Escherichia coli kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nature Communications* **7**, 13806 (2016).
62. Khodayari, A., Zomorrodi, A.R., Liao, J.C. & Maranas, C.D. A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data. *Metabolic Engineering* **25**, 50-62 (2014).
63. Stalidzans, E., Seiman, A., Peebo, K., Komasilovs, V. & Pentjuss, A. Model-based metabolism design: constraints for kinetic and stoichiometric models. *Biochemical Society Transactions*, BST20170263 (2018).
64. Piazza, I.*, et al.* A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **172**, 358-372.e323 (2018).
65. Reznik, E.*, et al.* Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. *Cell Reports* **20**, 2666-2677 (2017).
66. Moxley, J.F.*, et al.* Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. *Proc Natl Acad Sci U S A* **106**, 6477-6482 (2009).
67. Saa, P.A. & Nielsen, L.K. Fast-SNP: a fast matrix pre-processing algorithm for efficient loopless flux optimization of metabolic models. *Bioinformatics* **32**, 3807-3814 (2016).

68. Tanabe, M., Furumichi, M., Kanehisa, M., Kawashima, M. & Sato, Y. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research* **44**, D457-D462 (2015).
69. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**, 109-130 (2001).
70. Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences* **110**, 10039-10044 (2013).

Acknowledgements

The authors thank Schellenberger J's open-source Matlab package: Cobra, and Saa PA's open-source Matlab package: Fast_SNP for the integration of Decrem. This work was supported by the National Natural Science Foundation of China (61872418) and Natural Science Foundation of Jilin Province (20180101050JC).

Conflict of interest

The authors declare no conflict of interest.

Materials & Correspondence

Correspondence and requests for materials should be addressed to G.L., W.D. or H.C.

Supplementary information

Supplementary Fig. 1: The workflow of Decrem; **Supplementary Fig. 2**: The accumulation flux distribution of according the Yeast iMM904 model (red) and reconstructed decoupled model (blue) by the flux variability analysis; **Supplementary Fig. 3**: Comparison of activated reactions of TCA cycle among 13C MFA (A), Decrem (B) and FBA (C). Solid line indicate the activated reactions and dash line indicate the reactions having a zero flux; **Supplementary Fig. 4**: The workflow of kinetic regulation model of Decrem; **Supplementary Fig. 5**: The concentration profiles of 45 candidate regulating metabolites according to hierarchical clustering, under 24 mutant strains. The red and blue rectangles mark the identified two high-correlation groups;

890 **Supplementary Fig. 6**: The consistency of 32 genes between measured and predicted
891 expression levels by the PLSR methods; **Supplementary Fig. 7:** The growth rate distribution of
892 the 1030 metabolic genes knockout strains predicted by the conventional methods: FBA,
893 MOMA and Decrem, respectively, without any external flux constraints; **Material S1**: The data
894 processing in the section of "Prediction of metabolic fluxes in response to environmental
895 changes"; **Material S2**: The data processing in the section of "Prediction of metabolic fluxes
896 and growth rates in yeast knockout strains"; **Material S3**: The workflow and data processing in
897 the section of "Building a metabolic kinetic model incorporating metabolite-TF regulation";
898 **Material S4**: The workflow and data processing in the section of "Growth rate estimation for
899 E. coli genome-scale deletion strains with kinetic Decrem"; **Method S1**: The Decrem algorithm;
900 **Method S2**: The Parsimonious Decrem; **Method S3**: Identifying the high confidence regulators
901 by Bayesian model section method.
902
903 **Supplementary Table 1**. The annotation of reconstruction decoupled metabolic models,
904 reconstructed models can be found at Data availability.
905
906 **Supplementary Table 2**. Nutrient uptake reactions for three microorganisms under six growth
907 conditions.
908
909 **Supplementary Table 3**. The 13C MFA fluxes for three microorganisms under six growth
910 conditions.
911

912 **Supplementary Table 4**. The dataset used to predict the fluxes and growth rates for the Yeast
913 38 deletion strains.
914
915 **Supplementary Table 5**. The correlation statistic between the experimental and predicted
916 metabolic fluxes and growth rates for the Yeast 38 deletion strains.
917
918 **Supplementary Table 6**. The dataset used to build and validate the hierarchical regulation
919 model.
920
921 **Supplementary Table 7**. The dataset used to construct kinetic Decrem to predict the growth
922 rates on genome-scale deletion strains.
923