Supplementary Data

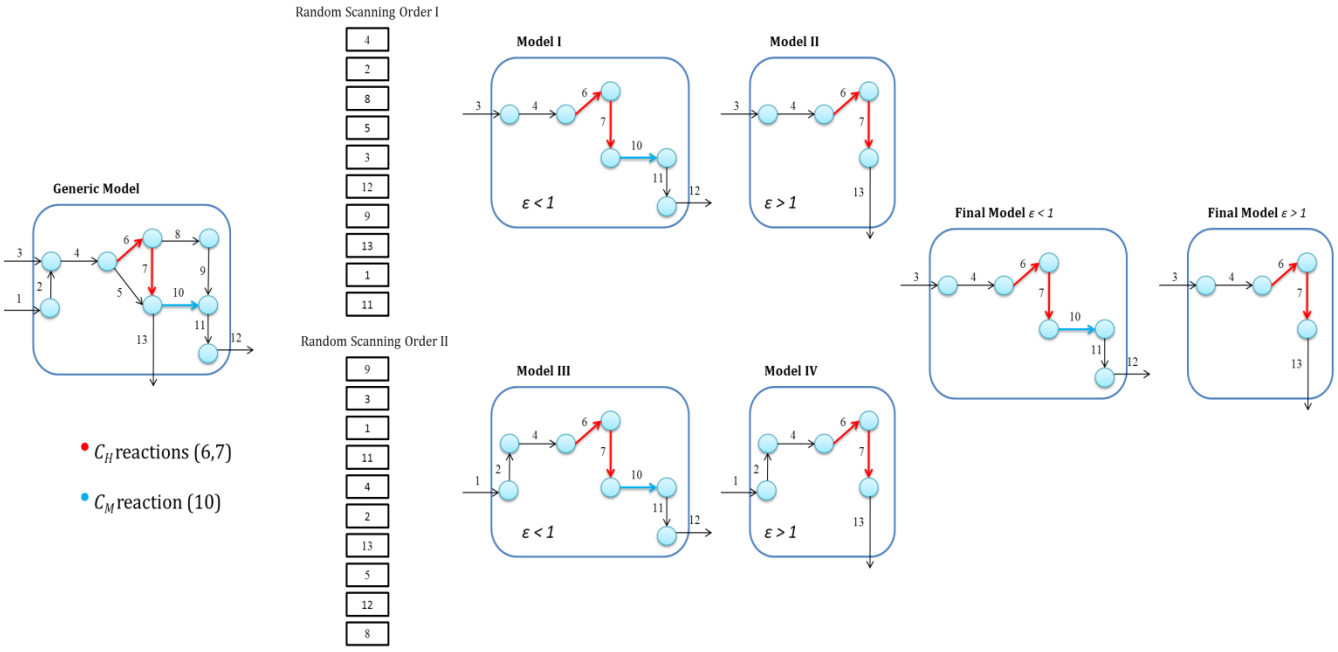

Figure 1 A simplified example that depicts the MBA workflow. Starting from a generic toy model and toy C_H and the C_M sets, a random scanning order aiming to delete non-core reactions is chosen. The resulting model in each such individual scanning procedure is affected, to some extent, by the scanning order and by the optimization criterion. Models I and III are constructed with $\varepsilon<1$, (i.e., it favors to preserve C_M reactions than removing non-core reactions) and therefore consist of reactions 10, 11, and 12, as opposed to models II and IV that are constructed with $\varepsilon>1$. The removal of reactions 5, 8, and 9 and the inclusion of reaction 4 are not affected either by the scanning order or by the optimization criterion. On the other hand, the inclusion of reactions 1, 2, and 3 depends on the scanning order, such that for 1/3 of the random scanning orders (i.e., scanning orders in which reaction 3 precedes both reaction 1 and reaction 2) the resulting model would consist of reactions 1 and 2, and for 2/3 of the random scanning orders (i.e., scanning orders in which either reaction 1 or 2 precede reaction 3) the resulting model would consist of reaction 3 instead of reactions 1 and 2. The final model used is hence produced using the aggregative approach described in the main text, which sums over the random selections to express the most likely reactions' set.

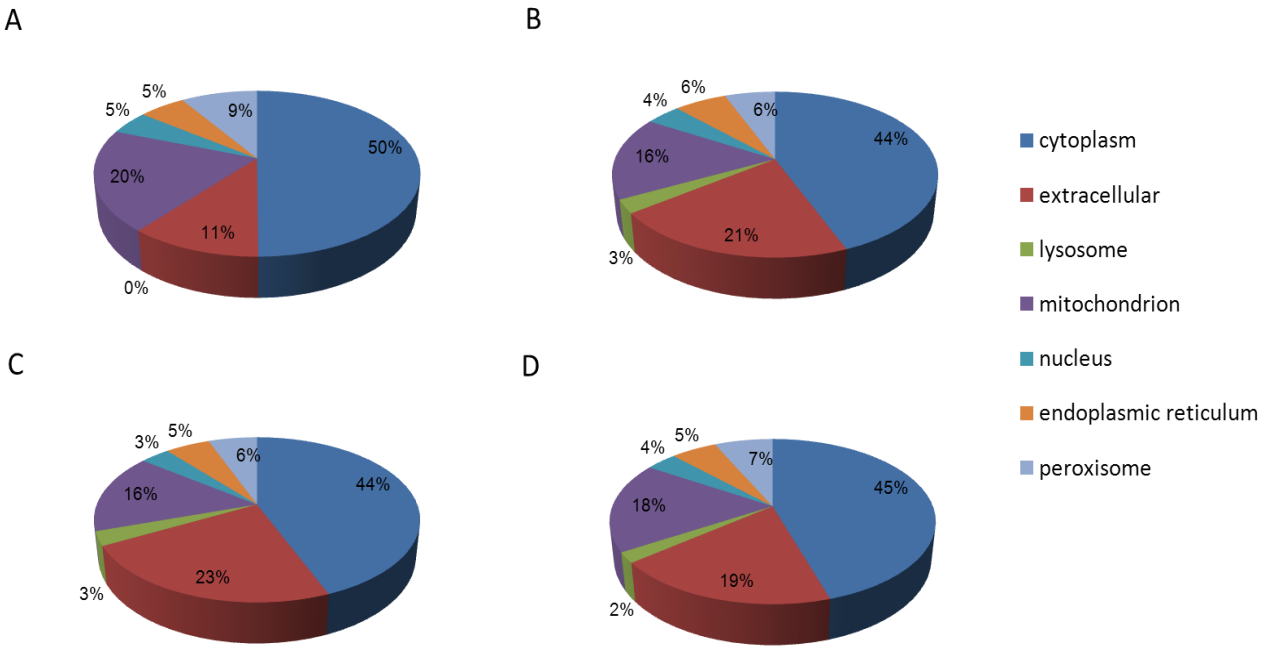

Figure 2 The compartmentalization of the metabolites comprising the input sets, that is, (A) the core and (B) the generic model, and the output sets, that is, (C) the non-core liver metabolites and (D) the liver model.

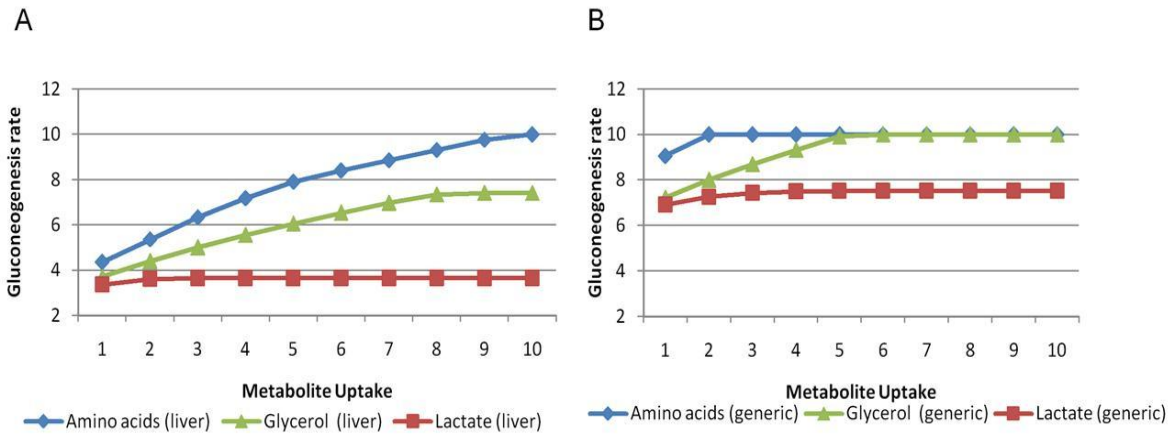

Figure 3 The gluconeogenesis rate obtained in the (A) liver model and the (B) generic model with an increasing uptake rate of glucogenic amino acids, lactate or glycerol.

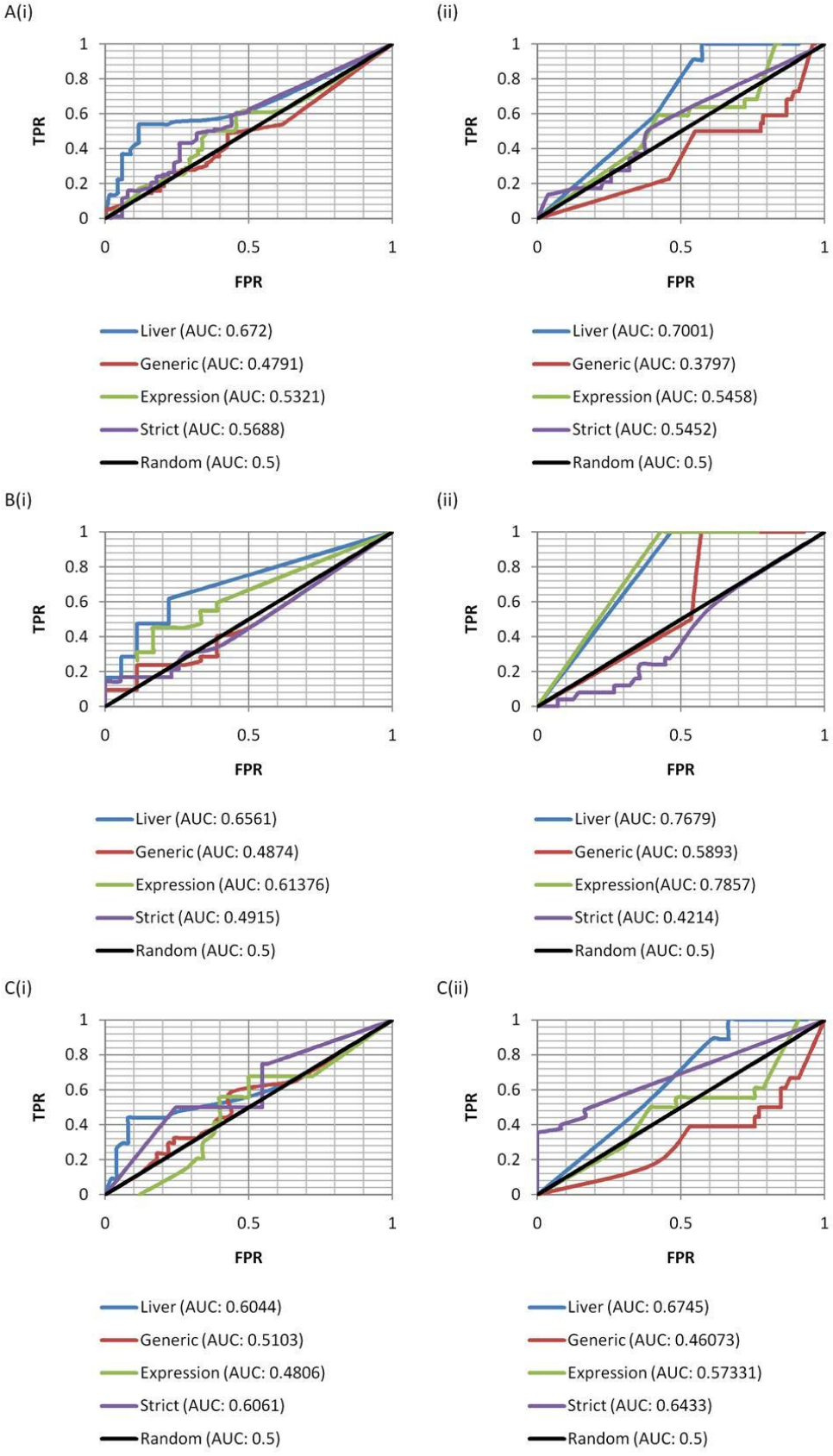

Figure 4 Prediction of experimental hepatic flux data. The ROC curves (and the resulting mean AUCs) of all classifiers are presented separately for predicting (i)

increasing and (ii) decreasing fluxes, for (A) all measured fluxes (using cross validation); (B) exchange fluxes (setting the internal fluxes to their measured values); (C) inner fluxes (setting the exchange fluxes to their measured values).

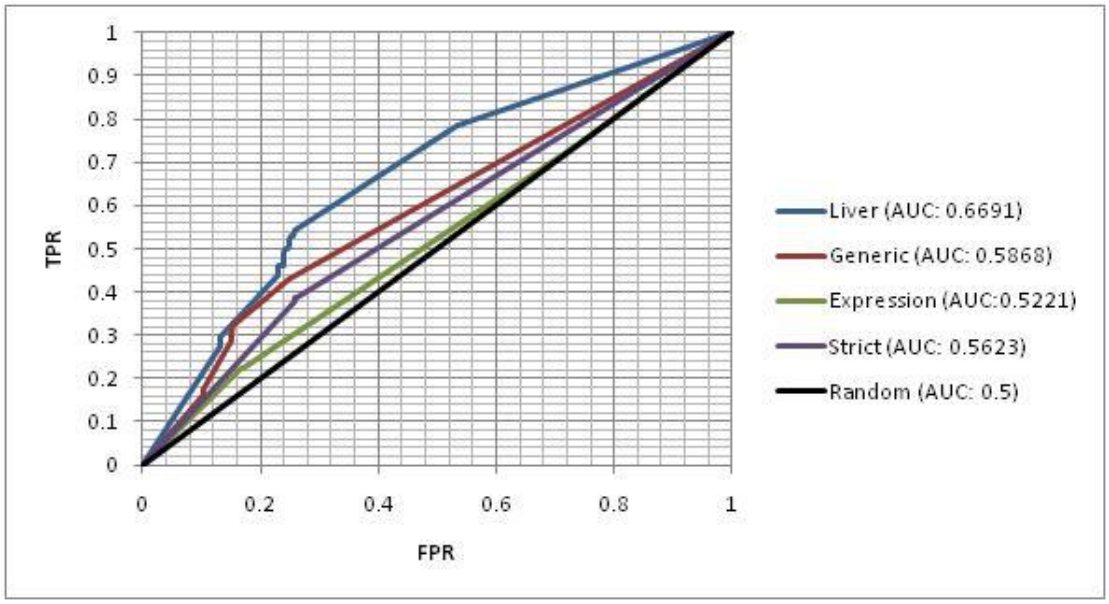

Figure 5 Prediction of metabolic biomarkers. The figure depicts the ROC curves (and the resulting AUCs) of all 5 classifiers, including the strict model.

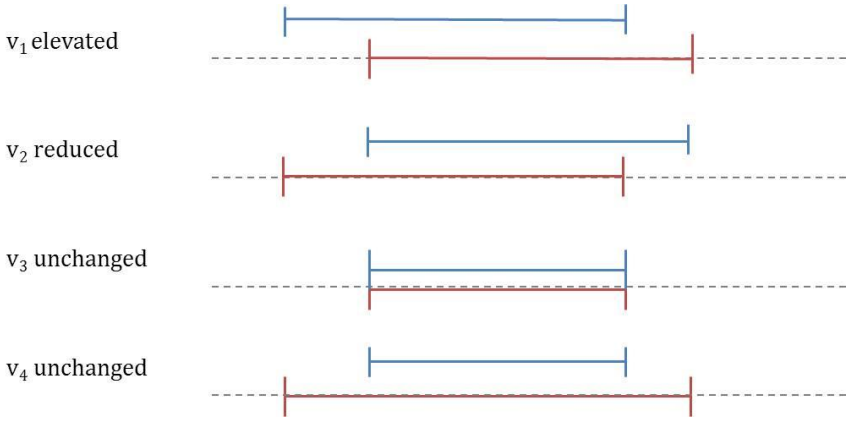

Figure 6 Interval comparisons: the flux interval of a reaction v_i, colored red, is compared to its reference interval, colored blue. v_1 (v_2) is elevated (reduced), since both the min and the max values are greater (lower) in comparison to the reference interval, and therefore *change*(v_1)>0 (*change*(v_1)<0). Both v_3 and v_4 are considered unchanged. Although the maximal value of v_4 increased, its minimal value decreased

just as much, such that they are canceled out (*change(v_1)=0*), and v_4 is considered unchanged.

Model	**Liver**			**Generic**		
Disorder	**OTC**	**ASL**	**ASS**	**OTC**	**ASL**	**ASS**
KO	2.56±6.2	1.89±4.26	1.06±2.85	0.05±93.61	4.26±22.68	2.83±21.4
Heterozygote	1.57±0.56	2.02±1.57	1.15±0.45	7.45±28.73	5.18±3.18	5.25±10.58
Normal homozygote	1.8±0.56	2.3±1.57	2.04±0.45	3.57±28.73	2.79±3.18	4.3±10.58

Table I The mean values ± the standard deviations of the urea secretion/glutamine uptake ratio of the healthy/pathologies metabolic profiles as defined by the generic and liver models.

Sensitivity Analysis

The parameter that weighs the optimization criterion, denoted as ε in the MBA formulation (see Methods) was set to 0.5 in the construction of the liver model (i.e., $|e_M| \leq 0.5 * |e_X|$). We performed a sensitivity analysis to examine the reliance of the resulting model on the optimization parameter by repeating the construction with 8 different ε values (0, 0.01, 0.1, 0.4, 0.8, 1, 2, and infinity). From a structural point of view, we compared the models by the content of their reactions. Notably, despite the varying optimization thresholds, the inclusion of many of the reactions in the resulting liver models is consistently predicted. The mean jacquard similarity coefficient is 0.8216.

From a functional perspective, we evaluated the effect of choosing various pruning thresholds on the predictive performance of the resulting model, by repeating the prediction of flux alterations and hepatic biomarkers for each of the eight models derived. The resulting prediction performances are given in Supplementary Figure 7 (enclosed below). The analysis reveals that for a wide range of pruning (optimization) thresholds, between 0.01 and 1, the overall predictive performance remains in a similar range. On the other hand, for extreme values of thresholds, such as zero (representing the strict model, where all C_M reactions are forced in the model or

infinite (representing a model that is constructed without the inclusion of C_M reactions, except for those added for gap filling), the prediction performances are lower. Hence, in summary, these results testify that the flexible MBA framework presented here is beneficial, and yet its performance is fairly robust and does not hinge upon a choice of a narrow range of threshold optimization values.

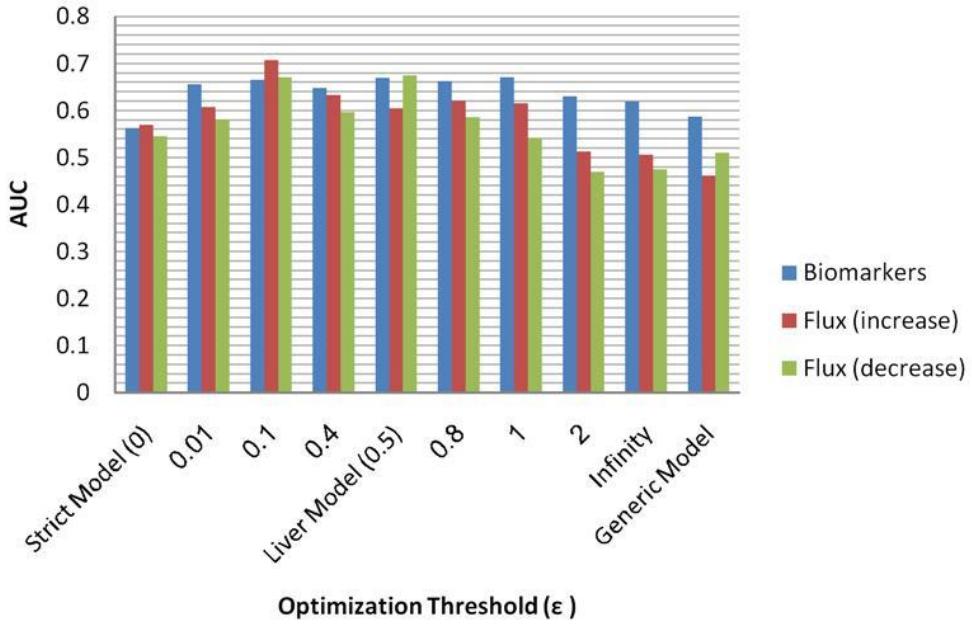

Figure 7 The performances of the sensitivity analysis models as well as the liver and the generic ones, quantified by the AUC of the ROC curves that depict their ability to predict biomarkers (blue), increasing (red) and decreasing (green) fluxes.

inactiveReactions = CheckModelConsistency(R_p, r)

Initialize
Remove reaction *r* from R_P
reactionsList = R_P
inactiveReactions = {r}
While *reactionsList* ≠ ∅
Step I – Maximize All
Solve a LP problem that maximizes the sum of fluxes through all of the reactions of *reactionsList*
Remove active reactions from *reactionsList*
Step II – Minimize All
Solve a LP problem that minimizes the sum of fluxes through all of the reversible reactions from *reactionsList*.
Remove active reactions from *reactionsList*
Step III – Single Test
If no reaction was removed from *reactionsList* in the current iteration
Let *i* be some randomely chosen reaction from *reactionsList*
Solve two LP probelms that maximize and minimize the flux through *i*
If *i* is inactive add it to *inactiveReactions*
Remove *i* from *reactionsList*

R_p- the *partial model's* reactions, that is, a subset of the generic model's reactions *(*$R_P \subset R_G$*)*; *r*- the reaction that is scanned for removal.

CheckModelConsistency determines which reactions cannot be activated due to the removal of *r* from R_P. *reactionsList* consists of reactions whose activity or inactivity is yet to be determined. Each iteration consists of three steps. In the first step, a LP problem that maximizes the sum of fluxes through all the reactions from *reactionsList* is solved. The reactions that were active (i.e., had a non-zero flux) in the obtained solution are removed from *reactionsList*. The second step is quiet similar to the first one, only that it minimizes the sum of fluxes through all of the reversible reactions from *reactionsList*. In the third step, if the previous steps did not determine the activity of any of the reactions of *reactionsList*, two other LP problems are solved in which the flux through some randomely chosen reactions *i* from *reactionsList* is individually maximized/minimized. If *i* cannot be activated it is added to *inactiveReactions*. Either way *i* would be removed from *reactionsList*. Take notice that in order to conclude that a reaction can be active, it is sufficient that it had a non-zero flux in one of the solutions that were obtained. On the other hand, in order to conclude that a reaction *i* is a dead-end reaction, a LP problem in which only the flux through *i* is maximized should be solved, as well as a LP problem in which *i* is minimized (if *i* is a reversible reaction).

Abbreviations

Argininosuccinate Lyase	ASL
Argininosuccinate Synthetase	ASS
Bioartificial Liver	BAL
Constraint-Based Modeling	CBM
Core High	C_H
Core Moderate	C_M
False Positive Rate	FPR
Flux Variability Analysis	FVA
Inborn Errors of Metabolism	IEMs
Linear Programming	LP
Modeling Building Algorithm	MBA
Ornithine Transcarbamylase	OTC
Receiver-Operator Curve	ROC
True Positive Rate	TPR