Extreme pathway analysis reveals the organizing rules of metabolic regulation

Yanping Xi[1,2,3], Fei Wang[1,2]*

1 Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai, China, **2** School of Computer Science and Technology, Fudan University, Shanghai, China, **3** Shanghai Ji Ai Genetics & IVF Institute, Obstetrics and Gynecology Hospital of Fudan University, Shanghai, China

* wangfei@fudan.edu.cn

Abstract

Cellular systems shift metabolic states by adjusting gene expression and enzyme activities to adapt to physiological and environmental changes. Biochemical and genetic studies are identifying how metabolic regulation affects the selection of metabolic phenotypes. However, how metabolism influences its regulatory architecture still remains unexplored. We present a new method of extreme pathway analysis (the minimal set of conically independent metabolic pathways) to deduce regulatory structures from pure pathway information. Applying our method to metabolic networks of human red blood cells and *Escherichia coli*, we shed light on how metabolic regulation are organized by showing which reactions within metabolic networks are more prone to transcriptional or allosteric regulation. Applied to a human genome-scale metabolic system, our method detects disease-associated reactions. Thus, our study deepens the understanding of the organizing principle of cellular metabolic regulation and may contribute to metabolic engineering, synthetic biology, and disease treatment.

Introduction

Organisms are constantly faced with variable internal physiological states and environmental conditions. The ability to rapidly shift between phenotypes to deal with these challenges is essential for the competitive fitness and survival of an organism [1–3]. It is well known that cellular response to internal and environmental perturbations is often reflected and/or mediated through changes in metabolism [3]. Such metabolic changes are often accomplished through both genetic and post-transcriptional controls, such as transcriptional regulation of gene expression and allosteric regulation of enzymes [4, 5]. Understanding the interaction between regulatory and metabolic processes is therefore a fundamental problem in biology.

The mechanisms controlling metabolic processes, including regulatory circuits and logics, have been intensively studied [6], especially since the emergence of massive 'omics' datasets in the post-genomic era [4]. Incorporating regulatory rules into constraint-based metabolic models allows researchers to more accurately predict the metabolic phenotype under different environmental and genetic perturbations [7–16]. However, these methods seldom provide insight

into the regulatory architecture, i.e. the key control reactions, of the metabolic networks. Understanding how the regulatory architecture is designed, how it evolves, and what role the structure of the metabolic network plays in regulatory processes is an ongoing research challenge [17–20].

Previously, some studies [4, 21–24] derived regulatory patterns of a metabolic pathway with limited scale and complexity by optimizing an objective function that incorporates multiple criteria of metabolic processes, such as benefit and cost, under given conditions. Some investegated the contribution of metabolic network connectivity in metabolic control or regulation [17, 19, 20, 25, 26]. Some integrated gene-expression data and topological information of the genome-scale metabolic network to uncover transcriptional regulation under certain perturbations [3]. The research above, although still in its infancy, has reported promising findings that the organizing principles of metabolic control may be deduced from the connection structure of a metabolic system.

Metabolic pathway analysis, such as elementary mode analysis and extreme pathway analysis, has provided a new outlet to understand topological structures of metabolic networks and pathway regulations [27, 28]. Elementary modes [29] consists of the minimum number of reactions that exist as a functional unit [30]. By introducing the term 'control-effective fluxes', elementary mode analysis succeeded in predicting the gene expression ratios of central carbon metabolism in the growth of *Escherichia coli* and *Saccharomyces cerevisiae* on two alternative substrates [31–33], estimating the significance of links between metabolic processes and levan biosythesis in *Halomonas smyrnensis* [34], and describing the behavior of folate-related processes in human placenta [35]. As another widely used and highly relative concept of network-based pathways, extreme pathways form the unique set of systemically independent and non-decomposable steady-state flux distributions based on the system's stoichiometry and thermo-dynamic constraints of a given metabolic network [36]. Extreme pathway analysis has already been used to hunt for regulation of metabolic systems by the approaches such as grouping and interpretation [37], Singular value decomposition (SVD) [38, 39], reaction participation analysis [40], feasible extreme pathway analysis [41] and alpha-specturm calculation [42, 43]. The approach of grouping and interpretation divides extreme pathways into groups based on some pre-set criteria and iterprete the metabolic and regulatory function of pathways in each group [37]. SVD produces eigenpathways by decomposing the extreme pathway matrix and shows that the eigenpathways correspond to the key control points in the network [38, 39]. Reaction participation analysis considers correlated reactions and the reactions that participate in a large number of extreme pathways as good targets for regulation [40]. Feasible extreme pathway analysis eliminates the extreme pathways that are inconsistent with regulatory constraints or physico-chemical constraints and shows that regulation forces a particular set of phenotypic behaviors to be expressed [41]. The alpha-spectrum defines the allowable range of extreme pathway contributions to a given steady state [42, 43].

However, the previous works also have their own shortcomings. Approaches of control-effective fluxes, feasible extreme pathways and alpha-spectrum are condition dependent. Namely, although they reveal important regulatory reactions in the given condition, they neglect regulatory reactions which function in other conditions. Grouping and interpretation is not available when the number of extreme pathway is large. Reaction participation analysis prefers the reactions with higher participation frequency, therefore it will miss out the reactions which participate less frequently in extreme pathways but still regulatory important. The approach of SVD is not intuitive enough since eigen pathways may not necessarily be biochemically feasible.

Above all, the role of regulatory reactions can be interpreted as reducing the uncertainty of the metabolic system from the perspective of information theory, because regulatory reactions put further constraints on metabolic system, reduce the space of steady sates and lead the

metabolic system to a objective state [41]. Therefore, in order to hunt for the potential regulators, it is crucial to measure the role on average a reaction plays in eliminating the uncertainty of the metabolic system. To the best of our knowledge, no approach of metabolic pathway analysis had attempted such a measurement. Here, we developed a new method to address the issue.

Since any steady-state flux vector describing a metabolic phenotype of the cell can be thought of as a non-negative combination of these extreme pathways [44], it is logical to assume that a cell 'switches' on and off its extreme pathways by metabolic control to reach a particular metabolic steady-state [36]. In other words, although the regulation of particular reactions is ongoing within a cell, the cell's ultimate aim may be to regulate the set of extreme pathways [36]. When the states of the extreme pathways are set, the information, what the target metabolic state should be, is passed from regulatory control to metabolism. Under this hypothesis, we evaluate the regulatory importance of a metabolic reaction by the role it plays in determining the "on/off" states of extreme pathways. This term takes into account both efficiency and flexibility, which are believed to be two major objectives of the evolutionary optimization process of metabolic regulation. Efficiency is the ability to fulfill the cellular demands of metabolic regulation at a minimal cost. For simplicity, supposing that equal investment is made to control each reaction in a given cell, then the cost of metabolic regulation can be estimated by the number of regulated reactions. Flexibility refers to the ability to maintain a quick and robust response against internal and enviromental perturbations. Generally, flexibility increases when more reactions are utilized for regulation. These two criteria impose challenges that are obviously contradictory; therefore, optimal regulatory architecture for metabolism needs to balance a trade-off between efficiency and flexibility.

In this study, we introduce a new method of extreme pathway analysis to deduce regulatory reactions of metabolic networks. Our method has the following advantages: 1) it is condition independent since all the extreme pathways consistent with mass balance constraints and reaction reversibility constraints are included in the analysis. 2) it can be applied on arbitrarily large number of extreme pathways as long as they can be enumerated. 3) it has no obvious preference for reaction participation. 4) rather than interpreting the regulatory effects by eigenpathways, we treated every metabolic reaction as a candidate regulator, which makes our method more intuitive. Our goal is to shed light on the universal organizing rules of regulated reactions in metabolic networks from the perspective of information theory and evolution.

We applied our method to the metabolic networks of the human red blood cell (hRBC) and *Escherichia coli* to study the architecture of allosterical and transcriptional regulation, respectively. The reactions of high regulatory importance show good agreement with findings from previous research. The results demonstrate a significant correlation between the topology of a metabolic network and its regulatory architecture. With the assumption that regulated reactions are likely associated with disease processes, we applied our method to a genome-scale human metabolic network to predict disease-associated reactions. Our computational results agree reasonably well with experimental results. Overall, extreme pathway analysis allows us to systematically investigate the organizing principle of metabolic regulation. Our findings suggest that the regulatory architecture of metabolism has evolved to put extreme pathways under more efficient and flexible controls. Our method may also be helpful in metabolic engineering and drug target discovering by recommending the reactions that are worthier of control.

Methods

Assigning exchange fluxes to a target subsystem

A *target subsystem* consists of a subset of internal reactions in the whole metabolic network. The complement of a target subsystem constitutes another subsystem called the *surrounding*

subsystem. The metabolite set of a subsystem contains all the metabolites that appear as substrates or products of at least one reaction in it.

The process to assign exchange fluxes starts by identifying the shared metabolites between the target and surrounding subsystems. For each shared metabolite $\mathcal{M}$, a duplicate metabolite $\mathcal{M}'$ is introduced to substitute $\mathcal{M}$ in reactions of the surrounding subsystem. Meanwhile, a bidirectional reaction $\mathcal{M} \leftrightarrow \mathcal{M}'$ is added to the original metabolic network and an exchange flux associated with $\mathcal{M}$ is introduced into the target subsystem. Obviously, the reaction $\mathcal{M} \leftrightarrow \mathcal{M}'$ describes the exchange of metabolite $\mathcal{M}$ between the target and surrounding subsystems. Thus, the flux of reaction $\mathcal{M} \leftrightarrow \mathcal{M}'$ should be equal to the exchange flux associated with M in the target subsystem. Next, the maximum and minimal fluxes of $\mathcal{M} \leftrightarrow \mathcal{M}'$ are determined through flux variability analysis [45] by setting the flux of reaction $\mathcal{M} \leftrightarrow \mathcal{M}'$ as the objective function. These values indicate the possible production and/or consumption of $\mathcal{M}$ in the target subsystem. Table 1 covers all 4 possible input and output combinations between the maximum and minimal values. Based on this information, flux constraints can be specified for each exchange reaction in the target subsystem. With all the exchange fluxes being properly added and constrained, the target subsystem is ready for extreme pathway computation.

Although adding a reaction $\mathcal{M} \leftrightarrow \mathcal{M}'$ is equivalent to specifying a compartment for each target subsystem, unlike the previous study [46], the procedure will not change the predicted behaviors of the whole metabolic network. Because the added reaction has only one substrate and one product, it is proven in Statement G in S1 Text that adding such a reaction does not change the constraint of the fluxes of steady states.

The process described above is illustrated by Fig A in S1 Text. This protocol works effectively for any metabolic network and any delimitation of the target subsystems.

Calculating extreme pathways

The extreme pathways of the hRBC metabolic network were derived straightforwardly from a published open-source bioinformatic software program, EXPA [47]. As for the target subsystems of *E. coli* and human metabolism, we applied the above protocol for the exchange flux assignment and then computed the extreme pathways with EXPA [47]. The extreme pathways of type I and II that comply with Kirchhoff's second law [48] were passed to the following analysis. Each pair of elements corresponding to a reversible internal reaction in an extreme pathway was merged together by subtracting the reverse flux from the forward one. The resulting vector is termed *compact extreme pathway*. Elements in a compact extreme pathway have a one-to-one correspondence to reactions in the metabolic network. It is proven in Statement D in S1 Text that converting from an extreme pathway to its compact form does not change the set of reactions it utilized.

Table 1. Logic table for determining the constraints of the exchange reaction that transports metabolite $\mathcal{M}$ in and out of the target subsystem based on flux variability analysis of the reaction $\mathcal{M} \leftrightarrow \mathcal{M}'$. The forward direction of $\mathcal{M} \leftrightarrow \mathcal{M}'$ is defined as from $\mathcal{M}$ to $\mathcal{M}'$. And the forward direction of the exchange reaction is defined as taking $\mathcal{M}$ away from the target subsystem. Min, minimum; Max, maximum.

Case	Exchange flux of a subnetwork		Available direction (Flux constraint)
	Min.	Max.	
1	< 0	> 0	Out and in (unconstrained)
2	$= 0$	> 0	Out (positive)
3	< 0	$= 0$	In (negative)
4	$= 0$	$= 0$	None (zero)

Sorting internal EqSets by their regulatory importance

A greedy algorithm was built to sort internal EqSets of internal reactions by their regulatory importance. Briefly, it begins with a pool $\mathbb{P}$ of all the EqSets and an empty queue of predicted regulatory EqSets, $\mathbb{Q}$. Then, it iteratively picks up the EqSet with the highest regulatory importance (defined in the following sections) from $\mathbb{P}$ and moves it to the end of $\mathbb{Q}$. The iteration stops when $\mathbb{P}$ is empty or the regulatory importance of any EqSet remaining in $\mathbb{P}$ equals zero. At last, the algorithm removes the EqSets of exchange reactions from $\mathbb{Q}$ and outputs $\mathbb{Q}$ as the resulting sequence. More details about the algorithm are provided in Algorithm A in S1 Text.

Shannon entropy and conditional entropy of reactions

Shannon entropy and conditional entropy are used to measure the role on average of a reaction in eliminating the uncertainty of a metabolic system. Suppose a metabolic network of n reactions, $\mathbb{X} = \mathcal{R}_1, \mathcal{R}_2, \cdots, \mathcal{R}_n$, has l compact extreme pathways, $[\mathbf{e}^1, \mathbf{e}^2, \cdots, \mathbf{e}^l]$. Then, a reaction $\mathcal{R}_i$ is employed by a compact extreme pathway $\mathbf{e}^j$ if and only if $e_i^j \neq 0$, where e_i^j is the ith element of $\mathbf{e}^j$, $1 \leq i \leq n$ and $1 \leq j \leq l$. We define $\mathbf{e}^j$ as an *employer extreme pathway* of $\mathcal{R}_i$. Further, we consider $\mathcal{R}_i$ to be 'on' if any of its employer extreme pathway is utilized by the metabolic system to form the current steady-state, and 'off' otherwise. Assume that each extreme pathway is utilized with equal opportunity. The probability of the on/off state of $\mathcal{R}_i$ is defined as follows: $P(\mathcal{R}_i \ is \ \text{on})$ equals the fraction of extreme pathways in which $\mathcal{R}_i$ is employed and $P(\mathcal{R}_i \ is \ \text{off})$ equals $1 - P(\mathcal{R}_i \ is \ \text{on})$. The joint probability distribution can be similarly derived by counting the number of compact extreme pathways that consist of certain reactions.

Based on the above probability distributions, the Shannon entropy $H(\mathcal{R}_i)$, combination entropy $H(\mathbb{X}_1)$ as well as conditional entropy $H(\mathbb{X}_1|\mathbb{X}_2)$ can be defined according to the information theory [49], where $\mathbb{X}_1$ and $\mathbb{X}_2$ are two subsets of $\mathbb{X}$. In particular, if reactions $\mathcal{R}_i$ and $\mathcal{R}_j$ satisfy that $H(\mathcal{R}_i|\mathcal{R}_j) = H(\mathcal{R}_j|\mathcal{R}_i) = 0$, they form an *equivalent reaction couple on Shannon entropy*.

Regulatory distance between reactions/EqSets

A weighted graph G was built for each metabolic system or subsystem in which individual reactions are nodes and the weight of the arc connecting two nodes, $\mathcal{R}_i$ and $\mathcal{R}_j$, equals the *local regulatory distance*, $d(\mathcal{R}_i, \mathcal{R}_j)$, between the two reactions.

$$d(\mathcal{R}_i, \mathcal{R}_j) = \begin{cases} 0 & \text{if } i = j, \\ \infty & \text{if } \mathbb{M}_c = \emptyset, \\ \min_{\mathcal{M} \in \mathbb{M}_c} (C(\mathcal{M})) & \text{otherwise,} \end{cases} \tag{1}$$

where $\mathbb{M}_c$ is the intersect of the metabolites of $\mathcal{R}_i$ and $\mathcal{R}_j$, and $C(\mathcal{M})$ equals the number of reactions that use $\mathcal{M}$ as a substrate. Note that any metabolite involved in a reversible reaction is considered to be both a substrate and a product. Without considering the change of enzyme activity, the flux variation of a reaction will change the fluxes of the adjacent reactions by alter the concentration of shared metabolites. If the concentration of the shared metabolite can also be altered by other reactions, then the flux change of the adjacent reaction will be diminished. Therefore, the local regulatory distance between two reactions measures the direct impact on the flux of one reaction caused by a flux change of the other.

The *global regulatory distance*, $D(\mathcal{R}_i, \mathcal{R}_j)$, between $\mathcal{R}_i$ and $\mathcal{R}_j$ equals the length of the shortest route between the corresponding nodes on the weighted graph, G. Intuitively, the global regulatory distance measures the indirect impact of flux change between two reactions.

The regulatory distance of two EqSets $D(\mathbb{X}_1, \mathbb{X}_2)$ is defined as the average of $D(\mathcal{R}_i, \mathcal{R}_j)$, where $\mathbb{X}_1$ and $\mathbb{X}_2$ are two EqSets, $\mathcal{R}_i \in \mathbb{X}_1$, and $\mathcal{R}_j \in \mathbb{X}_2$.

Equivalent reaction set on Shannon entropy

First, each exchange reaction forms an individual *equivalent reaction set on Shannon entropy* (EqSet). Second, an EqSet of k ($k \geq 1$) internal reactions $\mathbb{X} = \{\mathcal{R}_1, \mathcal{R}_2, \cdots, \mathcal{R}_k\}$ has the properties that

1. $\forall i, j = 1, \cdots, k, \mathcal{R}_i$ and $\mathcal{R}_j$ forms an equivalent reaction couple on Shannon entropy, i.e. $H(\mathcal{R}_i | \mathcal{R}_j) = H(\mathcal{R}_j | \mathcal{R}_i) = 0$;

2. $\forall i = 1, \cdots, k, \min\limits_{j=1,\cdots,k, j \neq i} D(\mathcal{R}_i, \mathcal{R}_j) \leq \rho_e$, where ρ_e is the effective radius within an EqSet that represents the maximum regulatory distance of any reaction to its nearest counterpart in the same EqSet.

Measuring the regulatory importance of an EqSet

Given a metabolic system with k EqSets, $\mathbb{P} = \{\mathbb{X}_1, \mathbb{X}_2, \cdots, \mathbb{X}_k\}$, and a queue of EqSets $\mathbb{Q} = [\mathbb{X}_1, \mathbb{X}_2, \cdots, \mathbb{X}_i]$ ($0 \leq i < k$) that are predicted to be regulatory, the regulatory importance of an EqSets is determined by the nearby EqSets within the distance of ρ_s because of the constraint that the regulatory influence of an EqSet $\mathbb{X}_j$ ($i < j \leq k$) exists locally, where ρ_s is the effective radius between EqSets that designates the maximum regulatory distance that an EqSet influences. We built two group of EqSets with the nearby EqSets of $\mathbb{X}_j$: First, the regulatory neighbor set $\mathbb{N}_j$ consists of at most τ nearby EqSets standing at last of the queue $\mathbb{Q}$, where τ is the size of a window sliding on $\mathbb{Q}$. And second, the competitor set $\mathbb{C}_j$ consists of the nearby EqSets that are not in $\mathbb{Q}$.

The regulatory importance, $val(\mathbb{X}_j)$ ($i < j \leq k$), of $\mathbb{X}_j$ is measured from two aspects: one is the average information it provides to the metabolic system conditioned on the regulatory neighbor set and the other is its non-substitutability among the competitor set. Formally, the former equals $H(\mathbb{X}_j | \widehat{\mathbb{N}}_j)$, and the latter equals $\min\limits_{\mathbb{X}_u \in \mathbb{C}_j}(H(\mathbb{X}_j | \widehat{\mathbb{N}}_j \cup \mathbb{X}_u))$, where $\widehat{\mathbb{N}}_j$ is the set of reactions participating in any regulatory neighbor EqSets of $\mathbb{X}_j$, i.e., $\widehat{\mathbb{N}}_j = \bigcup\limits_{\mathbb{X} \in \mathbb{N}_j} \mathbb{X}$. Lastly the summation of the two aspects is adjusted by the weight $1 + \mu(|\mathbb{X}_j| - 1)$, which gives more bonus to larger EqSets. The amount of bonus for an extra reaction in $\mathbb{X}_j$ is modulated by bonus ratio, μ. To sum up, $val(\mathbb{X}_j)$ is represented as

$$val(\mathbb{X}_j) = [1 + \mu(|\mathbb{X}_j| - 1)][H(\mathbb{X}_j | \widehat{\mathbb{N}}_j) + \min\limits_{\mathbb{X}_u \in \mathbb{C}_j} H(\mathbb{X}_j | \widehat{\mathbb{N}}_j \cup \mathbb{X}_u)]. \tag{2}$$

EqSet sequence evaluation and the corresponding *p*-value

The output EqSet sequence $\mathbb{Q}$ ($\mathbb{Q} = [\mathbb{X}_1, \mathbb{X}_2, \cdots, \mathbb{X}_s]$) of the above algorithm is evaluated by summing up the ranks of the EqSets that contain regulatory or disease-related (for the human metabolic network) reactions, where s is the length of $\mathbb{Q}$. If this sequence covers only a part of

the internal EqSets, a penalty is added to the rank summation, representing the missing EqSets that contain regulatory or disease-related reactions. The final evaluation score σ is represented as

$$\sigma = \sum_{\substack{i=1,\cdots,s \\ \mathbb{X}_i \cap \mathbb{R}^{bm} \neq \emptyset}} rank(\mathbb{X}_i) + d(s + \frac{l+1}{2}), \tag{3}$$

where $\mathbb{R}^{bm}$ is the set of biologically meaningful reactions, i.e., the regulatory or disease-related reactions, $rank(\mathbb{X}_i)$ is the rank of $\mathbb{X}_i$ in $\mathbb{Q}$, l is the number of lost internal EqSets and d is the number of the missing internal EqSets that contain biologically meaningful reactions. It is proven in Statement H in S1 Text that the penalty part $d(s + \frac{l+1}{2})$ equals the expectation of the rank summation of the lost EqSets that contain biologically meaningful reactions when all the missing EqSets are placed at the tail of the output sequence in a random order.

For an output sequence of the sorting algorithm whose evaluation score is σ_0, the corresponding p-value is defined as the probability that a randomly arranged sequence of internal EqSets has an evaluation score no higher than σ_0. Although the p-value can be estimated by stochastic simulation, we provide a dynamic algorithm in Algorithm B in S1 Text to meet the demands for precise computation.

Heuristically searching for the optimal setting of the parameters

A heuristic search algorithm was built in order to quickly find the nearly optimal values for the four parameters (i.e., μ, ρ_e, ρ_s and τ) used in the EqSet sorting algorithm. Assume that the ranges of μ, ρ_e, ρ_s and τ are limited and discrete, the algorithm starts from a random point $(\mu^0, \rho_e^0, \rho_s^0, \tau^0)$, which are randomly selected from their respective ranges. While keeping three of the parameters at fixed values, the search strategy sequentially updates one parameter at a time to its optimum value, which results in a sorted EqSet sequence with the lowest p-value. When all four parameters have been adjusted, the search algorithm moves on to the next iteration by updating from the first one again. If a local optimal setting of the parameters is found, the algorithm will start another try to search for the local optimal parameters from another randomly picked starting point. The algorithm will stop and return the best parameter values among all the tries when it reaches the maximum number of search steps. The search algorithm described above is illustrated in Fig D in S1 Text.

Evaluating the capability of the sorting algorithm for disease-related EqSet prediction

Since the EqSets of a certain metabolic network vary with the parameter ρ_e, we first divided the range of ρ_e into several separate subsets so that values in a same subset will result in exactly the same EqSets. In each subset, we randomly selected x% of the internal EqSets as a training set and left the other $(100 - x)$% as a test set, where $x \in \{10, 20, 30, 40, 50, 60, 70, 80, 90\}$. Given a sequence of internal EqSets, we extracted the subsequence that consists of the EqSets in the training set, on which the p-value for the training set is computed. We found the optimal setting of the parameters and the corresponding p-value for the training set. We repeated this process 100 times, which provided the mean of the p-values on the training sets. After that, we selected the subset of ρ_e with the lowest average p-values and counted the average rate of the disease-related EqSets being in the top 10% of the test set sequences. This rate shows the capability of the sorting algorithm for predicting disease-related EqSets.

Results

Reactions that play equal roles in determining the states of extreme pathways are merged as an equivalent reaction set of Shannon entropy (EqSet), in which each reaction pair satisfies the assumption that the conditional entropy of one reaction, given the other, equals zero. An EqSet is treated as a potential entity for metabolic regulation. The EqSet that contains at least one regulated reaction is defined as a regulatory EqSet. A greedy algorithm is used to organize the EqSets in a sequence according to their regulatory importance. The algorithm starts from an empty sequence. Under the assumption that the EqSets in the current sequence are regulatory, the algorithm iteratively picks up the EqSet from the rest that is most important for regulation and adds it to the end of the sequence. The EqSets ranked higher in the sequence are expected to have greater likelihood of containing strictly regulated reactions. The resulting EqSet sequence is evaluated by a *p*-value, which is the probability that the rank summation of the regulatory EqSets on a random sequence is lower than that on the sequence given by our algorithm. (See Methods and Statement I in S1 Text).

The regulatory importance of an EqSet is defined by regulatory efficiency and flexibility. The regulatory efficiency of an EqSet is measured by the conditional entropies that depend on the reactions' participation ratios on the extreme pathways. It characterizes the EqSet's average effect on determining the "on/off" state of each extreme pathway, as well as the degree to which its regulatory function cannot be replaced by other EqSets. The flexibility of an EqSet is modeled in two aspects: First, the regulatory distances between reactions and between EqSets are defined according to the topology of the metabolic network. And the regulatory influence of a reaction is restricted to a local scope in order to speed up the shift of a cell's metabolic states. Therefore, an EqSet is restriced further to include only reactions within certain regulatory distance given by a parameter, intra-EqSet effective radius (denoted as ρ_e). And the regulatory influence of an EqSet is restricted to other EqSets within the regulatory distance given by a parameter, inter-EqSet effective radius (denoted as ρ_s). Second, a sliding window of size τ is employed on the sequence to define the available regulatory EqSets. The Eqset that leaves the window is considered to be out of control, which simulates a disturbance in the regulatory system. Thus, the EqSet that best compensates for the disturbance will be considered to be of high flexibility in this case. Briefly, the values of ρ_e, ρ_s and τ have an inverse relation with the degree of flexibility, and consequently reflect a trade-off between efficiency and flexibility.

The regulatory importance of an EqSet calculated above is then adjusted according to the size of the EqSet for the reasons listed below: Assuming that each metabolic reaction has an equal probability of obtaining a regulatory function because of gene mutations, it is more likely that a mutation that brings a regulatory function to a reaction occurred earlier in a bigger EqSet than in a smaller one. If the reaction happened to be of regulatory importance, the mutation is more likely to increase in frequency within the population by means of natural selection. Thus, an EqSet that contains more members will gain some advantages in acquiring regulated reactions compared to smaller counterparts with reactions that have similar degrees of regulatory importance. Therefore, we give a bonus for each extra reactions in an EqSet. The amount of the bonus is controlled by a parameter, bonus ratio (denoted as μ).

We use the metabolic reconstructions of the human red blood cell (hRBC) [37], *E. coli* (iJR904) [50] and global human cells (H. sapiens Recon 1) [51] as examples to explore the utility of our method. These models were obtained from http://systemsbiology.ucsd.edu. As network properties and regulation demands differ between metabolic systems, the parameters ρ_e, ρ_s, τ and μ have to be optimized case by case.

Identifying allosterically regulated reactions in hRBC metabolism

The human erythrocyte is a complete cell with a simple metabolic system. It has been well studied, which makes it an attractive case for our research. The hRBC metabolic model used here consists of 39 metabolites and 51 reactions, of which 19 are exchange reactions (See Tables A and B in S1 Text). It contains four classical pathways: glycolysis, the pentose pathway, adenosine nucleotide metabolism, and the Rapoport-Luebering shunt [37]. The computation of the extreme pathways of this model resulted in 36 type I, 3 type II, and 16 type III extreme pathways. (See Figs F, G and H in S1 Text for type I and II extreme pathways.) Type III extreme pathways are reported to be thermodynamically infeasible [48], so they are neglected in the remaining analysis. Metabolism in the hRBC is mainly regulated by allosteric enzymes that control the cell's production of the cofactors necessary to maintain osmotic balance and electroneutrality, and to fight oxidative stresses [37, 52, 53]. We obtained 10 regulated reactions from the literature [54–63]. Lists of the reactions and regulatory mechanisms are detailed in Table C in S1 Text.

The relatively simple structure of the hRBC metabolic network ensures that the influence of any regulated reaction may quickly spread throughout the network. In addition, the human body employs huge numbers of erythrocytes to fulfill the task of transporting and exchanging oxygen and carbon dioxide; thus, cells with appropriately functioning regulatory processes could likely compensate for regulatory deficiencies in other cells. Both factors, a simple network structure and redundancy through a large quantity, may reduce the evolutionary demands on flexibility for the hRBC regulatory architecture. Therefore, we set parameters ρ_e, ρ_s, and τ to infinity so that the regulatory importance measurement focuses on the regulatory efficiency of each reaction.

For each pair of internal reactions, we calculated the conditional entropy of one participant's distribution given the other's and vice versa (Fig 1). The black blocks on the diagonal show 22 internal EqSets, of which 5 have more than one member reaction. There are 8 regulatory EqSets in the hRBC metabolic network. The entropies of the regulatory EqSets are higher than those of the non-regulatory EqSets (Wilcoxon rank sum test, p-value = 0.0121, Table D in S1 Text).

By plotting p-values of resulting sequences against the values taken by the parameter μ, we found that μ plays a minor role in the output sequence in general (Fig 2). However, the p-value decreases when μ increases from 0 to 0.125, and then turns back when the bonuses are overestimated, i.e., when μ continuously increases to 1. The relatively low p-values indicate that our estimation of the regulatory importance of the EqSets agrees well with estimations from previous research. The lowest p-value, i.e. the lowest rank summation σ, is achieved when μ is set between 0.075 and 0.125 ($\sigma = 44.5$, p-value $= 2.10 \times 10^{-4}$, Fig 3). The corresponding sequence contains 8 internal EqSets (Table 2), seven of which are regulatory EqSets. The reactions in the other internal EqSets are considered to be less important for regulation. The only missing regulated reaction in the sequence is 'HK', catalyzed by hexokinase. It is known that the subtypes of hexokinase in hRBCs are HK-I and HK-R [64, 65], which are inhibited by Glucose-6-phosphate dehydrogenase (G6PDH) and ADP [54]. However, the inhibition could be eliminated by a minimal amount of inorganic phosphate [54]. Thus, the control response of 'HK' in hRBC metabolism is not significant, which is consistent with our results.

As a comparison, we applied our method to elementary modes (Fig 4). The result shows that our method performed worse on elementary modes than on extreme pathways, no matter the value given to the parameter μ. The difference in performance may be caused by the fact that extreme pathways are systemically independent (i.e., it is impossible to describe any extreme pathway as the sum of the others) whereas elementary modes are not.
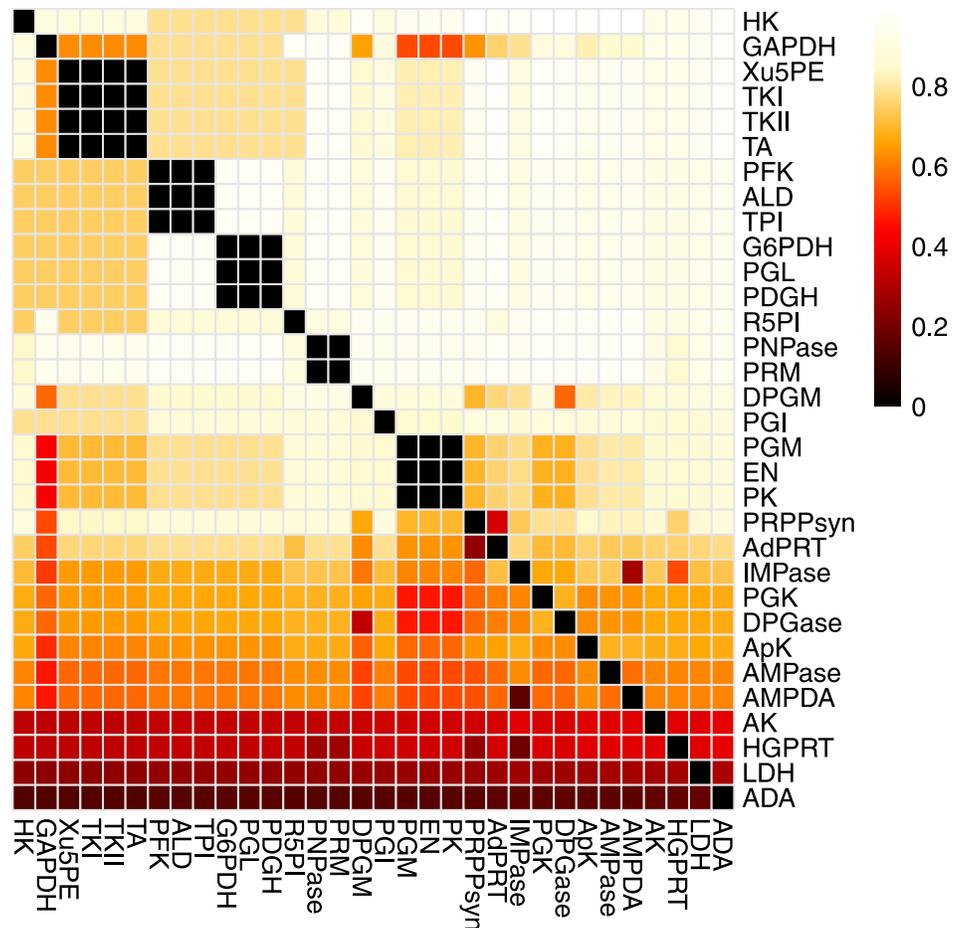
Fig 1. Heat map showing conditional entropy of internal reaction pairs of the hRBC metabolic network. The heat map colors represent the conditional entropy of each reaction at the beginning of each line, given the reaction listed at the bottom of each column. The black blocks on the diagonal represent the internal EqSets.

We further compared extreme pathways and artificial metabolic pathways, which are generated by adding up randomly selected extreme pathways (Fig 5 and Fig J in S1 Text). There are two parameters, p and t, that affect the number and complexity of the artificial pathways (see Methods and Algorithm C in S1 Text). Briefly, the higher the values of p and t, the greater are the number and complexity of the artificial pathways (Fig I in S1 Text). As p or t increase, the artificial pathways are less similar to the extreme pathways, and the p-values of the resulting EqSet sequences increase substantially.

In summary, the results imply that extreme pathways are likely to be the real targets of metabolic regulation.

Predicting transcriptionally regulated reactions in the *E. coli* metabolic network

The model iJR904 of *E. coli* accounts for 904 genes, 761 metabolites and 1,075 reactions, including 931 internal reactions and 143 exchange reactions [50]. Each reaction has a subsystem label indicating its metabolic function. A regulatory model originally designed for iJR904, namely iMC1010v1 [9], makes iJR904 a suitable object of our study.
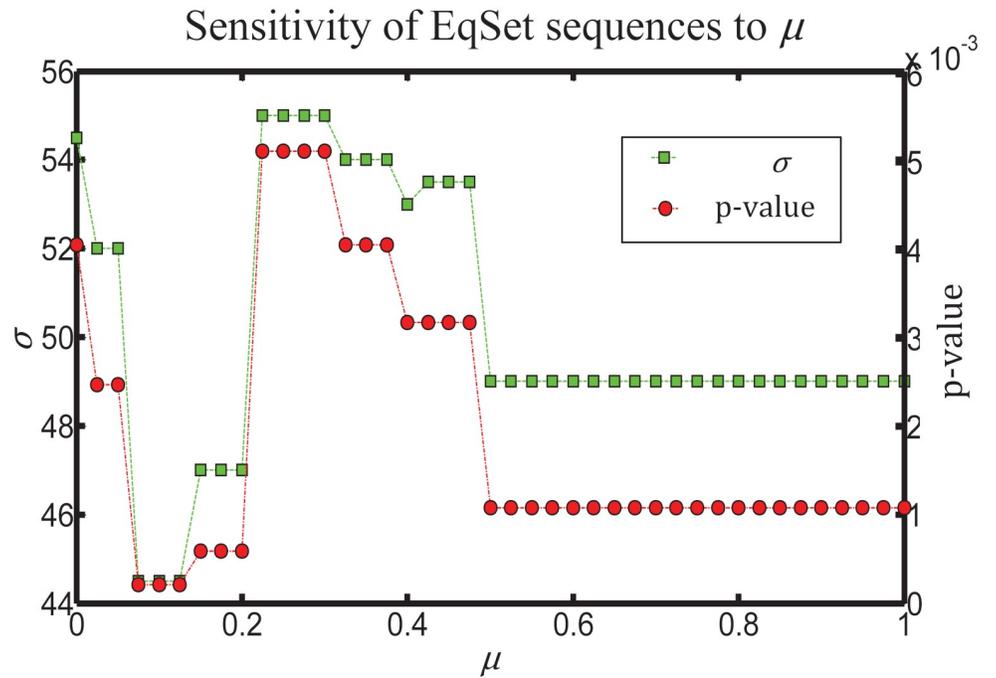
Sensitivity of EqSet sequences to μ

Fig 2. *P*-value (the line with circles) and evaluation score σ (the line with squares) of the resulting EqSet sequence of hRBC metabolic network as a function of the parameter μ. As μ ranges from 0 to 1 with an increasing step of 0.025, *p*-value varies between 2.10×10^{-4} and 5.1×10^{-3}, and σ varies between 44.5 and 55.

https://doi.org/10.1371/journal.pone.0210539.g002

Distribution of σ

$\delta = 44.5$

p-value $= 2.10 \times 10^{-4}$

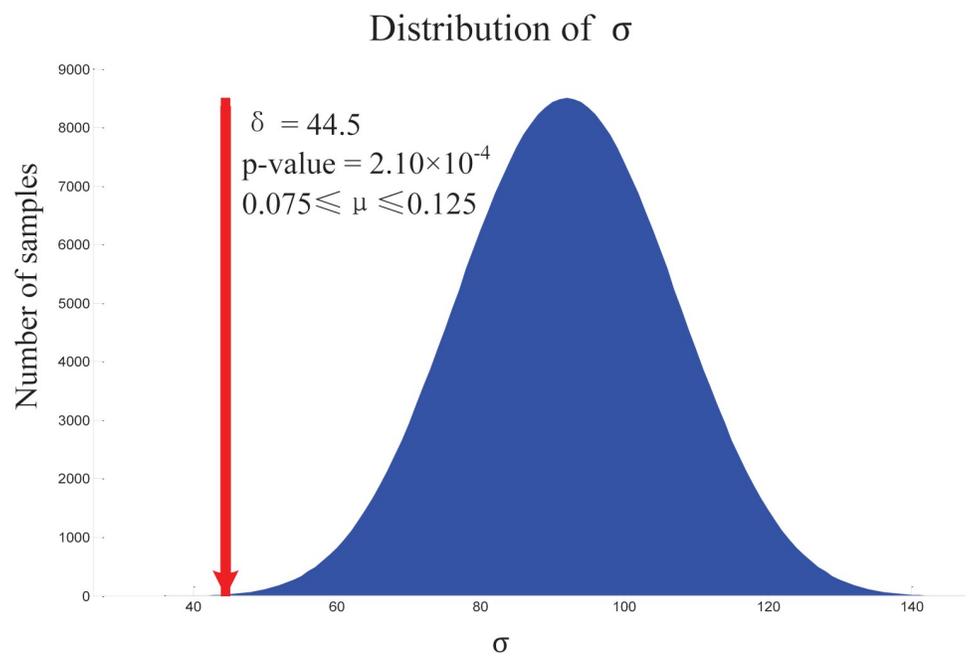$0.075 \leqslant \mu \leqslant 0.125$

Fig 3. Evaluation score distribution of randomly organized EqSet sequences of hRBC metabolic network. The arrow denotes the EqSet sequence shown in Table 2.

https://doi.org/10.1371/journal.pone.0210539.g003

Table 2. The internal EqSet sequence of hRBC metabolic network in descending order of regulatory importance.
The regulated reactions are denoted in boldface type. Full names for the abbreviations are listed in Table B in S1 Text.

EqSet	Rank
TKI, **TKII**, Xu5PE, TA	1
DPGM	2
PRPPsyn	3
PFK, ALD, TPI	4
G6PDH, **PDGH**, PGL	5
PK, PGM, EN	6
AMPase	7
AdPRT	8

https://doi.org/10.1371/journal.pone.0210539.t002

The metabolic network of *E. coli* is so complex that it is impossible to enumerate all the extreme pathways in a reasonable time period. Therefore, we defined three subnetworks that represent amino acid metabolism, hydrocarbon metabolism and lipid metabolism, respectively. The internal reactions of a subnetwork are composed of the reactions functionally related subsystems (Table 3). Some reactions participate in more than one subnetwork. Exchange reactions were assigned to each subnetwork and then the extreme pathways were computed (see Methods). Table 4 summarizes the number of internal, exchange reactions and extreme pathways in each subnetwork, as well as the proportion of regulated reactions. The regulated reactions are identified by combining the gene-protein-reaction association [50] and the gene's logical transcriptional regulatory rules of model iMC1010v1 [9], which resulted in 474 transcriptional regulated reactions (See S1 Table).
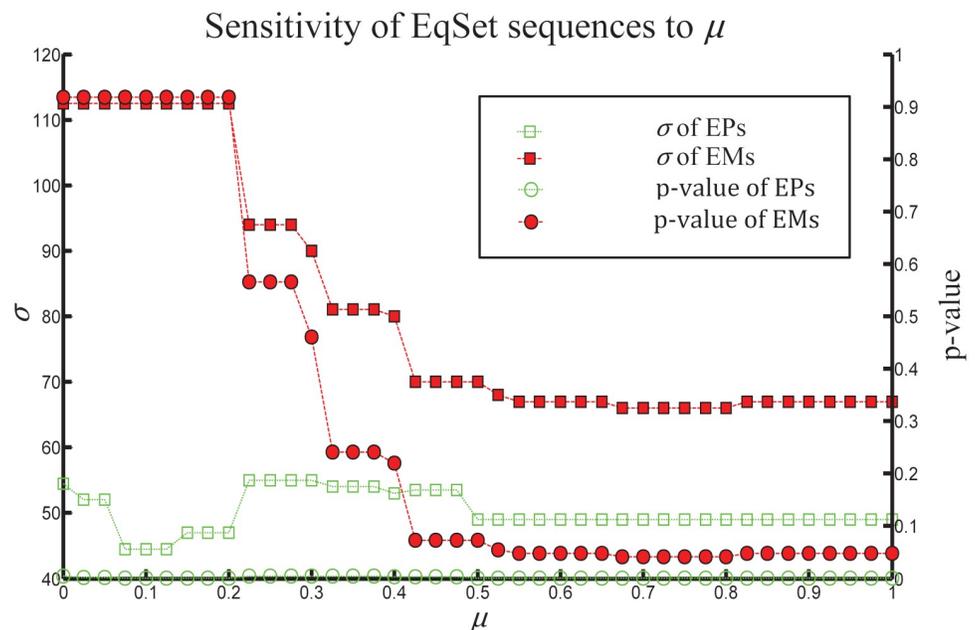

Fig 4. The resulting EqSet sequences calculated from extreme pathways (EPs; the lines with soft dots) versus those calculated from elementary modes (EMs; the lines with hard dots) in p-value (the lines with circles) and evaluation score σ (the lines with squares). As μ ranges from 0 to 1, *p*-value of the EqSet sequences calculated from EMs varies between 0.0408 and 0.9173, and the corresponding σ varies between 66 and 112.5. These values are always higher than those of the sequences calculated from EPs.

https://doi.org/10.1371/journal.pone.0210539.g004

The regulatory architecture of the metabolic network of *E. coli* is more flexible than that of hRBC. There are two reasons for this difference: first, the metabolic network of *E. coli* is much more complex so that a regulated reaction may have a relatively limited scope of influence. Second, in order to maintain survival, *E. coli* needs extra regulated reactions to compensate for the potential deficiency in metabolic control. This requires that the parameters ρ_e, ρ_s and τ take appropriate values that will properly characterize the evolutionary demands of flexibility.

According to the size of each subnetwork and the regulatory distance between pairs of reactions in the subnetwork(Table 4 and **Fig K in** S1 Text), we expand the possible ranges for μ, ρ_e, ρ_s and τ to a relatively wide ranges given in Table 5. Under the constraint that ρ_e is no higher than ρ_s, we built a candidate set that consists of all 210,600 different values of the parameters that were used to make the EqSet sequences. For each subnetwork, we found that regulatory EqSets ranked significantly higher ($p < 0.05$) on most sequences (Table 6), which suggests that the major factor affecting the architecture of metabolic regulation is the reaction participation ratio on extreme pathways rather than any of the parameters. Moreover, the target subnetworks show diverse preferences for the candidate parameter values. For example: 1) The *p*-value < 0.05 candidate parameter values for human hydrocarbon drops substantially when compared to the E.coli model performance. 2) Lower *p*-values are achieved at certain parameter values for each target subnetwork. Differences in network properties and regulatory demands may be one reason, and the other may be due to the different biases which are introduced when dividing the genome-scale metabolic networks into target subsystems and surrounding subsystems (see Discussion). The sequences of internal EqSets with the lowest *p*-values are shown in S3, S4 and S5 Tables for the above three subnetworks, respectively. Most EqSets in the top part of the sequences include at least one regulated reaction.

In order to take a closer look at the role of each parameter, we explored the distributions of the lowest 5% of the *p*-values when one of the parameters was fixed and the others were not (Figs 6, 7 and 8 for the subnetworks of amino acid metabolism, hydrocarbon metabolism and lipid metabolism, respectively). In general, the distribution of the lowest 5% of the *p*-values shows stable trends when the value of the fixed parameter ranges from the minimum to the maximum in the scope. The parameter μ has a slight effect on the *p*-values compared to the effects of parameters ρ_e, ρ_s and τ. For different subnetworks, the *p*-values drop to a low value at different values of ρ_e, ρ_s and τ, which implies that the evolutionary demands of metabolic regulation diverge slightly for different functional parts.

Predicting disease-associated reactions in the human metabolic network

Metabolic disorders have been linked with chronic disease processes, including heart disease, cancer, diabetes, and obesity [66]. Since the reactions of high importance for metabolic regulation play a key role in controlling a phenotype shift, their malfunction may induce a morbid state of the organism. So it is reasonable to infer that the reactions with the most regulatory importance are more likely to be associated with disease processes. We applied our methods to the genome-scale human metabolic network (Recon 1) [51] to predict disease-associated reactions.

The model H. sapiens Recon 1 of human metabolism accounts for 1,496 genes and 3,742 reactions, of which 431 are exchange reactions [51]. The disease-associated reactions were obtained from the public data of Lee et al [67]. The number of disease-associated reactions is 779, details are listed in S2 Table. As with our analysis of the *E. coli* metabolic network, we defined two subnetworks that respectively represent amino acid metabolism and hydrocarbon metabolism. The subsystems contained in each subnetwork are listed in Table 7 and other related information is summarized in Table 4.
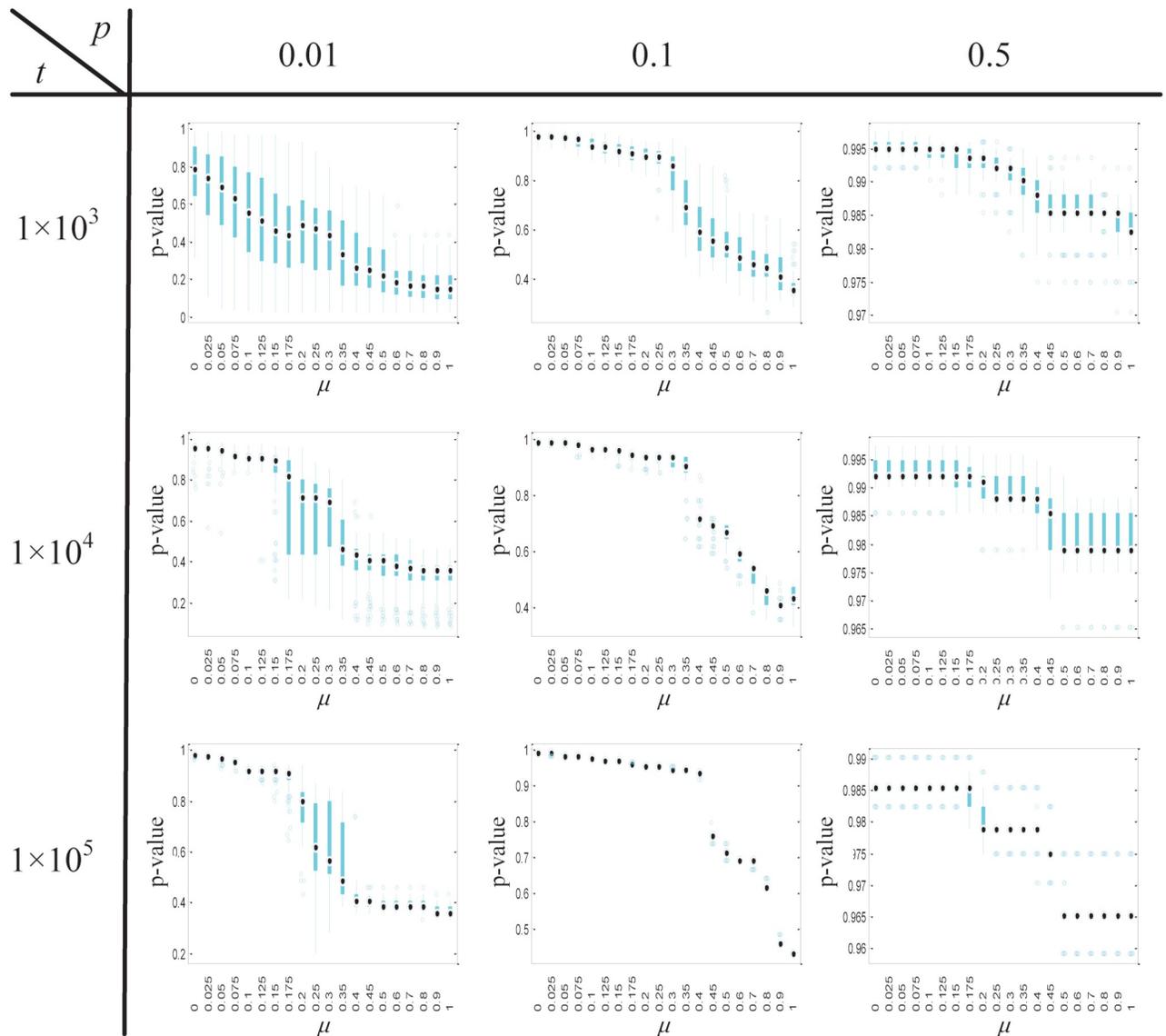
Fig 5. *P*-value distribution of the EqSet sequences calculated from artificial metabolic pathways of hRBC metabolic network. A set of artificial metabolic pathways are built as follows: (1) An artificial pathway is a summation of several randomly selected EPs. (2) The probability an EP being selected is *p*. (3) Altogether, *t* artificial pathways are generated, and the unique ones form the set. One hundred different artificial pathway sets are built for certain values of *p* and *t*. The distributions of the number of artificial pathways and the number of EPs contained in an artificial pathway in the sets generated at different values of *p* and *t* are shown in Fig I in S1 Text.

https://doi.org/10.1371/journal.pone.0210539.g005

We made sequences of EqSets with the same candidate set of parameter values as that of the *E. coli* model, and then calculated the corresponding *p*-values to see whether the disease-associated reactions tend to rank higher than the reactions that are not associated with disease processes. That tendency is significant for a large section of EqSet sequences (Table 6), especially for the subnetwork of human amino acid metabolism, of which most sequences get *p*-values less than 0.05. Similar to the situations we found for hRBCs and *E. coli*, the distributions of the lowest 5% of the *p*-values also showed stable trends with changes in each fixed parameter (Figs 9 and 10). The results suggest that disease-associated reactions may be discovered by evaluating their regulatory importance.

Table 3. Subsystems contained in each metabolic sub network of *E.coli*.

Subnetwork	Subsystems
Amino acid metabolism	Alanine and aspartate metabolism; alternate carbon metabolism; arginine and proline metabolism; cofactor and prosthetic group biosynthesis; cysteine metabolism; folate metabolism; glutamate metabolism; glycine and serine metabolism; histidine metabolism; methionine metabolism; threonine and lysine metabolism; tyrosine tryptophan and phenylalanine metabolism; unassigned; valine leucine and isoleucine metabolism
Hydrocarbon metabolism	Citric acid cycle; cofactor and prosthetic group biosynthesis; folate metabolism; glycolysis gluconeogenesis; glyoxylate metabolism; methylglyoxal metabolism; oxidative phosphorylation; pentose phosphate pathway; putative; unassigned
Lipid metabolism	Anaplerotic reactions; cell envelope biosynthesis; citric acid cycle; cofactor and prosthetic group biosynthesis; folate metabolism; membrane lipid metabolism; methylglyoxal metabolism; nitrogen metabolism; oxidative phosphorylation; pyruvate metabolism

Table 4. The number of internal reactions, exchange reactions, extreme pathways and regulatory or disease-associated reactions in each subnetwork. The values in the parentheses are the proportion of regulatory or disease-associated reactions among the internal reactions. EP, extreme pathway; RXN, reaction; Reg-, regulatory; Dis, disease-associated.

Subnetwork	No. internal RXNs	No. exchange RXNs	No. EPs	No. Reg- or Dis- RXNs (Rate)
Amino acid metabolism (*E. coli*)	255	141	385	167 (65.49%)
Hydrocarbon metabolism (*E. coli*)	114	87	396	54 (47.37%)
Lipid metabolism (*E. coli*)	127	83	239	65 (51.18%)
Hydrocarbon metabolism (human)	333	320	922	101 (30.33%)
Amino acid metabolism (human)	365	288	402	107 (29.32%)

Table 5. The span of the four parameters required in EqSet sequence calculation.

Parameter	Span
μ	0, 0.025, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9, 1
ρ_e	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, 22, 24, 28, 32, 36, 40, 44, 48, 52, ∞
ρ_s	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 16, 18, 20, 22, 24, 28, 32, 36, 40, 44, 48, 52, ∞
τ	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40

The sequences with the lowest *p*-values are shown in S6 and S7 Tables for the above two subnetworks, respectively, whose front parts are enriched with disease-associated reactions. After a literature study of the internal EqSets which stand in the top 30 of the sequences but are not disease-associated according to Lee's data [67], we found it worth mentioning that most of their member reactions were reported to be related to some diseases by other published papers (Table 8). Therefore, our approach of sorting reactions according to their

Table 6. Summary of the distribution of *p*-values corresponding to the EqSet sequences calculated on the candidate parameter values. The two right-most columns show the proportion of candidate parameter values that result in an EqSet sequence with *p*-value less than 0.05 or 0.01, respectively. Min, minimum; Max, maximum.

| Subnetwork	Min.	Max.	Median	Average	Variance	$p < 0.05$	$p < 0.01$
Amino acid metabolism (*E. coli*)	3.10×10^{-4}	0.9778	0.0266	0.1076	0.1924	62.07%	7.27%
Hydrocarbon metabolism (*E. coli*)	4.78×10^{-6}	0.9943	0.0232	0.1207	0.2389	78.94%	29.90%
Lipid metabolism (*E. coli*)	2.08×10^{-9}	0.7448	7.19×10^{-6}	0.0098	0.0456	96.13%	90.83%
Amino acid metabolism (human)	1.41×10^{-7}	0.7621	5.37×10^{-4}	0.0185	0.0047	91.93%	86.83%
Hydrocarbon metabolism (human)	7.33×10^{-6}	0.7892	0.0665	0.0830	0.0049	35.42%	5.73%
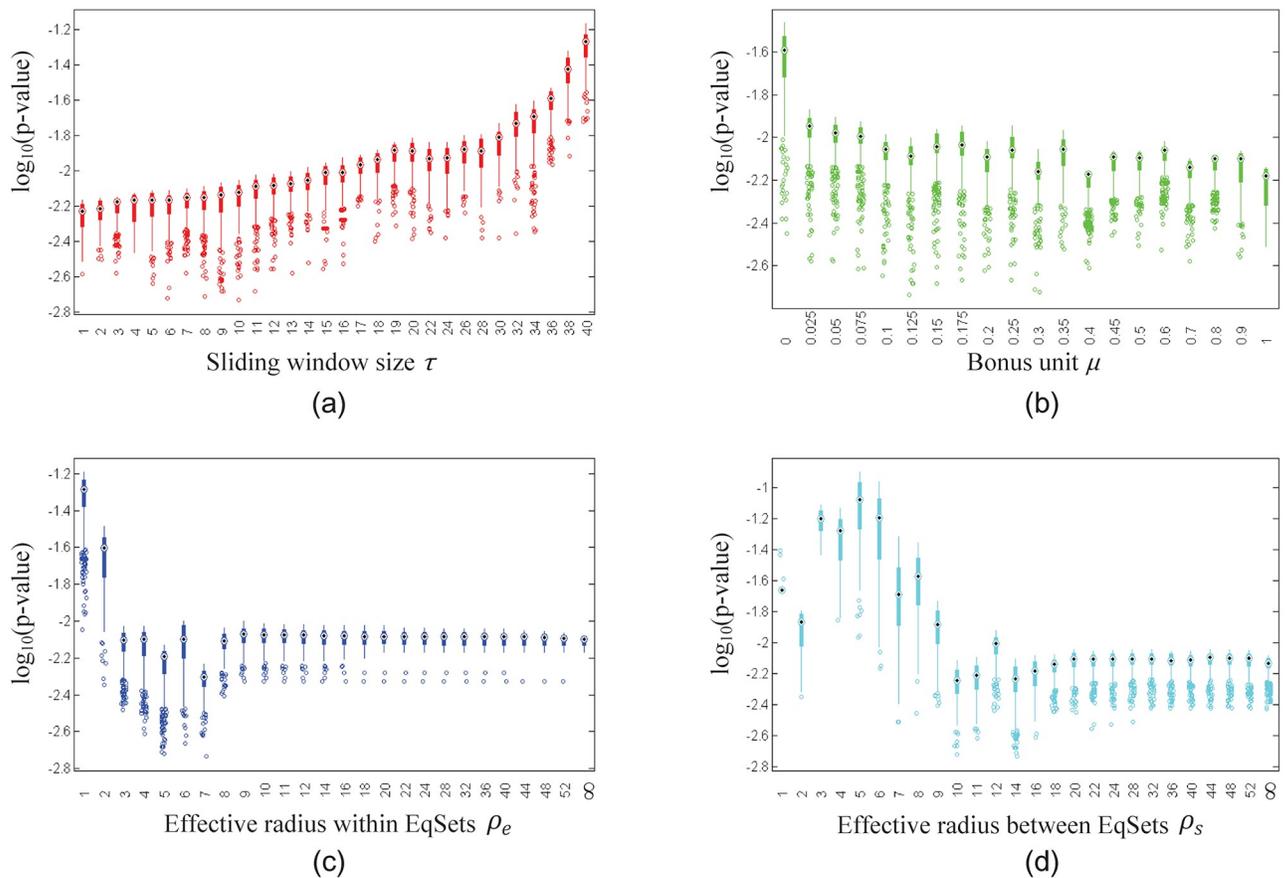
Fig 6. Box plots show the distribution of the lowest 5% of the *p*-values of the EqSet sequences of amino acid metabolism in *E. coli* when the parameter specified by the label of each subgraph is fixed to the value under the box. The ordinate axis is plotted in log scale.

regulatory importance is also capable of predicting the potential reactions which are related with some diseases.

In order to test whether the regulatory importance is a informative feature in detecting the disease-associated reactions, we divided the reactions of a subnetwork into a training set and a test set. The parameter values that fit the training set best, i.e., minimized the average *p*-value of the training set, were used to predict the disease-associated reactions in the test set. In the top 10% of the resulting sequences consisting of test EqSets, the average ratio of internal EqSets containing disease-associated reactions increased by 23% to 33% compared with that of random guessing, depending on the proportion of the EqSets used as training data (Fig 11). The accuracy of the prediction is significantly higher than random guessing, even if fewer disease-associated reactions are known (i.e., the proportion of the training set is low). Thus, the regulatory importance determined from extreme pathways can be a valuable new feature that contributes to the prediction of disease-associated reactions.

Quick search for suitable parameters

In the above examples, we tried over 200,000 different values of the parameters to find the best fit, which was so time-consuming that it may restrict the practical use of our methods. Therefore, an efficient approach to parameter optimization is extremely useful. Respecting the smooth and nearly single-trough distributions of the lowest *p*-values for each parameter,
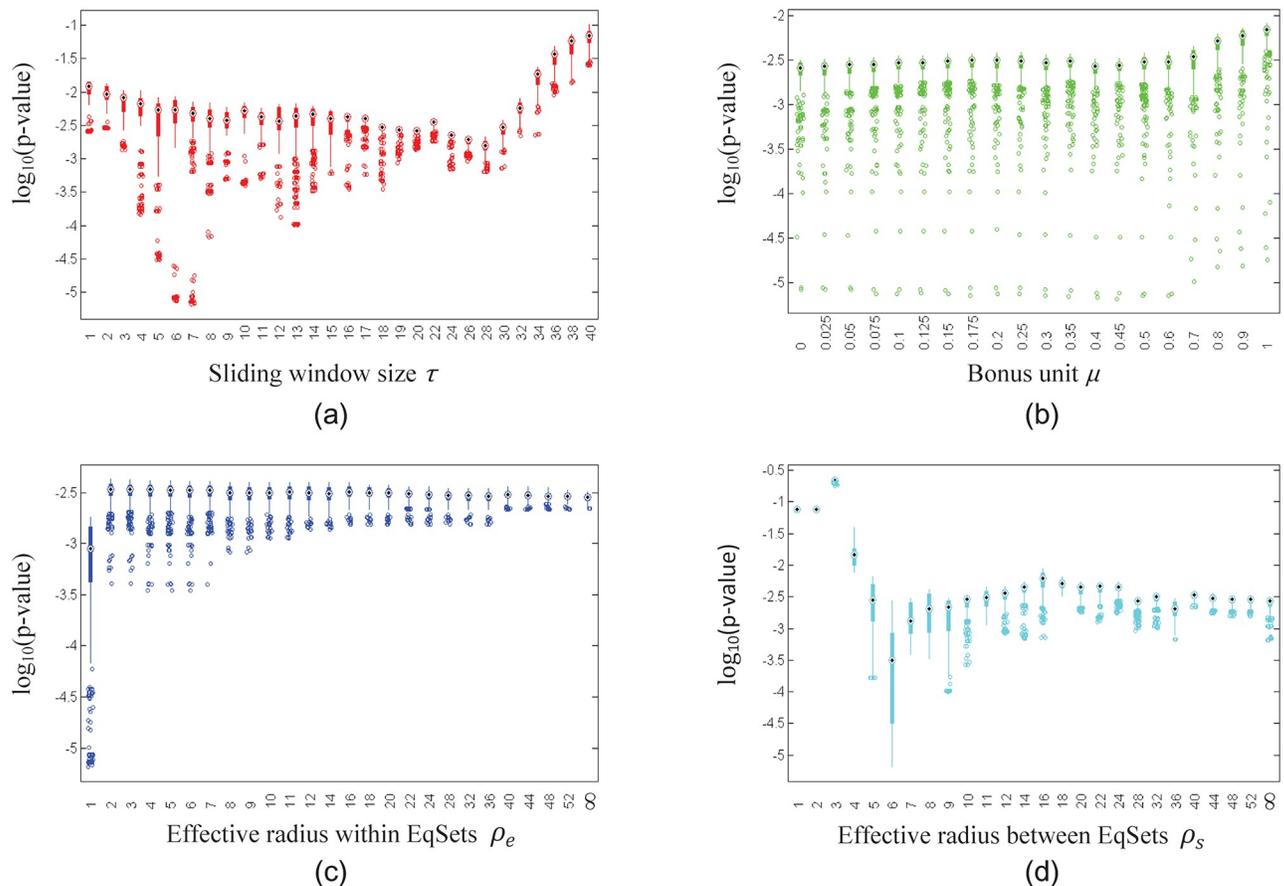
Fig 7. Box plots show the distribution of the lowest 5% of the *p*-values of the EqSet sequences of hydrocarbon metabolism of *E. coli* when the parameter specified by the label of each subgraph is fixed to the value under the box. The ordinate axis is plotted in log scale.

we improved the parameter search by use of a heuristic algorithm. The heuristic algorithm achieved the best or nearly best parameter values in the candidate set after trying 5,000 to 20,000 candidates (Fig 12); thereby providing a more than 90% time saving improvement compared with an exhaustive search.

Conclusion and discussion

Metabolism provides cells with energy and building blocks needed to produce biological structures, maintain the cell as well as carry out various cellular functions [66]. An elaborate regulatory structure enables the cell to adapt to a variety of internal and external perturbations. The organizing principle of the regulation of cellular metabolism remains a fundamental problem in biology. Previous research has focused on the regulatory patterns of metabolic reactions under certain perturbations of a cell's internal and environmental states. In this paper, we present a new approach for predicting the regulatory architecture of a cell's metabolism. The prediction is done by sorting the reactions according to their regulatory importance, which is measured based on the ratios of regulatory extreme pathways. By applying the method to the metabolic networks of the human erythrocyte and *E. coli*, we found that the regulation of metabolism prefers reactions that are efficient and flexible in controlling the extreme pathways.
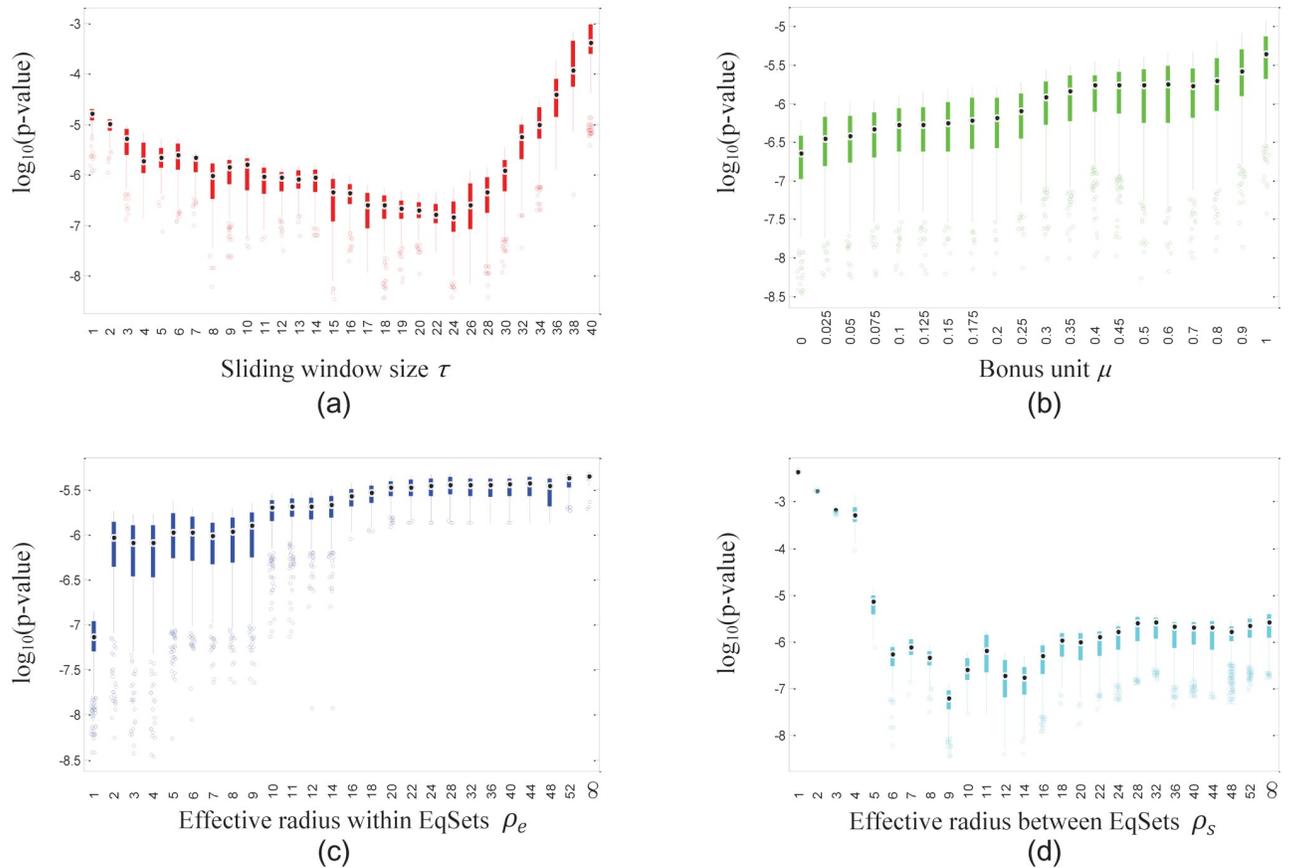
Fig 8. Box plots show the distribution of the lowest 5% of the *p*-values of the EqSet sequences of lipid metabolism of *E. coli* when the parameter specified by the label of each subgraph is fixed to the value under the box. The ordinate axis is plotted in log scale.

Table 7. Subsystems contained in each metabolic sub network of human.

| Subnetwork	Subsystems
Amino acid metabolism	Alanine and aspartate metabolism; aminosugar metabolism; arginine and proline metabolism; citric acid cycle; CoA biosynthesis; CoA catabolism; cysteine metabolism; D-alanine metabolism; folate metabolism; glutamate metabolism; glutathione metabolism; glycine, serine, and threonine metabolism; heme biosynthesis; heme degradation; histidine metabolism; lysine metabolism; methionine metabolism; phenylalanine metabolism; salvage pathway; taurine and hypotaurine metabolism; tetrahydrobiopterin; tryptophan metabolism; Tyr, Phe, Trp biosynthesis; tyrosine metabolism; urea cycle/amino group metabolism; valine, leucine, and isoleucine metabolism; vitamin b6 metabolism; beta-alanine metabolism
Hydrocarbon metabolism	Ascorbate and aldarate metabolism; biotin metabolism; CYP metabolism; citric acid cycle; CoA biosynthesis; CoA catabolism; folate metabolism; miscellaneous; fructose and mannose metabolism; NAD metabolism; glycolysis/gluconeogenesis; galactose metabolism; glyoxylate and dicarboxylate metabolism; hyaluronan metabolism; IMP biosynthesis; keratan sulfate degradation; N-glycan biosynthesis; unassigned; N-glycan degradation; oxidative phosphorylation; pentose phosphate pathway; pyruvate metabolism; pentose and glucuronate interconversions; putative; propanoate metabolism; riboflavin metabolism; salvage pathway; starch and sucrose metabolism; thiamine metabolism; oxidative phosphorylation; pentose phosphate pathway
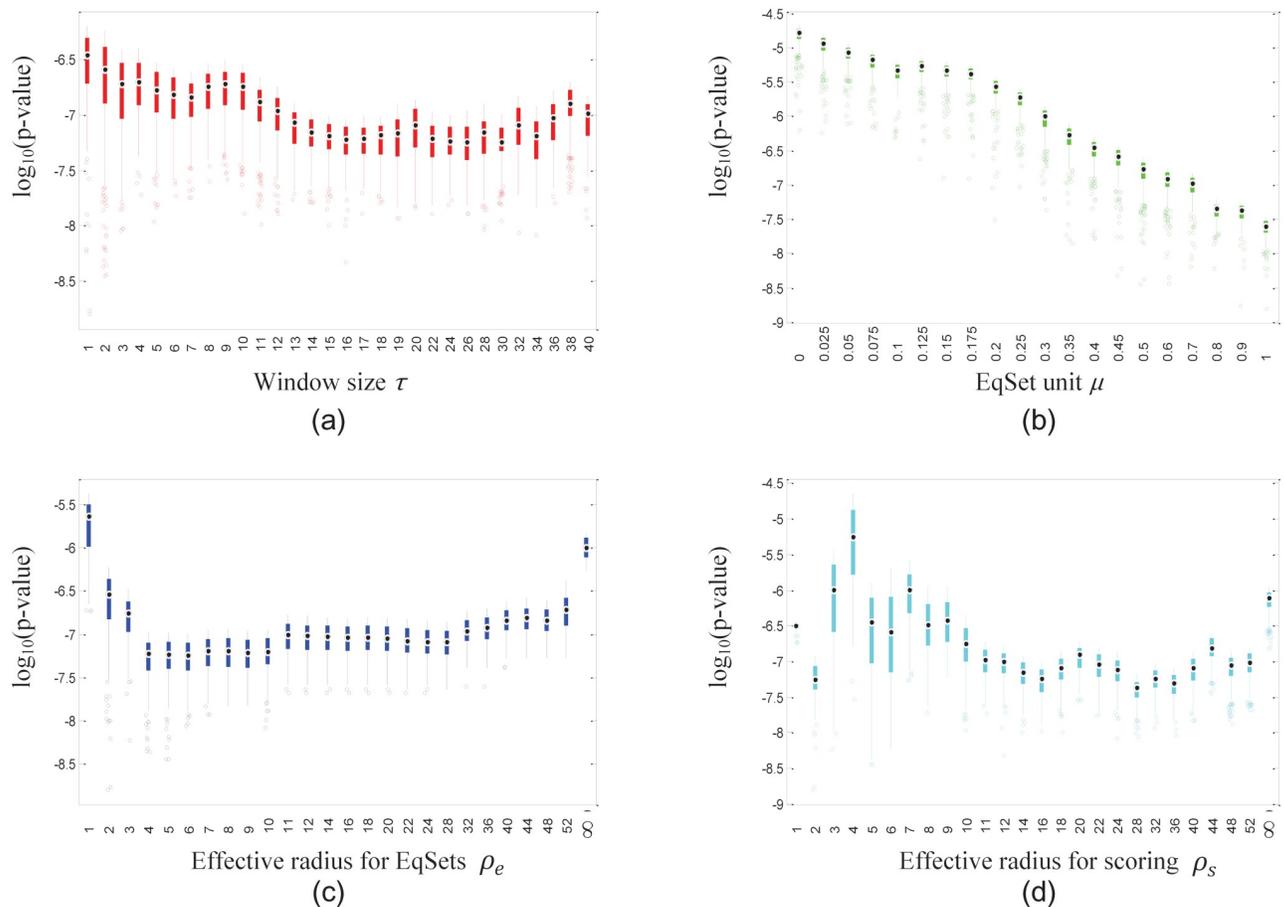
Fig 9. Box plots show the distribution of the lowest 5% of the *p*-values of the EqSet sequences of human amino acid metabolism when the parameter specified by the label of each subgraph is fixed to the value under the box. The ordinate axis is plotted in log scale.

https://doi.org/10.1371/journal.pone.0210539.g009

Our study sheds light on the possible mechanism whereby the regulatory architecture of a metabolic system is determined. It also supports the hypothesis that a cell transforms from one metabolic steady-state to another by switching "on/off" certain extreme pathways [36]. This implies that it is the extreme pathways, which are encoded in the stoichiometric relationship of a metabolic network, that are the real targets of metabolic regulation. In the process of evolution, metabolic reactions have acquired the ability to exert control by various means, such as the allosteric and transcriptional regulation of enzymes and genes, respectively. Reactions that play a key role in regulation are reserved as regulated reactions and form the evolutionary context in which the natural selection for the next regulated reaction occurs.

Our approach is context-free, meaning that it requires no information about the specific perturbations in the internal physiological states or environmental conditions of the cell. Rather, it concerns the whole space of valid metabolic steady-states and selects reactions that are important for regulation on average under various circumstances. Since the metabolic network and regulation strategies occurring in a cell are the result of an evolutionary process that arose under complicated and changeable internal and environmental conditions, our approach may provide general principles and closer approximations of the development of the regulatory structure of metabolism. The regulatory structure can be used as a blueprint, which provides the information about where metabolic regulation may occur, for context-dependent
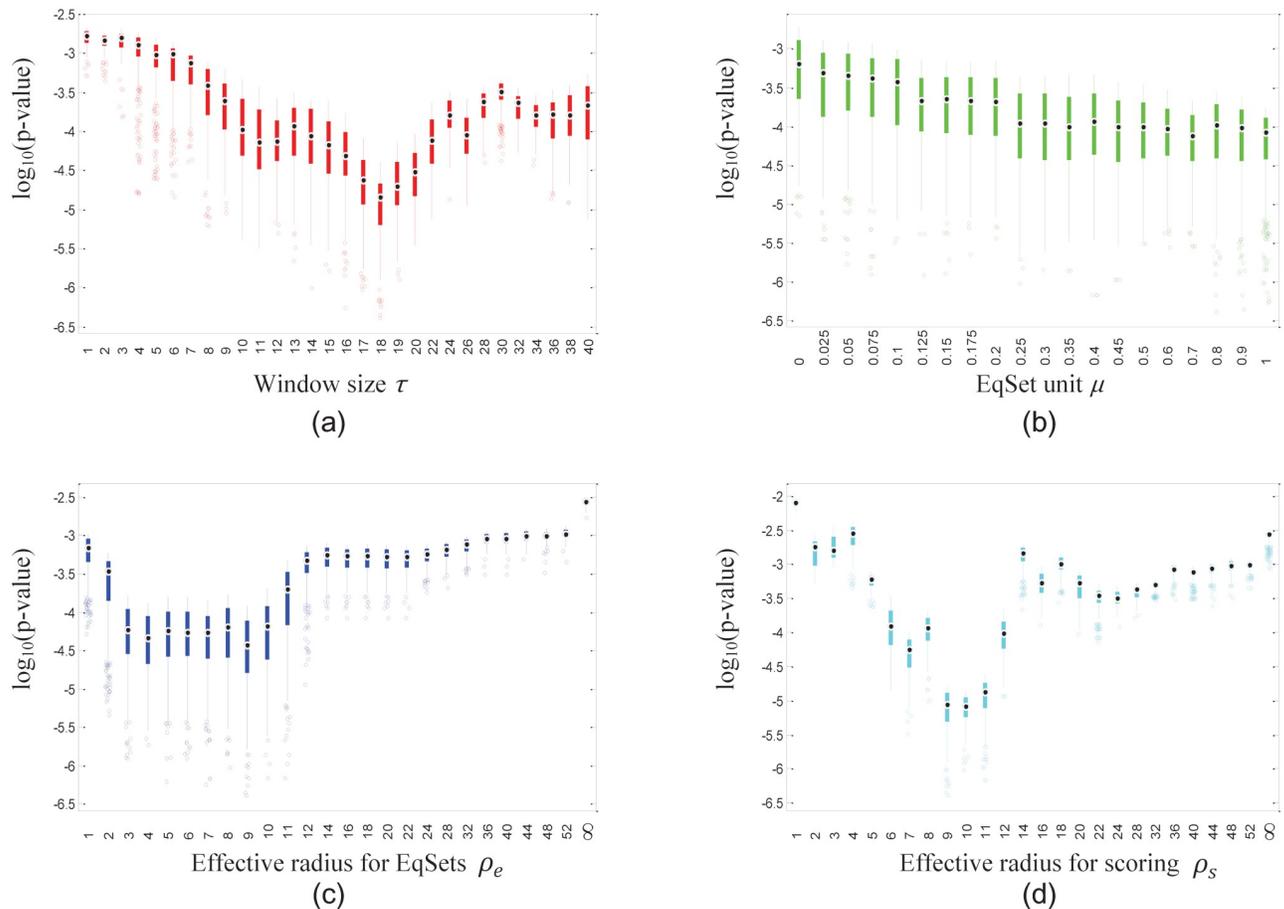
Fig 10. Box plots show the distribution of the lowest 5% of the *p*-values of the EqSet sequences of human hydrocarbon metabolism when the parameter specified by the label of each subgraph is fixed to the value under the box. The ordinate axis is plotted in log scale.

methods that aim to discover the regulatory strategies for certain perturbations. The combination of both context-free and context-dependent approaches may not only decrease the complexity of the problem, but also increase the quality of the answer.

Moreover, we demonstrated the ability of our approach to predict disease-associated reactions in human metabolic networks. According to the regulatory importance, we successfully detected the reactions whose mistuning or malfunction will lead to disease. In comparison with previous constraint-based reconstruction and analysis (COBRA) methods that predict lethal reactions, such as gene deletion analysis [96], the pathogenesis concerned in our study is not confined to the loss of enzymes, but also includes the loss of metabolic regulation. Gene deletion analysis requires the definitions of a cell's growing condition as well as its metabolic objective as represented by an objective function. The metabolic objective in many situations is unknown and moreover changes for different organisms, cell types, environmental conditions, or internal physiological states. Therefore, it is difficult to define the objective function, which prevents the practical use of gene deletion analysis. In contrast, our approach avoids that obstacle by using extreme pathway analysis. The prediction of disease-associated reactions can be used with other data, such as gene function and sequence variation, to screen out mutations closely related to diseases. In addition to its application to study disease processes, our approach can also be used in metabolic engineering and

Table 8. Non-disease-associated internal EqSets in top 30 of the sequences of the human amino acid or hydrocarbon metabolic networks with the lowest p-value. The reaction abbreviation and reaction name look-up table is listed in Table F in S1 Text.

EqSet	Rank	Disease-related reaction	Related diseases	References
Human amino acid metabolism				
3HAO, HKYNH, KYN3OX	13	3HAO	pellagra, olivopontocerebellar atrophy	[68, 69]
MTHFCm	24	MTHFCm	myelomeningocele, spina bifida	[70–72]
MTHFD2	26	MTHFD2	lung cancer, spina bifida, breast cancer	[73–76]
FKYNH, TRPO2	29	TRPO2	pellagra, haemophilus influenzae	[77]
GHMT2r	30	GHMT2r	adult acute lymphocytic leukemia, pediatric osteosarcoma	[78, 79]
Human hydrocarbon metabolism				
NMNS	9	NMNS	idiopathic recurrent pericarditis, gestational diabetes	[80–86]
NNDPR, EX_Sub_quln[c]	11	NNDPR	pellagra, follicular thyroid carcinoma	[87]
DOLGLCP_Lter, DOL_GPP_Ler, UDPDOLPT_L	12	-	-	-
DOLGLCP_Uter, DOL_GPP_Uer, UDPDOLPT_U	14	-	-	-
GAPD	19	GAPD	diffuse large B-cell lymphoma, obesity	[88, 89]
RBK, EX_Sub_rib-D[c]	27	-	-	-
FPGS4	28	FPGS4	Rheumatoid arthritis, psoriasis, colorectal cancer, non-small-cell lung cancer, non-Hodgkin lymphoma	[90–95]

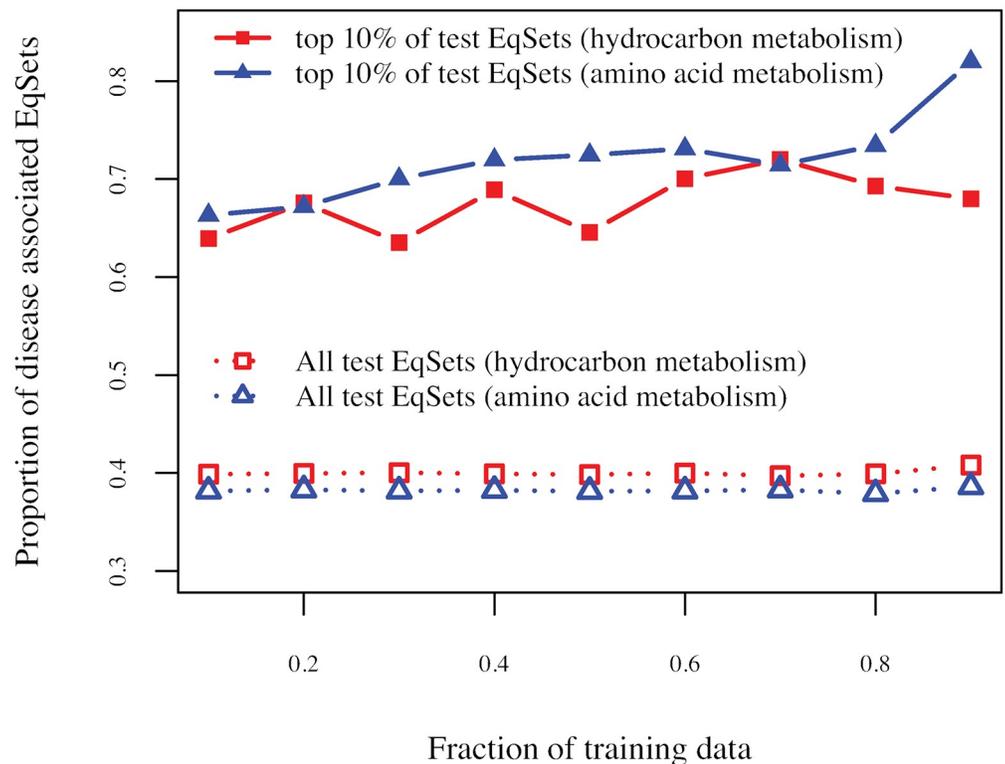https://doi.org/10.1371/journal.pone.0210539.t008

Fig 11. The average true positive rate versus the background proportion of a disease-associated EqSet. The average true positive rate equals the proportion of disease-associated EqSets among those that participate in the top 10% of the sequence composed of the test EqSets. The background proportion equals the ratio of disease-associated EqSets among all the test EqSets. The horizontal axis represents the fraction of EqSets that used as training data.
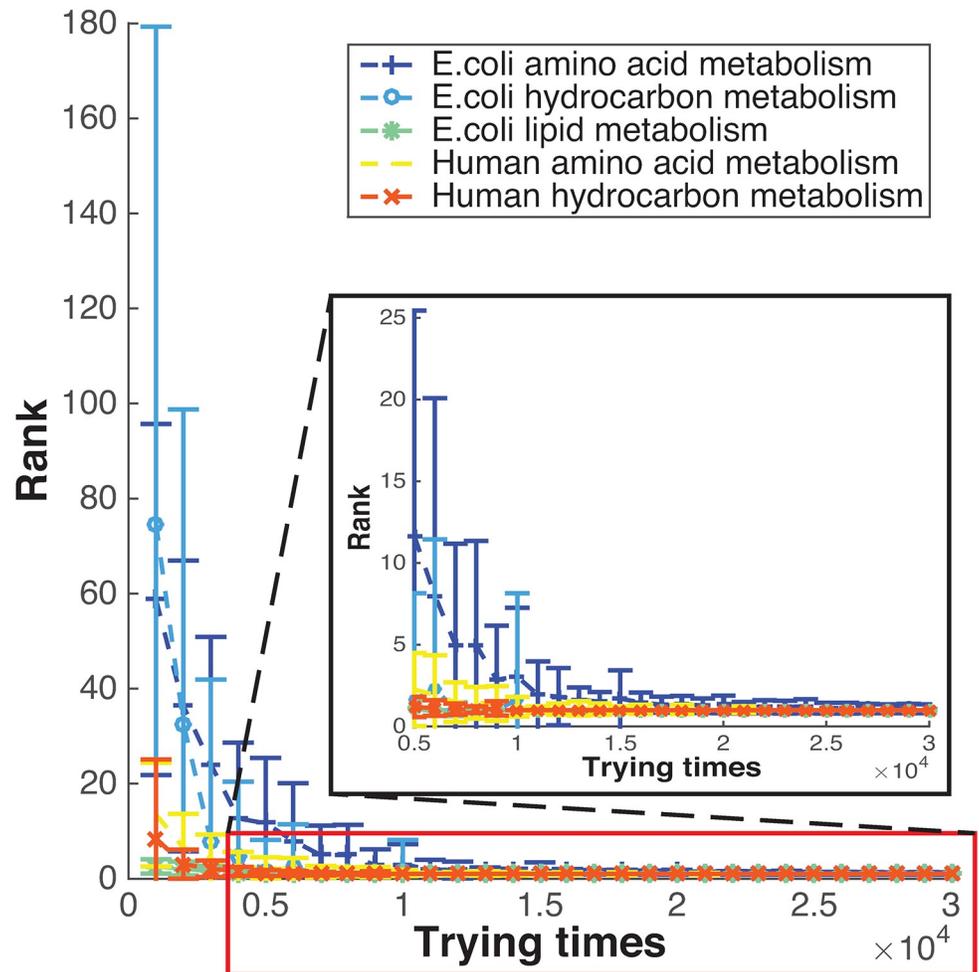
https://doi.org/10.1371/journal.pone.0210539.g011

Fig 12. The heuristic parameter searching algorithm quickly identified the best or nearly best among all the candidates after a few attempts. The horizontal axis represents the number of attempts in which an EqSet sequence is calculated on certain values of the 4 parameters. The number of attempts ranges from 1,000 to 50,000, with a step of 1,000. The vertical axis represents the number of the candidate parameter values for which the resulting *p*-values are no higher than those of the parameters suggested by the search algorithm. The algorithm was repeated 100 times at each attempt. The average ranks are shown by the line marks and standard deviations are shown with the error bars. The boxed off portion of the graph is zoomed in as a separate one to show more details.

synthetic biology to inform the optimal control nodes (reactions of most regulatory importance) in metabolic networks.

It is worth to note that the consideration of only subsystems of genome-scale metabolic networks is a trade-off between computational costs and accuracy. Although dividing the global metabolic network to a target subsystem and a surrounding subsystem does not change the space of steady states of the global system, constraining the analysis to the target subsystem does impact the extreme pathways detected [97, 98]. For example, some of the possible pathways of the global metabolic network will be missed [97] and interdependencies between pairs of fluxes may be lost [98]. Therefore the frequency of reaction participation in extreme pathways may be changed and some EqSets may be disrupted. As a result the predicted regulatory importance of reactions will be biased. In the previous study [97], Kaleta et al. suggested using another valuable concept of elementary flux patterns, which explicitly takes into account possible steady-state fluxes through the global metabolic network, instead of elementary modes or

extreme pathways. However elementary flux patterns can not solve our problem because the factor that 'each elementary flux pattern can be mapped to at leaset one elementary mode' in the global system [97] may introduce even more dramatical bias on the frequency of reaction participation in metabolic pathways. In this study, we reduce the above bias by optimizing the partition of the target subsystem. Our practice shows that it is feasible to divide the global metabolic network by reactions' functions. On the contrary, a target subsystem consisting of randomly selected reactions always resulted poor predictions of reactions' regulatory importance. However, an obvious disadvantage of function based partition is that the resulted subsystems are usually biased by the partitioner's knowledge of each reaction. An important part of the future work is to develop an algorithm to produce a reasonable partition of a genome-scale metabolic network. The aim of the algorithm is to introduce as few exchange reactions as possible while ensuring the complexity, i.e., the number of reactions, of the target subnetwork since abstracting the surrounding subsystem into exchange reactions is the main cause of the bias of extreme pathways detected in the target subsystem [97, 98].

Another challenge for the future work is to integrate omics data, such as transcriptomic and proteomic data, to refine our evaluation of the regulatory cost of each reaction, which will help improve the prediction quality of our approach. Valuable improvements will include determining the most possible regulated reactions in an EqSet, and increasing the agreement between the predicted EqSet sequence and the real sequence that occurs in nature. Furthermore, a quantitative estimation of the probability or confidence level that a reaction is regulatory or disease-associated is also in demand.

Supporting information

S1 Text. Supplementary methods and results.
(PDF)

S1 Table. The reaction abbreviations and reaction names of E.coli metabolic network and the information whether a reaction is regulatory or not.
(XLSX)

S2 Table. The reaction abbreviations and reaction names of human metabolic network and the information whether a reaction is disease-associated or not.
(XLSX)

S3 Table. The sequence of internal EqSets with the lowest p-value for the sub network of E. coli amino acid metabolism. P-value and the paramters for the sequence are listed at the end of the table.
(XLSX)

S4 Table. The sequence of internal EqSets with the lowest p-value for the sub network of E. coli hydrocarbon metabolism. P-value and the paramters for the sequence are listed at the end of the table.
(XLSX)

S5 Table. The sequence of internal EqSets with the lowest p-value for the sub network of E. coli lipid metabolism. P-value and the paramters for the sequence are listed at the end of the table.
(XLSX)

S6 Table. The sequence of internal EqSets with the lowest p-value for the sub network of human amino acid metabolism. P-value and the paramters for the sequence are listed at the

end of the table.
(XLSX)

S7 Table. The sequence of internal EqSets with the lowest p-value for the sub network of human hydrocarbon metabolism. P-value and the paramters for the sequence are listed at the end of the table.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Fei Wang.

Data curation: Yanping Xi.

Formal analysis: Yanping Xi.

Investigation: Yanping Xi.

Methodology: Yanping Xi.

Software: Yanping Xi.

Supervision: Fei Wang.

Validation: Fei Wang.

Visualization: Yanping Xi.

Writing – original draft: Yanping Xi.

Writing – review & editing: Fei Wang.

References

1. Mitchell A, Pilpel Y. A mathematical model for adaptive prediction of environmental changes by microorganisms. Proc Natl Acad Sci U S A. 2011; 108(17):7271–6. https://doi.org/10.1073/pnas.1019754108 PMID: 21487001

2. Mitchell A, Romano GH, Groisman B, Yona A, Dekel E, Kupiec M, et al. Adaptive prediction of environmental changes by microorganisms. Nature. 2009; 460(7252):220–4. https://doi.org/10.1038/nature08112 PMID: 19536156

3. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102 (8):2685–2689. https://doi.org/10.1073/pnas.0406811102 PMID: 15710883

4. Higuera C, Villaverde AF, Banga JR, Ross J, Morán F. Multi-Criteria Optimization of Regulation in Metabolic Networks. PLoS ONE. 2012; 7(7):e41122. https://doi.org/10.1371/journal.pone.0041122 PMID: 22848435

5. Chubukov V, Gerosa L, Kochanowski K, Sauer U. Coordination of microbial metabolism. Nat Rev Microbiol. 2014; 12(5):327–40. https://doi.org/10.1038/nrmicro3238 PMID: 24658329

6. Metallo CM, Vander Heiden MG. Understanding metabolic regulation and its influence on cell physiology. Mol Cell. 2013; 49(3):388–98. https://doi.org/10.1016/j.molcel.2013.01.018 PMID: 23395269

7. Price ND, Reed JL, Palsson BO. Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol. 2004; 2(11):886–97. https://doi.org/10.1038/nrmicro1023 PMID: 15494745

8. Aarash B, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. Nature Reviews Genetics. 2014; 15(2):107–20. https://doi.org/10.1038/nrg3643

9. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. Nature. 2004; 429(6987):92–96. https://doi.org/10.1038/nature02456 PMID: 15129285

10. Covert MW, Xiao N, Chen TJ, Karr JR. Integrating metabolic, transcriptional regulatory and signal transduction models in Escherichia coli. Bioinformatics. 2008; 24(18):2044–2050. https://doi.org/10.1093/bioinformatics/btn352 PMID: 18621757

11. Lee JM, Gianchandani EP, Eddy JA, Papin JA. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. Plos Computational Biology. 2008; 4(5):e1000086. https://doi.org/10.1371/journal.pcbi.1000086 PMID: 18483615

12. Koh R, Dunlop M. Modeling suggests that gene circuit architecture controls phenotypic variability in a bacterial persistence network. BMC Systems Biology. 2012; 6(1):47. https://doi.org/10.1186/1752-0509-6-47 PMID: 22607777

13. Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences. 2010; 107(41):17845–17850. https://doi.org/10.1073/pnas.1005139107

14. Watson E, MacNeil LT, Arda HE, Zhu LJ, Walhout AJ. Integration of metabolic and gene regulatory networks modulates the C. elegans dietary response. Cell. 2013; 153(1):253–66. https://doi.org/10.1016/j.cell.2013.02.050 PMID: 23540702

15. Österlund T, Nookaew I, Bordel S, Nielsen J. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. BMC Systems Biology. 2013; 7(1):1–10.

16. Stanford NJ, Lubitz T, Smallbone K, Klipp E, Mendes P, Liebermeister W. Systematic Construction of Kinetic Models from Genome-Scale Metabolic Networks. Plos One. 2013; 8(11):1845–1858. https://doi.org/10.1371/journal.pone.0079195

17. Liu YY, Slotine JJ, Barabasi AL. Controllability of complex networks. Nature. 2011; 473(7346):167–73. https://doi.org/10.1038/nature10011 PMID: 21562557

18. Lim WA, Lee CM, Tang C. Design principles of regulatory networks: searching for the molecular algorithms of the cell. Mol Cell. 2013; 49(2):202–12. https://doi.org/10.1016/j.molcel.2012.12.020 PMID: 23352241

19. Jia T, Liu YY, Csoka E, Posfai M, Slotine JJ, Barabasi AL. Emergence of bimodality in controlling complex networks. Nat Commun. 2013; 4:2002. https://doi.org/10.1038/ncomms3002 PMID: 23774965

20. Zelezniak A, Sheridan S, Patil KR. Contribution of network connectivity in determining the relationship between gene expression and metabolite concentration changes. Plos Computational Biology. 2014; 10(4):e1003572. https://doi.org/10.1371/journal.pcbi.1003572 PMID: 24762675

21. Wessely F, Bartl M, Guthke R, Li P, Schuster S, Kaleta C. Optimal regulatory strategies for metabolic pathways in Escherichia coli depending on protein costs. Mol Syst Biol. 2011; 7:515. https://doi.org/10.1038/msb.2011.46 PMID: 21772263

22. Chubukov V, Zuleta IA, Li H. Regulatory architecture determines optimal regulation of gene expression in metabolic pathways. Proceedings of the National Academy of Sciences. 2012; 109(13):5127–5132. https://doi.org/10.1073/pnas.1114235109

23. Gilman A, Ross J. Genetic-algorithm selection of a regulatory structure that directs flux in a simple metabolic model. Biophysical journal. 1995; 69(4):1321–1333. https://doi.org/10.1016/S0006-3495(95)79999-4 PMID: 8534802

24. Link H, Christodoulou D, Sauer U. Advancing metabolic models with kinetic information. Curr Opin Biotechnol. 2014; 29:8–14. https://doi.org/10.1016/j.copbio.2014.01.015 PMID: 24534671

25. Kiran Raosaheb P, Jens N. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(8):2685–2689. https://doi.org/10.1073/pnas.0406811102

26. Sajitz-Hermstein M, Nikoloski Z. Structural control of metabolic flux. PLoS Comput Biol. 2013; 9(12):e1003368. https://doi.org/10.1371/journal.pcbi.1003368 PMID: 24367246

27. Machado Daniel, Herrgård Markus J. Co-evolution of strain design methods based on flux balance and elementary mode analysis. Metabolic Engineering Communications. 2015; 2:85–92. https://doi.org/10.1016/j.meteno.2015.04.001

28. Kim WJ, Ahn JH, Kim HU, Kim TY, Lee SY. Metabolic engineering of Mannheimia succiniciproducens for succinic acid production based on elementary mode analysis with clustering Biotechnol J. 2017; 12(2):1600701. https://doi.org/10.1002/biot.201600701

29. Schuster S, Hlgetag C. On elementary flux modes in biochemical reaction systems at steady state. Journal of Biological Systems. 1994; 2:165–182. https://doi.org/10.1142/S0218339094000131

30. Orman MA, Mattick J, Androulakis IP, Berthiaume F, Ierapetritou MG. Stoichiometry based steady-state hepatic flux analysis: Computational and experimental aspects. Metabolites. 2012; 2:268–291. https://doi.org/10.3390/metabo2010268 PMID: 24957379

31. Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. Metabolic network structure determines key aspects of functionality and regulation. Nature. 2002; 420(6912):190–3. https://doi.org/10.1038/nature01166 PMID: 12432396

32. Cakir T, Kirdar B, Ulgen KO. Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks. Biotechnol Bioeng. 2004; 86(3):251–60. https://doi.org/10.1002/bit.20020 PMID: 15083505

33. Cakir T, Kirdar B, Onsan ZI, Ulgen KO, Nielsen J. Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in Saccharomyces cerevisiae. BMC Syst Biol. 2007; 1:18. https://doi.org/10.1186/1752-0509-1-18 PMID: 17408508

34. Ates O, Arga KY, Oner ET. The stimulatory effect of mannitol on levan biosynthesis: Lessons from metabolic systems analysis of Halomonas smyrnensis AAD6(T.). Biotechnol Prog. 2013; 29(6):1386–97. https://doi.org/10.1002/btpr.1823 PMID: 24123998

35. Dotsenko VA, Obolenskaya MY. Mathematical modeling of folate-related processes in human placenta. Biopolymers & Cell. 2014; 30(2):149–156. https://doi.org/10.7124/bc.00088E

36. Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol. 2000 Apr 7; 203(3):229–48. https://doi.org/10.1006/jtbi.2000.1073 PMID: 10716907

37. Wiback SJ, Palsson BO. Extreme pathway analysis of human red blood cell metabolism. Biophysical Journal. 2002; 83(2):808–818. https://doi.org/10.1016/S0006-3495(02)75210-7 PMID: 12124266

38. Price ND, Reed JL, Papin JA, Famili I, Palsson BO. Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. Biophys J. 2003 Feb; 84(2 Pt 1):794–804. https://doi.org/10.1016/S0006-3495(03)74899-1 PMID: 12547764

39. Price ND, Reed JL, Papin JA, Wiback SJ, Palsson BO. Network-based analysis of metabolic regulation in the human red blood cell. J Theor Biol. 2003 Nov 21; 225(2):185–94. https://doi.org/10.1016/S0022-5193(03)00237-6 PMID: 14575652

40. Papin JA, Price ND, Palsson BO. Extreme pathway lengths and reaction participation in genome-scale metabolic networks. Genome Res. 2002 Dec; 12(12):1889–900. https://doi.org/10.1101/gr.327702 PMID: 12466293

41. Covert MW, Palsson BO. Constraints-based models: regulation of gene expression reduces the steady-state solution space. J Theor Biol. 2003 Apr 7; 221(3):309–25. https://doi.org/10.1006/jtbi.2003.3071 PMID: 12642111

42. Wiback SJ, Mahadevan R, Palsson BØ. Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. J Theor Biol. 2003 Oct 7; 224(3):313–24. https://doi.org/10.1016/S0022-5193(03)00168-1 PMID: 12941590

43. Llaneras F, Picó J. An interval approach for dealing with flux distributions and elementary modes activity patterns. J Theor Biol. 2007 May 21; 246(2):290–308. Epub 2007 Jan 5. https://doi.org/10.1016/j.jtbi.2006.12.029 PMID: 17292923

44. Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO. Comparison of network-based pathway analysis methods. Trends Biotechnol. 2004; 22(8):400–5. https://doi.org/10.1016/j.tibtech.2004.06.010 PMID: 15283984

45. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metabolic Engineering. 2003; 5(4):264–276. https://doi.org/10.1016/j.ymben.2003.09.002 PMID: 14642354

46. Klitgord N, Segre D. The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. Genome Inform. 2010. 22: p. 41–55. PMID: 20238418

47. Bell SL, Palsson BO. Expa: a program for calculating extreme pathways in biochemical reaction networks. Bioinformatics. 2005; 21(8):1739–40. https://doi.org/10.1093/bioinformatics/bti228 PMID: 15613397

48. Price ND, Famili I, Beard DA, Palsson BO. Extreme pathways and Kirchhoff's second law. Biophys J. 2002; 83(5):2879–2882. https://doi.org/10.1016/S0006-3495(02)75297-1 PMID: 12425318

49. Cover TM, Thomas JA. Elements of information theory. Wiley-interscience; 2012.

50. Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biology. 2003; 4(9):R54. https://doi.org/10.1186/gb-2003-4-9-r54 PMID: 12952533

51. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci USA. 2007; 104:1777–1782. https://doi.org/10.1073/pnas.0610772104 PMID: 17267599

52. Schuster SHC, Jh W, Da F. Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. Journal of Mathematical Biology. 2002; 45(2):153–81. https://doi.org/10.1007/s002850200143 PMID: 12181603

53. Schuster S, Kenanov D. Adenine and adenosine salvage pathways in erythrocytes and the role of S-adenosylhomocysteine hydrolase. A theoretical study using elementary flux modes. Febs Journal. 2005; 272(20):5278–5290(13). https://doi.org/10.1111/j.1742-4658.2005.04924.x PMID: 16218958

54. Nelson DL, Cox MM. Lehninger principles of biochemistry (fourth edition). 4th ed. W.H. Freeman; 2005.

55. Lubert S, Mark BJ, L TJ. Biochemistry (Sixth edition). San Francisco: W.H. Freeman; 2007.

56. Dunaway GA, Kasten TP, Sebo T, Trapp R. Analysis of the phosphofructokinase subunits and isoenzymes in human tissues. Biochem J. 1988; 251(3):677–83. https://doi.org/10.1042/bj2510677 PMID: 2970843

57. Luzzatto L. Regulation of the activity of glucose-6-phosphate dehydrogenase by NADP+ and NADPH. Biochim Biophys Acta. 1967; 146(1):18–25. https://doi.org/10.1016/0005-2744(67)90069-1 PMID: 4383500

58. Yoshida A, Lin M. Regulation of glucose-6-phosphate dehydrogenase activity in red blood cells from hemolytic and nonhemolytic variant subjects. Blood. 1973; 41(6):877–91. PMID: 4145828

59. Rippa M, Giovannini PP, Barrett MP, Dallocchio F, Hanau S. 6-Phosphogluconate dehydrogenase: the mechanism of action investigated by a comparison of the enzyme from different species. Biochim Biophys Acta. 1998; 1429(1):83–92. https://doi.org/10.1016/S0167-4838(98)00222-2 PMID: 9920387

60. Takeuchi T, Nishino K, Itokawa Y. Purification and characterization of, and preparation of an antibody to, transketolase from human red blood cells. Biochim Biophys Acta. 1986; 872(1-2):24–32. https://doi.org/10.1016/0167-4838(86)90143-3 PMID: 3089282

61. Lonsdale D. Three case reports to illustrate clinical applications in the use of erythrocyte transketolase. Evidence Based Complementary and Alternative Medicine. 2007; 4(2):247–250. https://doi.org/10.1093/ecam/nel089 PMID: 17549243

62. Crespillo J, Llorente P, Argomaniz L, Montero C. APRT from erythrocytes of HGPRT deficient patients: kinetic, regulatory and thermostability properties. Mol Cell Biochem. 2003; 254(1-2):359–63. https://doi.org/10.1023/A:1027323521969 PMID: 14674717

63. Arnold WJ, Kelley WN. Adenine phosphoribosyltransferase. Methods Enzymol. 1978; 51:568–74. https://doi.org/10.1016/S0076-6879(78)51079-3 PMID: 692402

64. Murakami K, Piomelli S. Identification of the cDNA for human red blood cell-specific hexokinase isozyme. Blood. 1997; 89(3):762–6. PMID: 9028305

65. Murakami K, Blei F, Tilton W, Seaman C, Piomelli S. An isozyme of hexokinase specific for the human red blood cell (HKR). Blood. 1990; 75(3):770–5. PMID: 2297576

66. Palsson B. Systems biology: properties of reconstructed networks. Cambridge Univ Pr; 2006.

67. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabasi AL. The implications of human metabolic network topology for disease comorbidity. Proc Natl Acad Sci U S A. 2008; 105(29):9880–5. https://doi.org/10.1073/pnas.0802208105 PMID: 18599447

68. Decker RH, Kang HH, Leach FR, Henderson LM. Purification and properties of 3-hydroxyanthranilic acid oxidase. Journal of Biological Chemistry. 1961; 236(236):3076–82. PMID: 13884755

69. Brkić H. Human 3-hydroxyanthranilate 3,4-dioxygenase (3HAO) dynamics and reaction, a multilevel computational study. Molecular Biosystems. 2015; 11(3):898–907. https://doi.org/10.1039/c4mb00668b PMID: 25588817

70. Hum DW, Bell AW, Rozen R, Mackenzie RE. Primary structure of a human trifunctional enzyme. Isolation of a cDNA encoding methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase. Journal of Biological Chemistry. 1988; 263(31):15946–15950. PMID: 3053686

71. Rozen R, Barton D, Du J, Hum D W, Mackenzie R E, Francke U. Chromosomal localization of the gene for the human trifunctional enzyme, methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase-formyltetrahydrofolate synthetase. American Journal of Human Genetics. 1989; 44(6):781–6. PMID: 2786332

72. Watkins D, Schwartzentruber JA, Ganesh J, Orange JS, Kaplan BS, Nunez LD, et al. Novel inborn error of folate metabolism: identification by exome capture and sequencing of mutations in the MTHFD1 gene in a single proband. Journal of Medical Genetics. 2011; 48(9):590–592. https://doi.org/10.1136/jmedgenet-2011-100286 PMID: 21813566

73. Hosgood HD, Menashe I, Shen M, Yeager M, Yuenger J, Rajaraman P, et al. Pathway-based evaluation of 380 candidate genes and lung cancer susceptibility suggests the importance of the cell cycle pathway. Carcinogenesis. 2008; 29(10):1938–1943. https://doi.org/10.1093/carcin/bgn178 PMID: 18676680

74. Shaw GM, Lu W, Zhu H, Yang W, Briggs FB, Carmichael SL, et al. 118 SNPs of folate-related genes and risks of spina bifida and conotruncal heart defects. Bmc Medical Genetics. 2009; 10(1):49–49. https://doi.org/10.1186/1471-2350-10-49 PMID: 19493349

75. Laura L, Kirsi K, Rami M, John-Patrick M, Miro V, Olli K, et al. High-throughput RNAi screening for novel modulators of vimentin expression identifies MTHFD2 as a regulator of breast cancer cell migration and invasion. Oncotarget. 2012; 4(1):48–63.

76. Liu F, Liu Y, He C, Tao L, He X, Song H, et al. Increased MTHFD2 expression is associated with poor prognosis in breast cancer. Tumour Biology the Journal of the International Society for Oncodevelopmental Biology & Medicine. 2014; 35(9):8685–8690. https://doi.org/10.1007/s13277-014-2111-x

77. Comings DE, Muhleman D, Dietz G, Sherman M, Forest GL. Sequence of Human Tryptophan 2,3-Dioxygenase (TDO2): Presence of a Glucocorticoid Response-like Element Composed of a GTT Repeat and an Intronic CCCCT Repeat. Genomics. 1995; 29(2):390–396. https://doi.org/10.1006/geno.1995.9990 PMID: 8666386

78. Rao NA, Ambili M, Jala VR, Subramanya HS, Savithri HS. Structure–function relationship in serine hydroxymethyltransferase. Biochimica Et Biophysica Acta. 2003; 1647(1-2):24–9. https://doi.org/10.1016/S1570-9639(03)00043-8

79. Stover P, Schirch V. Serine hydroxymethyltransferase catalyzes the hydrolysis of 5,10-methenyltetrahydrofolate to 5-formyltetrahydrofolate. Journal of Biological Chemistry. 1990; 265(24):14227–33. PMID: 2201683

80. Kadoglou NPE, Tsanikidis H, Kapelouzou A, Vrabas I, Vitta I, Karayannacos PE, et al. Effects of rosiglitazone and metformin treatment on apelin, visfatin, and ghrelin levels in patients with type 2 diabetes mellitus. Metabolism Clinical & Experimental. 2010; 59(3):373–9. https://doi.org/10.1016/j.metabol.2009.08.005

81. Saddi-Rosa P, Oliveira CS, Giuffrida FM, Reis AF. Visfatin, glucose metabolism and vascular disease: a review of evidence. Diabetol Metab Syndr. 2010; 2:21. https://doi.org/10.1186/1758-5996-2-21 PMID: 20346149

82. Gonzalez-Gay MA, Vazquez-Rodriguez TR, Garcia-Unzueta MT, Berja A, Miranda-Filloy JA, de Matias JM, et al. Visfatin is not associated with inflammation or metabolic syndrome in patients with severe rheumatoid arthritis undergoing anti-TNF-alpha therapy. Clinical and Experimental Rheumatology. 2010; 28(1):56–62. PMID: 20346239

83. Mara G, Frédéric VG, Anthony R, Fabienne A, Oberdan L. The nicotinamide phosphoribosyltransferase: a molecular link between metabolism, inflammation, and cancer. Cancer Research. 2010; 70(1):8–11. https://doi.org/10.1158/0008-5472.CAN-09-2465

84. Paschou P, Kukuvitis A, Yavropoulou MP, Dritsoula A, Giapoutzidis V, Anastasiou O, et al. Genetic variation in the visfatin (PBEF1 / NAMPT) gene and type 2 diabetes in the Greek population. Cytokine. 2010; 51(1):25–7. https://doi.org/10.1016/j.cyto.2010.04.006 PMID: 20451405

85. Ünlütürk U, Harmancı A, Yıldız BO, Bayraktar M. Dynamics of Nampt/visfatin and high molecular weight adiponectin in response to oral glucose load in obese and lean women. Clinical Endocrinology. 2010; 72(4):469–474. https://doi.org/10.1111/j.1365-2265.2009.03674.x PMID: 19650786

86. Chiao-Po H, Nirmala H, Alcendor RR, Shinichi O, Junichi S. Nicotinamide phosphoribosyltransferase regulates cell survival through autophagy in cardiomyocytes. Autophagy. 2009; 5(8):1229–31. https://doi.org/10.4161/auto.5.8.10275

87. Liu H, Kcatton Woznica G, Crawford A, Botting N, Naismith J. Structural and Kinetic Characterization of Quinolinate Phosphoribosyltransferase (hQPRTase) from Homo sapiens. Journal of Molecular Biology. 2007; 373(3):755–63. https://doi.org/10.1016/j.jmb.2007.08.043 PMID: 17868694

88. Ercolani L, Florence B, Denaro M, Alexander M. Isolation and complete sequence of a functional human glyceraldehyde-3-phosphate dehydrogenase gene. Journal of Biological Chemistry. 1988; 263 (30):15335–15341. PMID: 3170585

89. Mazzola JL, Sirover MA. Subcellular localization of human glyceraldehyde-3-phosphate dehydrogenase is independent of its glycolytic function. Biochimica Et Biophysica Acta. 2003; 1622(1):50–56. https://doi.org/10.1016/S0304-4165(03)00117-X PMID: 12829261

90. vdS R J H M, Wessels JAM, Vries-Bouwstra D Jeska K, Goekoop-Ruiterman YPM, Allaart CF, Judith B, et al. Exploratory analysis of four polymorphisms in human GGH and FPGS genes and their effect in methotrexate-treated rheumatoid arthritis patients. Pharmacogenomics. 2007; 8(2):141–50. https://doi.org/10.2217/14622416.8.2.141

91. Prabha R, Robert C, Sharon M, Ami M, Scott-Horton TJ, Richard B, et al. Methotrexate (MTX) pathway gene polymorphisms and their effects on MTX toxicity in Caucasian and African American patients with rheumatoid arthritis. Journal of Rheumatology. 2008; 35(4):572–579.

92. Warren RB, Smith RL, Campalani E, Eyre S, Smith CH, Barker JNWN, et al. Outcomes of methotrexate therapy for psoriasis and relationship to genetic polymorphisms. British Journal of Dermatology. 2009; 160(2):438–441. https://doi.org/10.1111/j.1365-2133.2008.08898.x PMID: 19016697

93. Figueiredo JC, Levine AJ, Lee WH, Conti DV, Poynter JN, Campbell PT, et al. Genes involved with folate uptake and distribution and their association with colorectal cancer risk. Cancer Causes & Control. 2010; 21(4):597–608. https://doi.org/10.1007/s10552-009-9489-6

94. Adjei AA, Mandrekar SJ, Dy GK, Molina JR, Adjei AA, Gandara DR, et al. Phase II trial of pemetrexed plus bevacizumab for second-line therapy of patients with advanced non-small-cell lung cancer: NCCTG and SWOG study N0426. Journal of Clinical Oncology Official Journal of the American Society of Clinical Oncology. 2010; 28(4):614–9. https://doi.org/10.1200/JCO.2009.23.6406 PMID: 19841321

95. Unhee L, Wang SS, Patricia H, Wendy C, Kelemen LE, Stephen C, et al. Gene-nutrient interactions among determinants of folate and one-carbon metabolism on the risk of non-Hodgkin lymphoma: NCI-SEER case-control study. Blood. 2007; 109(7):3050–9.

96. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol. 2012 Feb 27; 10(4):291–305. https://doi.org/10.1038/nrmicro2737 PMID: 22367118

97. Kaleta C, de Figueiredo LF, Schuster S. Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. Genome Res. 2009. 19(10): p. 1872–83. https://doi.org/10.1101/gr.090639.108 PMID: 19541909

98. Marashi SA, David L, Bockmayr A. On flux coupling analysis of metabolic subsystems. J Theor Biol. 2012. 302: p. 62–9. https://doi.org/10.1016/j.jtbi.2012.02.023 PMID: 22406036