To the Editorial Board of Cell Reports:

Thank you for your reviews of our manuscript, “Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity.” by Ed Reznik, Dimitris Christodoulou, Elad Noor, and colleagues. After carefully going through the reviewers' comments, we felt that the points they raised were fair and thoughtful. We have addressed each of the reviewers’ concerns, and find that these changes have strengthened the arguments and findings of the work. In particular, we have:

- Augmented the section on metabolic control analysis to include reversible and multi-substrate reactions. The SI Text now includes a comprehensive theoretical treatment of small molecule regulation, which highlights (among other things) the fundamental trade-off between regulation and enzyme activity.
- Clarified several confusing points regarding the potential role of thermodynamics, futile cycle regulation, and feedback inhibition from biosynthesis, in selection for small molecule regulation.
- Identified a set of “high-confidence” edges in the small-molecule regulatory network of *E. coli*, which can be used as prioritized candidates for putative functional regulators in future studies.

We believe that the combination of a first-of-its-kind (open-source) genome-scale reconstruction of a small molecule regulatory network, the novel theoretical insights derived from metabolic control analysis, and the evaluation of metabolic design principles that shape the regulatory network, will have a substantial impact on the study of metabolic regulation. In this regard, we are excited for the opportunity to have our work disseminated to the interdisciplinary readership of Cell Reports.

Sincerely,
Elad Noor

Reviewer #1:

This is an interesting article that analyses more objectively a number of ideas that have been floating around for some years in metabolic circles. For example are reactions with high delta G always regulated? I was glad to read that the answer is no. I have a couple of issues that I feel should be addressed and some commentary:

1 Action: Page 5 ('Unnesessary waste): Negative feedback loops in amino acid biosynthesis are likely to be there so that protein synthesis has control over the supply. One could loosly relate this to the idea of 'waste' but its better to think of it as supply control. We can quantify supply control but I'm not sure how to quantify waste. Might be worth mentioning this, see Hofmeyr etc on supply/demand control.
We agree with the reviewer and have referenced this work accordingly (**see Page 6**). Although we wanted to apply the supply/demand theory to amino acid biosynthetic pathways, the information we have available is not sufficient to fully parameterize kinetic models of these pathways and thus gain their supply and demand control coefficients. However, we felt we could improve our analysis by communicating our hypothesis regarding amino acid economy more quantitatively. Therefore, we used our SMRN in combination with amino acid concentrations from literature and calculated both (1) the elasticities of the biosynthetic enzymes of 4 amino acids (these are the only ones for which we have enough information) in 3 different experimental conditions and (2) the cost of production of these amino acids (**see Page 8 of main text, Figure S9**). In addition, we also plotted in **Figure S9** the metabolic cost of producing the four amino acids (glycine, alanine, aspartate, glutamate) which do not inhibit their own biosynthesis. The results quantitatively demonstrate the gap in metabolic cost between the 4 amino acids not engaged in feedback inhibition and 4 amino acids which are. Although we do not have adequate statistical power and therefore cannot make any strong claims, our result does suggest that cells may control the supply of expensive amino acids more tightly than those that are less costly.

2. Commentary: Section on "Quantifying the metabolic response to small molecules across different conditions and the trade off etc". I am surprised this even has to be discussed but I suppose given the very poor level of understanding most biochemists and molecular biologists have of metabolism it would seem to be necessary to describe this idea.

2 Action: One thing however I am not quite sure about. The authors use the irreversible non-product inhibited elasticity expression, 1 - s/(s + Km). However product inhibition can have a profound effect on the substrate elasticities. I think at minimum I would use the product inhibition expression: v = Vm S /(S + Km (1 + P/Kp)), the elasticity for S is Km(Kp + P)/(Km(Kp+P) + Kp S), and for P is -(Km P)/(Km (Kp + P) + Kp S). This would give a fairer picture of where enzymes are on their response curves. If one can't get hold of the reverse Km then the irreversible substrate elasticities will give an upper limit to the elasticity which still probably matches the conclusions of the paper. If the authors only use the irreversible form then this should be mentioned. I am not sure how product inhibition affects the substrate elasticities for MWC models, don't think it has been looked at closely.
This is indeed an important comment, and we agree that substrate elasticities could be decreased by product competitive inhibition. We added the derivation of the substrate and product elasticities in this case to the SI Text **(see Section 3.3.4)**. We also now clearly state (**see Page 6, Page 7, and Pages 10-11)** that we use the pure irreversible form as an approximation when we calculate elasticity values in our work (only very few enzymes have Km values for both substrate and product). We also elaborate on the consequences of assuming irreversible kinetics in light of other assumptions (*e.g.* monosubstrate kinetics) and concepts in metabolic control analysis (**Pages 10-1)**.

3.. Action The final and most troublesome part is related to the use of BRENDA as a source of kinetic data. Personally I don't believe most of the kinetic data on BRENDA either because its not well curated or more common is that the original papers most likely measured the kinetic constants under non-physiological conditions. Its been noted a few times that building reliable kinetics models requires the kinetics constants to be remeasured under physiological conditions.

Indeed, we are aware of the challenges one must face when mining these databases. To address these concerns, we include in our results **(see SI Table S1)** an information field indicating the number of independent literature references for each independent SMRN edge. One way to use this information is to assign higher confidence to SMRN edges which are independently reported in 2 or more literature references. We did so (**see Pages 3 and 10, Fig S10),** and found that 325 (~20% of the total) edges in our SMRN are reported in at least 2 distinct publications. While this analysis does not address the possibility that these interactions are observed under non-physiological conditions in

each independent literature report, they do provide a rationale for *prioritizing* edges for testing either in the wet lab or *in silico* in kinetic models.

Furthermore, while we agree with the reviewer regarding the risks of using BRENDA, we feel that the benefits of doing so outweigh the risks. We would point out that BRENDA and BioCyc have been used as the only data sources for numerous highly cited publications evaluating metabolic regulation in the past, such as Bar-Even et al. 2011 (Biochemistry); Davidi et al. 2016 (PNAS), and indeed the SIMMER algorithm mentioned by Reviewer 3, Hackett et al. 2016 (Science). BRENDA is the most comprehensive and up-to-date source of experimentally collected and manually curated kinetic parameters and allosteric interactions, and there is no better alternative at the moment. Being aware of the challenges of using these data sources, we put great effort in avoiding any biases, and validated our findings using multiple approaches (e.g. testing the thermodynamic hypothesis both in central carbon metabolism and across the whole network). Thus, we believe that the risk of using the data from BRENDA and BioCyc is outweighed by the opportunity to integrate much broader parts of *E. coli*'s metabolic network (as well as other organisms besides *E. coli*).

Reviewer #2:

The authors provide here an interesting bioinformatics study of the E. coli Small Molecule Regulatory Network (SMRN), i.e. the network of small molecules interacting with enzymes to regulate metabolic flux. The authors provide a framework (along with code publicly shared in github) to extract interactions between enzymes (E.C. numbers) and metabolites from the BRENDA and Biocyc databases. The topic is of interest in systems biology and metabolic engineering, since metabolites and allosteric regulation appear to be crucial in regulating metabolic fluxes. The provided open-source code is an excellent addition to the library of tools in systems biology.

The analysis of this data reveals an interesting trend: a trade-off between effector elasticity and the activity reduction caused by the interaction with the enzyme. Hence, using a small molecule effector to regulate a flux comes at the cost of lowering enzyme activity. Furthermore, this work provides a useful resource to prioritize enzyme/metabolite interactions to study flux control, and to study other system-level questions about metabolism.

The study is interesting, a good addition to the literature, and will certainly server the authors well in their future endeavors. However, this reviewer does not find anything groundbreaking about it. It would probably find a better readership in a journal such as BMC Bioinformatics.
We thank the reviewer for the positive commentary.

We would highlight that one of our main results, the genome-scale reconstruction of the SMRN of E. coli, has the potential to attract some interest and lay the foundation for many following studies, and can have updated versions in the future. We compare this result to the first publication of a genome-scale stoichiometric network reconstruction [Edwards and Palsson, 2000 (PNAS)] which has been cited almost 1000 times. Furthermore, the computational pipeline which we have built and provide alongside this publication (**https://github.com/eladnoor/small-molecule-regulation**) automatically reconstructs SMRNs for any organism with a suitable genome-scale metabolic model (~25 species now have such manually curated models: http://bigg.ucsd.edu/models).

Reviewer #3:

4. This work summarizes the development of a regulatory network to account for the allosteric regulations of enzyme activity by metabolite pools. The authors have constructed and analyzed this network for E. coli by extracting information from existing databases as well as compared it to other organisms. The work is in principle interesting but the authors need to explain how the inferred networks can be used in conjunction with kinetics models for improved predictions.
This is an excellent point. We have now included a discussion of possible application to kinetic modeling in the **discussion on Page 10**.

5 Major Concerns:

The authors report that irreversible reactions associated with a large reduction in ΔG are not preferred regulation targets. However, this contradicts literature findings and the data reported in the manuscript (in Figure 3) in which all irreversible reactions in central metabolism are primary targets of metabolite-based regulation.
This is a very interesting point that highlights the importance of our non-biased approach to the thermodynamic question. Indeed from Figure 3 (and as most people in the community think of it) all the irreversible reactions in central carbon metabolism (CCM) seem to be regulated. This is not entirely true, as the enzyme pgl (encoding 6-phosphogluconolactonase) is not known to be regulated, but it is true that the majority of reactions with large delta G in CCM have a small molecule acting as regulator. However, even if pgl is found in the future to be regulated by a small molecule, one must also consider the converse argument: many (19) reversible reactions in CCM that are in fact regulated by small molecules.

We decided to investigated the hypothesis that irreversible reactions are more likely to be regulated using a complementary approach to the Mann-Whitney U test, namely Gene Set Enrichment Analysis (GSEA). Using GSEA, we find a weak but statistically significant overrepresentation of regulation in reactions that have large reduction in ΔG, when focusing only in CCM (p-value: 0.1). We have now amended the manuscript on **Page 5 and SI Text Pages 3-4** to describe these results. Interestingly, when we repeat the analysis for the *whole* of metabolism, we find no statistically significant overrepresentation of regulation in reactions that have large reduction in ΔG (p-value: 0.25). If we were to focus only on CCM (or, for that matter, only on irreversible reactions) we would have risked biasing our interpretation of the answer to the thermodynamic question. Here, we put great effort to avoid that and to perform the analyses both in CCM *and* in the whole metabolism, taking also into account both reversible and irreversible reactions.

6 Furthermore, conservation of resources is rejected as a design principle but proposed as a motivator of feedback inhibition of amino acid biosynthetic pathways. In light of these conflicting arguments, the section on design principles needs to be amended to become consistent with the results.
This is not an inconsistency, but rather a confusion introduced by us when explaining two separate paradigms: futile cycle regulation (where we did not find a significant correlation to small molecule regulation) and feedback inhibition from amino acid biosynthesis pathways (where we did find one). It seems that our phrasing for the "second hypothesis" in the design principles section was not clear enough. We have clarified the distinction between futile cycles and amino acid biosynthesis and amended the manuscript on **Page 5**. We believe that this change will make the two separate outcomes clear.

7 In page 4 paragraph 4, the authors admit that the available literature data on allosteric regulations is focused on central carbon metabolism, due to lack of information in peripheral parts of metabolism. This implies that the inferred regulatory motifs are incomplete with conclusions probably limited to central metabolism only. The same source of bias also propagates in the species comparisons. The authors may want to comment on this.
We thank the reviewer for the comment. In fact, we made an effort where possibly to always repeat our analysis for both genome-scale metabolism as well as central carbon metabolism (CCM). Doing so ensured our results were both robust to possible biases imposed when focusing on CCM, but also offered readers an intuitive window on our findings. In particular, we focused on central carbon metabolism (CCM) to bridge our results with the expertise of our readers and to zoom in on the genome-scale result, but none of our conclusions are based solely on the analysis of CCM. Historically, most hypotheses about SMRN design were derived from looking only at this limited section of metabolism, and it was instrumental to test them first on the same small network, before generalizing it for the full network. We comment on this issue in the discussion on **Page 10**.

8 The Metabolic Control Analysis (MCA) does not consider enzyme kinetics of bi-substrate reactions, which typically have a rate law described as v=(v_max*(S_1/(K_M^S1))*(S_2/(K_M^S2)))/(1+S_1/(K_M^S1)+S_2/(K_M^S2)) as opposed to v= (v_max*S)/(K_M+S) for a mono-substrate reaction implying that elasticity parameter ϵ_S^v depends on the concentration of both substrates. Since such reactions are primary regulation targets (Figure 3) and the corresponding rate laws do not conform to the generalized rate law structures in the supplementary text. Furthermore, the authors need to better highlight novel insights gained as many of the discussed observations have already been mentioned before [1, 2].
Indeed, many of the observation we make about saturation levels are known from previous studies, such as the ones suggested by the reviewer. We now have made reference where appropriate (e.g. including but not limited to on **Page 4**).

Due to their frequent occurrence in biochemical reactions, we added the analysis of bi-substrate reactions and their elasticities to the supplementary text and conclude that in such cases, the formula for the elasticity of a single substrate would overestimate the actual elasticity value (**new SI section 3.4**). This result fits well with our general claim that a high elasticity is required (but not sufficient) for having a high control on the reaction flux.

9 It is unclear from the manuscript why a genome-scale metabolic model is required in this study as most analyses pertain to central metabolism. A comprehensive comparison with the results from SIMMER [3] which relies on omics datasets to identify novel regulatory interactions would be helpful to assess the efficacy of the approach.
As mentioned above, our analysis is done on both central carbon metabolism as well as on the full genome-scale network. We use this approach to make our results more intuitive to readers who are likely most familiar with central carbon metabolism. We agree that a comparison with the observations from SIMMER would be a valuable addition to this work; unfortunately, SIMMER has only been applied on yeast metabolism, and therefore cannot be directly compared to our results in *E. coli*.

In order to evaluate the efficacy of our approach, we prioritized interactions in the SMRN based on the number of independent literature reports supporting the interactions. Doing so, we computed a list of 325 (~20% of the entire SMRN) "high-confidence" regulatory interactions in the *E. coli* SMRN supported by 2 or more independent literature reports (**see Page 3 and 10, Figure S10).**

To further evaluate the efficacy of our approach specifically in *E. coli*, we compared our results to those from Link et al [4], where the importance of different small molecule enzyme interactions in *E. coli* was evaluated based on how well they could explain dynamic data, when included in a mechanistic kinetic model of glycolysis. A complete comparison between the results is not possible, because (1) the work by Link et al was focused only on glycolysis, and (2) we are still lacking information regarding most K_I values in glycolysis. However, we do find agreement between the most important interaction in that work (inhibition of *pfk* by PEP when shifting from pyruvate to glucose) and our own (high elasticity of *pfk* by PEP in gluconeogenic conditions, including pyruvate) (**see Page 8**) .

Minor Concerns
The authors should move the MCA and other procedure descriptions from the supplementary section into the method section as these are central to results presented in the manuscript.
Unfortunately, in order to qualify for the limitations on manuscript length imposed by Cell Press, we cannot move the full MCA and other procedures appearing in the SI to the main text or the methods section. However, we do repeat the most important findings from MCA in the results section, specifically regarding the trade-off between elasticity and enzyme activity, on **Page 7.**

Figure 3 contains several abbreviations that are not defined (2pglyc, ara5p, oxa). The inhibitions are not explicitly categorized into competitive, non-competitive, or uncompetitive.
Supplementary Table 1 lists all the small molecule regulators but does not indicate the type of inhibition (competitive/uncompetitive/non-competitive). The absence of this information creates an ambiguity in the interpretation of regulation of enzyme activity by small molecules.
We augmented the abbreviation list, to include the cases we had missed. The information regarding the way a small molecule inhibits an enzyme (e.g. competitive inhibition) is usually not provided for the majority of the interactions. As an alternative, we added this information to the **SI table S1** (for the interactions where the data was available in EcoCyc).

In page 6 paragraph 2, the authors should replace "…calculated a saturation…" with "…calculated the saturation level…"
Done

In Figure S8, the authors should add metabolite and enzyme names.
Done

References

1. Kochanowski, K., et al., Functioning of a metabolic flux sensor in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 2013. 110(3): p. 1130-1135.

2. Fenton, A.W. and G.D. Reinhart, Disentangling the Web of Allosteric Communication in a Homotetramer: Heterotropic Inhibition in Phosphofructokinase from Escherichia coli. Biochemistry, 2009. 48(51): p. 12323-12328.
3. Hackett, S.R., et al., Systems-level analysis of mechanisms regulating yeast metabolic flux. Science, 2016. 354(6311).
4. Link, H., Kochanowski, K. & Sauer, U., 2013. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. Nature Biotechnology, 31(4), pp.357–361.