

# SCIENTIFIC REPORTS



OPEN

## Alleles of a gene differ in pleiotropy, often mediated through currency metabolite production, in *E. coli* and yeast metabolic simulations

Deya Alzoubi, Abdelmoneim Amer Desouki & Martin J. Lercher 

A major obstacle to the mapping of genotype-phenotype relationships is pleiotropy, the tendency of mutations to affect seemingly unrelated traits. Pleiotropy has major implications for evolution, development, ageing, and disease. Except for disease data, pleiotropy is almost exclusively estimated from full gene knockouts. However, most deleterious alleles segregating in natural populations do not fully abolish gene function, and the degree to which a polymorphism reduces protein function may influence the number of traits it affects. Utilizing genome-scale metabolic models for *Escherichia coli* and *Saccharomyces cerevisiae*, we show that most fitness-reducing full gene knockouts of metabolic genes in these fast-growing microbes have pleiotropic effects, *i.e.*, they compromise the production of multiple biomass components. Alleles of the same metabolic enzyme-encoding gene with increasingly reduced enzymatic function typically affect an increasing number of biomass components. This increasing pleiotropy is often mediated through effects on the generation of currency metabolites such as ATP or NADPH. We conclude that the physiological effects observed in full gene knockouts of metabolic genes will in most cases not be representative for alleles with only partially reduced enzyme capacity or expression level.

A gene is pleiotropic if it affects more than one phenotypic trait<sup>1,2</sup>. A classic example is phenylketonuria, a human disease that is caused by a single gene defect but which affects multiple systems, with symptoms ranging from lighter skin color to mental disorders<sup>3</sup>. Pleiotropic effects can cause alleles to affect fitness differentially at different ages, a phenomenon believed to be a major cause of aging<sup>4–6</sup>; indeed, alleles contributing to increased longevity often show reduced fertility and stress tolerance<sup>7</sup>. Similar antagonistic epistasis may underlie other important biological phenomena such as speciation<sup>8</sup> and cooperation<sup>9</sup>. Understanding the factors that contribute to pleiotropy is of fundamental importance in genetics<sup>10–12</sup>, evolution<sup>13–16</sup>, development<sup>17,18</sup>, as well as in disease<sup>19,20</sup> and ageing<sup>4</sup>. In comparison to its fundamental importance, empirical knowledge of the prevalence and especially on the causal mechanisms of pleiotropy is scarce<sup>2,21</sup>.

Pleiotropy may be classified according to the types of traits considered<sup>22</sup>. Molecular gene pleiotropy refers to the number of functions of a gene and its products, *e.g.*, the number of reactions catalyzed by a single enzyme. Developmental pleiotropy describes the genetic and evolutionary interdependence of phenotypic aspects. Finally, selectional pleiotropy refers to the number of separate fitness components affected by mutations to a gene. In this study, we focus on the latter type of pleiotropy.

Experimental studies generally assess pleiotropy through the analysis of gene knockouts<sup>23–25</sup>. The degree of pleiotropy is then defined as the number of traits affected when a gene becomes fully non-functional. Wang *et al.*<sup>25</sup> analyzed phenotypes of large numbers of yeast, nematode, and mouse mutants. They found that pleiotropy is widespread: on average, yeast gene knockouts affect 8% of the examined traits; for the nematode, the corresponding number is 10%, for the mouse 3% (see also<sup>24</sup>). The distributions of the degree of pleiotropy appear rather similar across very different study systems, from the skeletal features of mice<sup>25</sup> to metabolic systems<sup>26–28</sup>. Moreover, pleiotropy was found to be modular, such that sets of genes tend to affect the same sets of traits<sup>25</sup>.

Institute for Computer Science and Department of Biology, Heinrich Heine University, Universitätsstraße 1, Düsseldorf, D-40221, Germany. Correspondence and requests for materials should be addressed to M.J.L. (email: [martin.lercher@hhu.de](mailto:martin.lercher@hhu.de))

In *E. coli*, 36% of metabolic reactions are catalyzed by enzymes also involved in other reactions; the same is true for 27% of metabolic reactions in the yeast *Saccharomyces cerevisiae*<sup>28</sup>. Pleiotropic effects of mutations that affect enzyme activity can be simulated from genome-scale metabolic models using constraint-based modeling techniques such as flux balance analysis (FBA)<sup>29,30</sup>. The functional pleiotropy of a metabolic gene can then be defined as the number of biomass components whose maximal production is affected by the gene's knockout<sup>26</sup>. Previous studies using this definition found that a metabolic gene's functional pleiotropy is related to its propensity to form negative epistatic interactions with other metabolic genes<sup>26,27</sup>.

While full gene knockouts are easily examined experimentally, they may not be representative of the effects of deleterious alleles segregating in natural populations: individual mutations may affect only a subset of all traits influenced by the gene<sup>31</sup>. Thus, it is important to distinguish between the pleiotropy of the gene and the pleiotropy of individual mutations, especially in evolutionary and clinical contexts. For example, while 4.6% of human SNPs implicated in complex non-Mendelian phenotypes show pleiotropic effects, most of these do not fully abolish protein function<sup>32</sup>. Experimental studies indicate that mutational pleiotropy tends to be smaller than gene pleiotropy<sup>31</sup>.

Genome-scale metabolic models allow us to dissect the relationship between gene and mutational pleiotropy in quantitative detail, without being hampered by the detection limits of experimental assays. Does the degree of pleiotropy depend on how severely a given allele of a metabolic gene reduces protein activity, *i.e.*, are the same number of functions affected when protein function or expression is reduced only partially? How modular is metabolic pleiotropy? Currency metabolites, such as ATP and NADPH, are used as cofactors in many otherwise unrelated reactions; it thus appears highly likely that a substantial fraction of metabolic pleiotropy is due to effects on the production of currency metabolites. Is such an effect of currency metabolites on patterns of pleiotropy confirmed by simulated data?

Below, we address these questions by analyzing the metabolic networks of a representative bacterial model system, *Escherichia coli*, and a corresponding eukaryotic system, the baker's yeast *Saccharomyces cerevisiae*. We find that most gene knockouts that impact fitness do so by affecting the production of multiple biomass components, and that the number of affected biomass components typically increases with increasing mutation severity. Pleiotropy is rarely a consequence of multiple molecular gene functions, but is an emergent property of the metabolic network. For many genes, pleiotropy is indeed mediated through their involvement in the generation of currency metabolites.

## Results

### Estimating pleiotropy from contributions to biomass components within the wildtype flux distribution.

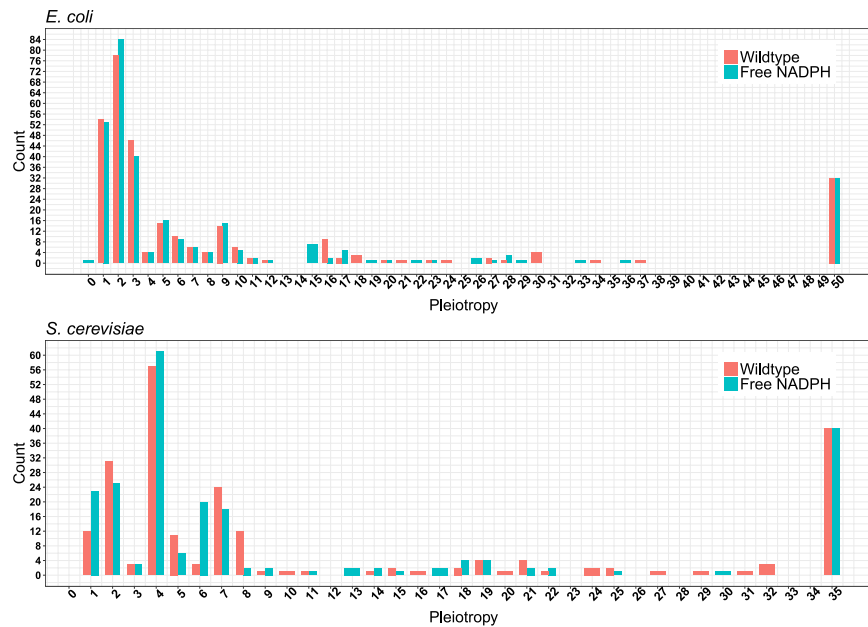
We first estimated wildtype flux distributions in the default growth condition for the genome-scale metabolic model of *E. coli*<sup>33</sup> and the yeast *S. cerevisiae*<sup>34</sup> (obtained from <https://sourceforge.net/projects/yeast/files/>). The maximal biomass production rates were estimated using flux balance analysis (FBA)<sup>29,30</sup>. For both model systems, we identified the flux distribution compatible with maximal biomass production that had the smallest sum of absolute fluxes, a strategy often termed parsimonious FBA (pFBA), which approximates optimal utilization of limited cellular protein resources<sup>35</sup>.

To simulate mutations that cause different reductions of protein function or expression and correspond to different deleterious alleles of a metabolic gene, we restricted the maximal flux through all reactions requiring this gene to a fixed percentage of the estimated wildtype flux<sup>36</sup>, starting from 100% (the wildtype) down to 0% (a full gene knockout) in steps of 0.5%. For each flux reduction, we defined the degree of pleiotropy (referred to simply as “pleiotropy” below) as the number of biomass components whose production was reduced by at least 0.01% compared to the maximal (wildtype) production. Note that with this definition, only genes with pleiotropy  $\geq 2$  are pleiotropic, while genes with pleiotropy 0 (no affected biomass component) or pleiotropy 1 (one affected biomass component) are non-pleiotropic.

Flux distributions at maximal biomass production rate are usually not unique<sup>35</sup>, and so in many cases a flux restriction through one reaction may be compensated by a rerouting of fluxes through alternative pathways. Such rerouting would require the upregulation of the corresponding genes. While it has been observed experimentally that cells can survive many gene deletions in central metabolism without drastic changes in gene expression<sup>37</sup>, the necessary upregulation of protein expression will not occur spontaneously at least for some pathways<sup>38</sup>. More importantly, if we are interested in the *de facto* contribution of a given gene to the production of biomass components, then it is of no consequence if alternative pathways *could* take over part of this functionality. Thus, when calculating the maximal (wildtype) production rate of individual biomass components as well as when simulating the effects of mutations to a given metabolic gene, we did not allow the redistribution of fluxes to alternative pathways: we allowed only decreases, not increases, of the absolute value of any flux compared to the wildtype flux distribution obtained with pFBA.

Note that experimental studies often employ a pragmatic working definition of pleiotropy that lies somewhere between the definitions of pleiotropy proposed here based on the wild type flux distribution on the one hand and a quantitative measure of essentiality based on an analogous calculation that allows the free redistribution of fluxes. In these studies, pleiotropy is typically estimated as the number of traits with observable phenotypic changes after the gene knockout, but before allowing the strain to adapt to its new genotype. In this case, some fluxes may be rerouted due to enzymes and transporters that are expressed either spuriously or because of other roles they play in wildtype physiology, while other fluxes that require the upregulation of the corresponding enzymes and transporters will not yet be active. Thus, our definition of pleiotropy describes a “worst case scenario”, providing an upper limit on experimentally measured pleiotropy.

**Many genes affect the production of multiple biomass components.** Pleiotropy varies widely between different genes. Mutations to the majority of genes affect no biomass components in the minimal growth



**Figure 1.** Most complete gene knockouts of fitness-relevant genes have pleiotropic effects, *i.e.*, they affect the production of multiple biomass components. For some genes, pleiotropy is reduced when NADPH is made freely available (cyan bars). For other freely available currency metabolites, see Supplementary Figure S2.

		<i>E. coli</i> <sup>c</sup>	<i>S. cerevisiae</i> <sup>c</sup>
<b>Number of biomass components</b>		<b>50</b>	<b>35</b>
Standard model	Pleiotropy <sup>a</sup>	10.0 (3)	12.1 (5)
	Essentiality <sup>a</sup>	4.6 (2)	9.3 (2)
Free NADPH <sup>b</sup>	Pleiotropy <sup>a</sup>	9.9 (3)	11.1 (4.5)
	Essentiality <sup>a</sup>	4.3 (2)	8.0 (1)

**Table 1.** Average number of biomass components whose production is affected by a full gene knockout. <sup>a</sup>In the FBA calculations, fluxes are either constrained to not exceed the wildtype (WT) fluxes to estimate the *de facto* contribution of gene products to biomass production (Pleiotropy), or they are allowed to vary freely to assess the number of biomass components for which gene products are essential even after allowing the mutant strain to adapt (Essentiality). <sup>b</sup>Solution when allowing unlimited conversion of NADPH to NADP<sup>+</sup> <sup>c</sup>Mean (median) number of affected biomass components.

medium assayed, independent of mutation severity (*E. coli*: 1,067 genes or 78.1%; *S. cerevisiae*: 687 genes or 75.6%). Among genes contributing to biomass production—and thus fitness—in the wildtype, non-pleiotropic cases are rare: in *E. coli*, only 54 full-gene knockouts (out of 299 knockouts with fitness contributions, 18.1%) affect exactly one biomass component, while the same is true for only 12 knockouts (out of 222 knockouts with fitness contributions, 5.4%) in *S. cerevisiae*. Conversely, knockouts of 32 genes in *E. coli* (10.7% of knockouts with fitness contributions) and 40 genes in *S. cerevisiae* (18.1% of knockouts with fitness contributions) affect the production of *all* biomass components. Many of the remaining genes show low degrees of pleiotropy, affecting the production of only a few biomass components; on average, full gene knockouts of fitness-relevant genes affect the production of 20% of biomass components in *E. coli* and 34% of biomass components in *S. cerevisiae* (Fig. 1, Table 1).

These percentages reflect functional pleiotropy, the *de facto* contribution of gene products to biomass component production. If, instead, we are interested in the phenotypic effects of gene knockouts after allowing the mutant strain to adapt its physiology to its altered gene content, we must allow free redistributions of fluxes after the gene knockouts. Corresponding simulations show that after adaptation, genes with fitness contributions are, on average, essential for the production of 9.2% of *E. coli* biomass components and of 26.6% of *S. cerevisiae* biomass components (Table 1, Supplementary Figure S1). The degree of gene pleiotropy for yeast is substantially higher than previous experimental estimates, which are around 2 (corresponding to 2–11% of considered traits depending on the types of traits analyzed)<sup>25</sup>; however, experimental estimates of gene pleiotropy tend to be downwardly biased due to experimental detection limits<sup>2,22</sup>.

**Pleiotropy is an emergent property of the metabolic network.** Pleiotropy can be classified by its origin into type I pleiotropy, caused by multiple molecular functions of a gene product, and type II pleiotropy, caused by multiple physiological consequences of a single molecular function<sup>2</sup>. Similar distinctions have been

made previously using the terms “horizontal” vs. “vertical”<sup>10</sup> and “mosaic” vs. “relational”<sup>39</sup> pleiotropy. Our model allows us to quantify the relative contributions of these two pleiotropy types. 41.7% of *E. coli* genes and 40.5% of yeast genes in our metabolic models catalyze multiple reactions. To what extent does this functional diversity cause functional pleiotropy as measured in the number of biomass components affected by a gene knockout? To answer this question, we compared the gene pleiotropy (Fig. 1) to the pleiotropy of individual reactions catalyzed by the gene product. For example, fully abolishing all functions of the *purB* (b1131) gene, whose gene product catalyzes two distinct biochemical reactions, reduced the production of 18 biomass components. In contrast, blocking only one of the catalyzed reactions results in a pleiotropy estimate of 16, while blocking only the other reaction results in a pleiotropy of 10. Thus, the pleiotropy of the b1131 gene is largely of type II, and is only in small part due to its multiple molecular functions.

This pattern is typical: the maximal pleiotropy arising from blocking only a single out of several reactions catalyzed by the same protein accounts for over 97% of the gene pleiotropy (*E. coli* 97.4%, yeast 97.6%). These numbers drop only marginally when we consider only gene products that are essential for multiple reactions, to 92.2% in *E. coli* and to 94.5% in yeast (Supplementary Figure S3). We conclude that the vast majority of metabolic epistasis is of type II, *i.e.*, is an emergent property of the metabolic network rather than a consequence of multiple molecular functions. This finding is consistent with the previous observation that the degree of pleiotropy in yeast is not significantly correlated with the number of molecular gene functions<sup>40</sup>.

**Metabolic networks show significant but low modularity.** The relationship between genes and biomass components (traits) can be represented as a bipartite graph, with links connecting genes with affected biomass components. Modules are defined as sets of genes and traits with significantly more within-module than between-module links<sup>25</sup>. A high degree of modularity thus indicates that pleiotropic genes tend to affect groups of related traits (*e.g.*, chemically related biomass components) rather than random sets of traits. Supplementary Figure S4 shows heatmaps that illustrate the modularity of both metabolic pleiotropy networks. To quantitatively assess the modularity, we used the LP&BRIM algorithm<sup>41</sup>, resulting in raw modularities of  $Q = 0.235$  for *E. coli* and  $Q = 0.197$  for *S. cerevisiae*. Both networks show highly statistically significant modularity: in each case, the modularity of 10,000 randomly rewired networks was always lower than observed for the real pleiotropy network (*i.e.*,  $P < 0.0001$ ; Supplementary Figure S5).

Following ref.<sup>25</sup>, we then defined a *z*-score for modularity (or “scaled modularity”)<sup>42</sup> as the difference between the observed modularity and the mean modularity of randomly rewired networks, measured in number of standard deviations. The *E. coli* pleiotropy network exhibits a scaled modularity of 9.1, while the *S. cerevisiae* network has a scaled modularity of 4.9, *i.e.*, the modularity of metabolic pleiotropy is about 9 and 5 standard deviations higher than for corresponding random gene-trait networks. These values are surprisingly low: for five different experimental study systems and trait definitions, Wang *et al.* found a median scaled modularity of 37 (range 34–238). Thus, metabolic pleiotropy networks are less modular than other pleiotropy networks, suggesting that the underlying metabolic network shows more interconnections between the pathways producing different sets of biomass components than the genetic networks underlying other types of traits. Our findings on modularity may be related to the role of currency metabolites, which crosslink the diverse metabolic pathways (see below).

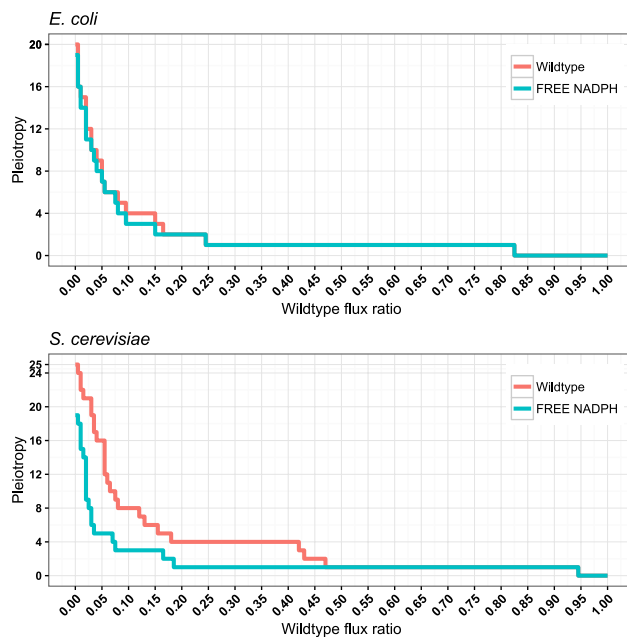
**Pleiotropy typically increases with increasing mutation severity.** We next examined the pleiotropy of alleles with small-effect mutations, *i.e.*, mutations that reduce enzyme capacity without fully abolishing enzyme function. About 20% of *E. coli* genes with fitness contributions have constant pleiotropy: small-effect mutations of these genes affect the same number of biomass components as full gene knockouts. In comparison, only 7.7% of yeast genes contributing to fitness exhibit constant pleiotropy.

All other genes contributing to fitness affect an increasing number of biomass components for increasingly deleterious alleles. Figure 2 shows this stepwise increase in pleiotropy for the example of *Lipoamide dehydrogenase* (gene names: *E. coli* b0116, *S. cerevisiae* YFL018C; for additional examples, see Supplementary Figure S6. In both organisms, pleiotropy typically increases in about a dozen steps from weakly to strongly deleterious alleles (Fig. 3; mean number of steps: *E. coli* 11.6, *S. cerevisiae* 12.6).

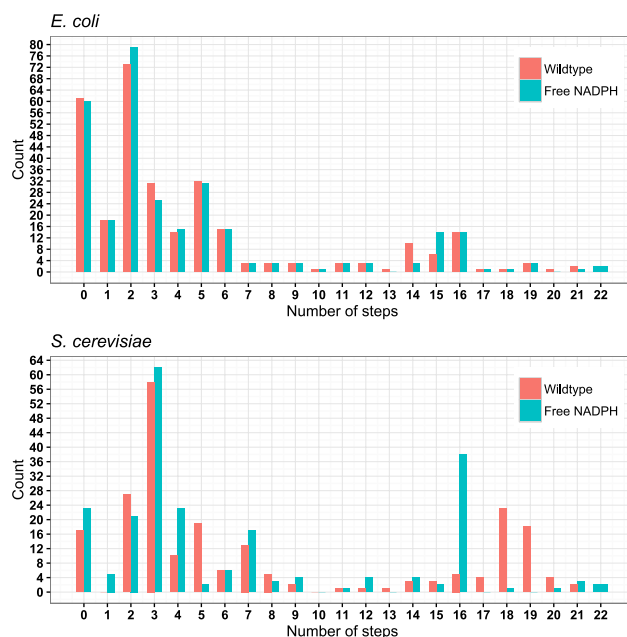
The pleiotropy of the full gene knockout constitutes an upper limit to the number of stepwise increases in pleiotropy. If there was otherwise no systematic relationship between maximal pleiotropy and the number of steps, we would expect the numbers of steps to be uniformly distributed between zero and the pleiotropy of the full knockout. However, the correlation between the number of steps and pleiotropy at full knockout was much stronger than expected from such a relationship (Supplementary Figure S7, Spearman’s  $\rho = 0.926$  (*E. coli*) and  $\rho = 0.986$  (*S. cerevisiae*),  $P < 10^{-6}$  in each case from randomizations; see Methods). Thus, genes whose full knockout showed higher metabolic pleiotropy also showed more stepwise increases in pleiotropy for increasingly debilitating mutations.

All genes whose mutations affect the production of at least one biomass component must also affect the overall production of biomass (*i.e.*, in the common interpretation of FBA, fitness). The reverse is not true: a mutation to a gene may affect the maximal production of biomass, but not the production of any individual biomass component. This is a consequence of the algorithm employed to estimate production capabilities for individual biomass components. If we maximize the production of a single compound, then pathways usually concerned with the production of other biomass components can be diverted to the production of this compound. While we find no such genes for *S. cerevisiae*, this is indeed the case for 3 essential *E. coli* genes, which encode transporters for acetate (b4067), magnesium/nickel/cobalt (b3816), and calcium/sodium (b3196, an antiporter).

**The pleiotropy of most genes is mediated by currency metabolites.** We can conceptually partition internal metabolites into currency metabolites—those involved in many reactions, *e.g.*, to provide energy or redox equivalents<sup>43</sup>—and primary metabolites. A deleterious allele may affect the production of a given biomass



**Figure 2.** Pleiotropy for the *Lipoamide dehydrogenase* gene increases for increasingly deleterious alleles. Pleiotropy is reduced when NADPH is made freely available (cyan curves). For additional examples, see Supplementary Figure S6.

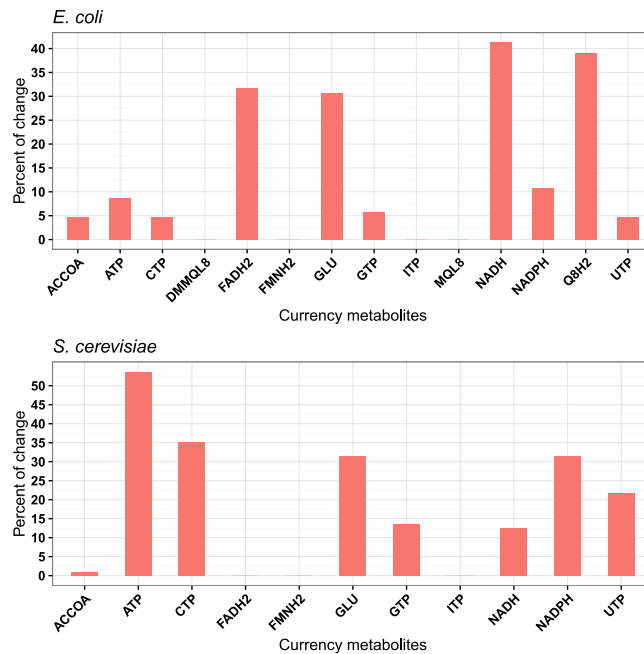


**Figure 3.** For the majority of genes contributing to biomass production, pleiotropy increases for increasingly deleterious alleles in multiple steps. Histograms for the number of pleiotropy steps in *E. coli* and the yeast *S. cerevisiae*. Cyan bars reflect the reduced numbers of pleiotropy increases when making NADPH freely available.

component because the mutated gene catalyzes a reaction in a pathway of primary metabolites that directly leads to the component's production. Conversely, a deleterious allele may affect not the primary metabolites, but the currency metabolites utilized in the component's production. A list of 14 currency metabolites was obtained from ref.<sup>43</sup>. Excluding exchange reactions, 753 out of a total of 2,251 reactions (33.5%) in *E. coli* and 310 out of 3,324 reactions (9.3%) in *S. cerevisiae* involved at least one of these metabolites.

A substantial fraction of pleiotropy is indeed associated with the generation of currency metabolites: 87.4% of previously pleiotropic genes show reduced pleiotropy when we make metabolites such as ATP, UTP, or NADPH freely available in yeast (Fig. 4). The free availability of ATP alone reduces the degree of pleiotropy of over half





**Figure 4.** Many genes show reduced pleiotropy when currency metabolites are made freely available. The bar chart shows the percentage of previously pleiotropic genes with reduced pleiotropy in response to the free availability of different currency metabolites. Abbreviations: Adenosine triphosphate (ATP); Cytidine triphosphate (CTP); Guanosine triphosphate (GTP); Uridine triphosphate (UTP); Inosine triphosphate (ITP); Nicotinamide adenine dinucleotide (NADH); Nicotinamide adenine dinucleotide phosphate (NADPH); Flavin adenine dinucleotide reduced (FADH<sub>2</sub>); Reduced flavin mononucleotide (FMNH<sub>2</sub>); Ubiquinol-8 (Q8H<sub>2</sub>); Menaquinol 8 (MQL8); 2-Demethylmenaquinol 8 (DMMQL8); Acetyl-CoA (ACCOA); L-Glutamate (GLU).

of pleiotropic yeast genes. The influence of currency metabolite production on pleiotropy is weaker, yet still substantial in *E. coli*: here, 55.3% of pleiotropic genes are affected, with NADH making the biggest contribution (over 40%) (Fig. 4).

Involvement in currency metabolite production is an important determinant of the number of biomass components for which a gene knockout is essential even after allowing the mutant strain to adapt its protein expression to the altered gene content of its genome. This contribution is particularly striking in yeast: for over half of the tested currency metabolites, free availability reduces the number of biomass components for which a gene is essential for almost half of the genes (Supplementary Figure S8).

## Discussion

Using constraint-based simulations of the metabolic models for *E. coli* and the yeast *S. cerevisiae*, we have characterized the distributions of pleiotropy. Consistent with earlier computational<sup>26–28</sup> and experimental<sup>23–25</sup> studies, we found that the knockout of a majority of genes that contribute to fitness has pleiotropic effects. The vast majority of this gene pleiotropy is not caused by multiple molecular functions of the gene product (type I), but is an emergent property of the metabolic network (type II). Pleiotropy is modular, but to a lower degree than estimated experimentally for non-metabolic systems<sup>25</sup>.

For most pleiotropic genes, pleiotropy increases strongly for alleles with increasingly debilitating effects. Thus, standard measures of pleiotropy based on gene knockout studies are more likely to reflect the maximal degree of mutational pleiotropy of a given gene<sup>2,31</sup>. Alleles that only knock down protein activity (by reducing enzyme/transporter function or expression level) often affect only a subset of phenotypic traits, with additional traits affected progressively as alleles become more deleterious. Thus, the physiological effect of the full gene knockout will in most cases not be representative for the effects of deleterious alleles that retain some level of enzyme function. This type of effect is also evident from individual medical observations of pleiotropy. For example, some small-effect mutations affecting human SOX9 expression lead to minor skeletal malformations, while the consequences of large-effect mutations can include sex reversals<sup>44</sup>.

How can we understand the dependence of pleiotropy on the degree to which an allele reduces protein activity? For increasingly deleterious alleles, more and more metabolic resources must be channeled into the compensation of the compromised pathway; as a consequence of this increasing drain of resources, more and more other pathways are affected. Not surprisingly<sup>27</sup>, we found that the pleiotropy of many genes is mediated through the generation of currency metabolites such as ATP, NADPH, or FADH<sub>2</sub>. This is true for more than half of the pleiotropic genes in *E. coli*, and for 87% of pleiotropic genes in yeast.

While the overall patterns of pleiotropy appear qualitatively similar between *E. coli* and yeast, we found a number of quantitative differences. Compared to *E. coli* genes, yeast genes (i) showed generally higher pleiotropy and were rarely of pleiotropy 1; (ii) were less likely to have constant pleiotropy; and (iii) were more likely to show

reduced pleiotropy when supplied with currency metabolites. Moreover, (iv) the yeast pleiotropy network exhibited lower modularity. In part, these differences may be related to network size. The *E. coli* metabolic network encompasses substantially more genes overall than the yeast network. However, we constrained network usage to reactions active in the wildtype. In contrast to total network sizes, the *active* metabolic network of yeast (755 reactions and 184 metabolites) is substantially larger than the *active* metabolic network of *E. coli* (462 reactions and 103 metabolites); this difference is consistent with the notion that yeast metabolism is more complex, yet less flexible than *E. coli* metabolism. While the average number of reactions per metabolite is similar between *E. coli* (4.49) and yeast (4.10), the lower yeast modularity indicates that reactions more often connect otherwise distant network parts. A role of currency metabolites in such connecting reactions would be consistent with the larger effect of currency metabolite supply on pleiotropy in yeast. In sum, the higher interconnectedness of the yeast pleiotropic network, combined with the larger active metabolic network size, appears to provide more potential for pleiotropic effects.

Pleiotropy is a complex phenomenon: it is not constant, but varies between different alleles of the same gene, and its causes are often indirect. Thus, experimental as well as computational analyses of pleiotropy should move away from focusing on full gene knockouts, and instead consider explicitly the degree to which mutations reduce protein activity. The necessity of a corresponding nuanced view of pleiotropy may be particularly evident in studies of medically relevant mutations, where full knockouts are often lethal, while small-effect mutations may segregate at appreciable frequencies in the human population<sup>45</sup>.

## Materials and Methods

**Metabolic models.** To simulate *Escherichia coli* metabolism, we used the metabolic reconstruction iJO1366<sup>33</sup>, encompassing 1,366 metabolic genes associated with 2,251 reactions. For the yeast *S. cerevisiae*, we used the yeast7.6 model (<https://sourceforge.net/projects/yeast>)<sup>34</sup>, accounting for 909 metabolic genes associated with 3,324 reactions. The published models were used without any modifications. The *E. coli* model contains a growth-independent maintenance energy consumption term (the ATPM reaction), which enforces a minimal ATPase activity of 3.15 mmol/gDW/h. We utilized the default biomass reactions for *E. coli* (Ec\_biomass\_iJO1366\_core\_53p95M), which comprises 50 essential biomass components (Supplementary Table S1), considering only “substrates” of the biomass reactions and excluding inorganic ions and H<sub>2</sub>O. For *S. cerevisiae*, we used the “yeast 5 biomass pseudoreaction”, which comprises 35 essential biomass components (Supplementary Table S2), again considering only “substrates” and excluding inorganic ions and H<sub>2</sub>O.

**Flux distribution constraints derived from wildtype simulations.** In order to approximate the *de facto* contribution of individual metabolic proteins to the production of individual biomass components *in vivo*, we should only consider flux distributions that are naturally active during growth (biomass production) in the nutritional environment studied, and fluxes should not exceed these wildtype fluxes. We thus first estimate the wildtype flux distribution  $\mathbf{v}^{\text{WT}}$ , by running a flux balance analysis (FBA) with the biomass reaction as the objective function, followed by a minimization of the sum of absolute fluxes at the previously determined maximal biomass production rate (parsimonious FBA<sup>35</sup>).

When simulating the production of individual biomass components, we constrained all fluxes  $v_i$  to values between zero and the wildtype flux  $v_i^{\text{WT}}$  for this reaction, *i.e.*,

$$\begin{aligned} 0 \leq v_i \leq v_i^{\text{WT}} \text{ for } v_i^{\text{WT}} \geq 0 \\ 0 \geq v_i \geq v_i^{\text{WT}} \text{ for } v_i^{\text{WT}} < 0 \end{aligned} \quad (1)$$

**Estimating pleiotropy.** For each essential biomass component (Supplementary Tables S1 and S2, respectively), we added a new exchange reaction representing its secretion<sup>46</sup>. As some biomass components may be coupled through the biomass reaction, we allowed the free excretion of all other biomass components when maximizing the production of one selected biomass component (*i.e.*,  $v_j \geq 0$  for all added exchange reactions  $j$ ).

We then calculated the maximum production of each biomass component by maximizing its exchange reaction flux while enforcing the wildtype flux distribution constraints (Eq. 1). For each metabolic gene, we compared this unperturbed maximal production with the maximal production rate of alleles with increasingly reduced protein activity, simulated by restricting the flux through all reactions catalyzed by the gene to a fixed fraction of the wildtype flux<sup>36</sup>, which we reduced from 100% to 0% in steps of 0.5%. The flux through a specific reaction was constrained in this way only if the gene-protein-reaction (GPR) mapping contained the affected gene either alone or only in an “AND” relationship (*i.e.*, as an essential part of a protein complex); if the GPR listed the affected gene in an “OR” relationship (*i.e.*, as one of multiple isoenzymes or alternative transporters), the reaction was not affected.

We defined pleiotropy as the number of biomass components whose maximal production was reduced by at least 0.01% compared to the unperturbed state (WT) for the allele considered<sup>26</sup>. Thus, an allele not involved in the maximal production of any essential biomass component is considered to have pleiotropy 0; an allele that affects the production of exactly one essential biomass component has pleiotropy 1.

Our estimate of pleiotropy reflects the actual contribution of a gene product to biomass formation, based on estimated enzyme and transporter activities in the wildtype. If instead, one is interested in a quantitative measure of essentiality, defined as the number of biomass components affected by a deleterious allele *after* the mutant strain has been allowed to adapt its physiology to the gene deletion, a different algorithm is more appropriate. In this case, one needs to allow the free redistribution of fluxes after the simulated activity reduction of the protein encoded by the gene in question.

**Statistical test for the relationship between pleiotropy at full gene knockout and number of steps.** The full knockout (maximal) pleiotropy sets an upper limit to the possible number of pleiotropy steps at decreasing enzyme activity. The null hypothesis is that the number of steps is uniformly distributed between zero and maximal pleiotropy; *i.e.*, the null hypothesis assumes that apart from the upper limit, there is no systematic relationship between maximal pleiotropy and number of steps. We tested this through a randomization protocol, where we constructed  $10^6$  datasets with the same maximal pleiotropies, but with step numbers drawn from the corresponding uniform distributions. All random datasets for both the *E. coli* and the yeast data had Spearman rank correlation coefficients between pleiotropy at full knockout and number of steps that were lower than the observed correlation coefficients. Thus, the empirical *P*-value was  $<10^{-6}$  for both data sets.

**Currency metabolites.** In additional analyses, we made several cofactors freely available to study how pleiotropy is associated with the generation of currency metabolites. We did this by adding a balanced biochemical reaction that interconverts the activated and inactivated versions of the cofactor and allowing unlimited flux of this reaction in both directions. For example, to simulate free NADPH, we added the following reversible reaction:



A list of currency metabolites was obtained from ref.<sup>43</sup>. Supplementary Table S3 lists the currency metabolites and the corresponding exchange reactions as well as the number of reactions utilizing each currency metabolite.

**Software used.** All simulations were performed in R<sup>47</sup> using *sybil*, a computer library optimized for efficient constraint-based modeling of metabolic networks<sup>48</sup>. We used IBM ILOG CPLEX as the linear solver, connected to *sybil* via the *cplexAPI* R package.

To calculate network modularities, we used the LP&BRIM algorithm (Label Propagation with Bipartite Recursively Induced Modules) implemented in Matlab<sup>49</sup>.

## Data Availability

All input files, R scripts, and raw data used to generate the results and figures can be found on github at <https://github.com/deyazoubi/pleiotropy-git>. An overview over the individual files is given in the Readme file.

## References

1. Stearns, F. W. One Hundred Years of Pleiotropy: A Retrospective. *Genetics* **187**, 355–355 (2011).
2. Wagner, G. P. & Zhang, J. The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat Rev Genet* **12**, 204–213 (2011).
3. Paul, D. A double-edged sword. *Nature* **405**, 515–515 (2000).
4. Williams, G. C. Pleiotropy, Natural-Selection, and the Evolution of Senescence. *Evolution* **11**, 398–411 (1957).
5. Zwaan, B. J. The evolutionary genetics of ageing and longevity. *Heredity* **82**, 589–597 (1999).
6. Moorad, J. A. & Promislow, D. E. L. What can genetic variation tell us about the evolution of senescence? *Proceedings of the Royal Society B-Biological Sciences* **276**, 2271–2278 (2009).
7. Paaby, A. B. & Schmidt, P. S. Dissecting the genetics of longevity in *Drosophila melanogaster*. *Fly (Austin)* **3**, 29–38 (2009).
8. Slatkin, M. Pleiotropy and Parapatric Speciation. *Evolution* **36**, 263–270 (1982).
9. Foster, K. R., Shaulsky, G., Strassmann, J. E., Queller, D. C. & Thompson, C. R. Pleiotropy as a mechanism to stabilize cooperation. *Nature* **431**, 693–696 (2004).
10. Tyler, A. L., Asselbergs, F. W., Williams, S. M. & Moore, J. H. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. *Bioessays* **31**, 220–227 (2009).
11. Wright, S. *Evolution and the genetics of populations: a treatise in four volumes*. (University of Chicago Press, 1968).
12. Barton, N. H. Pleiotropic Models of Quantitative Variation. *Genetics* **124**, 773–782 (1990).
13. Fisher, R. A. *The genetical theory of natural selection*. (Clarendon Press, 1930).
14. Orr, H. A. Adaptation and the cost of complexity. *Evolution* **54**, 13–20 (2000).
15. Waxman, D. & Peck, J. R. Pleiotropy and the preservation of perfection. *Science (New York, N.Y.)* **279**, 1210–1213 (1998).
16. Otto, S. P. Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc Biol Sci* **271**, 705–714 (2004).
17. Hodgkin, J. Seven types of pleiotropy. *Int J Dev Biol* **42**, 501–505 (1998).
18. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
19. Albin, R. L. Antagonistic pleiotropy, mutation accumulation, and human genetic disease. *Genetica* **91**, 279–286 (1993).
20. Brunner, H. G. & van Driel, M. A. From syndrome families to functional genomics. *Nat Rev Genet* **5**, 545–551 (2004).
21. Zhang, J. & Wagner, G. P. On the definition and measurement of pleiotropy. *Trends Genet* **29**, 383–384 (2013).
22. Paaby, A. B. & Rockman, M. V. The many faces of pleiotropy. *Trends Genet* **29**, 66–73 (2013).
23. Dudley, A. M., Janse, D. M., Tanay, A., Shamir, R. & Church, G. M. A global view of pleiotropy and phenotypically derived gene function in yeast. *Molecular Systems Biology* **1**, 2005.0001 (2005).
24. Wagner, G. P. *et al.* Pleiotropic scaling of gene effects and the ‘cost of complexity’. *Nature* **452**, 470–472 (2008).
25. Wang, Z., Liao, B. Y. & Zhang, J. Z. Genomic patterns of pleiotropy and the evolution of complexity. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 18034–18039 (2010).
26. Szappanos, B. *et al.* An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat Genet* **43**, 656–662 (2011).
27. Bajic, D., Moreno-Fenoll, C. & Poyatos, J. F. Rewiring of genetic networks in response to modification of genetic background. *Genome Biol Evol* **6**, 3267–3280 (2014).
28. Wang, Z. & Zhang, J. Z. Abundant Indispensable Redundancies in Cellular Metabolic Networks. *Genome Biology and Evolution* **1**, 23–33 (2009).
29. Watson, M. R. Metabolic Maps for the Apple-II. *Biochem Soc T* **12**, 1093–1094 (1984).
30. Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nature Biotechnology* **28**, 245–248 (2010).
31. Stern, D. L. Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079–1091 (2000).
32. Sivakumaran, S. *et al.* Abundant Pleiotropy in Human Complex Diseases and Traits. *American Journal of Human Genetics* **89**, 607–618 (2011).
33. Orth, J. D. *et al.* A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism–2011. *Mol Syst Biol* **7**, 535 (2011).



34. Aung, H. W., Henry, S. A. & Walker, L. P. Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Ind Biotechnol (New Rochelle N Y)* **9**, 215–228 (2013).
35. Holzhutter, H. G. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European Journal of Biochemistry* **271**, 2905–2922 (2004).
36. Xu, L., Barker, B. & Gu, Z. L. Dynamic epistasis for different alleles of the same gene. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 10420–10425 (2012).
37. Ishii, N. *et al.* Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science (New York, N.Y.)* **316**, 593–597 (2007).
38. Shlomi, T., Berkman, O. & Ruppin, E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci USA* **102**, 7695–7700 (2005).
39. Hadorn, E. *Developmental genetics and lethal factors*. (Methuen; Wiley, 1961).
40. He, X. & Zhang, J. Toward a molecular understanding of pleiotropy. *Genetics* **173**, 1885–1891 (2006).
41. Liu, X. & Murata, T. In 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (IEEE, Milan, Italy, 2009).
42. Wang, Z. & Zhang, J. Z. In search of the biological significance of modular structures in protein networks. *Plos Computational Biology* **3**, 1011–1021 (2007).
43. Fritzscheier, C. J., Hartleb, D., Szappanos, B., Papp, B. & Lercher, M. J. Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS Comput Biol* **13**, e1005494 (2017).
44. Cameron, F. J. & Sinclair, A. H. Mutations in SRY and SOX9: Testis-determining genes. *Human Mutation* **9**, 388–395 (1997).
45. McKusick-Nathans Institute of Genetic Medicine - Johns Hopkins University (Baltimore MD). *Online Mendelian Inheritance in Man (OMIM)*, <http://www.ncbi.nlm.nih.gov/omim/>.
46. Shlomi, T. *et al.* Systematic condition-dependent annotation of metabolic genes. *Genome Research* **17**, 1626–1633 (2007).
47. The R Foundation. The R Project for Statistical Computing, <https://www.r-project.org>.
48. Gelius-Dietrich, G., Desouki, A. A., Fritzscheier, C. J. & Lercher, M. J. Sybil-efficient constraint-based modelling in R. *BMC Syst Biol* **7**, 125 (2013).
49. Flores, C. O., Poisot, T., Valverde, S. & Weitz, J. S. BiMat: a MATLAB package to facilitate the analysis of bipartite networks. *Methods Ecol Evol* **7**, 127–132 (2016).

## Acknowledgements

We thank Balazs Papp and Jonathan Fritzscheier for helpful discussions. This work was supported financially through fellowships by the German Academic Exchange Service (DAAD) to DA and AAD, and through German Research Foundation (DFG) grants to MJL (CRC 680 and CRC 1310).

## Author Contributions

D.A. and M.J.L.: conceptualization, methodology, writing. D.A.: investigation, software, validation. A.A.D.: assistance with metabolic modeling and software implementation. M.J.L.: funding acquisition, supervision.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-35092-1>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018