

RESEARCH ARTICLE

Selection on Network Dynamics Drives Differential Rates of Protein Domain Evolution

Brian K. Mannakee¹, Ryan N. Gutenkunst^{2*}

1 Division of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, Arizona, United States of America, **2** Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona, United States of America

* rgutenk@email.arizona.edu



 OPEN ACCESS

Citation: Mannakee BK, Gutenkunst RN (2016) Selection on Network Dynamics Drives Differential Rates of Protein Domain Evolution. *PLoS Genet* 12 (7): e1006132. doi:10.1371/journal.pgen.1006132

Editor: Jianzhi Zhang, University of Michigan, UNITED STATES

Received: January 5, 2016

Accepted: May 27, 2016

Published: July 5, 2016

Copyright: © 2016 Mannakee, Gutenkunst. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by the National Science Foundation, via Graduate Research Fellowship grant DGE-1143953 to BKM. BKM was also supported by an Achievement Rewards for College Scientists scholarship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

The long-held principle that functionally important proteins evolve slowly has recently been challenged by studies in mice and yeast showing that the severity of a protein knockout only weakly predicts that protein's rate of evolution. However, the relevance of these studies to evolutionary changes within proteins is unknown, because amino acid substitutions, unlike knockouts, often only slightly perturb protein activity. To quantify the phenotypic effect of small biochemical perturbations, we developed an approach to use computational systems biology models to measure the influence of individual reaction rate constants on network dynamics. We show that this dynamical influence is predictive of protein domain evolutionary rate within networks in vertebrates and yeast, even after controlling for expression level and breadth, network topology, and knockout effect. Thus, our results not only demonstrate the importance of protein domain function in determining evolutionary rate, but also the power of systems biology modeling to uncover unanticipated evolutionary forces.

Author Summary

Different proteins evolve at dramatically different rates. To understand this variation, it is necessary to determine which characteristics of proteins are visible to natural selection and how the strength of selection depends on those characteristics. One protein characteristic that is evidently visible to natural selection is expression level; more highly expressed proteins are subject to stronger purifying selection and evolve more slowly. Theory and intuition suggest another such characteristic should be some measure of functional importance, but studies of various measures of functional importance, such as knockout essentiality or knockout growth rate, have shown at best weak correlations with evolutionary rate. Here we develop a novel measure of functional importance, *dynamical influence*, which quantifies the importance of a protein or protein domain to the dynamics of the network of proteins in which it functions. Using 18 biochemically-detailed systems biology models, we compute dynamical influences for each protein domain in each model. We

find that dynamical influence is indeed visible to natural selection and that within networks protein domains with higher dynamical influence evolve more slowly.

Introduction

Over evolutionary time, every protein accumulates amino acid changes at its own characteristic rate, which Zuckerkandl and Pauling likened to the ticking of a molecular clock [1]. Remarkably, this evolutionary rate varies by orders of magnitude among proteins. Understanding the determinants of this variation is a fundamental goal in molecular evolution research [2–5]. Early theoretical work suggested that functional constraints within proteins [1] and the functional importance of each protein to the organism [6, 7] would be key factors in determining evolutionary rates. Yet, empirical studies using knockouts have observed only weak effects. In bacteria [8, 9], yeast [10, 11], and mammals [12] knockout studies conclude that essential proteins evolve only slightly more slowly than non-essential proteins. Moreover, among non-essential genes in yeast, there is little to no correlation between the effect of a protein knockout on growth rate, in a wide range of conditions, and that protein's evolutionary rate [11, 13, 14], particularly when controlling for expression level [15]. This poor correlation between knockout effects and rates of protein evolution has led some researchers to conclude that function-specific selection plays little role in determining evolutionary rates [4, 5]. This conclusion, however, contradicts theoretical expectations, the intuition of most molecular biologists, and the reasoning behind much of comparative genomics [16], motivating our search for an alternative measure of protein function.

We reasoned that knockouts do not mimic evolutionarily relevant mutations, which often have small or moderate effects [17]. In particular, most amino-acid changes do not completely destroy a protein's function, but rather alter its biochemical activity to a greater or lesser extent [18]. The ideal experiment would thus measure the functional effects of many random mutations on many proteins, but such experiments remain challenging [19]. To overcome this experimental limitation, we undertook a computational approach, using biochemically-detailed systems biology models to predict the effects that small perturbations to protein activities will have on the dynamics of the networks in which they function (Fig 1). We ascribed high and low *dynamical influence* to protein domains for which amino acid substitutions were predicted to have respectively large or small effects on network dynamics. We hypothesized that network dynamics is a synthetic phenotype that is likely subject to natural selection. To test this hypothesis, we compared our predictions of dynamical influence in functionally and structurally conserved intracellular signaling and biosynthetic networks with genomic data on protein domain evolutionary rates in both vertebrates and yeast. We found that, within these networks, dynamical influence is as strongly correlated with evolutionary rate as many previously known correlates. Moreover, dynamical influence remains predictive when knockout phenotype, expression, and network topology are controlled for. Dynamical influence thus offers new insight into selective constraint within dynamical protein networks.

Results and Discussion

Dynamical influence quantifies the network consequences of small-effect mutations

Biochemically-detailed systems biology models encapsulate vast amounts of molecular biology knowledge in a form that can be used for *in silico* experimentation [20, 21]. In particular, they

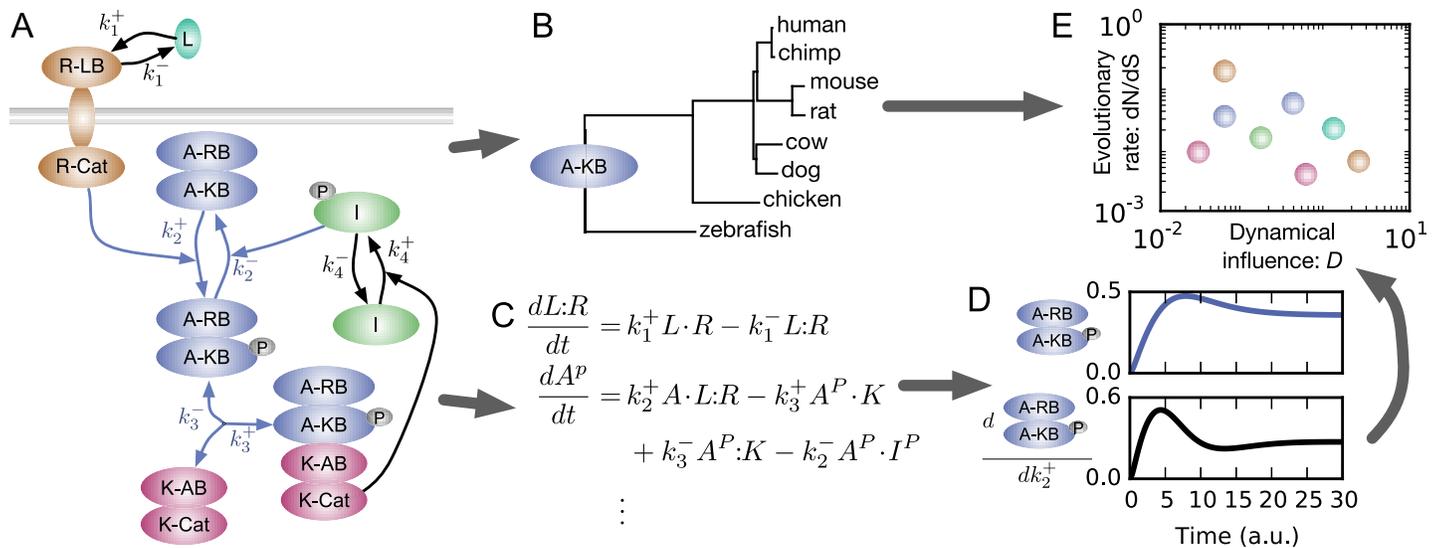


Fig 1. Overview of analysis. A: Illustrative hypothetical signaling network. The dynamical influence of the activator kinase-binding domain (A-KB) is calculated from the influences of the rate constants of the reactions in which it is involved (highlighted in blue): phosphorylation, k_2^+ ; dephosphorylation, k_2^- ; kinase-binding, k_3^+ ; and kinase-unbinding, k_3^- . B: Illustrative phylogenetic analysis of the kinase-binding domain of the activator protein. C: Partial list of ordinary differential equations that model the dynamics of this network. Here all reactions are assumed to be mass-action, but that is not the case in all models analyzed. D: Dynamics of phosphorylated activator protein levels and sensitivity of those dynamics to changes in rate constant k_2^+ , following addition of ligand L. Small increases in k_2^+ hasten the peak of phosphorylated activator protein and increase its steady-state level. The dynamical influence of rate constant k_2^+ is calculated by summing such sensitivities for all molecular species in the network. E: Illustrative plot comparing dynamical influence and evolutionary rate for all domains in the network. A single multi-domain protein can contribute multiple data points.

doi:10.1371/journal.pgen.1006132.g001

enable simulation of the dynamics of molecular species (e.g., proteins, metabolites, modified forms, and complexes) concentrations under a variety of conditions. In these models, protein biochemical activities are quantified by reaction rate constants k [22]. To assess the phenotypic effects of small changes in protein activity caused by mutations, we first calculated the dynamical influence of each reaction rate constant (Eq 1, Materials and Methods). To do so, we calculated how a differential perturbation to that constant would change the concentration time course of each molecular species in the network (Fig 1D), for biologically-relevant stimuli. We then normalized those changes and integrated the squared changes over time. Lastly, we summed over all molecular species in the network. The dynamical influence of a rate constant is thus the total effect that small changes in that rate constant would have on network dynamics.

The dynamical influence of each reaction rate constant quantifies its importance to network dynamics, but there is little data on evolutionary divergence of reaction rate constants to which we can compare. To compare with the abundant genomic data detailing sequence divergence at the domain level, we aggregated the influences of reaction rate constants for all reactions in which a given protein domain is involved. Whenever possible, we analyzed at the domain level, because that is the level at which distinct functions can be assigned to distinct regions of protein sequence [23]. Thus, we defined the dynamical influence D of a domain to be the geometric mean of the dynamical influences of the reaction rate constants for reactions in which it participates (Fig 1A, Eq 2). In general, any mutation in a domain will cause a multidimensional perturbation of all rate constants associated with that domain. Furthermore, different mutations in a domain will differ in the overall magnitude of that perturbation and its relative effect on different parameters [24]. Unfortunately, little systematic data exists about the distributions of such perturbations. The geometric average we took is an approximation to the more

complex averaging that occurs as various mutations arise over evolutionary time. As more systematic data is generated about mutation effects on biochemical activities of different domains [19], the geometric average may be replaced by domain-specific distributions of perturbations.

Dynamical influence within networks is correlated with protein domain evolutionary rate

To test whether dynamical influence is informative about protein evolution, we analyzed dynamic protein network models from BioModels [25], a database which not only collects such models but also annotates them with links to other bioinformatic databases [26, 27]. We considered only models with experimental validation that were formulated in terms of molecular species and reactions, were runnable as ordinary differential equations, and contained at least eight distinct UniProt protein annotations. In total, we studied 12 vertebrate [28–39] and 6 yeast [40–45] signaling and biosynthesis models. We further annotated these models to connect molecular species and reactions with particular protein domains (S1 Dataset). For each model, we calculated dynamical influences for each reaction rate constant using the stimulation conditions considered in the model's original publication (S1 Text).

Using this novel method, we determined protein domain dynamical influence and evolutionary rate for 18 conserved signaling and metabolic networks (Fig 2). We quantified the strength of the relationship between dynamical influence and evolutionary rate using Spearman rank correlations (ρ), and in 10 of 12 vertebrate networks and 6 of 6 yeast networks, we found a negative correlation. This is consistent with the expectation that most sequences and networks evolve primarily under purifying selection [46], in which natural selection is primarily acting to remove deleterious mutations from the population. Mutations in protein domains with high dynamical influence are predicted to have greater phenotypic effect and thus, in general, be more deleterious. So mutations in those domains are more efficiently removed, and those domains evolve more slowly. Demonstrating the strength of our approach, the two exceptional vertebrate models with a positive correlation, visual signal transduction and interleukin 6 (IL-6) signaling, were recently identified as undergoing network-level adaptation in humans using population genetic data [47]. Positively selected molecular changes in rhodopsin associated with changes in absorption wavelength have been shown to affect dose-response behavior in visual signal transduction [48, 49], suggesting that network-level adaptation may compensate for changes in rhodopsin. As part of the innate immune system, IL-6 and its receptor evolve under strong diversifying selection, so downstream proteins may evolve to maintain signal fidelity. Moreover, viruses are known that directly interfere with proteins downstream of IL-6 [50, 51], potentially driving additional adaptation. Dynamical influence is thus predictive not only about purifying selection but also about adaptive selection.

The strength of the correlation between dynamical influence and protein domain evolutionary rate varies considerably among networks (Fig 3A, S1 and S2 Tables). Dynamical influence quantifies the relative effects of perturbations within a single network. Directly comparing dynamical influences among networks would require assumptions about the heterogeneous relative fitness impact of those networks, such as EGF/NGF signaling versus cell cycle progression. Pooling of heterogeneous data can lead to biased estimates of overall correlations [53, 54]. To avoid such bias, we did not pool data across networks but rather applied meta-analysis.

Because selection may act differently on networks with different functions, we considered a random-effects meta-analysis. Thus the sampled networks were assumed to represent a population of networks, among which the correlation between domain evolutionary rate and dynamical influence varies. The meta-analysis seeks to estimate the distribution of those correlations. We applied the random-effects meta-analysis method of Hunter and Schmidt [52],

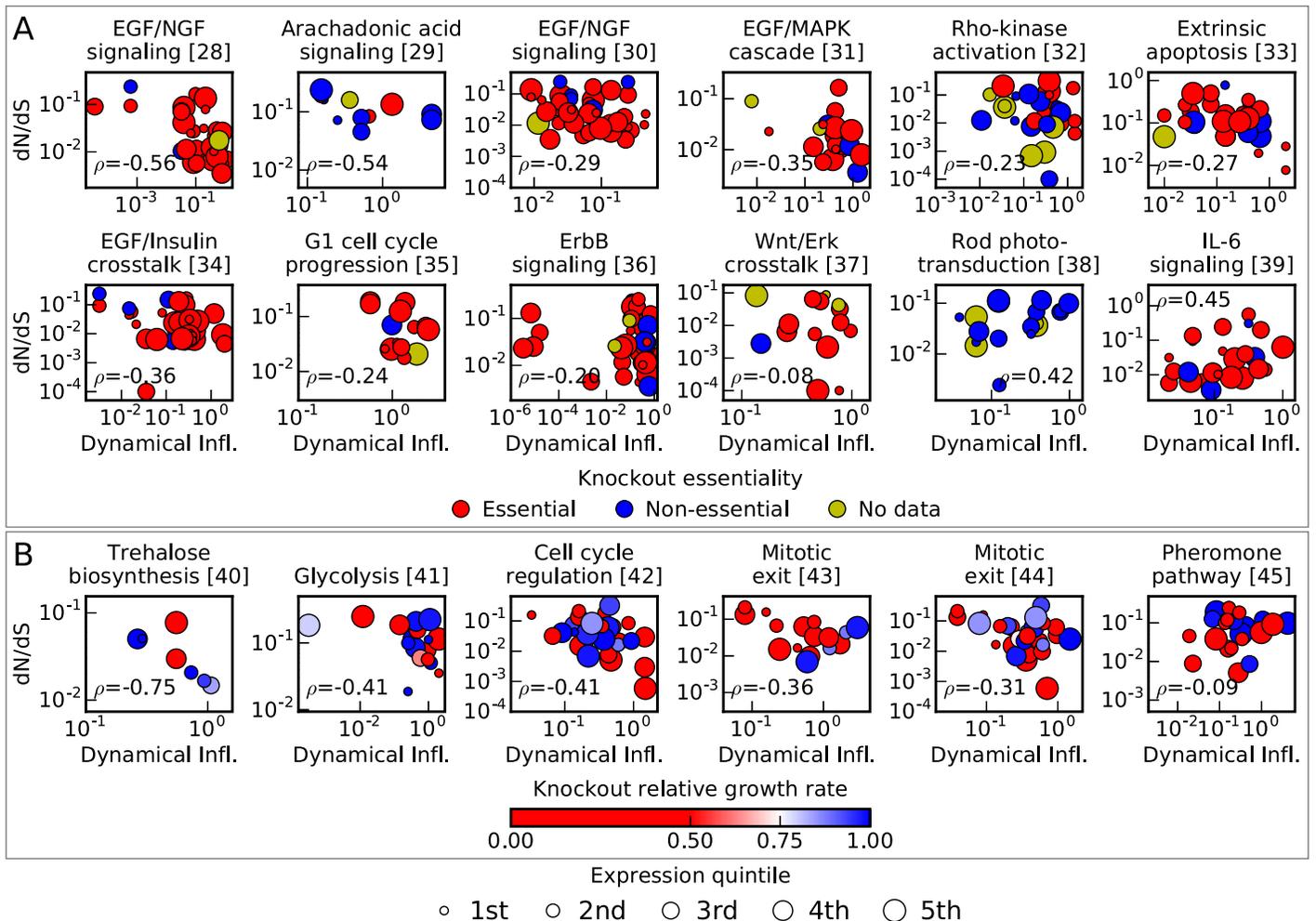


Fig 2. Domain evolutionary rate is correlated with dynamical influence in signaling and biosynthetic networks. Each point represents a protein domain, plotted given its evolutionary rate dN/dS and dynamical influence. Spearman rank correlations ρ between dynamical influence and evolutionary rate are generally negative, indicative of widespread purifying selection on network dynamics. Expression level is represented by marker size and is weakly correlated with evolutionary rate but not significantly correlated with dynamical influence (Table 1). A: Vertebrate networks. Knockout essentiality is represented by color and is not significantly correlated with evolutionary rate or dynamical influence (Table 1). B: Yeast networks. Knockout growth rate is represented by color, with red indicating a more severe phenotype. Knockout growth rate is not significantly correlated with evolutionary rate or dynamical influence (Table 1).

doi:10.1371/journal.pgen.1006132.g002

because simulation studies suggest that it provides an accurate estimate of the mean correlation, particularly when that correlation is modest [55, 56]. The estimated distribution of correlations between domain evolutionary rate and dynamical influence is wide, but the 95% confidence interval for the mean correlation excludes zero (Fig 3A and Table 1). This suggests that negative correlation is more common than positive correlation, consistent with the expectation that purifying selection is more common than adaptive selection [46].

As a complementary approach to evaluating the relationship between domain evolutionary rate and dynamical influence, we also performed a permutation test for dependence between them. In this test, we compared the estimated mean correlation from the real data with a null distribution of mean correlations calculated from scrambled data (Materials and Methods). This test rejected the null hypothesis that domain evolutionary rate and dynamical influence

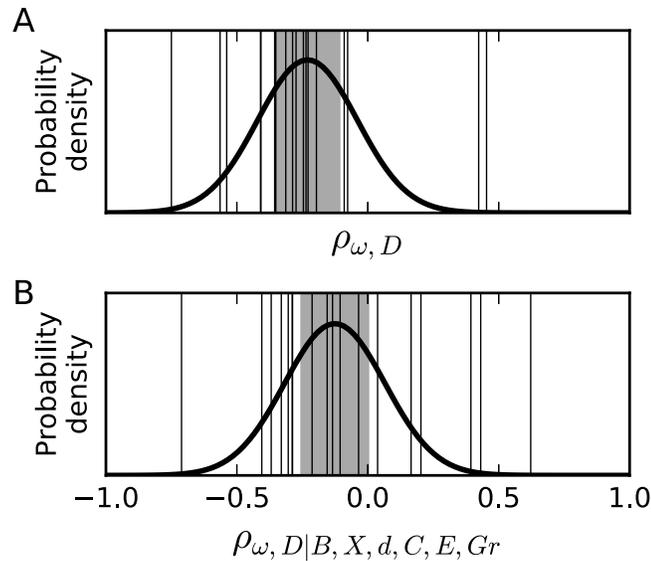


Fig 3. Meta-analysis of correlations between protein domain evolutionary rate and dynamical influence. A: The measured correlation between protein domain evolutionary rate and dynamical influence varies among sampled networks (thin vertical lines). The meta-analysis estimates the distribution of the correlation within the population of networks (thick normal curve), and the shaded region indicates the 95% confidence interval for the mean of that distribution. The mean is significantly less than zero, indicating that evolutionary rate and dynamical influence are negatively correlated in most networks, consistent with predominant purifying selection on network dynamics. B: As in A, but controlling for numerous other factors known to correlated with evolutionary rate, as defined in [Table 1](#).

doi:10.1371/journal.pgen.1006132.g003

Table 1. Overall correlations between domain evolutionary rate, dynamical influence, and other explanatory variables. Spearman rank (ρ) and rank biserial (rb) correlation coefficients for variables evolutionary rate dN/dS (ω), dynamical influence (D), expression breadth (B , for vertebrates only), expression level (X), interaction degree (d), interaction betweenness centrality (C), knockout essentiality (E), and knockout growth rate (Gr , for yeast only). Mean population correlations and their confidence intervals were estimated from all analyzed models using the random-effects meta-analysis approach of Hunter and Schmidt [52]. For a complementary test of the null hypothesis of independence between variables, two-sided p-values were calculated via permutation ([Materials and Methods](#)). Dynamical influence is independently predictive of evolutionary rate, as shown by the negative and statistically significant mean partial correlation after controlling for all other variables.

	estimated population mean correlation (95% C.I.)	permutation p-value
$\rho_{\omega,D}$	-0.23 (-0.35, -0.11)	<0.0001
$\rho_{\omega,B}$	-0.25 (-0.34, -0.15)	0.0011
$\rho_{\omega,X}$	-0.14 (-0.29, -0.00)	0.0191
$\rho_{\omega,d}$	-0.21 (-0.31, -0.11)	0.0004
$\rho_{\omega,C}$	-0.18 (-0.26, -0.10)	0.0024
$rb_{\omega,E}$	-0.12 (-0.28, +0.04)	0.2064
$\rho_{\omega,Gr}$	-0.02 (-0.15, +0.11)	0.8528
$\rho_{D,B}$	+0.17 (+0.03, +0.31)	0.0076
$\rho_{D,X}$	+0.09 (-0.03, +0.21)	0.0838
$\rho_{D,d}$	+0.07 (-0.02, +0.16)	0.1667
$\rho_{D,C}$	+0.06 (-0.04, +0.15)	0.2789
$rb_{D,E}$	+0.08 (-0.13, +0.29)	0.3498
$\rho_{D,Gr}$	+0.11 (-0.10, +0.31)	0.2475
$\rho_{\omega,D B,X,d,C,E,Gr}$	-0.13 (-0.25, +0.00)	0.0125

doi:10.1371/journal.pgen.1006132.t001

are independent within networks, consistent with the confidence interval analysis that excludes zero correlation (Table 1).

Dynamical influence calculation is robust to modeling uncertainties

We measured dynamical influence using hand-built systems biology models; what effect do uncertainties in these models have on our analysis? To be agnostic about what aspects of network dynamics are critical to fitness, in calculating dynamical influence we summed over the integrated sensitivities of all molecular species in the network. It is, however, often evident that the builders of each model had specific molecular species in which they were most interested. If we restricted our dynamical influence calculation to those species (S1 Text), we found very similar correlation with domain evolutionary rate (Fig 4A). Our results are thus not strongly sensitive to which aspects of network function are assumed to be subject to natural selection.

Given a network model, substantial uncertainty can exist about the values of the rate constants k [22], because they are difficult to measure directly and are thus often fit to experimental data on network behavior [57, 58]. To account for this rate constant uncertainty, an ensemble of rate constant sets consistent with the experimental data and the model can be built [59, 60], but this has unfortunately been done for only a small number of models. To assess the importance of rate constant uncertainty to our results, we used an ensemble of 2000 sets of rate

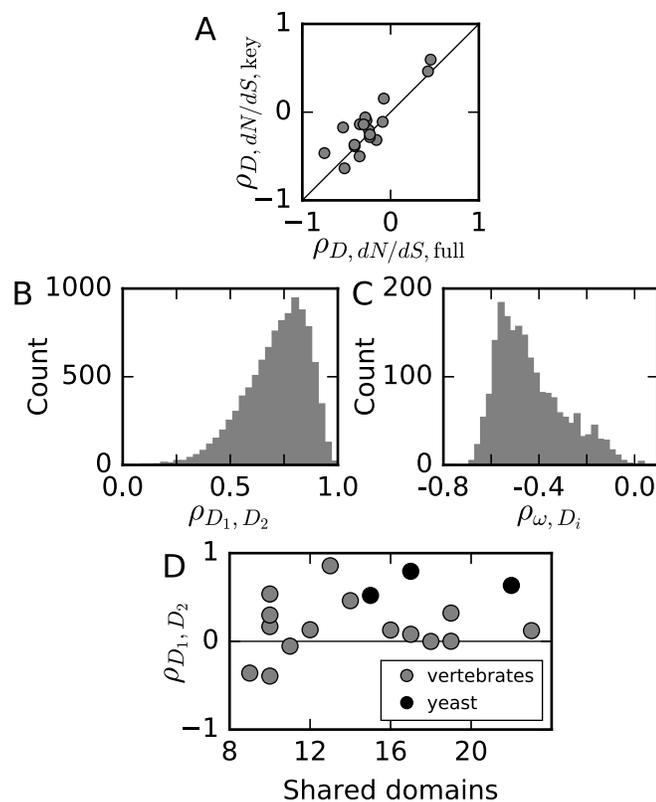


Fig 4. Effects of model uncertainty on dynamical influence. A: The correlation between dynamical influence and evolutionary rate (dN/dS) is similarly strong in all models if dynamical influence is evaluated using all model molecular species (“full”) or only those deemed most important by the authors of the original study (“key”). B: Dynamical influences are strongly correlated between biologically-plausible rate constant sets for a model of EGF/NGF signaling [28]. C: For that model and ensemble of rate constant sets, the correlation between dynamical influence and evolutionary rate varies in magnitude but is consistent in sign. D: Models with overlapping domains produce positively rank-correlated estimates of dynamical influence.

doi:10.1371/journal.pgen.1006132.g004

constants [61] that were previously identified as consistent with experimental data for one of our models of EGF/NGF signaling [28]. This ensemble was built by Markov Chain Monte Carlo sampling of the posterior distribution when fitting the model to data from 14 systems biology experiments. Rate constant values in the resulting ensemble vary dramatically, with many values varying by more than four orders of magnitude, but all sets of rate constants reproduce the experimentally-measured network dynamics. We calculated the dynamical influence of all protein domains in the network using all these sets of rate constants. Comparing 10,000 randomly chosen pairs of sets of dynamical influences to each other, we found that they were highly correlated (Fig 4B), with a median rank correlation of 0.74. Over the ensemble of plausible rate constant sets, the correlation between domain dynamical influence and evolutionary rate varied in magnitude, but 99.8% of rate constant sets yielded a negative correlation (Fig 4C). Together these analyses suggest that, while rate constants themselves vary dramatically over the ensemble for this model, relative dynamical influence varies much less, such that rate constant uncertainty does not affect the sign of the observed correlation, although it may affect its magnitude. We could not build rate constant ensembles for the other models in our analysis without access to the original data used to fit those models, but the universality of the “sloppy” pattern of sensitivities in systems biology models [61, 62] suggests that similar results would be found using rate constant ensembles for the other models in our analysis.

In addition to rate constant uncertainty, different modelers may also make different assumptions when studying the same network regarding forms of interaction, which molecular players to include, or which conditions to consider. We assessed the effect of these assumptions using the models in our data which consider overlapping protein domains. The rank correlation between dynamical influences calculated for the same domains using different models varied considerably and was stronger for pairs of models with larger numbers of overlapping domains (Fig 4D and S3 Table). Weighting correlations from different comparisons as in our meta-analysis, we found a mean correlation of 0.26. For comparison, the correlation between different research groups in measurements of gene expression in log-phase growth of budding yeast is roughly 0.62 [63], while for degree in protein-protein interaction data, the correlation is 0.11 [64]. Thus model uncertainty plays a strong but not dominant role in our analysis, and it is comparable to variables that have previously been found to be informative about evolutionary rate.

We defined dynamical influence in terms of differential perturbations to reaction rate constants (Eq 1), but mutations introduce finite perturbations. Dynamical influence values calculated using finite perturbations of $\pm 25\%$ to rate constants were, however, almost perfectly correlated with values calculated using differential perturbations (S4 Table). Our results thus also apply to mutations of moderate effect.

Dynamical influence predicts evolutionary rates independently from previously known factors

Dynamical influence captures the phenotypic effect within dynamical networks of small perturbations to protein domain activity, but how does it relate to factors previously linked to evolutionary rate? In many cases, previously known factors were discovered and validated using genome-wide analyses. The set of protein sequences for which dynamical influence can be calculated is smaller and potentially biased. First, dynamical influence considers effects on network dynamics after stimulus, so it is not applicable to proteins that do not respond to any stimuli. Second, we calculated dynamical influence from mathematical models, and such models exist for only some systems. Lastly, we calculated dynamical influence at the domain level, so we did not consider the evolution of linker sequences between domains. Because the

previously known factors we consider are defined at the whole-protein level, they can never fully explain evolutionary rate at the domain level. Nevertheless, we used correlation analysis to understand how dynamical influence compares with previously known predictors of evolutionary rate, for the set of networks and protein domains represented in our study.

In multicellular organisms, proteins that are expressed in more cell types (i.e., have higher expression breadth) evolve more slowly [65], and this is true in the vertebrate networks we study (Table 1). The significant positive correlation between dynamical influence and expression breadth (Table 1) suggests that protein domains with key roles in these networks exert their effects across multiple tissues, providing a functional explanation for the observed correlation between expression breadth and evolutionary rate.

Expanding from expression breadth, the strongest known correlate with protein evolutionary rate is expression level. Proteins with greater expression evolve more slowly in both yeast [66] and vertebrates [67], which may reflect the costs of protein mis-folding [15, 68–70] or mis-interaction [71]. In our networks, we found the expected negative correlation between evolutionary rate and expression level (Table 1). That estimated mean correlation is weaker than that between evolutionary rate and dynamical influence (Table 1), although the confidence intervals overlap. Indeed, dynamical influence is not significantly correlated with expression level (Table 1), indicating that dynamical influence reveals previously unanticipated evolutionary pressures beyond the strongest previously known correlate.

A significant advantage of our approach is that it captures how molecular inputs are integrated into functional phenotypic outcomes that may be selected upon. One aspect of network biology that has been previously considered in determining protein evolution is topology. Specifically, proteins with more interaction partners (i.e., greater degree) [72] or more central locations within networks (greater betweenness centrality) [73] evolve more slowly. Consistent with this previous work, we find that domain evolutionary rate has a significant negative correlation with both protein degree and betweenness centrality (Table 1). But, intriguingly, dynamical influence of protein domains is not significantly correlated with degree or betweenness centrality (Table 1) of the corresponding proteins. Why is the influence of topology not captured in our dynamics-based analysis of evolutionary rate? Network topology is a crude measure of function; networks with the same topology can have different dynamics and thus different functions [74]. Thus, our focus on network dynamics rather than topology provides novel insight into protein domain evolution by directly quantifying system output.

Expression level and network topology are also represented in the models themselves, by the total abundance of the molecular species that represent each protein and by the reactions that connect them. The correlations of dynamical influence and evolutionary rate with these model-derived quantities (S5 Table) are similar to those with experimentally-derived expression and topology (Table 1). In fact, the partial correlation between dynamical influence and evolutionary rate controlling for expression and topology is stronger when using model-derived values than when using experimental values. The relationship we find between dynamical influence and evolutionary rate is thus not driven by hidden co-variation between dynamical influence and abundance or topology within the models.

These assessments of dynamical influence relative to known contributors to protein evolution clearly indicate that our approach has uncovered previously unappreciated constraints on protein evolution. Is this new insight sufficient to explain the conundrums raised by knockout experiments? In our data, we found that the correlation between evolutionary rate and knockout measures of function was so weak as to be nonsignificant (Table 1), consistent with prior work [11, 12]. Strikingly, across the eleven vertebrate networks that include both essential and non-essential proteins and the six yeast networks (for which knockout growth rate data is available), we find no statistical correlation between dynamical influence and essentiality or

knockout growth rate (Table 1). Thus, the highly significant correlation between dynamical influence and evolutionary rate (Fig 2, Table 1) provides a new perspective on the influence of protein function on evolutionary rate.

But, evolutionary rates are complex and likely integrate selection on multiple processes [2–5]. To assess the power of our approach in comparison with alternative integrative analyses, we used partial correlation analysis [12]. Across all our networks, we find that when expression, network topology, and knockout effect are controlled for, the mean correlation between protein domain evolutionary rate and dynamical influence remains statistically significant (Fig 3 and Table 1). Because the predictive power of dynamical influence cannot be explained by other factors, it provides novel and previously inaccessible insight into evolutionary rates within protein networks.

The existence of overlapping protein domains might inflate statistical significance in our analyses across models. To account for this, for all domains that appeared in more than one model, we randomly kept each domain in one of the models and deleted it from the others. We did this randomization one thousand times and repeated our correlation analysis each time, obtaining distributions of mean correlations and permutation p-values. There was a tail of large p-values for the fully controlled correlation between dynamical influence and evolutionary rate, corresponding to randomizations in which many domains happened to be retained in the few models with a positive correlation (S1 Fig). The median results we found were, however, similar to our analysis using all the data (S6 Table), suggesting that overlapping domains do not substantially affect our statistics.

Dynamical systems biology models offer great promise for developing and testing evolutionary hypotheses [21, 75]. Previous topological and flux-balance analysis of networks has offered insight into protein evolution [76–78], but dynamical models contribute substantial biological detail not previously captured by these approaches. We have shown that incorporating that detail can, for domains within dynamical networks, explain the previous lack of correlation between protein function and evolutionary rate. Dynamical models have previously been used to predict the phenotypic effects of mutations [79] and to assess the correlation between network sensitivity and protein evolution in phototransduction [80] and in pyrimidine biosynthesis [81]. Here we consider many networks to reveal a previously unexplored and general link between dynamical influence and protein domain evolutionary rate within networks. Given the rapid pace of progress in systems biology modeling [82], the anticipated advances in model scope and validation will provide even more robust data sets to uncover previously unanticipated factors influencing evolutionary processes.

Materials and Methods

Dynamical influence

We defined the dynamical influence κ_i of reaction rate constant k_i by

$$\kappa_i^2 = \sum_{\text{stimulation conditions } c} \sum_{\text{molecular species } y} \frac{1}{T_c} \int_0^{T_c} \left(\frac{dy_c(t)}{dk_i} \frac{k_i}{y_{\max}} \right)^2 \Big|_{k=k^*} dt. \quad (1)$$

Here $y_c(t)$ is the time course of molecular species y in condition c , evaluated using the rate constant values k^* from the original publication. The derivative $dy_c(t)/dk_i$ of the time course with respect to rate constant k_i measures how sensitive that molecular species or metabolite is to changes in that rate constant. To make relative comparisons, we normalized these sensitivities by the value k_i of the rate constant and the maximum y_{\max} of molecular species y over all stimulation conditions. We normalized by y_{\max} rather than using a control coefficient $\frac{dy(t)}{dk} \frac{k}{y(t)} = \frac{d \ln y(t)}{d \ln k}$ [83]

because many molecular species in signaling models begin with zero concentration, so the control coefficient would be undefined. We found the total effect of changes in k_i by squaring these normalized sensitivities, integrating over the time course of each stimulation condition, and summing over all molecular species and stimulation conditions.

We defined the dynamical influence D_d of protein domain d to be the geometric mean of the influences κ of the N_d reaction rate constants for reactions in which that domain participates:

$$D_d = \left(\prod_{r=1}^{N_d} \kappa_r \right)^{\frac{1}{N_d}}. \quad (2)$$

We took a geometric mean because rate constant sensitivities range over orders of magnitude [61].

Stochastic noise plays an important role in many cellular networks [84, 85]. In those cases, networks are not well-modeled by ordinary differential equations (ODEs), although parameter sensitivities can be defined and calculated [86]. To minimize the complications introduced by stochasticity, we focused on models of signaling and biosynthesis in which concentrations of molecular species were sufficiently large to justify a continuous approximation to probabilistic biochemical reaction rates [57]. Moreover, because sources and levels of gene expression noise are known to vary according to initial conditions [87], we restricted our analysis to models which were fit to experimental data arising from multiple initial conditions and measuring multiple reporters.

We downloaded systems biology models in Systems Biology Markup Language (SBML) format [88] from the Feb. 8, 2012 release of BioModels [25]. We calculated dynamical influence for all protein-related biological parameters in each model, using SloppyCell [89] and simulating under the conditions considered in each model's original paper (S1 Text). These parameters represent a variety of biological phenomena, such as binding and catalytic constants and rates of diffusion and production. We considered only those parameters representing rates of biochemical reactions that depend on protein structure, because we expected constraint on those reactions to have the strongest effect on evolutionary rates. Given the dynamical influences κ for each reaction constant, we reviewed the literature to determine the protein domain or domains at which the reaction occurs, and we assigned those influences to that domain or domains (S1 Dataset).

Evolutionary rates

UniProt protein ID's were acquired from the BioModels annotation in the SBML file for each model and converted to NCBI Protein IDs for vertebrates or open reading frame (ORF) numbers for yeast. Some models specified more than one Uniprot ID for a single protein, in cases where there is more than one transcript identified and both appear to perform the same function (for example, MEK1 and MEK2). Where more than one Uniprot ID was specified, we reviewed the model publication and the protein network literature to select a single transcript. In the case of metabolic flux models that track metabolites rather than proteins, we used the names of the enzymes involved in the reactions to find the appropriate protein identifier.

Vertebrate homologous protein alignments were downloaded from the NCBI Homologene database [90], and for each protein in the alignment, nucleotide sequence was downloaded from NCBI Entrez [90]. These nucleotide sequences were then used as a template to back-translate the Homologene protein alignments to nucleotide alignments. Yeast gene information for the 7 species in the tree in S2 Fig was downloaded from the Saccharomyces Genome

Database [91] on Nov. 19, 2012. These gene sequences were translated to protein amino acid sequences using Biopython [92], aligned using ClustalW [93], and then back-translated to aligned nucleotide sequence using the gene sequence as a template.

Protein domain annotation was done manually using literature review, based on information for the human protein in vertebrate models or the *Sa. cerevisiae* protein in yeast. Evolutionary rates were calculated using codeML from PAML Version 4.4b [94], with one dN/dS ratio per tree (model 0), the F3x4 codon substitution model, and a rooted tree, as in [95]. The Mgene = 3 setting of codeml was used to estimate a single dN/dS ratio per annotated protein domain. We required a minimum of 4 homologs to include a gene in the analysis, and for each gene any species with more than one homologue was excluded. Because instability is a concern when estimating multiple dN/dS ratios for a single protein sequence, we iterated each codeml run until we acquired three models for which the log-likelihood was within 0.01 of the lowest log-likelihood obtained and then used the model with the lowest log-likelihood.

Gene expression and specificity

Vertebrate gene expression and tissue specificity data was compiled from the mouse GNF1M dataset [96], downloaded from <http://bioGPS.org/downloads>. The data consist of microarray probes for a number of tissue types, with each probe's name including the corresponding gene name, which we mapped to Ensembl gene IDs using Ensembl BioMart [97]. We restricted our analysis to normal adult tissues as in Fig S2 of [95]. To calculate the expression level corresponding to each microarray probe, we took the arithmetic average over replicates of the same tissue and then took the geometric average over tissues. To calculate the expression level of each gene, we then took the arithmetic average of the probe expression levels corresponding to that gene.

Yeast expression data [98] was obtained from http://younglab.wi.mit.edu/pub/data/orf_transcriptome.txt and used without modification.

Protein abundance within each model was calculated as the sum of initial conditions for all molecular species corresponding to a given protein, including modified forms and complexes. None of the models we considered included transcription or translation, so total levels of all proteins were constant throughout the simulations.

Gene essentiality and dispensibility

We downloaded mouse knockout phenotype data from the Mouse Genome Informatics database [99] at <http://www.informatics.jax.org/phenotypes.shtml> on July 11, 2011. We assembled phenotype information for homozygous knockouts and coded a gene as essential if it resulted in one of the following phenotypes: abnormal reproductive system physiology, prenatal lethality, perinatal lethality, postnatal lethality, premature death, abnormal reproductive system morphology, lethality at weaning, preweaning lethality, partial lethality, and all sub-phenotypes of these phenotypes. If homozygous knockout of a gene did not cause one or more of these phenotypes we coded it as non-essential. To validate our parsing of this data, we compared against the results of [12].

Data for yeast knockout growth rate on YPD media were obtained from the file `Regression_Tc1_hom.txt` downloaded from the Stanford YDPM database http://www-deletion.stanford.edu/YDPM/YDPM_index.html on March 13, 2013.

Network degree and centrality

We downloaded protein-protein interaction data for both humans and yeast from the Interologous Interaction Database [100] on April 20, 2012. These data take the form of a list of

interactions between two proteins, and the dataset from which the interaction was curated. Because we were interested in experimentally verified interactions we restricted our analysis to the HPRD, BIND, IntAct, and INNATEDB datasets for humans and the Krogan_Core, Yu_GoldStd, YeastHigh, YeastLow, and BIND datasets for yeast. We used the python package NetworkX [101] to load these lists of interactions and compute each protein's degree and its betweenness centrality, which is the fraction of all of the shortest paths between protein pairs in the network that pass through that protein. Model-derived network degree and centrality were calculated from a graph with nodes for each domain in the model and edges between any pairs of domains that participate in a reaction together.

Statistical analyses

Dynamical influence and evolutionary rate are defined at the domain level, but all other factors in Table 1 are defined at the protein level, so in our statistical analyses these other factors were assumed to be equal for all domains within a given protein. We used partial correlation analysis to assess the degree to which these factors account for the observed relationship between dynamical influence and domain evolutionary rate. To do so, within each model we fit linear models for dynamical influence and evolutionary rate as a function of all the other factors and then computed the product-moment correlation between the residuals from these models. Only domains for which all variables were measured were included in these partial correlations.

To analyze correlations and partial correlations across models (Table 1, Fig 3, and S5 Table), we applied the random-effects meta-analysis approach of Hunter and Schmidt [52]. In this approach, the mean correlation ρ_0 in the population is estimated by the average of the observed correlations r of the sampled models, weighted by the sample size n of domains with relevant data in each model. So the estimated mean correlation is

$$\hat{\rho}_0 = \frac{\sum n_i r_i}{\sum n_i}, \tag{3}$$

where sums here and below are over models (Eq. 3.1 in [52]). The variance across samples from the population σ_r^2 is the sum of the variance in population correlations σ_ρ^2 and the variance due to sampling error σ_e^2 :

$$\sigma_r^2 = \sigma_\rho^2 + \sigma_e^2. \tag{4}$$

The variance across samples can be estimated as [52]

$$\hat{\sigma}_r^2 = \frac{\sum n_i (r_i - \hat{\rho}_0)^2}{\sum n_i}. \tag{5}$$

The sampling variance can be estimated as (Eq. 3.5 in [52])

$$\hat{\sigma}_e^2 = \frac{\sum \frac{n_i(1-r_i^2)^2}{n_i-1}}{\sum n_i}. \tag{6}$$

The standard deviation σ_ρ of the population correlations sets the width of the distribution curves in Fig 3, and it can be estimated by solving Eq 4 for σ_ρ and substituting in the estimates $\hat{\sigma}_e$ and $\hat{\sigma}_r$. The standard error $SE(\hat{\rho}_0)$ of the estimated mean correlation $\hat{\rho}_0$ depends on the number of population samples K , which is 18 here (Eq. 5.1 in [52]):

$$SE(\hat{\rho}_0) = \hat{\sigma}_r / \sqrt{K}. \tag{7}$$

The 95% confidence intervals reported in [Table 1](#) and [S5 Table](#) are thus $\hat{\rho}_0 \pm 1.96 SE(\hat{\rho}_0)$. The most popular alternative approach for random-effects meta-analysis of correlations, developed by Hedges and colleagues [102], estimates the population mean correlation ρ_0 and the standard error of that estimate $SE(\hat{\rho}_0)$ using more complex weightings based on Fisher's *r*-to-*Z* transform. We adopted the Hunter and Schmidt approach because simulation studies suggest that it produces more accurate estimates of the population mean correlation and more accurate confidence intervals when variation in the population is large [56].

In our permutation tests, our null model was that dynamical influence or evolutionary rate was uncorrelated with other protein domain properties ([Table 1](#)). To generate null distributions of correlations, we permuted dynamical influences, evolutionary rates, and all other factors within each model. Because domains share reactions, their influences are not independent, and thus we could not simply permute them to simulate our null model. Instead, we permuted the influences of reaction parameters, which are the most basic unit of our analysis, and we then recalculated the influence for each domain based on the new sets of parameter influences. Evolutionary rates are defined at the domain level, so we simply permuted them within each model. The other factors are defined at the protein level, and we permuted them at the protein level, so that domains within the same protein would still always, for example, have the same expression level. To generate the null distribution for the partial correlation between domain evolutionary rate and dynamical influence, controlling for other variables, we permuted the residuals from the linear models used to calculate the partial correlation [103, 104]. This approach disrupts any relationship between domain evolutionary rate and dynamical influence while preserving all other relationships between network variables. We permuted variables or residuals separately within each model and then used [Eq 3](#) to calculate mean correlations across models for each permutation. After carrying out 10,000 permutations, the two-sided p-values we report ([Table 1](#)) are the quantiles of the real data absolute mean correlations among the permuted absolute mean correlations. In all cases, the permutation test results are compatible with the 95% confidence intervals ([Table 1](#)); smaller p-values correspond to confidence intervals that more strongly exclude zero. Note that permutation tests based on correlation or partial correlation coefficients are strictly tests of the null hypothesis that the two variables are independent [105]. Thus our permutation test cannot reject the possibility that the considered variables are dependent but uncorrelated.

Supporting Information

S1 Text. Details of analysis for each model.

(PDF)

S1 Table. Correlations in vertebrate models. Spearman rank (ρ) and rank biserial (rb) correlation coefficients for variables evolutionary rate dN/dS (ω), dynamical influence (D), expression breadth (B), expression level (X), interaction degree (d), interaction betweenness centrality (C), and knock-out essentiality (E). Domains with missing values for any correlate were dropped prior to calculating correlations, and N represents the number of domains used in the analysis. For correlations, p-values were calculated via permutation of the data as described in the main text. For partial correlations, p-values were calculated by permutation of the residuals from the linear models [103, 104].

(PDF)

S2 Table. Correlations in yeast models. As in [S1 Table](#), but for yeast models, which have knock-out growth rate (Gr) data.

(PDF)

S3 Table. Between-model correlations between protein domain dynamical influences. For each pair of models with at least four overlapping domains, shown is the Spearman rank correlation and number of overlapping domains.

(PDF)

S4 Table. Dynamical influence calculated with finite versus differential perturbations.

Dynamical influence was recalculated using finite perturbations $\frac{\Delta y}{\Delta k} = \frac{y(k^+) - y(k^-)}{k^+ - k^-}$ instead of differential perturbations $\frac{dy}{dk}$ (Eq 1), with k^+ and k^- being $\pm 25\%$ perturbations of each parameter. Tabulated are the rank correlations between domain influences calculated using finite and differential perturbations. The almost perfect correlations suggest that our dynamical influence analysis also applies to mutations of moderate effect.

(PDF)

S5 Table. Overall correlations calculated with model-derived expression and topology data.

Meta-analysis confidence intervals and permutation p-values as in Table 1. Spearman rank (ρ) correlation coefficients for variables evolutionary rate dN/dS (ω), dynamical influence (D), model-derived expression (MX), model-derived interaction degree (Md), model-derived interaction centrality (MC), expression breadth (B), knock-out essentiality (E), and knock-out growth rate (Gr).

(PDF)

S6 Table. Overall correlations with overlapping domains removed. Meta-analysis mean correlations and permutation p-values as in Table 1, but without overlapping domains between models. Values are the 50th (5th, 95th) quantiles of correlations and p-values calculated from 1000 runs in which each domain that appears in multiple models was considered for only a single randomly chosen model.

(PDF)

S1 Fig. Distributions of overall correlation and p-value when overlapping domains are removed.

Shown are the full distributions that are summarized in the last row of S6 Table. A: The partial correlation between domain dynamical influence and evolutionary rate is negative for all randomizations. B: The distribution of p-values is strongly concentrated below 0.05. The tail of larger p-values is generated by randomizations that concentrate domains in the few models with a positive correlation.

(PDF)

S2 Fig. Phylogenetic trees for species used in this study. A: Vertebrates. B: Yeasts. In both, branch lengths represent amino acid divergence.

(PDF)

S1 Dataset. Complete model and protein domain annotation, including covariate data for each domain. Each model corresponds to two sheets. The first sheet contains the reaction parameters, their dynamical influences, and the reactions they correspond to. The second sheet contains the protein and domain data, including assignment of reactions to domains and corresponding references (as PubMed IDs).

(XLSX)

Acknowledgments

We thank Tricia Serio for helpful comments on the manuscript, and Edward Bedrick for consultation regarding meta-analysis and permutation testing.

Author Contributions

Conceived and designed the experiments: BKM RNG. Performed the experiments: BKM RNG. Analyzed the data: BKM RNG. Contributed reagents/materials/analysis tools: BKM RNG. Wrote the paper: BKM RNG.

References

1. Zuckerkandl E, Pauling L. Evolutionary Divergence and Convergence in Proteins. *Evol Genes Proteins*. 1965;p. 97–165.
2. Pál C, Papp B, Lercher MJ. An integrated view of protein evolution. *Nat Rev Genet*. 2006 may; 7(5):337–48. doi: [10.1038/nrg1838](https://doi.org/10.1038/nrg1838) PMID: [16619049](https://pubmed.ncbi.nlm.nih.gov/16619049/)
3. Koonin EV, Wolf YI. Evolutionary systems biology: links between gene evolution and function. *Curr Opin Biotech*. 2006 oct; 17(5):481–7. doi: [10.1016/j.copbio.2006.08.003](https://doi.org/10.1016/j.copbio.2006.08.003) PMID: [16962765](https://pubmed.ncbi.nlm.nih.gov/16962765/)
4. Alvarez-Ponce D. Why proteins evolve at different rates: the determinants of proteins' rates of evolution. In: Fares MA, editor. *Natural Selection: Methods and Applications*. CRC Press; 2014. p. 126–178.
5. Zhang J, Yang JR. Determinants of the rate of protein sequence evolution. *Nat Rev Genet*. 2015; 16(7):409–420. doi: [10.1038/nrg3950](https://doi.org/10.1038/nrg3950) PMID: [26055156](https://pubmed.ncbi.nlm.nih.gov/26055156/)
6. Kimura M, Ota T. On some principles governing molecular evolution. *Proc Natl Acad Sci U S A*. 1974 jul; 71(7):2848–52. doi: [10.1073/pnas.71.7.2848](https://doi.org/10.1073/pnas.71.7.2848) PMID: [4527913](https://pubmed.ncbi.nlm.nih.gov/4527913/)
7. Wilson AC, Carlson SS, White TJ. Biochemical evolution. *Annu Rev Biochem*. 1977 jan; 46:573–639. doi: [10.1146/annurev.bi.46.070177.003041](https://doi.org/10.1146/annurev.bi.46.070177.003041) PMID: [409339](https://pubmed.ncbi.nlm.nih.gov/409339/)
8. Jordan IK, Rogozin IB, Wolf YI, Koonin EV. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res*. 2002 may; 12(6):962–968. doi: [10.1101/gr.87702](https://doi.org/10.1101/gr.87702) PMID: [12045149](https://pubmed.ncbi.nlm.nih.gov/12045149/)
9. Rocha EPC, Danchin A. An analysis of determinants of amino acids substitution rates in bacterial proteins. *Mol Biol Evol*. 2004 jan; 21(1):108–16. doi: [10.1093/molbev/msh004](https://doi.org/10.1093/molbev/msh004) PMID: [14595100](https://pubmed.ncbi.nlm.nih.gov/14595100/)
10. Hurst LD, Smith NG. Do essential genes evolve slowly? *Curr Biol*. 1999 jul; 9(14):747–50. doi: [10.1016/S0960-9822\(99\)80334-0](https://doi.org/10.1016/S0960-9822(99)80334-0) PMID: [10421576](https://pubmed.ncbi.nlm.nih.gov/10421576/)
11. Wang Z, Zhang J. Why is the correlation between gene importance and gene evolutionary rate so weak? *PLoS Genet*. 2009 jan; 5(1):e1000329. doi: [10.1371/journal.pgen.1000329](https://doi.org/10.1371/journal.pgen.1000329) PMID: [19132081](https://pubmed.ncbi.nlm.nih.gov/19132081/)
12. Liao BY, Scott NM, Zhang J. Impacts of gene essentiality, expression pattern, and gene compactness on the evolutionary rate of mammalian proteins. *Mol Biol Evol*. 2006 nov; 23(11):2072–80. doi: [10.1093/molbev/msl076](https://doi.org/10.1093/molbev/msl076) PMID: [16887903](https://pubmed.ncbi.nlm.nih.gov/16887903/)
13. Hirsh AE, Fraser HB. Protein dispensability and rate of evolution. *Nature*. 2001 jun; 411(6841):1046–9. doi: [10.1038/35082561](https://doi.org/10.1038/35082561) PMID: [11429604](https://pubmed.ncbi.nlm.nih.gov/11429604/)
14. Pál C, Papp B, Hurst LD. Rate of evolution and gene dispensability. *Nature*. 2003; 421(January):496–497. PMID: [12556881](https://pubmed.ncbi.nlm.nih.gov/12556881/)
15. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A*. 2005 oct; 102(40):14338–43. doi: [10.1073/pnas.0504070102](https://doi.org/10.1073/pnas.0504070102) PMID: [16176987](https://pubmed.ncbi.nlm.nih.gov/16176987/)
16. Hardison RC. Comparative genomics. *PLoS Biol*. 2003 nov; 1(2):e58. doi: [10.1371/journal.pbio.0000058](https://doi.org/10.1371/journal.pbio.0000058) PMID: [14624258](https://pubmed.ncbi.nlm.nih.gov/14624258/)
17. Soskine M, Tawfik DS. Mutational effects and the evolution of new protein functions. *Nat Rev Genet*. 2010 aug; 11(8):572–82. doi: [10.1038/nrg2808](https://doi.org/10.1038/nrg2808) PMID: [20634811](https://pubmed.ncbi.nlm.nih.gov/20634811/)
18. Guo HH, Choe J, Loeb LA. Protein tolerance to random amino acid change. *Proc Natl Acad Sci U S A*. 2004 jun; 101(25):9205–10. doi: [10.1073/pnas.0403255101](https://doi.org/10.1073/pnas.0403255101) PMID: [15197260](https://pubmed.ncbi.nlm.nih.gov/15197260/)
19. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence-function relationships. *Nat Methods*. 2010 sep; 7(9):741–6. doi: [10.1038/nmeth.1492](https://doi.org/10.1038/nmeth.1492) PMID: [20711194](https://pubmed.ncbi.nlm.nih.gov/20711194/)
20. Di Ventura B, Lemerle C, Michalodimitrakis K, Serrano L. From in vivo to in silico biology and back. *Nature*. 2006 oct; 443(7111):527–33. doi: [10.1038/nature05127](https://doi.org/10.1038/nature05127) PMID: [17024084](https://pubmed.ncbi.nlm.nih.gov/17024084/)
21. Loewe L. A framework for evolutionary systems biology. *BMC Syst Biol*. 2009 jan; 3:27. doi: [10.1186/1752-0509-3-27](https://doi.org/10.1186/1752-0509-3-27) PMID: [19239699](https://pubmed.ncbi.nlm.nih.gov/19239699/)
22. Gunawardena J. Models in systems biology: the parameter problem and the meaning of robustness. In: Lodhi HM, Muggleton SH, editors. *Elements of Computational Systems Biology*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010. p. 19–47.

23. Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA. Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol.* 2004; 14(2):208–216. doi: [10.1016/j.sbi.2004.03.011](https://doi.org/10.1016/j.sbi.2004.03.011) PMID: [15093836](https://pubmed.ncbi.nlm.nih.gov/15093836/)
24. Starita LM, Young DL, Islam M, Kitzman JO, Gullingsrud J, Hause RJ, et al. Massively parallel functional analysis of BRCA1 RING domain variants. *Genetics.* 2015; 200(2):413–422. doi: [10.1534/genetics.115.175802](https://doi.org/10.1534/genetics.115.175802) PMID: [25823446](https://pubmed.ncbi.nlm.nih.gov/25823446/)
25. Le Novère N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, et al. BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* 2006 jan; 34(Database issue):D689–91. doi: [10.1093/nar/gkj092](https://doi.org/10.1093/nar/gkj092) PMID: [16381960](https://pubmed.ncbi.nlm.nih.gov/16381960/)
26. Li C, Donizelli M, Rodriguez N, Dharuri H, Endler L, Chelliah V, et al. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst Biol.* 2010 jan; 4(1):92. doi: [10.1186/1752-0509-4-92](https://doi.org/10.1186/1752-0509-4-92) PMID: [20587024](https://pubmed.ncbi.nlm.nih.gov/20587024/)
27. Chelliah V, Juty N, Ajmera I, Ali R, Dumousseau M, Glont M, et al. BioModels: ten-year anniversary. *Nucleic Acids Res.* 2014; 43(November 2014):D542–D548. doi: [10.1093/nar/gku1181](https://doi.org/10.1093/nar/gku1181) PMID: [25414348](https://pubmed.ncbi.nlm.nih.gov/25414348/)
28. Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, et al. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Phys Biol.* 2004 dec; 1(3–4):184–95. doi: [10.1088/1478-3967/1/3/006](https://doi.org/10.1088/1478-3967/1/3/006) PMID: [16204838](https://pubmed.ncbi.nlm.nih.gov/16204838/)
29. Yang K, Ma W, Liang H, Ouyang Q, Tang C, Lai L. Dynamic simulations on the arachidonic acid metabolic network. *PLoS Comput Biol.* 2007 mar; 3(3):e55. doi: [10.1371/journal.pcbi.0030055](https://doi.org/10.1371/journal.pcbi.0030055) PMID: [17381237](https://pubmed.ncbi.nlm.nih.gov/17381237/)
30. Sasagawa S, Ozaki YI, Fujita K, Kuroda S. Prediction and validation of the distinct dynamics of transient and sustained ERK activation. *Nat Cell Biol.* 2005 apr; 7(4):365–73. doi: [10.1038/ncb1233](https://doi.org/10.1038/ncb1233) PMID: [15793571](https://pubmed.ncbi.nlm.nih.gov/15793571/)
31. Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol.* 2002 apr; 20(4):370–5. doi: [10.1038/nbt0402-370](https://doi.org/10.1038/nbt0402-370) PMID: [11923843](https://pubmed.ncbi.nlm.nih.gov/11923843/)
32. Maeda A, Ozaki Yi, Sivakumaran S, Akiyama T, Urakubo H, Usami A, et al. Ca²⁺-independent phospholipase A₂-dependent sustained Rho-kinase activation exhibits all-or-none response. *Genes Cells.* 2006 sep; 11(9):1071–83. doi: [10.1111/j.1365-2443.2006.01001.x](https://doi.org/10.1111/j.1365-2443.2006.01001.x) PMID: [16923126](https://pubmed.ncbi.nlm.nih.gov/16923126/)
33. Albeck JG, Burke JM, Aldridge BB, Zhang M, Lauffenburger DA, Sorger PK. Quantitative analysis of pathways controlling extrinsic apoptosis in single cells. *Mol Cell.* 2008 apr; 30(1):11–25. doi: [10.1016/j.molcel.2008.02.012](https://doi.org/10.1016/j.molcel.2008.02.012) PMID: [18406323](https://pubmed.ncbi.nlm.nih.gov/18406323/)
34. Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, et al. Systems-level interactions between insulin-EGF networks amplify mitogenic signaling. *Mol Syst Biol.* 2009 jan; 5(256):256. doi: [10.1038/msb.2009.19](https://doi.org/10.1038/msb.2009.19) PMID: [19357636](https://pubmed.ncbi.nlm.nih.gov/19357636/)
35. Haberichter T, Mädge B, Christopher Ra, Yoshioka N, Dhiman A, Miller R, et al. A systems biology dynamical model of mammalian G1 cell cycle progression. *Mol Syst Biol.* 2007 jan; 3(84):84. doi: [10.1038/msb4100126](https://doi.org/10.1038/msb4100126) PMID: [17299420](https://pubmed.ncbi.nlm.nih.gov/17299420/)
36. Birtwistle MR, Hatakeyama M, Yumoto N, Ogunnaike BA, Hoek JB, Kholodenko BN. Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. *Mol Syst Biol.* 2007 jan; 3:144. doi: [10.1038/msb4100188](https://doi.org/10.1038/msb4100188) PMID: [18004277](https://pubmed.ncbi.nlm.nih.gov/18004277/)
37. Kim D, Rath O, Kolch W, Cho KH. A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK pathways. *Oncogene.* 2007 jul; 26(31):4571–9. doi: [10.1038/sj.onc.1210230](https://doi.org/10.1038/sj.onc.1210230) PMID: [17237813](https://pubmed.ncbi.nlm.nih.gov/17237813/)
38. Dell'Orco D, Schmidt H, Mariani S, Fanelli F. Network-level analysis of light adaptation in rod cells under normal and altered conditions. *Mol Biosyst.* 2009 oct; 5(10):1232–46. doi: [10.1039/b908123b](https://doi.org/10.1039/b908123b) PMID: [19756313](https://pubmed.ncbi.nlm.nih.gov/19756313/)
39. Singh A, Jayaraman A, Hahn J. Modeling regulatory mechanisms in IL-6 signal transduction in hepatocytes. *Biotechnol Progr.* 2006 dec; 95(979):850–862.
40. Smallbone K, Malys N, Messiha HL, Wishart JA, Simeonidis E. Building a kinetic model of trehalose biosynthesis in *Saccharomyces cerevisiae*. *Methods Enzym.* 2011 jan; 500(null):355–70. doi: [10.1016/B978-0-12-385118-5.00018-9](https://doi.org/10.1016/B978-0-12-385118-5.00018-9)
41. Ralser M, Wamelink MM, Kowald A, Gerisch B, Heeren G, Struys Ea, et al. Dynamic rerouting of the carbohydrate flux is key to counteracting oxidative stress. *J Biol.* 2007 jan; 6(4):10. doi: [10.1186/jbiol61](https://doi.org/10.1186/jbiol61) PMID: [18154684](https://pubmed.ncbi.nlm.nih.gov/18154684/)

42. Chen KCKC, Calzone L, Csikasz-Nagy A, Cross FR, Novak B, Tyson JJ. Integrative analysis of cell cycle control in budding yeast. *Mol Biol Cell*. 2004 aug; 15(8):3841. doi: [10.1091/mbc.E03-11-0794](https://doi.org/10.1091/mbc.E03-11-0794) PMID: [15169868](https://pubmed.ncbi.nlm.nih.gov/15169868/)
43. Queralt E, Lehane C, Novak B, Uhlmann F. Downregulation of PP2A(Cdc55) phosphatase by separase initiates mitotic exit in budding yeast. *Cell*. 2006 may; 125(4):719–32. doi: [10.1016/j.cell.2006.03.038](https://doi.org/10.1016/j.cell.2006.03.038) PMID: [16713564](https://pubmed.ncbi.nlm.nih.gov/16713564/)
44. Vinod PK, Freire P, Rattani A, Ciliberto A, Uhlmann F, Novak B. Computational modelling of mitotic exit in budding yeast: the role of separase and Cdc14 endocycles. *J Royal Soc Interface*. 2011 aug; 8(61):1128–41. doi: [10.1098/rsif.2010.0649](https://doi.org/10.1098/rsif.2010.0649)
45. Kofahl B, Klipp E. Modelling the dynamics of the yeast pheromone pathway. *Yeast*. 2004 jul; 21(10):831–50. doi: [10.1002/yea.1122](https://doi.org/10.1002/yea.1122) PMID: [15300679](https://pubmed.ncbi.nlm.nih.gov/15300679/)
46. Haldane JBS. The Effect of Variation of Fitness. *Am Nat*. 1937; 71(735):337–349. doi: [10.1086/280722](https://doi.org/10.1086/280722)
47. Daub JT, Hofer T, Cutivet E, Dupanloup I, Quintana-Murci L, Robinson-Rechavi M, et al. Evidence for polygenic adaptation to pathogens in the human genome. *Mol Biol Evol*. 2013 jul; 30(7):1544–58. doi: [10.1093/molbev/mst080](https://doi.org/10.1093/molbev/mst080) PMID: [23625889](https://pubmed.ncbi.nlm.nih.gov/23625889/)
48. Imai H, Kefalov V, Sakurai K, Chisaka O, Ueda Y, Onishi A, et al. Molecular properties of rhodopsin and rod function. *J Biol Chem*. 2007 mar; 282(9):6677–84. doi: [10.1074/jbc.M610086200](https://doi.org/10.1074/jbc.M610086200) PMID: [17194706](https://pubmed.ncbi.nlm.nih.gov/17194706/)
49. Yokoyama S, Tada T, Zhang H, Britt L. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. *Proc Natl Acad Sci U S A*. 2008 sep; 105(36):13480–5. doi: [10.1073/pnas.0802426105](https://doi.org/10.1073/pnas.0802426105) PMID: [18768804](https://pubmed.ncbi.nlm.nih.gov/18768804/)
50. Chatterjee M, Osborne J, Bestetti G, Chang Y, Moore PS. Viral IL-6-induced cell proliferation and immune evasion of interferon activity. *Science*. 2002 nov; 298(5597):1432–5. doi: [10.1126/science.1074883](https://doi.org/10.1126/science.1074883) PMID: [12434062](https://pubmed.ncbi.nlm.nih.gov/12434062/)
51. Harker JA, Dolgote A, Zuniga EI. Cell-intrinsic IL-27 and gp130 cytokine receptor signaling regulates virus-specific CD4+ T cell responses and viral control during chronic infection. *Immunity*. 2013 sep; 39(3):548–59. doi: [10.1016/j.immuni.2013.08.010](https://doi.org/10.1016/j.immuni.2013.08.010) PMID: [23993651](https://pubmed.ncbi.nlm.nih.gov/23993651/)
52. Schmidt FL, Hunter JE. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 3rd ed. Los Angeles: SAGE Publications; 2014.
53. Almeida-de Macedo MM, Ransom N, Feng Y, Hurst J, Wurtele ES. Comprehensive analysis of correlation coefficients estimated from pooling heterogeneous microarray data. *BMC Bioinformatics*. 2013; 14(1):214. doi: [10.1186/1471-2105-14-214](https://doi.org/10.1186/1471-2105-14-214) PMID: [23822712](https://pubmed.ncbi.nlm.nih.gov/23822712/)
54. Hassler U, Thadewald T. Nonsensical and biased correlation due to pooling heterogeneous samples. *J R Stat Soc Ser D Stat*. 2003; 52(3):367–379. doi: [10.1111/1467-9884.00365](https://doi.org/10.1111/1467-9884.00365)
55. Field AP. Meta-analysis of correlation coefficients: a Monte Carlo comparison of fixed- and random-effects methods. *Psychol Methods*. 2001; 6(0):161–180. doi: [10.1037/1082-989X.6.2.161](https://doi.org/10.1037/1082-989X.6.2.161) PMID: [11411440](https://pubmed.ncbi.nlm.nih.gov/11411440/)
56. Field AP. Is the meta-analysis of correlation coefficients accurate when population correlations vary? *Psychol Methods*. 2005; 10(4):444–467. PMID: [16392999](https://pubmed.ncbi.nlm.nih.gov/16392999/)
57. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol*. 2006 nov; 8(11):1195–203. doi: [10.1038/ncb1497](https://doi.org/10.1038/ncb1497) PMID: [17060902](https://pubmed.ncbi.nlm.nih.gov/17060902/)
58. Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. Systems biology: parameter estimation for biochemical models. *FEBS J*. 2009 feb; 276(4):886–902. doi: [10.1111/j.1742-4658.2008.06844.x](https://doi.org/10.1111/j.1742-4658.2008.06844.x) PMID: [19215296](https://pubmed.ncbi.nlm.nih.gov/19215296/)
59. Brown KS, Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. *Phys Rev E*. 2003 aug; 68(2):1–9. doi: [10.1103/PhysRevE.68.021904](https://doi.org/10.1103/PhysRevE.68.021904)
60. Mannakee BK, Ragsdale AP, Transtrum MK, Gutenkunst RN. Sloppiness and the Geometry of Parameter Space. In: Geris L, Gomez-Cabrero D, editors. *Uncertainty in Biology*. Switzerland: Springer International; 2016. p. 271–299.
61. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*. 2007 oct; 3(10):e189. doi: [10.1371/journal.pcbi.0030189](https://doi.org/10.1371/journal.pcbi.0030189)
62. Transtrum MK, Machta BB, Brown KS, Daniels BC, Myers CR, Sethna JP. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *J Chem Phys*. 2015; 143(1):010901. doi: [10.1063/1.4923066](https://doi.org/10.1063/1.4923066) PMID: [26156455](https://pubmed.ncbi.nlm.nih.gov/26156455/)
63. Csárdi G, Franks A, Choi DS, Airoidi EM, Drummond DA. Accounting for experimental noise reveals That mRNA Levels, amplified by post-transcriptional processes, largely determine steady-state

- protein levels in yeast. *PLoS Genet.* 2015; 11(5):e1005206. doi: [10.1371/journal.pgen.1005206](https://doi.org/10.1371/journal.pgen.1005206) PMID: [25950722](https://pubmed.ncbi.nlm.nih.gov/25950722/)
64. Plotkin JB, Fraser HB. Assessing the determinants of evolutionary rates in the presence of noise. *Mol Biol Evol.* 2007 may; 24(5):1113–21. doi: [10.1093/molbev/msm044](https://doi.org/10.1093/molbev/msm044) PMID: [17347158](https://pubmed.ncbi.nlm.nih.gov/17347158/)
 65. Duret L, Mouchiroud D. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol.* 2000 jan; 17(1):68–70. doi: [10.1093/oxfordjournals.molbev.a026239](https://doi.org/10.1093/oxfordjournals.molbev.a026239) PMID: [10666707](https://pubmed.ncbi.nlm.nih.gov/10666707/)
 66. Pál C, Papp BB, Hurst LD, Pal C. Highly expressed genes in yeast evolve slowly. *Genetics.* 2001 aug; 158(2):927–931. PMID: [11430355](https://pubmed.ncbi.nlm.nih.gov/11430355/)
 67. Subramanian S, Kumar S. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. *Genetics.* 2004 sep; 168(1):373–81. doi: [10.1534/genetics.104.028944](https://doi.org/10.1534/genetics.104.028944) PMID: [15454550](https://pubmed.ncbi.nlm.nih.gov/15454550/)
 68. Drummond DA, Wilke CO. The evolutionary consequences of erroneous protein synthesis. *Nat Rev Genet.* 2009 oct; 10(10):715–24. doi: [10.1038/nrg2662](https://doi.org/10.1038/nrg2662) PMID: [19763154](https://pubmed.ncbi.nlm.nih.gov/19763154/)
 69. Yang JR, Zhuang SM, Zhang J. Impact of translational error-induced and error-free misfolding on the rate of protein evolution. *Mol Syst Biol.* 2010 oct; 6(421):421. doi: [10.1038/msb.2010.78](https://doi.org/10.1038/msb.2010.78) PMID: [20959819](https://pubmed.ncbi.nlm.nih.gov/20959819/)
 70. Geiler-Samerotte KA, Dion MF, Budnik BA, Wang SM, Hartl DL, Drummond DA. Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proc Natl Acad Sci U S A.* 2011; 108:680–685. doi: [10.1073/pnas.1017570108](https://doi.org/10.1073/pnas.1017570108) PMID: [21187411](https://pubmed.ncbi.nlm.nih.gov/21187411/)
 71. Yang JR, Liao BY, Zhuang SM, Zhang J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc Natl Acad Sci U S A.* 2012 mar;p. 831–840. doi: [10.1073/pnas.1117408109](https://doi.org/10.1073/pnas.1117408109)
 72. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW. Evolutionary rate in the protein interaction network. *Science.* 2002 apr; 296(5568):750–2. doi: [10.1126/science.1068696](https://doi.org/10.1126/science.1068696) PMID: [11976460](https://pubmed.ncbi.nlm.nih.gov/11976460/)
 73. Hahn MW, Kern AD. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. *Mol Biol Evol.* 2005 apr; 22(4):803–6. doi: [10.1093/molbev/msi072](https://doi.org/10.1093/molbev/msi072) PMID: [15616139](https://pubmed.ncbi.nlm.nih.gov/15616139/)
 74. Mangan S, Alon U. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A.* 2003 oct; 100(21):11980–5. doi: [10.1073/pnas.2133841100](https://doi.org/10.1073/pnas.2133841100) PMID: [14530388](https://pubmed.ncbi.nlm.nih.gov/14530388/)
 75. Soyer OS, editor. *Evolutionary Systems Biology.* Springer; 2012.
 76. Hartl D, Dykhuizen D, Dean A. Limits of adaptation: the evolution of selective neutrality. *Genetics.* 1985;p. 655–674. PMID: [3932127](https://pubmed.ncbi.nlm.nih.gov/3932127/)
 77. Alvarez-Ponce D, Aguadé M, Rozas J. Comparative genomics of the vertebrate insulin/TOR signal transduction pathway: a network-level analysis of selective pressures. *Genome Biol Evol.* 2011 jan; 3:87–101. doi: [10.1093/gbe/evq084](https://doi.org/10.1093/gbe/evq084) PMID: [21149867](https://pubmed.ncbi.nlm.nih.gov/21149867/)
 78. Vitkup D, Kharchenko P, Wagner A. Influence of metabolic network structure and function on enzyme evolution. *Genome Biol.* 2006 jan; 7(5):R39. doi: [10.1186/gb-2006-7-5-r39](https://doi.org/10.1186/gb-2006-7-5-r39) PMID: [16684370](https://pubmed.ncbi.nlm.nih.gov/16684370/)
 79. Loewe L, Hillston J. The distribution of mutational effects on fitness in a simple circadian clock. In: *Proceedings of the 6th International Conference on Computational Methods in Systems Biology.* Berlin: Springer-Verlag; 2008. p. 156–175.
 80. Invergo BM, Montanucci L, Bertranpetit J. Dynamic sensitivity and nonlinear interactions influence the system-level evolutionary patterns of phototransduction proteins. *Proc R Soc B.* 2015; 282:20152215. doi: [10.1098/rspb.2015.2215](https://doi.org/10.1098/rspb.2015.2215) PMID: [26631565](https://pubmed.ncbi.nlm.nih.gov/26631565/)
 81. Hermansen RA, Mannakee BK, Knecht W, Liberles Da, Gutenkunst RN. Characterizing selective pressures on the pathway for de novo biosynthesis of pyrimidines in yeast. *BMC Evol Biol.* 2015; 15(1):232. doi: [10.1186/s12862-015-0515-x](https://doi.org/10.1186/s12862-015-0515-x) PMID: [26511837](https://pubmed.ncbi.nlm.nih.gov/26511837/)
 82. Karr JR, Sanghvi JC, MacKlin DN, Gutschow MV, Jacobs JM, Bolival B, et al. A whole-cell computational model predicts phenotype from genotype. *Cell.* 2012; 150(2):389–401. doi: [10.1016/j.cell.2012.05.044](https://doi.org/10.1016/j.cell.2012.05.044) PMID: [22817898](https://pubmed.ncbi.nlm.nih.gov/22817898/)
 83. Kacser H, Burns JA. The control of flux. *Symp Soc Exp Biol.* 1973 jan; 27:65–104. PMID: [4148886](https://pubmed.ncbi.nlm.nih.gov/4148886/)
 84. Elowitz MB. Stochastic gene expression in a single cell. *Science.* 2014; 1183(2002):1183–1187.
 85. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A.* 2002; 99(20):12795–12800. doi: [10.1073/pnas.162041399](https://doi.org/10.1073/pnas.162041399) PMID: [12237400](https://pubmed.ncbi.nlm.nih.gov/12237400/)
 86. Komorowski M, Costa MJ, Rand Da, Stumpf MPH. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proc Natl Acad Sci U S A.* 2011 may; 108(21):8645–50. doi: [10.1073/pnas.1015814108](https://doi.org/10.1073/pnas.1015814108) PMID: [21551095](https://pubmed.ncbi.nlm.nih.gov/21551095/)

87. Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, et al. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*. 2005; 437(7059):699–706. doi: [10.1038/nature03998](https://doi.org/10.1038/nature03998) PMID: [16170311](https://pubmed.ncbi.nlm.nih.gov/16170311/)
88. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*. 2003 mar; 19(4):524–531. doi: [10.1093/bioinformatics/btg015](https://doi.org/10.1093/bioinformatics/btg015) PMID: [12611808](https://pubmed.ncbi.nlm.nih.gov/12611808/)
89. Myers CR, Gutenkunst RN, Sethna JP. Python unleashed on systems biology. *Comput Sci Eng*. 2007; 9(3):34–37. doi: [10.1109/MCSE.2007.60](https://doi.org/10.1109/MCSE.2007.60)
90. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2014 jan; 42(Database issue):D7–17. doi: [10.1093/nar/gkt1146](https://doi.org/10.1093/nar/gkt1146) PMID: [24259429](https://pubmed.ncbi.nlm.nih.gov/24259429/)
91. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res*. 2012 jan; 40(Database issue):D700–5. doi: [10.1093/nar/gkr1029](https://doi.org/10.1093/nar/gkr1029) PMID: [22110037](https://pubmed.ncbi.nlm.nih.gov/22110037/)
92. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009 jun; 25(11):1422–1423. doi: [10.1093/bioinformatics/btp163](https://doi.org/10.1093/bioinformatics/btp163) PMID: [19304878](https://pubmed.ncbi.nlm.nih.gov/19304878/)
93. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics*. 2007 nov; 23(21):2947–2948. doi: [10.1093/bioinformatics/btm404](https://doi.org/10.1093/bioinformatics/btm404) PMID: [17846036](https://pubmed.ncbi.nlm.nih.gov/17846036/)
94. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007 aug; 24(8):1586–91. doi: [10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088) PMID: [17483113](https://pubmed.ncbi.nlm.nih.gov/17483113/)
95. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell*. 2008 jul; 134(2):341–52. doi: [10.1016/j.cell.2008.05.042](https://doi.org/10.1016/j.cell.2008.05.042) PMID: [18662548](https://pubmed.ncbi.nlm.nih.gov/18662548/)
96. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A*. 2004 apr; 101(16):6062–7. doi: [10.1073/pnas.0400782101](https://doi.org/10.1073/pnas.0400782101) PMID: [15075390](https://pubmed.ncbi.nlm.nih.gov/15075390/)
97. Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*. 2011 jan; 2011:bar030. doi: [10.1093/database/bar030](https://doi.org/10.1093/database/bar030) PMID: [21785142](https://pubmed.ncbi.nlm.nih.gov/21785142/)
98. Holstege FCP, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, et al. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*. 1998 nov; 95(5):717–728. doi: [10.1016/S0092-8674\(00\)81641-4](https://doi.org/10.1016/S0092-8674(00)81641-4) PMID: [9845373](https://pubmed.ncbi.nlm.nih.gov/9845373/)
99. Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res*. 2011 jan; 39(Database issue):D842–8. doi: [10.1093/nar/gkq1008](https://doi.org/10.1093/nar/gkq1008) PMID: [21051359](https://pubmed.ncbi.nlm.nih.gov/21051359/)
100. Brown KR, Jurisica I. Online predicted human interaction database. *Bioinformatics*. 2005 may; 21(9):2076–82. doi: [10.1093/bioinformatics/bti273](https://doi.org/10.1093/bioinformatics/bti273) PMID: [15657099](https://pubmed.ncbi.nlm.nih.gov/15657099/)
101. Hagberg AA, Schult DA, Swart PJ. Exploring network structure, dynamics, and function using NetworkX. In: *Proceedings of the 7th Python in Science Conference*. Pasadena, CA USA; 2008. p. 11–15.
102. Hedges LV, Vevea JL. Fixed- and random-effects models in meta-analysis. *Psychol Methods*. 1998; 3(4):486–504. doi: [10.1037/1082-989X.3.4.486](https://doi.org/10.1037/1082-989X.3.4.486)
103. Freedman D, Lane D. A nonstochastic interpretation of reported significance levels. *J Bus Econ Stat*. 1983; 1(4):292–298. doi: [10.1080/07350015.1983.10509354](https://doi.org/10.1080/07350015.1983.10509354)
104. Anderson MJ, Legendre P. An empirical comparison of permutation methods for tests of partial regression coefficients in a linear model. *J Stat Comput Simul*. 1999; 62(February 2015):271–303. doi: [10.1080/00949659908811936](https://doi.org/10.1080/00949659908811936)
105. DiCiccio CJ, Romano JP. Robust permutation tests for correlation and regression coefficients; Stanford University Department of Statistics technical report 2015-15; 2015.