

Title:

Being noisy in a crowd: differential selective pressure on gene expression noise in model gene regulatory networks

Nataša Puzović<sup>1\*</sup>, Tanvi Madaan<sup>1</sup>, Julien Y. Dutheil<sup>1,2</sup>

<sup>1</sup> Molecular Systems Evolution Research Group, Max Planck Institute for Evolutionary Biology, Plön, Schleswig-Holstein, Germany

<sup>2</sup> Institut des sciences de l'évolution, Montpellier, Languedoc-Roussillon, France

\* Corresponding author

Email: [puzovic@evolbio.mpg.de](mailto:puzovic@evolbio.mpg.de)

## Abstract

Expression noise, the variability of the amount of gene product among isogenic cells grown in identical conditions, originates from the inherent stochasticity of diffusion and binding of the molecular players involved in transcription and translation. It has been shown that expression noise is an evolvable trait and that central genes exhibit less noise than peripheral genes in gene networks. A possible explanation for this pattern is increased selective pressure on central genes since they propagate their noise to downstream targets, leading to noise amplification. To test this hypothesis, we developed a new gene regulatory network model with inheritable stochastic gene expression and simulated the evolution of gene-specific expression noise under constraint at the network level. Stabilizing selection was imposed on the expression level of all genes in the network and rounds of mutation, selection, replication and recombination were performed. We observed that local network features affect both the probability to respond to selection, and the strength of the selective pressure acting on individual genes. In particular, the reduction of gene-specific expression noise as a response to stabilizing selection on the mean expression is higher in genes with higher centrality metrics. Furthermore, global topological structures such as network diameter, centralization and average degree affect the average expression variance and average selective pressure acting on constituent genes. Our results demonstrate that selection at the network level leads to differential selective pressure at the gene level, and local and global network characteristics are an essential component of gene-specific expression noise evolution.

## Introduction

Living beings are complex systems constituted of many genes that interact with each other and the environment to create an organism. From prokaryotes with a few hundred essential genes, to eukaryotes with possibly several thousands, cells require many gene products to work together to perform housekeeping functions and to replicate. Fine-tuned molecular processes, generally referred to as *gene expression*, ensure how, where and when these products are generated. However, gene expression is an inherently noisy process (1; 2), which involves many steps where molecules participating in the expression machinery diffuse and bind to target molecules. Additionally, these molecules are often present in small copy numbers, increasing the susceptibility of gene expression to stochastic events. Consequently, there is a variation in gene expression levels among cells, even if they are isogenic and grown in a homogeneous environment, and this inevitable variation has been termed *gene expression noise*. Organisms have to express hundreds of genes, each one of which is noisy – raising the question of how they evolved to cope with this inevitable noise.

The expression noise level of a particular gene may be decomposed into two components, called *extrinsic* and *intrinsic*. Extrinsic noise affects all genes equally and results from the sharing of key molecules, such as RNA polymerases and ribosomes, by all genes in the expression process, as well as, for instance, differences in cell size and phase in the cell cycle. Intrinsic noise is gene-specific and results from different chromatin states, cis-regulatory elements and kinetic parameters of transcription and translation of each gene (3). Minor sequence mutations can have a significant effect on the level of expression noise. For example, a small number of single-nucleotide changes in a transcription factor binding site were reported to have a large effect on the expression noise level (4). Since (i) there is variation in the level of intrinsic noise of genes, and (ii) intrinsic noise is genetically determined – and, therefore, heritable – gene expression noise can be shaped by natural selection.

Evidence of selection on expression noise was first seen in the fact that dosage-sensitive genes (5) and essential genes exhibit lower levels of expression noise (6; 7). Intrinsic noise was also reported to

correlate with the strength of selection acting on the encoded protein. Namely, proteins with a lower ratio of non-synonymous over synonymous substitution rate ( $K_a/K_s$ ) have a lower level of expression noise (8). Changes in the expression noise of a single gene may be either beneficial or deleterious, depending on how far its mean expression is from the optimal expression level (9). Expression noise is deleterious if the mean expression level is close to the optimal, as higher variation, in this case, generates a larger number of less fit individuals, reducing the population fitness. Conversely, expression noise can be beneficial if the mean expression level is far from the optimum, as noisy genes are more likely to generate cells with an expression level closer to the optimum. Noisy gene expression can thus be part of a bet-hedging strategy and was observed in genes involved in immune and environmental response (10). The fitness cost of changes in the level of expression noise in the fitness landscapes of  $\approx 30$  yeast genes have been shown to be on the same order as fitness costs of changes in mean expression level (11). Since the fitness effect of different levels of expression noise can be as detrimental as different mean expression levels, which are thought to be extensively under selection (12), it can be assumed that expression noise is extensively under selection genome-wide. Prevalent selection on expression noise has been demonstrated in naturally segregating promoter variants of *E. coli* (13).

The phenotype (and, therefore, the fitness) of an organism depends on the interaction of many genes. As a result, genes do not evolve independently, and the selective pressure acting on a gene's intrinsic noise depends on its interactions with other genes. Understanding the evolution of gene expression noise requires accounting for such gene-to-gene interactions, commonly depicted by a gene network. The propagation of noise from gene to gene in the network was established both theoretically and experimentally (14; 15). Genes with many connections propagate their noise to a more substantial extent than genes with fewer connections and, therefore, contribute more to the global noise levels of the network. Gene networks are robust to variation in the expression level of their system components to some degree, but at a critical point the global noise of the network becomes too high and leads to network collapse. Selection against noise at the network level was, therefore, hypothesized to result in stronger constraints on the intrinsic noise of highly connected genes (8). Moreover, the topological structure of the network has been shown to affect the pattern of noise propagation (16), suggesting that the topology of the network might impose additional selective constraints on the constituent genes.

Here, we test the hypothesis that expression noise of highly connected genes in gene networks is under stronger selective pressure than expression noise in peripheral genes using an *in silico* evolutionary experiment. We introduce a new gene regulatory network evolution model, which includes an evolvable component of stochastic gene expression, and use it to evolve thousands of network topology samples over 10,000 generations. These simulations showed that highly connected genes have a more constrained intrinsic expression noise. They further revealed that not all genes might evolve in response to network-level selection, and the probability that they do so depends on local network properties. Lastly, the average selective pressure acting on genes in a network is affected by topological features such as network diameter, centralization and average degree.

## Materials and methods

We introduce a new gene regulatory network model that incorporates intrinsic expression noise. We then use this model within a forward simulation framework to simulate the evolution of populations of networks with mutable levels of intrinsic expression noise. These simulations allow us to study how the selective pressure acting on expression noise varies within the regulatory network.

## A gene regulatory network model with stochastic gene expression

To investigate the evolution of stochastic gene expression in gene regulatory networks, we first extend Wagner's gene network model (17) to integrate gene-specific expression noise.

We model a network of  $n$  genes ( $n = 40$  in this study) defined by a regulatory matrix  $W = (w_{ij})_{1 \leq i \leq n, 1 \leq j \leq n}$ , and a vector of intrinsic, gene-specific noise  $\{\eta_i^{\text{int}}\}_{1 \leq i \leq n}$ . Each element  $w_{ij}$  of the regulatory matrix  $W$  defines the regulatory effect of gene  $j$  on gene  $i$ . The value of  $w_{ij}$  is a real number and is referred to as regulatory strength of gene  $j$  on gene  $i$ . In case  $w_{ij} > 0$ , gene  $j$  is an activator of gene  $i$  and increases its expression level. Conversely, when  $w_{ij} < 0$ , gene  $j$  is a repressor of gene  $i$  and decreases its expression level. Lastly, if  $w_{ij} = 0$ , gene  $i$  is not regulated by gene  $j$  and gene  $j$  has no effect on expression level of gene  $i$ . Two genes  $i$  and  $j$  are connected by an edge in the network if at least one of  $w_{ij}$  and  $w_{ji}$  is non-null. The intrinsic noise vector  $\{\eta_i^{\text{int}}\}_{1 \leq i \leq n}$  defines the gene-specific expression noise of each gene in the network. The regulatory matrix and the intrinsic noise vector together constitute a unique genotype in this modeling framework (Fig 1A).

The phenotype (the expression level of each gene) in the model is represented by a state vector  $\{S_i\}_{1 \leq i \leq n} = \{s_1, s_2, \dots, s_n\}$ , which describes the expression level of each gene. The state vector at  $t_0$  is set to an arbitrary basal expression level value ( $\{S_i^0\}_{1 \leq i \leq n} = \{S_i^{\text{basal}}\}_{1 \leq i \leq n} = \{20, \dots, 20\}$  in this study). In every time step  $t$  ( $1 \leq t \leq T_r$ , with  $T_r = 50$  in this study), the expression level of each gene is recomputed. The cumulative effect of all transcription factors in the expression level of each gene is considered to be additive, *i.e.* we assume there is no cooperative or competitive binding of transcription factors to transcription factor binding sites. The activation rate  $a_i(t)$  is defined as the sum of all effects the regulators of gene  $i$  have on its expression level at time step  $t$ :

$$a_i(t) = \sum_{j=1}^n w_{ij} \cdot s_j(t), \quad (1)$$

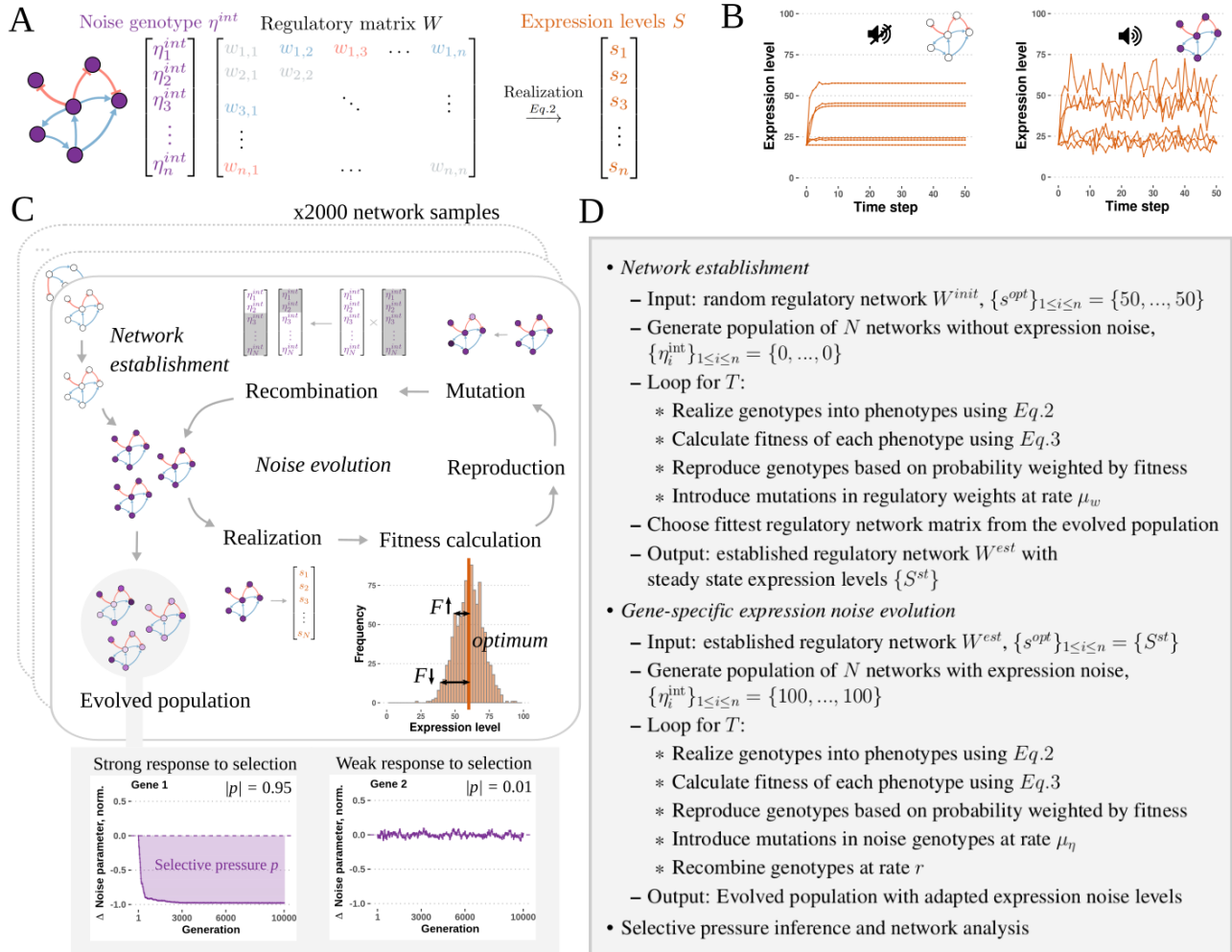
in which case the dynamic equation for the expression level of each gene in the following time step is:

$$s_i(t+1) \sim \mathcal{N}\left(s_i^{\text{basal}} + a_i(t), \eta_i^{\text{int}}\right). \quad (2)$$

In every time step the expression level of a gene is drawn from a normal distribution, where the mean equals the sum of basal expression level ( $s_i^{\text{basal}}$ ) and activation rate ( $a_i(t)$ ), and the variance equals the gene noise genotype ( $\eta_i^{\text{int}}$ ). If the expression level value drawn from the normal distribution is below the minimal ( $s_{\text{min}} = 0$ ) or above the maximal expression level ( $s_{\text{max}} = 100$ ), it is set to the minimal or maximal expression level, respectively. The expression levels of all genes are synchronously updated in each time step. The steady state expression levels are invariant to whether the expression levels of each gene are updated synchronously or asynchronously (SI Section 1.2, Fig S1A and Fig S2). Similarly, mean expression level, expression variance, CV, noise and Fano factor are invariant to the updating mode (Fig S1B-F). The model may be realized as stochastic or deterministic, depending on the noise parameter values (Fig 1B).

## Forward-in-time simulation of expression noise evolution

To investigate how gene-specific expression noise of constituent genes responds to stabilizing selection at the network level, we used the newly introduced model to perform forward-in-time evolutionary



**Fig 1. The evolution of gene-specific expression noise was simulated using populations of model gene regulatory networks with mutable levels of gene-specific expression noise under selective and non-selective conditions. A - Gene regulatory network model.** The genotype consists of the intrinsic noise vector  $\eta^{int}$  and regulatory matrix  $W$ . The intrinsic noise vector defines the gene-specific expression variance of each gene in the network. The regulatory matrix defines the regulatory interactions in the network. The genotype is realized into the phenotype using the dynamical equation described in the main text. The phenotype is given by the state vector  $S$ , which represents the expression level of each gene in the network. **B - Deterministic (left) and stochastic (right) realizations of the model.** **C - Steps of the evolutionary simulation process.** Each established network configuration was used as a founding network for the network populations used in the noise evolution simulation. In every generation, genotypes are realized and phenotypes (expression levels) are sampled from the last time step. Fitness is calculated from the expression levels. If the populations are evolved under selection, fitness is calculated as the distance of the expression level of each gene from the optimal expression level. Genotypes are reproduced based on their relative fitness and mutations in the intrinsic noise vectors are introduced. Noise genotype vectors are recombined by randomly choosing individuals for recombination and shuffling their noise vectors. The process is repeated for 10,000 generations. **D - Algorithm overview.**

simulations in which we allow the gene-specific noise levels to mutate. An *in silico* evolutionary process consisting of rounds of mutation, selection, recombination and replication events of a population of  $N$  ( $N = 1,000$  in this study) individuals was performed for  $T$  ( $T = 10,000$ ) generations (Fig 1C).

We first generated network topologies that would serve as the founding network for the populations in our simulations. We generated 2,000 random (Erdős–Rényi model) network topologies of 40 nodes with regulatory strength values drawn from a uniform distribution  $\mathcal{U}(-3, 3)$ . The network density was  $d = 0.05$ . Only connected network graphs were used, meaning there is only one component and there are no disconnected subgraphs. In order to assess the effect of the topology structure on the evolution of expression noise, we also generated an additional 1,000 scale-free (Barabási–Albert model) and 1,000 small-world (Watts–Strogatz model) network topologies with the same size and density. Both random and small-world networks are characterized by a Poisson degree distribution and short mean shortest path length, but random networks have a low clustering coefficient, while small-world networks have a high clustering coefficient. Scale-free networks are characterized by a degree distribution that follows a power law. Real-world networks exhibit degree distributions similar to power-law distributions, high clustering and short path lengths. As such, real-world networks have features of both scale-free and small-world networks (18).

In the simulation of expression noise evolution the regulatory interactions were immutable and the values of the noise genotype vectors were allowed to mutate. Stabilizing selection was imposed on all constituent genes by setting the value of optimal expression level as the mean equilibrium expression level of each gene. The fitness  $F(s)$  of a phenotype  $s$  was calculated as in Laarits et al. (19), where fitness is defined as the distance from the optimal expression state vector  $\{s_i^{opt}\}_{1 \leq i \leq n}$ , weighted by the selective pressure given by  $\{\rho_i\}_{1 \leq i \leq n}$ :

$$F(s) = e^{-\sum_{i=1}^n |s_i^{opt} - s_i| / (n\rho_i)} \quad (3)$$

The selective pressure parameters  $\{\rho_i\}_{1 \leq i \leq n}$  define the contribution of each gene to the fitness of the phenotype. In this study, the strength of the imposed selective pressure is set to be identical for all constituent genes ( $\forall i \rho_i = 1$ ). Since the strength of the selective pressure we impose on all genes is identical, any differences in the evolutionary outcome we observe after removing the effect of drift will be due to gene differences in their network interactions. Individuals were reproduced into the next generation with a probability equal to their relative phenotype fitness. The fitness of all phenotypes in populations evolved under non-selective conditions was set to an equal constant value, regardless of gene expression levels. Mutations were introduced at a rate  $\mu_\eta$  ( $\mu_\eta = 0.01$ ) per gene per replication event. The values for noise genotype mutations were drawn from a normal distribution  $\mathcal{N}(100, 40)$ . Recombination was implemented by choosing a random offspring individual at a rate  $r$  ( $r = 0.05$ ) and introducing a random break point in the linear genome. The genotype values in the genome segment defined by the break point were then exchanged with another randomly chosen individual from the offspring population. A constant population size  $N$  ( $N = 1,000$ ) was maintained. To account for the effect of genetic drift, the noise evolution simulations of each founding network population were replicated 10 times under selection and 10 times under neutrality.

We found that the expression level of most genes in networks with random configurations converge to either  $s_{min}$  or  $s_{max}$  under a deterministic realization. The measurement of variance of genes that are either not expressed at all or expressed at the maximal level would be impaired since their expression range is constrained by the lower and upper expression level boundary. Since the study of expression variance is our main focus, we added a network establishment step before the noise evolution simulations, in which we subject the network regulatory matrix to mutation and selection for intermediate expression levels. During the network establishment step networks are realized



deterministically, *i.e.* the intrinsic noise genotype of all genes is 0. Networks with intermediate steady state expression levels were established through the evolutionary process by imposing a target expression level ( $\{s_i^{opt}\}_{1 \leq i \leq n} = \{50, \dots, 50\}$ ) for all genes and allowing the strength of regulatory interactions to mutate. Mutations were introduced at a rate  $\mu_w$  ( $\mu_w = 0.05$ ) in non-zero entries in the regulatory matrix, preserving the network topology structure (Erdős–Rényi, Barabási–Albert, or Watts–Strogatz model). The values for regulatory strength mutations were drawn from a normal distribution  $\mathcal{N}(0, 2)$ . Recombination was not implemented at this stage. Fitness of each individual was computed as the distance of the phenotype to the optimal expression state vector using Eq.1. Individuals were reproduced with a probability equal to the relative fitness and the population size kept constant. Network regulatory configurations in which the expression level of all genes would not converge to a fixed point and would oscillate were removed by setting their fitness value to 0. Expression level dynamics were termed oscillating if the sum of the differences between expression level in the last time step and previous  $\tau$  time steps ( $\tau = 10$ ) was higher than  $\epsilon$  ( $\epsilon = 10^{-6}$ ). A stable, *i.e.* non-oscillating, expression level dynamics satisfied the following criterion (19):

$$\Phi(S(t)) = \frac{1}{\tau} \sum_{\theta=t-\tau}^t D(S(\theta), S(t)) < \epsilon \quad (4)$$

where  $D$  is the distance between two vectors  $D(S^1, S^2) = \sum_{i=1}^n |S_i^1 - S_i^2|/n$ .

The network establishment process consisting of rounds of mutation, selection and reproduction of a population of  $N$  ( $N = 1,000$ ) individuals was performed for  $T$  ( $T = 10,000$ ) generations, for each network topology. The network regulatory configuration with the highest fitness was chosen from the evolved population and this network configuration was used to generate the starting network population for the noise evolution simulations.

The gene network model and evolutionary simulations were implemented in C++ and the source code is available at [https://gitlab.gwdg.de/molsysevol/supplementarydata\\_expressionnoise/cpp](https://gitlab.gwdg.de/molsysevol/supplementarydata_expressionnoise/cpp).

## Analysis of simulation results: expression noise and network centrality measures

The evolutionary outcomes (*i.e.* the change of phenotypes and genotypes) were measured as change of expression noise and selective pressure for each network, respectively. Expression noise in the first and last generation in each evolved population was measured as the variance of the population expression level states for each gene. The change of expression noise (phenotypic evolution) between the first and last generation was measured as the relative change of expression noise, calculated as the difference of expression variance between the first and last generation divided by their sum  $(\sigma_{gen1}^2 - \sigma_{gen10k}^2)/(\sigma_{gen1}^2 + \sigma_{gen10k}^2)$ .

The selective pressure (genotypic evolution) acting on each gene was measured as the average change of noise genotype in every second generation relative to the starting level (Fig 1C). To compare the effect of node centrality on the selective pressure acting on constituent genes, we computed node-level network centrality measures for each node in the networks. We focused our analysis on two local network centrality measures, node instrength and outstrength, but over 30 network centrality measures were analyzed (SI Section 2). Instrength of node  $i$  is measure of the strength and number of in-going links, *i.e.* how strongly a gene is being regulated:

$$\text{Instrength}(i) = \sum_j^n |w_{ij}|. \quad (5)$$

Conversely, the outstrength of node  $j$  is a measure of the strength and number of outgoing links, *i.e.* how strongly a gene regulates other genes downstream:

$$\text{Outstrength}(j) = \sum_i^n |w_{ij}|. \quad (6)$$

Further, we computed global graph-level metrics, such as mean graph distance and performed a principal component analysis to reduce the dimensionality (SI Section 2.2). The results were analysed in R 3.6.3 (20). Network analyses were performed using the `igraph` 1.2.4.2 (21) and `statnet` 2019.6 (22) packages. Principal component analysis was performed using the `ade4` 1.7.15 (23) package.

## Analysis of simulation results: linear modeling

We fitted linear mixed-effects models using network centrality measures as fixed effect variables and the network topology sample as a random effect variable, allowing for control of intra-network correlation in the response variable. We tested different transformations of the response and explanatory variables in order to improve linearity, and variance structures to account for heteroskedasticity of the residuals. A model where the residual variance was an exponential function of the node absolute instrength was shown to provide the best fit according to the minimal Akaike's Information criterion and was used for all subsequent models (SI Section 3). Two types of models were fitted: a logistic regression where the response variable was set to whether a gene answered to selection or not, and standard regressions that used expression variance, relative change of expression variance or selective pressure as response variables. Linear mixed-effect modelling was performed using the `nlme` 3.1.144 (24) and `lme4` 1.1.27.1 (25) packages. Marginal and conditional  $R^2$  values were computed using the `MuMIn` 1.43.17 (26) package. Network centrality measures used as explanatory variables in our linear models were correlated (Pearson's  $r = -0.17$ , p-value  $< 2.2 \times 10^{-16}$ , SI Fig 7B), so we computed the variance inflation factor (VIF) using the `car` 3.0.11 (27) package. The VIF of all linear models was less than 3; therefore, colinearity was considered to have negligible impact on the inferred statistical significance (28). To improve homoskedasticity of the residuals in the linear models, we also performed each model fit on two filtered datasets: one in which genes with zero values of instrength or outstrength were removed, and one in which only genes with zero values of instrength or outstrength were kept. The same pattern of effects and significance is observed in the filtered as in the main dataset, so we included the results of the complete dataset in the main text and reported the results of the reduced dataset in the supplementary information (SI Section 5).

Finally, since in some cases variable transformation, heterogeneous variance modeling and data filtering did not ensure normality and independence of the residuals, we assessed the amount of resulting bias in the estimation of p-values using a randomization test, in which we fitted a selected model on 10,000 permuted datasets. We chose the model of relative noise change (SI Section 3.2), as the corresponding residuals were significantly departing normality (Shapiro-Wilk test, p-value  $< 2.2 \times 10^{-16}$ ) and independence (Box-Ljung test, p-value  $= 8.9 \times 10^{-7}$ ). For each permutation, we shuffled the values of the response variable (relative change of variance) within each network topology, which removes the effect of network metrics on the change of noise, but preserves the distributions of each metric per network, as well as putative colinearity between explanatory variables. Using  $\alpha = 0.05$  as a significance cutoff value, we found a false discovery rate (FDR) of 6.0% for the effect of instrength and 6.7% for the effect of outstrength. While these values are above the expected 5%, the FDR inflation was found to be relatively low and we concluded that the non-normality of residuals did not affect our conclusions.



## Analysis of simulation results: information-based metrics

249

As an additional line of evidence, we calculated mutual information between the expression noise and centrality metrics using the `infotheo` 1.2.0 (29) package. Monte Carlo permutation tests with 10,000 permutations were used to compute p-values for the significance of the mutual information between each pair of tested variables. The simulation results data and the code necessary to reproduce all figures is available at <https://doi.org/10.5281/zenodo.6939845>, together with the the code necessary to generate all raw simulation files.

250  
251  
252  
253  
254  
255

## Results

256

To investigate the evolution of gene-specific expression noise in gene networks, we introduced a new gene regulatory model with stochastic gene expression. This model extends Wagner's model (17) by adding node-specific intrinsic noise parameters (Fig 1A-B). In this framework, the phenotype is represented by the expression level of each gene, and is the realization of a random distribution determined by the genotype. The fitness of an individual is further determined by its distance to an optimal phenotype, therefore, stabilizing selection is implemented as acting on the expression level. We used this model to simulate the evolution of populations of gene regulatory networks with mutable levels of gene-specific expression noise under selective and non-selective conditions (Fig 1C-D), and assessed how node properties affect the evolution of intrinsic noise.

257  
258  
259  
260  
261  
262  
263  
264  
265

## Expression noise propagates along the regulatory network

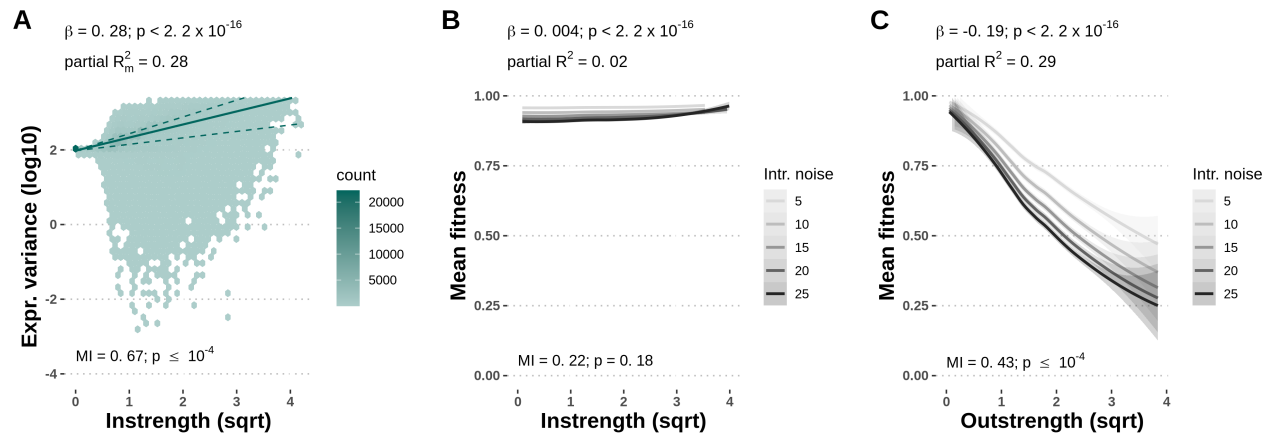
266

We first investigated how noise propagated in the model gene regulatory networks. It was shown that noise is additive in biological networks and, therefore, propagates from regulators to regulated genes (14; 15). To assess whether our model successfully captured this property, we generated a dataset of 2,000 realized random network topologies, and tested whether gene expression variance increased with the number of ingoing regulatory links. As expected, we found that the absolute instrength of a gene had a significant positive effect on gene expression variance (linear mixed-effects model with coefficient  $\beta = 0.28$ , p-value  $< 2.2 \times 10^{-16}$ ) (Fig 2A), indicating that noise propagation was captured in our model. Furthermore, the mutual information between gene expression variance and absolute instrength was significant (MI = 0.67, p-value  $\leq 10^{-4}$ , permutation test). High node instrength increases expression noise, in line with the experimental evidence that the noisiness of promoters increases with the number of regulatory inputs (30).

267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277

We then looked at fitness costs associated with high expression noise in regulators and regulated genes. In a dataset of 1,000 random network topologies, we assessed the mean fitness of the clonal populations of 1,000 individuals under stabilizing selection on the expression level. Each gene was imposed 5 different levels of intrinsic noise, while the intrinsic noise of the rest of the network was kept at 0. We found that increasing the level of expression noise of a single gene decreased the mean fitness of the network (linear mixed-effects model with coefficient  $\beta = -0.002$ , p-value  $< 2.2 \times 10^{-16}$ ), as expected. However, the strength of this effect depended on the gene centrality. The reduction of fitness due to gene-specific expression noise was significantly, but marginally, affected by instrength (linear model with coefficient  $\beta = 0.004$ , p-value  $< 2.2 \times 10^{-16}$ , Fig 2B). The mutual information between mean fitness of the population and absolute instrength was not significant (MI = 0.22, p-value = 0.18, permutation test). However, the mean fitness significantly decreased with node outstrength (linear model with coefficient  $\beta = -0.19$ , p-value  $< 2.2 \times 10^{-16}$ , Fig 2C). The mutual information between mean fitness of the population and absolute outstrength was significant (MI =

278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290



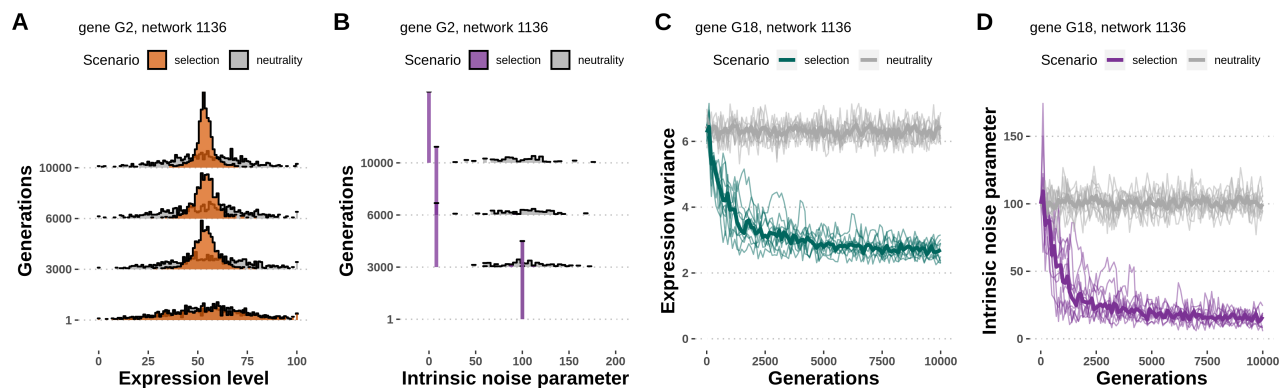
**Fig 2. Noise propagation is captured by the gene regulatory network model. A -**

Gene-specific expression variance increases with the absolute instrength of the node, indicating noise propagation is reflected in the gene regulatory network model. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **B, C** - Gene-specific expression variance decreases fitness in gene networks under stabilizing selection on gene expression level. Increasing the level of gene-specific expression noise reduces the mean fitness of the clonal population. The mean fitness of the population is significantly, but marginally, increased by noise in genes with higher node instrength (B), and significantly decreased by noise in genes with higher node outstrength (C). Lines represent the smoothed conditional means and grey bands represent the 95% confidence interval bands. Coefficients, p-values and partial marginal  $R^2$  measures are estimated using linear mixed-effects models with expression variance or mean fitness as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were computed with a permutation test with 10,000 permutations.

0.43, p-value  $\leq 10^{-4}$ , permutation test). Higher fitness cost of expression noise in gene with high outstrength suggests there is a differential selective pressure acting on genes based on their centrality in the gene regulatory network, which we explore in the next section using an *in silico* evolutionary experiment.

## Gene expression noise is reduced under a stabilizing selection regime

To investigate how gene-specific expression noise responds to stabilizing selection at the network-level, we simulated the evolution of 2,000 random network topologies with and without selection on the gene expression level. We observed that gene expression variance decreased throughout evolution under selective conditions (Fig 3A), and the distribution of intrinsic noise parameters in the population shifted towards lower noise genotype values (Fig 3B), indicating that low-noise alleles conferred a fitness increase to the network. Conversely, gene expression variance remained constant throughout evolution under neutral conditions, and the distribution of noise genotypes reflected only the distribution of random mutations. Replicating the simulations for each network topology sample yielded similar reduction of gene expression variance (Fig 3C) and median noise parameter in the population (Fig 3D). As the initial networks were at their optimal expression level,



**Fig 3. Gene-specific expression noise evolves in a model with selection.** **A** - The distribution of expression levels of an example gene throughout evolution in populations evolved under stabilizing selection on gene expression level and under neutrality. The variance of gene expression level is reduced under selection, but not under neutrality. **B** - The distribution of intrinsic noise parameters of an example gene throughout evolution in populations evolved under selection and under neutrality. The median intrinsic noise parameter skews to lower values under stabilizing selection, but not under neutrality. **C, D** - Replicates of the simulations with the same input network and parameters. Replicates have different dynamics, but reach similar outcomes in terms of expression variance (**C**) and median intrinsic noise parameter (**D**) in the evolved populations. The evolution of each network topology sample was replicated 10 times under selection and 10 times under neutrality.

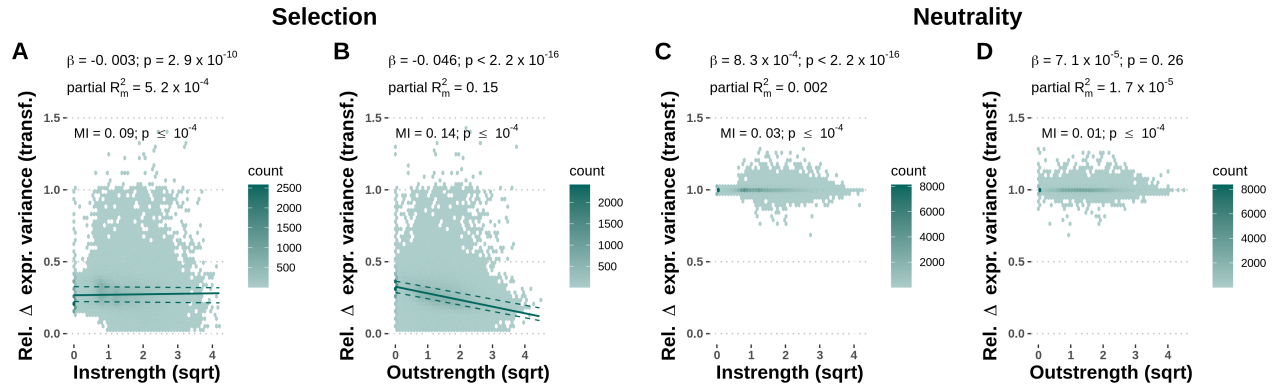
the mean expression level did not change during evolution and was highly correlated between the first and last generations (Pearson's  $r = 0.99$ ,  $p$ -value  $< 2.2 \times 10^{-16}$ , SI Section 1.4), confirming that selection acted only on the gene expression variance. Population size had a positive effect on the selective pressure acting on genes, as expected, selection being more efficient in large populations (SI Section 1.3, Fig S3). A population size of 1,000 individuals was chosen for the main simulations as the optimal population size in the trade-off between selecting mutations with small effects and reducing computational speed.

Next, we investigated how individual nodes within a network respond to selection, based on their centrality properties.

### Evolutionary change in phenotypes: regulators reduce their expression noise to a higher degree

We first analysed the phenotype change, *i.e.* the relative change in gene-specific expression variance after evolution. The variance of gene expression depends both on the intrinsic noise of the genes (that is, its genotype in our model) and the number and noise of the genes it is connected with.

We fitted linear models to assess the impact of the absolute instrength and outstrength measures on the relative change in expression variance for each node in each network. Under selection, both absolute instrength and absolute outstrength had a significant negative effect (linear mixed-effects model with coefficients  $\beta_{\text{instrength}} = -0.003$ ,  $p$ -value  $= 2.9 \times 10^{-10}$ , Fig 4A;  $\beta_{\text{outstrength}} = -0.046$ ,  $p$ -value  $< 2.2 \times 10^{-16}$ , Fig 4B), meaning that genes with more and stronger connections reduced their expression variance to a larger extent than less connected genes. The effect was notably



**Fig 4. Node-level network centrality measures affect the relative change of gene-specific expression variance under network-level selection.** For each gene, the relative change of expression variance before and after evolution (Rel.  $\Delta$  expr. variance) was averaged over all replicates. **A, B** - Absolute instrength (A) and absolute outstrength (B) have a significant negative effect on the relative change in gene expression variance in populations evolved under selection. A lower value of relative change of expression variance indicates a bigger reduction in expression variance between the first and last generation and a stronger response to selection. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **C, D** - Absolute instrength (C) and absolute outstrength (D) have a significant, but negligible, negative effect on the relative change in gene expression variance in the populations evolved under neutrality. The dataset consists of 2,000 populations with unique 40-gene random network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. Coefficients, p-values and partial marginal  $R^2$  measures were estimated using linear mixed-effects models with relative change of gene-specific variance as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were computed using 10,000 permutations.

stronger for outstrength (marginal  $R^2 = 0.15$ ) than for instrength (marginal  $R^2 = 5.2 \times 10^{-4}$ ). 328  
 Similarly, the mutual information was significant between the relative change in gene expression 329  
 variance under selection and absolute instrength (MI = 0.09, p-value  $\leq 10^{-4}$ , permutation test) and 330  
 absolute outstrength (MI = 0.14, p-value  $\leq 10^{-4}$ , permutation test). Genes with high outstrength 331  
 are strong regulators and their reduction of expression variance to a larger extent indicates that 332  
 high expression noise is more detrimental in regulators than in regulated genes. Under neutrality, 333  
 absolute instrength had a significantly positive effect (linear mixed-effects model with coefficient  $\beta$  334  
 $= 8.3 \times 10^{-4}$ , p-value  $< 2.2 \times 10^{-16}$ , Fig 4C) and absolute outstrength did not have a significant 335  
 effect on the relative change in gene expression variance (linear mixed-effects model with coefficient 336  
 $\beta = 7.1 \times 10^{-5}$ , p-value = 0.26, Fig 4D). The mutual information was significant between the relative 337  
 change in gene expression variance under neutrality and absolute instrength (MI = 0.03, p-value 338  
 $\leq 10^{-4}$ , permutation test) and absolute outstrength (MI = 0.01, p-value  $\leq 10^{-4}$ , permutation 339  
 test). These effects are much smaller and of opposite direction than the ones measured in selective 340  
 conditions, indicating that genetic drift did not cause the effect of centrality measures on expression 341  
 variance observed in selected populations. 342

343

## Evolutionary change in genotypes: regulators are more likely to respond – and display a stronger response – to selection

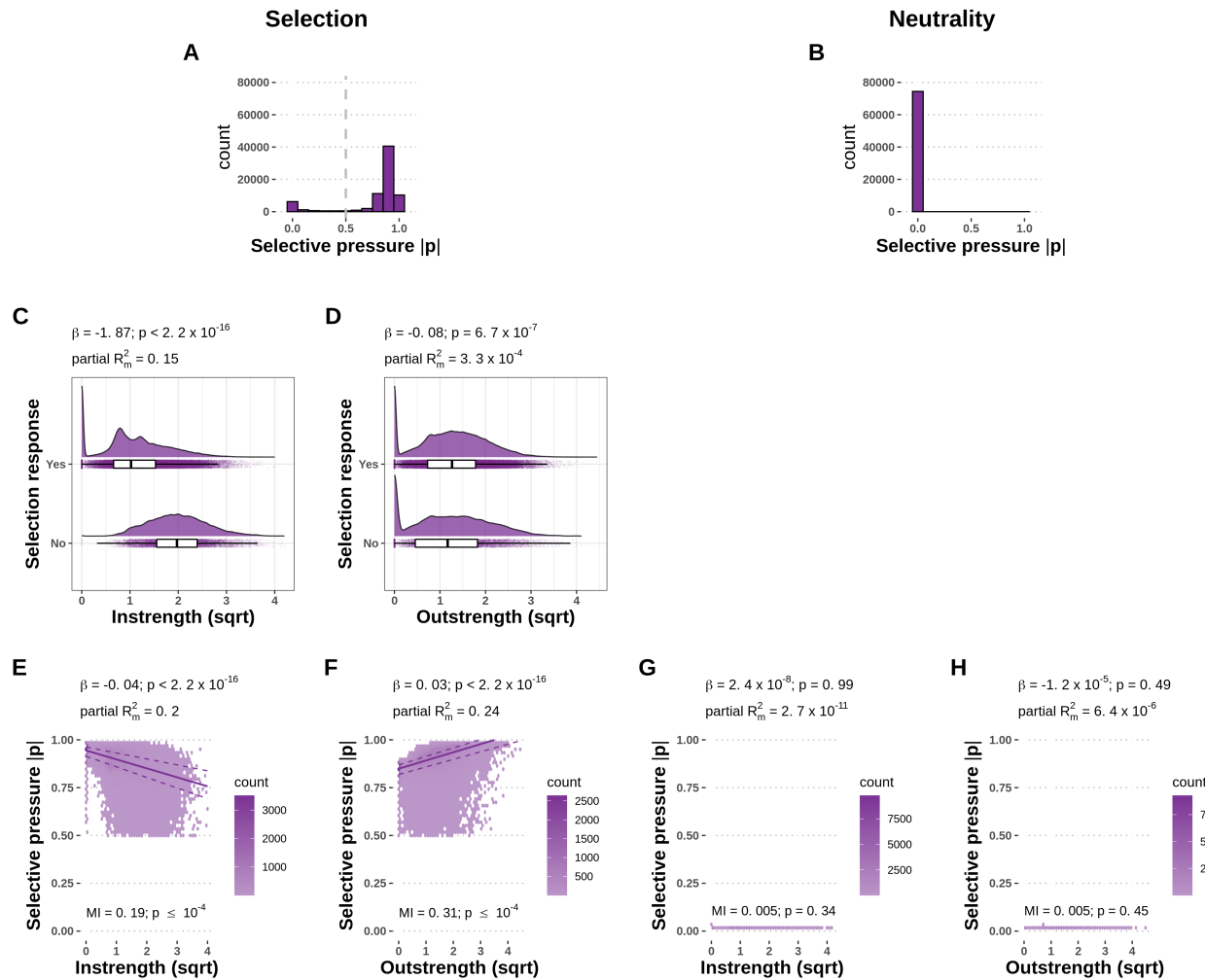
To investigate differential selective pressure acting on gene-specific expression noise, we analysed the change of intrinsic noise parameters in populations of gene regulatory networks evolved with or without stabilizing selection on the expression level. We measured the selective pressure acting on individual genes as the average reduction in the intrinsic noise parameter relative to the beginning of the evolutionary simulation (see Methods). The selective pressure on genes was found to be close to 0 in neutrally evolving populations, as expected (Fig 5B). In the presence of selection, however, the distribution of selective pressures was found to be bimodal (Fig 5A). Therefore, we binned genes in two categories according to whether they responded to selection (selective pressure  $> 0.5$ ) or not (selective pressure  $\leq 0.5$ ). We then separately analysed the probability to respond to selection and the strength of the response.

Absolute instrength had a significant and strongly negative effect (logistic regression with coefficient  $\beta = -1.87$ , p-value  $< 2.2 \times 10^{-16}$ , Fig 5C) on the probability of a gene to respond to selection, that is, genes with more and stronger incoming links are less likely to respond to selection. Absolute outstrength also had a significant effect on the probability of a gene to respond to selection (logistic regression with coefficient  $\beta = -0.08$ , p-value  $= 6.7 \times 10^{-7}$ , Fig 5D). However, this effect was small and was lost when the interaction terms between instrength and outstrength were included in the model (SI). For a qualitative analysis of the effect of network centrality on the selective pressure acting on individual genes, we fitted linear-mixed effects models on the set of genes that responded to selection, with selective pressure as the response variable. In the genes that responded to selection from the selected populations, absolute instrength had a significant negative effect (linear mixed-effects model with coefficient  $\beta = -0.04$ , p-value  $< 2.2 \times 10^{-16}$ , Fig 5E). Conversely, absolute outstrength had a significant positive effect (linear mixed-effects model with coefficient  $\beta = 0.03$ , p-value  $< 2.2 \times 10^{-16}$ , Fig 5F) on the selective pressure. In the selected populations, the mutual information was significant between the selective pressure and absolute instrength (MI = 0.19, p-value  $\leq 10^{-4}$ , permutation test) and absolute outstrength (MI = 0.31, p-value  $\leq 10^{-4}$ , permutation test).

The increased selective pressure in genes with high outstrength (strong regulators) can be explained by noise propagation to downstream elements. Namely, expression noise in regulators propagates to the genes they regulate, increasing the overall expression noise in the gene regulatory network. If gene expression levels in the network are under stabilizing selection, expression noise is deleterious. Therefore, regulator genes experience a comparatively higher selective pressure to reduce expression noise than regulated genes. In the neutral populations, neither absolute instrength nor absolute outstrength had a significant effect (linear mixed-effects model with coefficient  $\beta_{\text{instrength}} = 2.4 \times 10^{-8}$ , p-value = 0.99, Fig 5G;  $\beta_{\text{outstrength}} = -1.2 \times 10^{-5}$ , p-value = 0.49, Fig 5H) on the selective pressure. Similarly, the mutual information was not significant between the selective pressure and absolute instrength (MI = 0.005, p-value = 0.34, permutation test), nor absolute outstrength (MI = 0.005, p-value = 0.45, permutation test).

In contrast to regulator genes, we found that regulated genes, *i.e.* genes with high node instrength, are less likely to respond to selection and the selective pressure decreases with node instrength. Since the expression noise of genes is a sum of their intrinsic noise and noise propagated from upstream elements, the contribution of intrinsic noise to the total noise of the gene will be comparatively smaller in strongly regulated genes. The network can thus respond to selection either by reducing the intrinsic noise of the focal gene, or by reducing the intrinsic noise of any of the upstream elements, which would reduce propagated noise. As a result, there is a relaxation of selective pressure in





**Fig 5. Differential selective pressure is acting on genes based on their centrality. A, B** - Distributions of the measured selective pressure in selected (A) and neutral (B) populations. Genes with a selective pressure above 0.5 were categorized as responsive to selection. **C, D** - High instrength genes are less likely to respond to selection. Absolute instrength (C) has a strongly significant negative effect on the probability of selection response. Absolute outstrength (D) has a weak significant negative effect on the probability of selection response. **E, F** - In the subset of genes that responded to selection, high instrength (E) decreases the selective pressure, while high outstrength (F) increases the selective pressure acting on individual genes. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. **G, H** - Absolute instrength (G) and outstrength (H) have no significant effect on the selective pressure in the non-selected populations. The dataset consists of 2,000 populations with unique 40-gene random network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene is calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized as the mean over all replicates in each scenario. Coefficients, p-values and partial marginal  $R^2$  measures are estimated using logistic regression and linear mixed-effects models with selection responsiveness or selective pressure as the response variable, instrength and outstrength as fixed effect explanatory variables, and the network topology sample as the random effect explanatory variable. Mutual information (MI) p-values were using 10,000 permutations.



regulated genes, which is distributed on upstream genes. On the other hand, the same mechanism increases the selective pressure on upstream genes, *i.e.* regulators.

To check the robustness of our results, we performed the node-level network centrality analysis on two additional datasets with different topology structures: scale-free (Barabási–Albert) and small-world (Watts–Strogatz) topology models. We find consistent effects (direction and significance) of local network centrality metrics on the selective pressure acting on gene-specific noise across topology models, showing that our findings are robust to the topology model used (SI Section 4). However, the effect size of network centrality metrics differed between the topology models, pointing at an effect of the topology model on noise propagation and the evolution of gene-specific expression noise, which we investigate in the next section.

## Global network properties affect the evolvability of expression noise and selective pressure on constituent genes

Lastly, we analysed how topological structures and graph-level network properties affect the expression noise response of constituent genes to selection on a joint dataset of random (Erdős–Rényi), scale-free (Barabási–Albert) and small-world (Watts–Strogatz) network topologies. Jointly analysing genes from all three topology types with linear models, we observed statistically significant interactions between instrength and outstrength and network topology types on both the probability to respond to selection and the selective pressure acting on gene-specific expression noise (Table 1). We found that genes in scale-free networks have a significantly higher probability of responding to selection than genes in random networks. These results are in agreement with previous studies reporting a higher evolvability of scale-free networks (31; 32). Conversely, genes in small-world networks have a significantly lower probability of responding to selection than genes in random networks. Furthermore, there are significant effects of interactions between instrength and outstrength with the topology type on the selective pressure on constituent genes.

To investigate which global topological features of the three network models affect expression noise evolution, we performed a principal component analysis (PCA) on 12 graph-level measures. The first two dimensions of the PCA expressed 85.4% of the total dataset inertia (Fig S8A), so we used the first two principal components (PCs) as synthetic explanatory variables in linear mixed-effects models. The loading of the first synthetic variable (PC1) is dominated by negative loadings of diameter and mean path distance, and the centralization measures, namely positive loadings of outdegree and closeness centralization and negative loadings of indegree and betweenness centralization. The diameter of a network is defined as the longest shortest path between any two nodes. Centralization is a measure of the extent to which a network is centered around a single node and can be computed from different centrality metrics. The loading of the second synthetic variable (PC2) is dominated by the negative loading of the average degree, average indegree and average outdegree measures (Fig S8B). For a more intuitive interpretation, the signs of both PCs have been switched in the statistical analysis. Therefore, PC1 shown in the results is dominated by positive loadings of diameter, mean path distance, indegree centralization and negative loadings of outdegree centralization, and PC2 is dominated by positive loadings of average degree. We refer to PC1 and PC2 as synthetic network diameter and centralization and synthetic average degree, respectively.

The average expression variance per network is significantly negatively affected by synthetic network diameter and centralization (linear model with synthetic network diameter and centralization coefficient  $\beta = -6.19$ , p-value  $< 2.2 \times 10^{-16}$ ) and significantly positively affected by the synthetic average degree (linear model with synthetic average degree coefficient  $\beta = 13.26$ , p-value  $< 2.2 \times 10^{-16}$ ). The mutual information was significant between the average expression variance per network and

synthetic network diameter and centralization (MI = 0.21, p-value  $\leq 10^{-4}$ , permutation test) and synthetic average degree (MI = 0.21, p-value  $\leq 10^{-4}$ , permutation test). This finding means that global network properties affect the amplification of noise through noise propagation between the genes. Specifically, networks with a lower diameter, mean path distance, indegree centralization, and higher outdegree centralization and average degree, had higher average gene expression variance. In the selected populations, the average selective pressure per network was significantly negatively affected by both synthetic network diameter and centralization and the synthetic average degree (linear model with synthetic network diameter and centralization coefficient  $\beta = -0.003$ , p-value =  $4.9 \times 10^{-11}$ , Fig 6A; synthetic average degree coefficient  $\beta = -0.009$ , p-value  $< 2.2 \times 10^{-16}$ , Fig 6B). The mutual information was significant between the average selective pressure per network and synthetic network diameter and centralization (MI = 0.27, p-value  $\leq 10^{-4}$ , permutation test) and synthetic average degree (MI = 0.26, p-value  $\leq 10^{-4}$ , permutation test). This result shows that the average selective pressure acting on gene-specific expression noise in networks decreases with an increase of network diameter, mean path distance, indegree centralization and average degree

**Table 1. Network topology type affects the probability of responding to selection and selective pressure on gene-specific expression noise under stabilizing selection on gene expression level.**

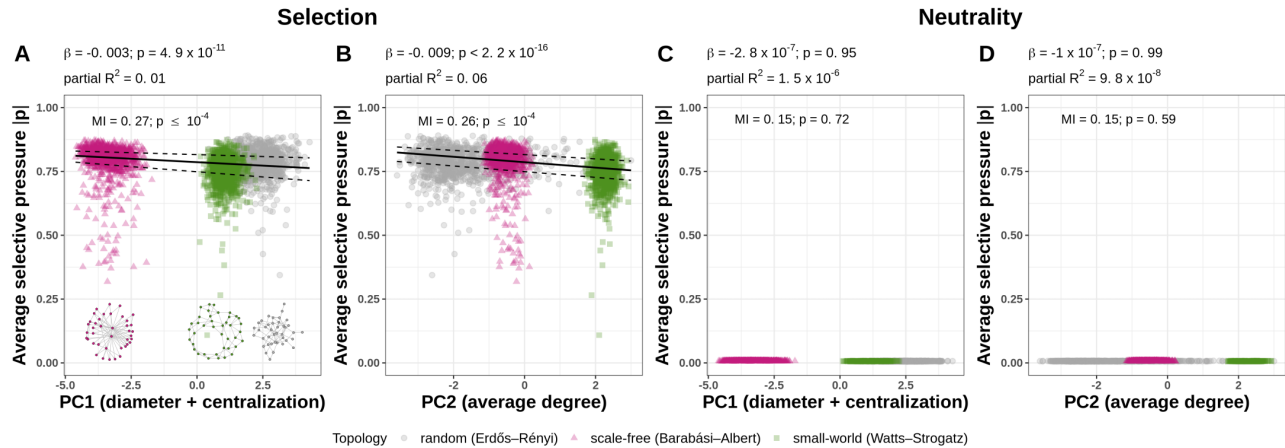
Response	Explanatory variable	Beta	SE	p-value <sup>1</sup>	
Probability of responding to selection	Instrength	-1.9270	0.0284	$< 2.2 \times 10^{-16}$	****
	Outstrength	-0.0829	0.0226	$< 2.6 \times 10^{-4}$	***
	Scale-free (BA) topology <sup>2</sup>	0.9209	0.1075	$< 2.2 \times 10^{-16}$	****
	Small-world (WS) topology <sup>3</sup>	-0.2684	0.0945	0.0045	**
	Instrength:BA <sup>4</sup>	0.0120	0.0516	0.8159	n.s.
	Instrength:WS	0.0006	0.0401	0.9873	n.s.
	Outstrength:BA	-0.2947	0.0252	$< 2.2 \times 10^{-16}$	****
	Outstrength:WS	-0.0728	0.0333	0.0287	*
Gene-specific selective pressure	Instrength	-0.0377	0.0004	$< 2.2 \times 10^{-16}$	****
	Outstrength	0.0347	0.0003	$< 2.2 \times 10^{-16}$	****
	Scale-free (BA) topology	0.0019	0.0012	0.1404	n.s.
	Small-world (WS) topology	0.0222	0.0013	$< 2.2 \times 10^{-16}$	****
	Instrength:BA	0.0143	0.0007	$< 2.2 \times 10^{-16}$	****
	Instrength:WS	-0.0055	0.0006	$< 2.2 \times 10^{-16}$	****
	Outstrength:BA	-0.0151	0.0003	$< 2.2 \times 10^{-16}$	****
	Outstrength:WS	-0.0075	0.0005	$< 2.2 \times 10^{-16}$	****

<sup>1</sup> Coefficients and their significance were computed using linear mixed-effects models (see Methods). The dataset consisted of 3,000 populations with unique 40-gene random, scale-free and small-world network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene was calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized as the mean over all replicates in each scenario. Genes were termed responsive to selection if their selective pressure was above 0.5. Asterisks indicate statistical significance: n.s. - p-value  $> 0.05$ ; \* - p-value  $\leq 0.05$ ; \*\* - p-value  $\leq 0.01$ ; \*\*\* - p-value  $\leq 0.001$ ; \*\*\*\* - p-value  $\leq 0.0001$ .

<sup>2</sup> Barabási–Albert network model.

<sup>3</sup> Watts–Strogatz network model.

<sup>4</sup> Colons (':') indicate variable interactions.



**Fig 6. Global network properties affect the average selective pressure acting on gene expression noise under stabilizing selection on gene expression level.** A, B - Principal component variables consisting of the diameter and network centralization (A) and average degree (B) have a significant negative effect on the average selective pressure per network. The two synthetic variables were constructed by performing a principal component analysis on 12 graph-level network metrics. The lines indicate the 25% (lower dashed line), 50% (solid line), and 75% (upper dashed line) fitted quantiles. The dataset consisted of 3,000 populations with unique 40-gene random, scale-free and small-world network topology samples, which were independently evolved 10 times under selection and 10 times under neutrality. The selective pressure on each gene is calculated as the average normalized reduction of the intrinsic noise parameter during the evolutionary simulation and summarized over all replicates in each scenario. Coefficients and p-values are estimated using a linear model with average selective pressure as the response variable, and PC1 and PC2 as explanatory variables. Mutual information (MI) p-values were computed with permutation test with 10,000 permutations.

per network. Conversely, the average selective pressure increases with an increase of outdegree centralization (Fig 6A-B). In the populations evolved under neutrality, neither synthetic network diameter and centralization, nor synthetic average degree, had a significant effect on the average selective pressure per network (linear model with synthetic network diameter and centralization coefficient  $\beta = -2.8 \times 10^{-7}$ , p-value = 0.95; synthetic average degree coefficient  $\beta = -1 \times 10^{-7}$ , p-value = 0.99, Fig 6C-D). Similarly, the mutual information was insignificant between the average selective pressure per network and synthetic network diameter and centralization (MI = 0.15, p-value = 0.72, permutation test) and synthetic average degree (MI = 0.15, p-value = 0.59, permutation test).

## Discussion

To study the evolution of gene expression noise, we developed a new gene regulatory network model that incorporates stochastic gene expression, where the gene expression mean and variance are both heritable and, therefore, potentially subject to natural selection. We simulated the evolution of gene-specific expression noise in populations of model gene regulatory networks under selective and non-selective conditions. Under a stabilizing selection regime around optimal gene expression levels, we observed the evolution of noise reduction. However, individual genes responded differently to

this global selective pressure. First, only some genes reduced their intrinsic noise, and were marked as “responding to selection”. Second, these genes differed in the amount by which they reduced their intrinsic noise. We unravelled the factors that determine both the probability to respond to selection and the strength of the response using several complementary statistical approaches. We used generalized linear models with mixed effects (GLMM), allowing us to jointly test multiple putative factors. GLMMs, however, make several assumptions that might be violated by the data in some cases. To further validate our conclusions, we computed the mutual information (MI) between variables. The results of the MI analyses were consistent with those of the GLMM, demonstrating the robustness of our conclusions.

We found that genes of high centrality exhibit a stronger selective pressure to reduce gene-specific expression noise under stabilizing selection on the expression level and the genetic network structure affects the propagation and evolvability of gene-specific expression noise. In the following, we further discuss the implications of differential selective pressure acting on constituent genes in gene networks.

## Differential selective pressure on genes in gene networks

We found that central genes in gene regulatory networks had a comparatively higher impact on the network fitness and, therefore, gene expression noise in central genes was more detrimental than in peripheral genes. Distinguishing between node input (summarized by the node instrength) and output (summarized by the node outstrength) further allowed us to dissect the selective pressure acting on individual genes. Regulator genes, *i.e.* genes with high node outstrength, were more likely to respond and had a stronger response to selection than non-regulator genes. Regulator genes have been experimentally shown to propagate their expression noise to downstream elements (14), a property well captured by our model. This imposes a stronger selective pressure on them under stabilizing selection on gene expression acting at the network-level. In a genome-wide expression noise screen in *Drosophila melanogaster*, transcription factors were found to have lower expression variation (33). Suppression of expression noise can be attained through negative autoregulation (34; 35; 36), whereby a regulator acts as its own repressor. Incidentally, 40% of transcription factors in *E. coli* (37) and many eukaryotic transcription factors (38) have negative autoregulation, indicating a wide-spread control of expression noise in natural regulatory networks.

## Mechanisms of intrinsic noise reduction

In this study we abstracted and summarized the many determinants of intrinsic expression noise into a single parameter, which can be viewed as a modifier locus that can directly change the intrinsic noise of a given gene. This simplification permitted us to investigate the evolution of expression noise in gene networks with computationally feasible evolutionary simulations. In reality, multiple factors that affect gene expression variance in biological systems have been reported. These include epigenetic factors, such as chromatic dynamics (39) and presence of chromatin remodelling complexes (40). Other factors affect transcription directly and can, therefore, control expression noise: the promoter shape (33), presence of a TATA box (40), presence and number (4) of TF binding sites, TF binding dynamics (41), presence of TF decoy binding sites (42), and transcription rate. Factors affecting translation have also been shown to play a role in controlling noise: miRNA targeting (43), mRNA lifetime, translation rate, and post-translational modifications such as the protein degradation rate. Compartmentalization of proteins by phase separation has also been shown to reduce noise (44). We have demonstrated the existence of a general selective pressure acting on gene expression noise. Biological organisms may differ in the mechanisms used to respond to this selective

pressure, calling for further, data-driven, investigations.

508

## Global network structure impacts noise propagation and evolution

509

As instrength and outstrength are local centrality measures, we further assessed the impact of the global network structure on noise evolution by conducting our analyses on three different network topology models. While we found consistent qualitative results with the three models, there were quantitative differences between the network models, which we assessed by measuring graph-level characteristics. We found that network characteristics such as diameter, mean path distance, centralization and average degree affect the propagation of noise within the regulatory network. Given that there is a trade-off between the fitness advantage of reducing gene-specific expression noise at the gene level and its mechanistic cost (for instance, in terms of mRNA processing (45)), evolving the global network structure may offer an alternative way to reduce network-level noise. Several motifs recurrently found in regulatory networks have an impact on expression noise, such as negative (34; 35; 36) and positive autoregulation (38), feed-forward loops (38) and interlinked feed-forward loops (46).

510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521

It is important, however, to distinguish two aspects when considering the effect of the network structure on the expression dynamics of constituent genes: the network structure, *i.e.* the topology of the graph, and the strength of each of the regulatory interactions, both of which impact expression noise. The same network topology, but with different regulatory interactions strengths, can give rise to markedly different network behaviours. In the *gap* gene system, for example, it was shown that multiple subcircuits share the same regulatory structure, but yield different expression patterns because of their differences in active components and strength of regulatory interactions (47). It results that network models of gene expression noise must incorporate both graph topology and interaction strength between all constituent genes. The Wagner model constitutes a simple framework that fulfills these two conditions. However, it has its limitations. Namely, it is not fine-grained enough to capture the complex dynamics of real regulatory networks. Models that incorporate higher molecular detail, such as large systems of differential equations, are necessary to precisely capture in fine detail the expression dynamics of a real biological network, but they come with a cost in terms of high computation time (preventing their use in evolutionary simulations), low tractability and, often, the inability to model noise.

522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536

## Implications of selection on expression noise on the evolution of genomes and gene regulatory networks

537  
538

One mechanism by which networks and genomes evolve is gene duplication. Gene duplications are a major source of new genes and thought to be a primary source of evolutionary novelties. It has been long proposed that new functionality arises from duplicated genes by allowing the other gene copy to acquire new functions (neofunctionalization) or improve existing functions (subfunctionalization) by relaxing the selective pressure acting on a single gene through an additional redundant copy (48). However, most of the time the redundant copy is lost before new functionality can arise (49), either by genetic drift alone or because having the extra copy is deleterious. The redundant copy has a chance to evolve a new function or improve an existing one while it is evolving neutrally or reaches fixation in the population, or alternatively, if there is some fitness benefit of the additional copy that increases its frequency in the population. Some benefits of having additional gene copies have been shown, such as increased expression level for genes whose pre-duplication expression level was far from the optimum (50). Moreover, duplicating a gene reduces its expression noise (51; 52),

539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550



averaging the stochastic events over the two gene copies. The reduction of expression noise may, therefore, constitute another benefit of a gene duplication, increasing its chance of fixation in the population. As the gene number increases in bacterial genomes, the number of regulatory genes increases 4-fold (53), indicating a gene duplication is more likely to stay if the gene is a regulatory gene. We hypothesize that selection on expression noise, particularly on regulatory genes, could, therefore, be one of the forces driving the maintenance of duplicated genes.

## Applications of the model framework to study complex systems

In this study, we developed a new regulatory and evolutionary model to study expression noise in gene regulatory networks. The model represents key features of evolving gene regulatory networks, namely the non-independence of gene expression levels and fitness determined by the expression level of many or all genes in the network. Our results revealed that differential selective pressure act on intrinsic expression noise of constituent genes and that network-level topological properties affect noise propagation within the network.

Although our study focused on gene regulatory networks, our conclusions potentially apply to a broader range of systems. In particular, we posit that any system that fulfills two essential properties will exhibit a similar behavior: (i) the amount of product of each system components (here called “expression level”) are not independent and (ii) the performance (here termed “fitness”) is determined by the product level of one or several of the components of the system. There are many other complex systems that fulfill these criteria, such as biological metabolic networks, ecological food webs, neural networks, economies, transportation and other infrastructure networks, and social networks. We expect that the same constraints act on noise in elements of these systems, and that some of the conclusions on gene regulatory networks could be carefully applied to other complex systems.

## Conclusion

By modeling the interactions of genes in a regulatory network, we showed that the selective pressure acting on a system (gene network) resulted in differential pressures on individual components (genes). With this, we demonstrated that the position of the gene in the network and global network structure impact the evolution of intrinsic expression noise. Investigating how gene networks evolve to cope with expression noise will reveal mechanisms of how complex biological systems adapt to function with an inevitable molecular noise in their components. A better comprehension of these mechanisms is a prerequisite to understand the evolution of complexity in biological systems, from the first self-replicating RNA systems to modern eukaryotic cells expressing tens of thousands of genes.

## Supporting information

All additional analyses and their results, used parameters, their descriptions and biological meaning, and other information can be found as the supplementary PDF file. The source code for the gene network model and evolutionary simulations is available at [https://gitlab.gwdg.de/molsysevol/supplementarydata\\_expressionnoise/cpp](https://gitlab.gwdg.de/molsysevol/supplementarydata_expressionnoise/cpp). The R Notebook files and other scripts necessary to reproduce the complete analysis are available at [https://gitlab.gwdg.de/molsysevol/supplementarydata\\_expressionnoise/scripts](https://gitlab.gwdg.de/molsysevol/supplementarydata_expressionnoise/scripts). The simulation data necessary to reproduce all figures is available at <https://zenodo.org/X>.



<b>S1 Table.</b>	<b>Parameters used in the simulations.</b>	591
<b>S1 Fig.</b>	<b>Network realization is robust to synchronous or asynchronous expression level updating mode during network realization.</b>	592 593
<b>S2 Fig.</b>	<b>Examples of expression level dynamics in realizations of the same network with different expression level updating modes and noise levels.</b>	594 595
<b>S3 Fig.</b>	<b>Increasing the population size increases the selective pressure on genes under stabilizing selection on gene expression level.</b>	596 597
<b>S4 Fig.</b>	<b>Mean expression level does not change after noise evolution under stabilizing selection on gene expression levels.</b>	598 599
<b>S5 Fig.</b>	<b>Correlation matrix of node-level network centrality metrics and expression noise metrics.</b>	600 601
<b>S6 Fig.</b>	<b>Correlation matrix of graph-level network centrality metrics and expression noise metrics.</b>	602 603
<b>S7 Fig.</b>	<b>Correlations between node instrength and outstrength in unfiltered and filtered datasets.</b>	604 605
<b>S8 Fig.</b>	<b>Principal component analysis of the graph-level network centrality metrics</b>	606
<b>S9 Fig.</b>	<b>Diagnostics of linear mixed-effects models with expression variance as a response variable and different variance structures.</b>	607 608
<b>S2 Table.</b>	<b>Different variance structures do not affect the sign of the effect and significance in linear mixed-effects models with expression variance as a response variable.</b>	609 610
<b>S10 Fig.</b>	<b>Diagnostics of linear mixed-effects models with relative change of expression variance after selection as a response variable and different variance structures.</b>	611 612
<b>S3 Table.</b>	<b>Different variance structures do not affect the sign of the effect and significance in linear mixed-effects models with relative change of expression variance after selection as a response variable.</b>	613 614 615
<b>S11 Fig.</b>	<b>Diagnostics of linear mixed-effects models with selective pressure as a response variable and different variance structures.</b>	616 617
<b>S4 Table.</b>	<b>Different variance structures do not affect the sign of the effect and significance in linear mixed-effects models with selective pressure as a response variable.</b>	618 619

**S12 Fig. Diagnostics of linear model with average selective pressure per network as a response variable.** 620  
621

**S5 Table. The effects and significance of local network centrality metrics are consistent across different topological structures.** 622  
623

**S6 Table. Filtered and unfiltered datasets.** 624

## Acknowledgments

 625

This work is funded by the International Max Planck Research School for Evolutionary Biology and the Max Planck Society. The authors would like to thank Arnaud Le Rouzic for the beneficial discussions, Andreas Wagner and Gabriel Schweizer for the valuable input during the model development, Arne Traulsen and Tal Dagan for the helpful suggestions throughout the project, Nikhil Sharma for his study on additional aspects of the model, and Andrea Bours, Artemis Efstratiou and Carolina Peralta for the careful reading of the manuscript. 626  
627  
628  
629  
630  
631

## References

1. Elowitz MB. Stochastic Gene Expression in a Single Cell;297(5584):1183–1186. doi:10.1126/science.1070919.
2. Raser JM, O’Shea EK. Noise in Gene Expression: Origins, Consequences, and Control;309(5743):2010–2013. doi:10.1126/science.1105891.
3. Chalancon G, Ravarani CNJ, Balaji S, Martinez-Arias A, Aravind L, Jothi R, et al. Interplay between gene expression noise and regulatory network architecture;28(5):221–232. doi:10.1016/j.tig.2012.01.006.
4. Sharon E, van Dijk D, Kalma Y, Keren L, Manor O, Yakhini Z, et al. Probing the effect of promoters on noise in gene expression using thousands of designed sequences;24(10):1698–1706. doi:10.1101/gr.168773.113.
5. Lehner B. Selection to minimise noise in living systems and its implications for the evolution of gene expression;4(1):170. doi:10.1038/msb.2008.11.
6. Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise Minimization in Eukaryotic Gene Expression;2(6):e137. doi:10.1371/journal.pbio.0020137.
7. Wang Z, Zhang J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection;108(16):E67–E76. doi:10.1073/pnas.1100059108.
8. Barroso GV, Puzovic N, Dutheil JY. The Evolution of Gene-Specific Transcriptional Noise Is Driven by Selection at the Pathway Level;208(1):173–189. doi:10.1534/genetics.117.300467.
9. Duveau F, Hodgins-Davis A, Metzger BP, Yang B, Tryban S, Walker EA, et al. Fitness effects of altering gene expression noise in *Saccharomyces cerevisiae*;7:e37272. doi:10.7554/eLife.37272.

10. Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. Experimental evolution of bet hedging;462(7269):90–93. doi:10.1038/nature08504.
11. Schmedel JM, Carey LB, Lehner B. Empirical mean-noise fitness landscapes reveal the fitness impact of gene expression noise;10(1):3180. doi:10.1038/s41467-019-11116-w.
12. Gilad Y, Oshlack A, Rifkin SA. Natural selection on gene expression;22(8):456–461. doi:10.1016/j.tig.2006.06.002.
13. Vlková M, Silander OK. Gene regulation in *Escherichia coli* is commonly selected for both high plasticity and low noise; p. 1–15. doi:10.1038/s41559-022-01783-2.
14. Pedraza JM. Noise Propagation in Gene Networks;307(5717):1965–1969. doi:10.1126/science.1109090.
15. Blake WJ, Kærn M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression;422(6932):633–637. doi:10.1038/nature01546.
16. Hens C, Harush U, Haber S, Cohen R, Barzel B. Spatiotemporal signal propagation in complex networks;15(4):403–412. doi:10.1038/s41567-018-0409-0.
17. Wagner A. Does Evolutionary Plasticity Evolve?;50(3):1008–1023. doi:10.1111/j.1558-5646.1996.tb02342.x.
18. Newman M. Networks: An Introduction. Oxford University Press;. Available from: <https://oxford.universitypressscholarship.com/10.1093/acprof:oso/9780199206650.001.0001/acprof-9780199206650>.
19. Laarits T, Bordalo P, Lemos B. Genes under weaker stabilizing selection increase network evolvability and rapid regulatory adaptation to an environmental shift;29(8):1602–1616. doi:10.1111/jeb.12897.
20. Team RC. R: A Language and Environment for Statistical Computing;. Available from: <https://www.R-project.org/>.
21. Csardi G, Nepusz T. The igraph software package for complex network research;1695(5):1–9.
22. Hunter DR, Handcock MS, Butts CT, Goodreau SM, Morris M. ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks;24(3):nihpa54860.
23. Dray S, Dufour AB. The ade4 Package: Implementing the Duality Diagram for Ecologists;22:1–20. doi:10.18637/jss.v022.i04.
24. Pinheiro J, Bates D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models;. Available from: <https://CRAN.R-project.org/package=nlme>.
25. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4;67:1–48. doi:10.18637/jss.v067.i01.
26. Bartoń K. MuMIn: Multi-Model Inference;. Available from: <https://CRAN.R-project.org/package=MumIn>.

27. Fox J, Weisberg S. An R Companion to Applied Regression. 3rd ed. Sage;. Available from: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
28. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning: with Applications in R. Springer Texts in Statistics. Springer US;. Available from: <https://link.springer.com/10.1007/978-1-0716-1418-1>.
29. Meyer PE. infotheo: Information-Theoretic Measures;. Available from: <https://cran.r-project.org/package=infotheo>.
30. Urchueguía A, Galbusera L, Chauvin D, Bellement G, Julou T, Nimwegen Ev. Genome-wide gene expression noise in Escherichia coli is condition-dependent and determined by propagation of noise through the regulatory network;19(12):e3001491. doi:10.1371/journal.pbio.3001491.
31. Oikonomou P, Cluzel P. Effects of topology on network evolution;2(8):532–536. doi:10.1038/nphys359.
32. Greenbury SF, Johnston IG, Smith MA, Doye JPK, Louis AA. The effect of scale-free topology on the robustness and evolvability of genetic regulatory networks;267(1):48–61. doi:10.1016/j.jtbi.2010.08.006.
33. Sigalova OM, Shaeiri A, Forneris M, Furlong EE, Zaugg JB. Predictive features of gene expression variation reveal mechanistic link with differential expression;16(8). doi:10.15252/msb.20209539.
34. Becskei A, Serrano L. Engineering stability in gene networks by autoregulation;405(6786):590–593. doi:10.1038/35014651.
35. Dublanche Y, Michalodimitrakis K, Kümmerer N, Foglierini M, Serrano L. Noise in transcription negative feedback loops: simulation and experimental analysis;2(1):41. doi:10.1038/msb4100081.
36. Grönlund A, Lötstedt P, Elf J. Transcription factor binding kinetics constrain noise suppression via negative feedback;4(1):1864. doi:10.1038/ncomms2867.
37. Rosenfeld N, Elowitz MB, Alon U. Negative Autoregulation Speeds the Response Times of Transcription Networks;323(5):785–793. doi:10.1016/S0022-2836(02)00994-4.
38. Alon U. Network motifs: theory and experimental approaches;8(6):450–461. doi:10.1038/nrg2102.
39. Sun M, Zhang J. Chromosome-wide co-fluctuation of stochastic gene expression in mammalian cells;15(9):e1008389. doi:10.1371/journal.pgen.1008389.
40. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, et al. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise;441(7095):840–846. doi:10.1038/nature04785.
41. Azpeitia E, Wagner A. Short Residence Times of DNA-Bound Transcription Factors Can Reduce Gene Expression Noise and Increase the Transmission of Information in a Gene Regulation System;7:67. doi:10.3389/fmolb.2020.00067.

42. Dey S, Soltani M, Singh A. Enhancement of gene expression noise from transcription factor binding to genomic decoy sites;10(1):9126. doi:10.1038/s41598-020-65750-2.
43. Schmiedel JM, Klemm SL, Zheng Y, Sahay A, Blüthgen N, Marks DS, et al. MicroRNA control of protein expression noise;348(6230):128–132. doi:10.1126/science.aaa1738.
44. Klosin A, Oltsch F, Harmon T, Honigsmann A, Jülicher F, Hyman AA, et al. Phase separation provides a mechanism to reduce noise in cells; p. 6.
45. Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression;10(1):68. doi:10.1038/s41467-018-07391-8.
46. Chepyala SR, Chen YC, Yan CCS, Lu CYD, Wu YC, Hsu CP. Noise propagation with interlinked feed-forward pathways;6(1):23607. doi:10.1038/srep23607.
47. Verd B, Monk NA, Jaeger J. Modularity, criticality, and evolvability of a developmental gene regulatory network;8:e42832. doi:10.7554/eLife.42832.
48. Ohno S. Evolution by Gene Duplication. Springer Berlin, Heidelberg;.
49. Lynch M, Conery JS. The Evolutionary Fate and Consequences of Duplicate Genes;290(5494):1151–1155. doi:10.1126/science.290.5494.1151.
50. Riehle MM, Bennett AF, Long AD. Genetic architecture of thermal adaptation in *Escherichia coli*;98(2):525–530. doi:10.1073/pnas.98.2.525.
51. Rodrigo G, Fares MA. Intrinsic adaptive value and early fate of gene duplication revealed by a bottom-up approach;7:e29739. doi:10.7554/eLife.29739.
52. Chapal M, Mintzer S, Brodsky S, Carmi M, Barkai N. Resolving noise–control conflict by gene duplication;17(11):e3000289. doi:10.1371/journal.pbio.3000289.
53. Molina N, van Nimwegen E. The evolution of domain-content in bacterial genomes;3:51. doi:10.1186/1745-6150-3-51.