Nonlinear expression patterns and multiple shifts in gene network interactions underlie robust phenotypic change in Drosophila melanogaster selected for night sleep duration

Caetano Souto-Maior[1‡*], Yanzhu Lin[1§], Yazmin L. Serrano Negron[1], Susan T. Harbison[1*]

[1]Laboratory of Systems Genetics, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, MD, United States of America

***For correspondence:** caetanosoutomaior@proton.me; susan.harbison@nih.gov

Present address: [‡]Basque Center for Applied Mathematics BCAM, Bilbao, Bizkaia, Spain; [§]Eli Lilly and Co., Lilly Corporate Center, Indianapolis, IN, United States of America

Abstract All but the simplest phenotypes are believed to result from interactions between two or more genes forming complex networks of gene regulation. Sleep is a complex trait known to depend on the system of feedback loops of the circadian clock, and on many other genes; however, the main components regulating the phenotype and how they interact remain an unsolved puzzle. Genomic and transcriptomic data may well provide part of the answer, but a full account requires a suitable quantitative framework. Here we conducted an artificial selection experiment for sleep duration with RNA-seq data acquired each generation. The phenotypic results are robust across replicates and previous experiments, and the transcription data provides a high-resolution, time-course data set for the evolution of sleep-related gene expression. In addition to a Hierarchical Generalized Linear Model analysis of differential expression that accounts for experimental replicates we develop a flexible Gaussian Process model that estimates interactions between genes. 145 gene pairs are found to have interactions that are different from controls. Our method not only is considerably more specific than standard correlation metrics but also more sensitive, finding correlations not significant by other methods. Statistical predictions were compared to experimental data from public databases on gene interactions. Mutations of candidate genes implicated by our results affected night sleep, and gene expression profiles largely met predicted gene-gene interactions.

Introduction

Despite the plethora of modern and increasingly refined molecular biology assays – from DNA to metabolites and beyond – systematically uncovering the molecular bases of phenotypes remains one of the thorniest challenges in biology. "Omics" approaches allow whole genome, transcriptome, proteome, and other "omes" to be generated and candidate genes to be fished out of these high dimensional data, but understanding how these biomolecules interact even in the simplest pathways requires painstaking follow-on experimentation, construction of databases, and an im-

mense collective effort to make connections from disjointed assays into a coherent model. Despite the large amount of studies and data generated for many systems, identifying underlying processes is still very rare; this is clear indication that better methods are needed to obtain understanding of biological processes from data. For complex traits the task is even more difficult. Sleep is a complex phenotype the evolution of which remains a classic mystery in biology. Although sleep and sleep-like behavior is conserved among species, its main purpose is not completely understood, and hypotheses for its purpose span functions like conservation of resources (***Berger and Phillips, 1995***; ***Scharf et al., 2008***; ***Schmidt, 2014***), pruning of synapses and memory formation (***Krueger and Obál, 1993***; ***Tononi and Cirelli, 2014***; ***Joiner, 2016***; ***Ly et al., 2018***), and management of metabolite and waste products (***Xie et al., 2013***; ***Hill et al., 2020***). It is plausible that sleep is a manifestation of multiple functions, and that it involves the activity of many genes to regulate a complex higher-level function; indeed many genes have been implicated in sleep (***Harbison et al., 2017***, ***2013***; ***Laing et al., 2019***; ***Dashti et al., 2019***; ***Jones et al., 2016***; ***Jansen et al., 2019***; ***Lane et al., 2019***; ***Hammerschlag et al., 2017***; ***Diessler et al., 2018***; ***Joshi et al., 2019***; ***Boyle et al., 2017***). Assuming anything but the simplest possible model would therefore require a description that accounts for this complexity in the interactions of genes and gene products.

Artificial selection plus sequencing/resequencing is a powerful approach for identifying heritable variation in phenotypes and their underlying molecular bases (***Schlötterer et al., 2015***), typically assaying DNA or RNA expression in the initial and evolved populations and comparing them to controls (***Faria et al., 2015***, ***2016***). Coupling selection with gene expression identified candidate genes for diurnal preference (***Pegoraro et al., 2020***), olfactory behavior (***Brown et al., 2017***, ***2020***), food consumption (***Garlapow et al., 2017***), mating behavior (***Mackay et al., 2005***), resistance to parasitism (***Wertheim et al., 2011***), environmental stressors (***Telonis-Scott et al., 2009***; ***Sørensen et al., 2007***), ethanol tolerance (***Morozova et al., 2007***), and aggressive behavior (***Edwards et al., 2006***). Caveats of that method include often not having molecular data on the intermediate generations, and relying on traditional statistical methods to assess the significance of polymorphic variants. In the case of gene expression, RNA levels are often modeled for each gene individually using linear models, without further consideration of the processes involved or interactions between genes. Inferring interaction between genes (as opposed to individual changes) requires observations of how the genes covary in time. Correlation or information theory-based methods (and others, reviewed in ***Emmert-Streib et al.*** (***2012***); ***Villaverde and Banga*** (***2014***); ***Liu*** (***2015***)) could be applied to estimate the relationship between the genes when that information is present, but neither is time course data usually available, nor are these methods standard in artificial selection experiments.

In this work we have artificially selected *Drosophila melanogaster* for increased or decreased night sleep duration and sequenced the mRNA of the flies from each generation of selection. The selection procedure produced both long- and short-sleeping fly populations significantly deviant from unselected controls. The RNA sequence data, which consisted of expression levels as a function of time (measured in generations), was analyzed using a Multi-Channel Gaussian Process (***Melkumyan and Ramos, 2011***; ***Bonilla et al., 2008***) where each gene is described by one of these "channels", and their relationships are estimated by an underlying covariance structure in the model. We describe the expression of 85 genes that had significant changes in the artificial selection long or short schemes along generation common to both males and females. We used this model to infer the magnitude of all 3,570 possible pairwise interactions between all possible pairs of genes. Results from this analysis and comparison to unselected controls suggest that multiple shifts in interactions underlie the increase and decrease of night sleep duration, with 145 interactions not being observed in the controls. Further experiments revealed candidate genes that impact night sleep and confirm these interactions.

Methods and Materials

Construction of outbred population

We constructed an outbred population of flies – using ten lines from the *Drosophila* Genetic Reference Panel (DGRP) (***Mackay et al., 2012***; ***Huang et al., 2014***) with extreme night sleep phenotypes (***Harbison et al., 2013***). Five lines had the shortest average night sleep for both males and females combined in the population: DGRP_38, DGRP_310, DGRP_365, DGRP_808, DGRP_832. The other five lines had the longest average night sleep in the population: DGRP_235, DGRP_313, DGRP_335, DGRP_338, and DGRP_379. The ten lines were crossed in a full diallel design, resulting in 100 crosses. Two virgin females and two males from the F1 of each cross were randomly assigned into 20 bottles, with 10 males and 10 females placed in each bottle. At each subsequent generation, 20 virgin females and 20 males from each bottle were randomly mixed across bottles to propagate the next generation. The census population size was 800 for each generation of random mating. This mating scheme was continued for 21 generations, resulting in the Sleep Advanced Intercross Population, or SAIP (***Harbison et al., 2017***; ***Serrano Negron et al., 2018***). The SAIP was maintained by pooling the flies from each bottle together, then randomly assigning 20 males and 20 females to each bottle each generation.

Artificial selection procedure for night sleep

At generation 47 of the SAIP, we began the artificial selection procedure, which we defined as generation 0. We seeded six bottles with 25 males and 25 females mixed from all bottles of the outbred population. Two replicate bottles were designated for the short-sleeping protocol (S1 and S2), two for the long-sleeping protocol (L1 and L2), and two for a control (unselected) protocol (C1 and C2). Each generation, 100 virgin males and 100 virgin females were collected from each of the six population bottles. Virgins were maintained at 20 individuals to a same-sex vial for four days to control for the potential effects of social exposure on sleep (***Ganguly-Fitzgerald et al., 2006***). Flies were placed into Trikinetics (Waltham, MA) sleep monitors, and sleep and activity were recorded continuously for four days. We used an in-house C# program (R. Sean Barnes, personal communication) to calculate sleep duration, bout number, and average bout length during the night and day, as well as waking activity. We also calculated sleep latency, defined as the number of minutes prior to the first sleep bout after the incubator lights turn off. In addition, we computed the coefficient of environmental variation (CV_E) for each sleep trait as the product of the standard deviation in each replicate population (σ) divided by the mean (μ) $\times 100$ (***Mackay and Lyman, 2005***).

All sleep traits including night sleep duration were averaged over the four-day period. For the short (long)-sleeping populations, we chose the 25 males and 25 females in each replicate population having the lowest (highest) average night sleep as parents for the next generation. Any flies found dead were discarded, and the next shortest (longest)-sleeping fly was used in order to ensure that 25 females and 25 males were used as parents. For the control populations, we chose 25 males and 25 females at random to start the next generation. Flies were not mixed across replicate populations. We repeated this procedure for 13 generations.

Quantitative genetic analyses of selected and correlated phenotypic responses

We analyzed the differences in night sleep among selection populations as well as other potentially correlated sleep traits using a mixed analysis of variance (ANOVA) model:

$$\begin{aligned} Y = \mu &+ Sel + Rep(Sel) + Sex + Gen \\ &+ Sel \times Sex + Sel \times Gen + Rep(Sel) \times Sex + Rep(Sel) \times Gen + Sex \times Gen \\ &+ Sel \times Sex \times Gen + Rep(Sel) \times Sex \times Gen + \varepsilon \end{aligned}$$

where Y is the phenotype; μ is the overall phenotypic mean; Sel, Sex, and Gen are the fixed effects of selection scheme (short- or long-sleeper), sex, and generation, respectively; Rep is random effect of replicate population; and ε is the error term. The CV_E traits were assessed using the same model

with the replicate terms removed. A statistically significant Sel term indicates a response of the trait to selection for night sleep; a significant $Sel \times Sex$ term indicates a sex-specific response to selection. We repeated the analysis for sexes separately using the reduced model

$$Y = \mu + Sel + Rep(Sel) + Gen + Sel \times Gen + Rep(Sel) \times Gen + \varepsilon$$

where the terms are as defined above. We also analyzed the response to selection in each generation separately using the reduced model

$$Y = \mu + Sel + Rep(Sel) + Sex + Sel \times Sex + Rep(Sel) \times Sex + \varepsilon$$

and the reduced model

$$Y = \mu + Sel + Rep(Sel) + \varepsilon$$

for each sex separately per generation.

Finally, we analyzed the change in sleep parameters over generations in the control populations using the model

$$Y = \mu + Rep + Sex + Gen + Rep \times Sex + Rep \times Gen + Sex \times Gen + Rep \times Sex \times Gen + \varepsilon$$

where each factor is as defined above.

RNA extraction and sequencing

As described above, sleep was monitored in 100 virgin males and 100 virgin females each generation. Twenty-five flies of either sex were used as parents for the next generation, leaving 75 flies of each sex in each selection and control population. Four pools of 10 flies of each sex were chosen at random from these 75 flies and frozen for RNA extraction at 12:00 pm. RNA was extracted from two of these pools; the remaining two pools were kept as back-up samples and used if needed. Samples were collected for the initial generation (0), and all subsequent generations. RNA was extracted using Qiazol (Qiagen, Hilden, Germany), followed by phenol-chloroform extraction, isopropanol precipitation, and DNase digestion (Qiagen, Hilden, Germany). Qiagen RNeasy MinElute Cleanup kits (Qiagen, Hilden, Germany) were used to purify RNA according to the manufacturer's instructions. With the exception of generation 1, which had RNA that was degraded, RNA from all other generations was sequenced. This produced 312 RNA samples (6 populations $\times$ 13 generations $\times$ 2 sexes $\times$ 2 replicate RNA samples).

Poly-A selected stranded mRNA libraries were constructed from 1 μg total RNA using the Illumina TruSeq Stranded mRNA Sample Prep Kits (Illumina, San Diego, CA) according to manufacturer's instructions with the following exception: PCR amplification was performed for 10 cycles rather than 15 in order to minimize the risk of over-amplification. Unique barcode adapters were applied to each library. Libraries were pooled for sequencing. The pooled libraries were sequenced on multiple lanes of an Illumina HiSeq2500 using version 4 chemistry to achieve a minimum of 38 million 126 base read pairs. The sequences were processed using RTA version 1.18.64 and CASAVA 1.8.2.

RNA alignment of reads

Sequences were assessed for standard quality parameters using fastqc (0.11.4) (Babraham Institute, Cambridge, UK). Reads were aligned to the FB2015_04 Release 6.07 reference annotation of the *Drosophila melanogaster* genome using STAR (***Dobin et al., 2013***). Default parameters were used except that the minimum intron size was specified as 2, and the maximum intron size was specified as 268,107, consistent with the largest intron size in the *D. melanogaster* genome. STAR outputs aligned sequence to a SAM file format, which contains the code 'NH' (***Dobin et al., 2013***). An NH of 1 indicates a uniquely mapped read, while $NH > 1$ indicates that the read did not map uniquely. HTSeq was used to count only the uniquely mapped reads ($NH = 1$) (***Anders et al., 2015***).

Principal Component Analysis (PCA)

It was expected from previous studies of gene expression that there would be large differences in gene expression due to sex (***Jin et al., 2001***; ***Arbeitman et al., 2002***; ***Parisi et al., 2003***; ***Ranz et al., 2003***; ***Harbison et al., 2005***; ***Wayne et al., 2007***; ***Zhang et al., 2007***; ***Ayroles et al., 2009***; ***Huylmans and Parsch, 2014***; ***Huang et al., 2015***; ***Lin et al., 2016***). We performed Principal Component Analysis to assess those differences (Supplementary Figure S1). The principal components of the normalized RNA-seq count normalized matrix were computed, with each gene being treated as a different variable, and each sample a different observation. Samples were projected in the planes of the three first components, and clustering according to the experimental labels was inspected visually.

Gene normalization and filtering

The combined genic and intergenic counts were normalized by the expression of a pseudo-reference sample computed from the geometric mean of all samples, using the method described by ***Love et al.*** (***2014***). Filtering was performed by computing the 95th percentile of the distribution of normalized, base 2 logarithm, levels in the intergenic regions for males and females and using those values as cut-off level for the genic regions – i.e. any genes that did not have expression above this level for at least one sample were removed from further analyses (***Zhang et al., 2010***). The (linear scale) cutoff expression value for males was 48.6, and for females 102.

Generalized Linear Model analysis of expression data

Analysis of differential expression between selection schemes was initially performed for each gene independently. Given the separation of the expression levels by sex seen in the PCA analysis, analyses were conducted separately for the subsets of male or female flies.

We implemented a generalized linear model (GLM) with a hierarchical structure to account for non-independent, replicate-specific parameters. The description is similar to a generalized linear mixed model (GLMM), but uses a Bayesian formulation to specify the hyper-priors and is fully described below. Normalization factors for the RNA levels was performed using the scheme described by ***Love et al.*** (***2014***). A negative binomial likelihood was used and parameterized with the mean (given by the prediction of the linear model) and dispersion parameters; the number of samples (156 for each sex) allowed estimation of the latter together with model coefficients, dispensing with the need of other schemes applied when the number of samples is small, commonly implemented in some packages.

Bayesian inference was used and parameter priors were exploited to treat replicate effects in a hierarchical formulation (***Gelman et al., 2013***). Specifically, for each replicate-dependent parameter (say $\beta_{short,rep}$), two parameters were specified at the top-level (μ_{short} and σ_{short}), given (hyper-)priors, and estimated from the data together with all other parameters. Below that, both replicate-specific model parameters ($\beta_{short,1}$ and $\beta_{short,2}$) are given the same gaussian prior using top-level parameters (e.g. $\beta_{short,1} \sim \mathcal{N}(\mu_{short}, \sigma_{short})$ for that coefficient in replicate 1 as well as replicate 2). Under this formulation the full model for the expression of a gene j is given by $log\mu_j \propto sel_{rep} + gen + sel \times gen_{rep}$, where a relationship between each set of replicate-dependent parameters is enforced hierarchically through their higher level common parameters and hyperpriors. Explicitly, we have:

$$\begin{aligned}\eta_j &= log\mu_j \\ &= [\beta_1, \beta_2, \beta_{short,1}, \beta_{short,2}, \beta_{long,1}, \beta_{long,2}, \beta_{gen}, \beta_{short\times gen,1}, \beta_{short\times gen,2}, \beta_{long\times gen,1}, \beta_{long\times gen,2}]X\end{aligned}$$

where X is the design matrix, with binary $0/1$ variables indicating parameters that apply to specific treatments (e.g. the entries multiplying β_1,β_2, are present for all, that $\beta_{short,1}$, is present for short sleepers from replicate 1, etc.) except for parameters dependent on the gen variable which takes the value of the generation (e.g. 0 through 13 for the entries multiplying the β_{gen} parameter in all treatments, and for those multiplying $\beta_{short\times gen,1}$ for short sleepers from replicate 1, etc.). Table 1 lists all parameters, their descriptions, design matrix values associated to them, and priors.

Table 1. Parameter names, description, design values, and priors for Bayesian inference ($\bar{y_0}$ denotes the mean expression of all samples at generation zero).

Parameter	Description	Design values	Prior
$\mu_{control}$	Hyperprior on mean of β_{rep}	n/a	$\mathcal{N}(\bar{y_0}, 1)$
$\sigma_{control}$	Hyperprior on (square root of) variance of β_{rep}	n/a	$Cauchy(0, 1)$
μ_{short}	Hyperprior on mean of $\beta_{short,rep}$	n/a	$\mathcal{N}(0, 1)$
σ_{short}	Hyperprior on variance of $\beta_{short,rep}$	n/a	$Cauchy(0, 1)$
μ_{long}	Hyperprior on mean of $\beta_{long,rep}$	n/a	$\mathcal{N}(0, 1)$
σ_{long}	Hyperprior on variance of $\beta_{long,rep}$	n/a	$Cauchy(0, 1)$
$\mu_{short\times gen}$	Hyperprior on mean of $\beta_{short\times gen,rep}$	n/a	$\mathcal{N}(0, 1)$
$\sigma_{short\times gen}$	Hyperprior on variance of $\beta_{short\times gen,rep}$	n/a	$Cauchy(0, 1)$
$\mu_{long\times gen}$	Hyperprior on mean of $\beta_{long\times gen,rep}$	n/a	$\mathcal{N}(0, 1)$
$\sigma_{long\times gen}$	Hyperprior on variance of $\beta_{long\times gen,rep}$	n/a	$Cauchy(0, 1)$
β_1	Intercept for replicate 1	$0, 1$	$\mathcal{N}(\mu_{control}, \sigma_{control})$
β_2	Intercept for replicate 2	$0, 1$	$\mathcal{N}(\mu_{control}, \sigma_{control})$
$\beta_{short,1}$	Effect from short sleeper, replicate 1 treatment	$0, 1$	$\mathcal{N}(\mu_{short}, \sigma_{short})$
$\beta_{short,2}$	Effect from short sleeper, replicate 2 treatment	$0, 1$	$\mathcal{N}(\mu_{short}, \sigma_{short})$
$\beta_{long,1}$	Effect from long sleeper, replicate 1 treatment	$0, 1$	$\mathcal{N}(\mu_{long}, \sigma_{long})$
$\beta_{long,2}$	Effect from long sleeper, replicate 2 treatment	$0, 1$	$\mathcal{N}(\mu_{long}, \sigma_{long})$
β_{gen}	Treatment-independent generation effect	$0 - 13$	$\mathcal{N}(0, 2)$
$\beta_{short\times gen,1}$	Interaction short by generation, rep 1 effect	$0 - 13$	$\mathcal{N}(\mu_{short\times gen}, \sigma_{short\times gen})$
$\beta_{short\times gen,2}$	Interaction short by generation, rep 2 effect	$0 - 13$	$\mathcal{N}(\mu_{short\times gen}, \sigma_{short\times gen})$
$\beta_{long\times gen,1}$	Interaction long by generation, rep 1 effect	$0 - 13$	$\mathcal{N}(\mu_{long\times gen}, \sigma_{long\times gen})$
$\beta_{long\times gen,2}$	Interaction long by generation, rep 2 effect	$0 - 13$	$\mathcal{N}(\mu_{long\times gen}, \sigma_{long\times gen})$
α	Negative binomial dispersion	n/a	$Uniform(0, 10^9)$

Maximum a posteriori probability (MAP) estimates and confidence intervals were obtained using the Stan package (***Carpenter et al., 2017***). Significance was calculated using a likelihood ratio test comparing the point estimates from the full model to a reduced model not including the interaction terms (i.e. $log\mu_{j,rep} = sel_{rep} + gen$). Model p-values were corrected for multiple testing using the Benjamini-Hochberg method (***Benjamini and Hochberg, 1995***), with significance defined at the 0.001 level.

Calculation of non-parametric correlations between genes

The correlation coefficients (ρ) between any two pairs of genes can be computed directly from the data. Pearson correlation assumes the relationship between the two variables is linear, while Spearman correlation is rank-based and therefore accommodates non-linear relationships, although it still assumes the relationship is monotonically increasing or decreasing. We therefore computed Spearman correlations between genes that were found to be significant for both males and females in the GLM analysis –- one correlation coefficient was obtained for the data subset from each sex-selection combination. The significance of each correlation coefficient is tested using the null hypothesis that $\rho = 0$. Because the main interest is the interaction between genes in the selected populations that are different from controls we compare the coefficients by computing and comparing the confidence intervals for ρ_{sel} (where sel can be "short" or "long") and $\rho_{control}$ using the normal approximation to $arctanh(\rho)$ (***Ruscio, 2008***). We note that this is not exactly equivalent to the significance testing of the null hypothesis that $\rho_{sel} = \rho_{control}$ (***Austin and Hux, 2002***) (which relies on computing the confidence interval for $\rho_{sel} - \rho_{control}$ using the same method), since it overestimates the total variance (i.e., one would find fewer significant instances). Nevertheless, the approach is

valid and is more broadly applicable, in that it can be computed when a joint distribution with the two variables cannot be obtained – we use the term "significant" for either kind of difference, but explicitly state which one is used.

Gaussian Process regression

Gaussian Processes (GP) are an alternative function-space formulation to the well-known weight-space linear models of the form $y = f(x) + \varepsilon$; their use dates back to the 19th century and they have been covered extensively in the statistical and information theory literature (***MacKay, 2003***), becoming popular in machine learning applications (***Bishop, 2006***; ***Rasmussen and Williams, 2006***), and more recently implemented in less technical contexts like the life sciences (***Schulz et al., 2018***). We give a brief overview of their usefulness, motivate their use in this work, and point to the references above for formal description of the method.

The weight-space linear model expresses the observations in terms of explicit linear coefficients (or weights) of the independent variable, x, possibly with further basis function expansions (e.g. square, x^2, or higher order polynomials, x^n), for instance $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$, (where ε is normally distributed noise). Gaussian Processes describe the basis functions implicitly instead, with $y \sim \mathcal{N}(\mu, \boldsymbol{K})$; that is, a set y of N observations is distributed according to a multivariate normal distribution with mean given by the vector μ (of size N) and covariance between the values of x given by the matrix $\boldsymbol{K}$ (with dimension $N \times N$). The entries of this matrix in row i, column j are defined by some covariance function such that $k_{ij} = cov(x_i, x_j)$ – if the covariance function is linear in the values of x, for instance, the prediction for y is a straight line similar to $y = \beta_0 + \beta_1 x$. Formulating the model in terms of function-space enables the use of flexible sets of basis functions; this approach of only implicitly describing a basis function, thus avoiding specification of a potentially large basis is called the "kernel trick". Functions like the commonly used squared exponential kernel can be shown to be equivalent to an infinite number of basis functions (***Rasmussen and Williams, 2006***), and therefore cannot be incorporated in the explicit terms of the weight-space formulation.

While Gaussian Processes are a classic formulation in statistics, the recent surge in machine learning applications has popularized its use in the natural sciences. They have been used to analyze gene expression by using their flexible output in combination with ordinary differential equations (***Honkela et al., 2010***; ***Äijö et al., 2013***; ***Aalto et al., 2020***) , with clustering approaches (***McDowell et al., 2018***), within other regression models (***Kontio and Sillanpää, 2019***), or modeling spatial covariance (***Arnol et al., 2019***). In the context of our experimental design Gaussian Process Regression could be used as a flexible alternative to GLMs, with each selection scheme having a different mean function μ_{sel} and a squared exponential covariance function $k(x, x') = \sigma_f^2 c(x, x') = \sigma_f^2 exp\left(\frac{|x-x'|^2}{2l^2}\right)$ where x takes the values of the generations in our experiment. The exponentiated term gives the correlation $c(x, x')$ between a pair of time points, with parameter ℓ modulating the correlation level given a distance $r = x - x'$, and σ_f^2 being the signal variance of the data. Under this model, unlike with the GLM analysis, the change in RNA-seq counts is a function not of slope coefficients but of the signal variance σ_f^2. It is worth noting that the signal variance is a scalar constant for all terms in the covariance matrix, so it can also be written as $\boldsymbol{K} = \sigma_f^2 \boldsymbol{C}$, where $\boldsymbol{C}$ is analogous to $\boldsymbol{K}$ but with correlations instead of covariances, a notation that will be useful shortly.

Multi-channel Gaussian Processes

Despite the extensive use of Gaussian Processes, most applications in the life sciences have been restricted to single-channel GPs; that is, models that only describe one set of observations at a time (here the expression time series for a single gene). These models – in this aspect not unlike GLMs – describe expression of genes independently, i.e. they implicitly assume genes do not interact in any way. Gaussian Processes can however be extended to include covariance between two or more sets of observations, a formulation that seems to be underexploited in the biological literature (but see ***Velten et al.*** (***2020***) and ***Bahg et al.*** (***2020***)). The different dependent variables

Table 2. Parameter names, description, and priors for Gaussian Processes Bayesian inference.

Parameter	Description	Prior
s	Standard deviations of data (one for each channel)	n/a
$\hat{\sigma}_i^2, (V_{\sigma,i})$	Signal variance expectation (variance) from single-channel i model	n/a
$\hat{\ell}_i, (V_{\ell,i})$	Bandwidth expectation (variance) from single-channel i model	n/a
σ_{ii}^2	Signal variance for channels i	$\mathcal{N}(\hat{\sigma}_i, \sqrt{V_{\sigma,i}})$
σ_{ij}^2	Signal covariance between channels i and j	$\mathcal{N}(0, max(s))$
ℓ	Bandwidth parameters	$\mathcal{N}(\hat{\ell}_i, \sqrt{V_{\ell,i}})$
$\widetilde{f}$	Gaussian Process latent normal variates	$\mathcal{N}(0,1)$
ϕ	inverse of square of dispersion parameter ($\phi = 1/\alpha^2$)	$\mathcal{N}(0,1)$

y_i are sometimes called channels or tasks, and the resulting model is called a multi-task or multi-channel Gaussian Process. The details of the specification of this model can be found in ***Bonilla et al.*** (***2008***) and ***Melkumyan and Ramos*** (***2011***), which we summarize below. For an array of two genes only, for instance, instead of describing each vector y_1 and y_2 separately as multivariate gaussians of dimension N_1 and N_2, respectively, the concatenated vector $[y_1\ y_2]^T$ with $N_1 + N_2$ observations can be modeled as a single multivariate gaussian with a covariance matrix of K dimensions $(N_1+N_2)\times(N_1+N_2)$, or $[y_1\ y_2]^T \sim \mathcal{N}(\mu, K)$. The diagonal blocks of the covariance matrix with dimensions $N_1 \times N_1$ and $N_2 \times N_2$ are the same as above, and the off-diagonal blocks of dimensions $N_2 \times N_1$ and $N_1 \times N_2$ specify the correlations $c_{12ij}(x_{1i}, x_{2j}) = exp\left(\frac{|x_{1i}-x_{2j}|^2}{\ell_1^2+\ell_2^2}\right)$ between the two points ij from channels 1 and 2 (***Melkumyan and Ramos, 2011***). Finally, the signal variance for each of those blocks need to be specified, and the final matrix is given by $K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 C_{11} & \sigma_{12}^2 C_{12} \\ \sigma_{12}^2 C_{12} & \sigma_2^2 C_{22} \end{bmatrix}$ (***Bonilla et al., 2008***), and the mean of the multivariate gaussian is specified by a concatenated vector $\mu = [\mu_1\ \mu_2]^T$. The number of parameters is reduced by recognizing that the covariance matrix is symmetric so in this example $\sigma_{21}^2 = \sigma_{12}^2$, where we also dropped the subscript f. For this model, the variation in the RNA levels of say gene 1 is a function not only of σ_1^2, but also of $\sigma_{21}^2 = \sigma_{12}^2$. Therefore, fitting the data with this model infers interaction between genes from scratch without any external information not contained in the array of RNA-seq counts.

The model can be extended to any number of genes, although computational requirements for performing the necessary matrix operations on K also grow with its size and may be limiting – the computational and mathematical limitations of this approach are discussed in the appendix.

Bayesian MCMC inference of Gaussian Processes

Analogously to GLM models, we maintain the negative binomial likelihood for the Gaussian Process inference, but unlike the transition between linear models and their generalized versions, the incorporation of non-gaussian likelihoods is not as straightforward, and requires methods to approximate the underlying latent Gaussian Process model, leading to what is sometimes referred to as Gaussian Process Classification (***Rasmussen and Williams, 2006***). Because of the Bayesian inference implemented for this model we chose to infer the latent function via Markov Chain Monte Carlo sampling as these variables can be estimated jointly with the other parameters and have priors that by design are standard gaussian, and therefore are straightforward to specify. Table 2 gives the description of all parameters in the Multi-Channel Gaussian Process model and their priors.

The number of covariance parameters in a multi-channel Gaussian Process model with M channels is $(M^2 - M)/2$, and the total number of parameters scales roughly as $\mathcal{O}(M^2)$ as the number of channels becomes large. For 100 genes, for instance, that would result in about 5,000 covariances. Due to the statistical challenge of exploring a parameter space with a dimension of several

thousand, as well the computational demand of factorizing a large matrix at each MCMC step, the estimation of the signal covariance parameters between genes was not performed jointly. Instead, each pair of genes was fitted separately, with a single-channel Gaussian Process being first used to estimate the signal variance and bandwidth parameters for each gene and this estimate being used as a prior for the (pairwise) joint inference. This procedure effectively breaks down a Gaussian Process inference of any size into several smaller inference problems requiring factorization of a matrix of size $2N$, with a total number of parameters of the order of N, which are computationally much more manageable and can be run in parallel. Because the covariance parameters depend only on the relationship between two variables (here, genes), separate estimation does not affect inference of the parameters; in fact, it removes the constraint of positive-definiteness on the matrix of covariances of all genes (which instead applies to the matrix of two genes only, see Appendix I).

Eight parallel chains were run for each estimation with 40 thousand samples each; half were excluded as warm-up and 1 out of every 40 was kept for further calculations. Convergence was assessed using the $\hat{R}$ metric and observing the number of effective samples (ESS) (***Gelman et al., 2013***). The annotated model implemented in the Stan probabilistic language is made available in the supplementary material. Because inference was done separately for each selection scheme, differences between them were assessed by comparing the posterior distribution of the parameters of interest.

Confirmatory experiments

We tested *Minos* insertions putatively disrupting five genes from the significant Gaussian Process correlation table for their effect on sleep, along with their two background controls. We assayed phenotypes using the same procedure outlined above. Twenty-four flies per sex per line were assayed, and the experiment was replicated twice. Sleep was analyzed using the following ANOVA model:

$$Y = \mu + Genotype + Sex + Rep + Genotype \times Sex + Genotype \times Rep + Sex \times Rep + Genotype \times Sex \times Rep + \varepsilon$$

where Sex and Rep are as previously defined above and $Genotype$ refers to the *Minos* insertion line or control. We collected 10 flies per sex/line at the conclusion of sleep monitoring for RNA extraction. RNA was extracted as detailed above, with the following exception: ERCC spike-ins (ThermoFisher Scientific, Waltham, MA) were added to the RNA after the extraction procedure. A total of 28 samples (7 lines × 2 sexes × 2 replicates) were collected and processed. We then sequenced these samples and processed them as detailed above. Note that we discarded one sample due to failed quality control during library preparation. We compared normalized gene expression in the mutants to their respective controls using a Kruskal-Wallis non-parametric test. We conducted RNA-Seq on the *Minos* insertion lines and their controls. Expression ratios were computed between knock down genotypes and their wild-type controls (w^{1118} or y^1w^{67c23}) for individual genes predicted by the Gaussian Processes to be significantly correlated with the knocked down genes in the relevant sex-selection scheme combination (Table S12), henceforth candidate genes). Expression ratios were also computed for sets of 1000 genes chosen at random, each set matching the genetic backgrounds (knockdown and controls) as well as the sex (henceforth random sets); a distribution of expression ratios was generated for each random set.

Results

Phenotypic response to artificial selection

The selection procedure for night sleep was very effective. Long-sleeper and short-sleeper populations had significant differences in night sleep across all generations ($P_{Sel} = 0.0003$); in fact, night sleep was different for the two selection schemes for each generation considered separately except for generations 0 and 1 (Supplementary Tables S1 and S2). Both males and females responded

equally to the selection procedure. Figure 1A shows the phenotypic response to 13 generations of selection for night sleep. At generation 13, the long-sleeper populations averaged 642.2 ± 3.83 and 667.8 ± 2.97 minutes of night sleep for Replicate 1 and Replicate 2, respectively. The short-sleeper populations averaged 104.3 ± 6.71 and 156.2 ± 8.76 minutes of night sleep for Replicate 1 and Replicate 2, respectively. The average difference between the long- and short-sleeper lines was 537.9 minutes for Replicate 1, and 511.6 minutes for Replicate 2. In contrast, the two control populations did not have differences in their night sleep after 13 generations of random mating ($P_{Gen} = 0.7083$; Supplementary Table S3). In the initial generation, night sleep was 519.6 ± 10.57 minutes in the Replicate 1 control and 567.9 ± 7.63 minutes in the Replicate 2 control. At generation 13, night sleep was 563.4 ± 7.62 and 542.3 ± 7.91 in Replicates 1 and 2, respectively, a difference of only 43.8 and 25.6 minutes. These negligible changes in night sleep in the control population suggest that little inbreeding depression occurred over the course of the experiment (***Falconer and Mackay, 1996***). Selection was asymmetric, with a greater phenotypic response in the direction of reduced night sleep. Note also that night sleep is bounded from 0 to 720 minutes, and the initial generation had 515.39 minutes of night sleep on average across all populations, a fairly long night sleep phenotype. This high initial sleep may explain why the response to selection for short night sleep was more effective. Night sleep is sexually dimorphic (***Harbison and Sehgal, 2008***; ***Harbison et al., 2009***, ***2013***); yet both males and females responded to the selection protocol equally ($P_{Sel \times Sex} = 0.9492$; Supplementary Table S1). Thus, we constructed a set of selection populations with nearly 9 hours difference in night sleep.

In an artificial selection experiment, some amount of inbreeding will necessarily take place. Only a subset of the animals are selected each generation as parents; thus phenotypic variance is expected to decrease as selection proceeds (***Falconer and Mackay, 1996***).

However, this is not the case for all artificial selection experiments (***Falconer and Mackay, 1996***). We calculated the coefficient of environmental variation (CV_E) (***Mackay and Lyman, 2005***) and evaluated its trajectory across time in order to determine whether the populations were becoming more or less variable over time. As Figure 1B shows, night sleep CV_E increased over time in the short sleepers, and decreased over time in the long sleepers ($P < 0.0001$; Table S4). The increase in CV_E in short sleepers was largely due to a decrease in the population mean as the standard deviation also decreased over time, indicating that the phenotypic variance decreased (Figure S2). Likewise, the standard deviation decreased in the long sleepers over time, even as the mean night sleep increased, indicating decreased variability in these populations as well. These changes in CV_E mimic previous observations in populations artificially selected for sleep (***Harbison et al., 2017***). Regressions of the cumulated response on the cumulated selection differential were used to estimate heritability (h^2). Long-sleeper population h^2 ($\pm SE$ of the coefficient of regression) were estimated as 0.145 ± 0.021 and 0.141 ± 0.014 (all $P < 0.0001$) for Replicates 1 and 2, respectively (Figure 1C); short-sleeper population h^2 were 0.0169 ± 0.013 and 0.183 ± 0.019 (all $P < 0.0001$) for Replicates 1 and 2 (Figure 1D). In contrast, estimated regression coefficients for the control population were non-significant and with high standard errors associated to the regression estimates: 0.405 ± 0.695 ($P = 0.57$) and -0.078 ± 0.487 ($P = 0.88$) for Replicates 1 and 2, respectively (Figure 1E).

Correlated response of other sleep traits to selection for night sleep

Traits that are genetically correlated with night sleep might also respond to selection for long or short night sleep (***Falconer and Mackay, 1996***). Indeed, some sleep and activity traits have been previously shown to be phenotypically and genetically correlated (***Harbison and Sehgal, 2008***; ***Harbison et al., 2009***, ***2013***). We examined the other sleep and activity traits for evidence of a correlated response to selection. Night and day average bout length ($P = 0.0008$ and $P = 0.0391$, respectively) and sleep latency ($P = 0.0023$) exhibited a correlated response to selection for night sleep across generations $0 - 13$, while night and day bout number, day sleep, and waking activity did not (Figure S2; Supplementary Table S1). In the case of day average bout length, the correlated response was sex-specific to males ($P = 0.0140$) (Supplementary Table S1). Significant correlated responses for

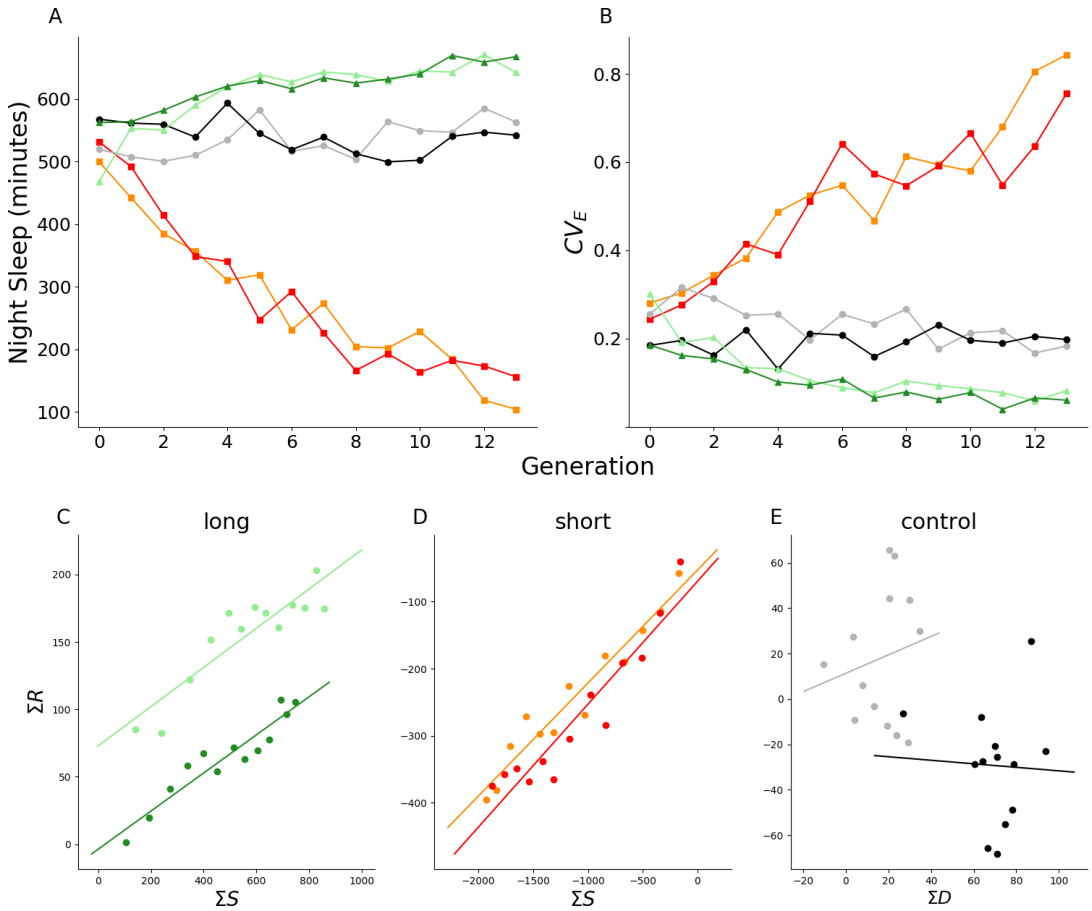


Figure 1. (A) Mean and (B) coefficient of environmental variation of night sleep. Plot and regression lines of cumulated selection differential (ΣS) against cumulated selection response (ΣR) for (C) long- and (D) short-sleeping populations, and against cumulated differential ΣD for (E) controls. Light green, Replicate 1 long-sleeper population; Dark green, Replicate 2 long-sleeper population; Orange, Replicate 1 short-sleeper population; Red, Replicate 2 short-sleeper population; Gray, Replicate 1 control population; Black, Replicate 2 control population.

night and day average bout length and sleep latency did not occur in all generations (Supplementary Table S2).

Night average bout length responded to selection for night sleep in most generations, while day average bout length responded in only four of the last six generations. Sleep latency responded to selection after the second generation. In addition, we observed significant differences between the long-sleeping and short-sleeping populations for the CV_E of all sleep traits except waking activity CV_E (Figure S2; Table S4). However, the pattern of the CV_E for each trait appeared to be more random across time.

Phenotypes in flies used for RNA-Seq

Every generation, we harvested RNA from flies chosen at random from the 200 measured for sleep in each selection population, with the exception of the flies chosen as parents for the next generation. We extracted RNA from two replicates of 10 flies each per sex and selection population. Since these flies amount to only 20% of the flies measured for sleep each generation, their sleep may or may not be representative of the group as a whole. We therefore correlated the mean night sleep for each generation in the flies harvested for RNA with the mean night sleep of all flies measured to determine how similar night sleep was to the total in the group (Figure S3). The correlations were very high for the selected populations: long-sleeper flies harvested for RNA were very well

correlated with the total measured in each population [$r^2 = 0.99$ and 0.96 (all $P < 0.0001$) for Replicate 1 and 2 respectively], as were short-sleepers [$r^2 = 0.99$ for Replicate 1 and 0.97 for Replicate 2 (all $P < 0.0001$)]. The control populations, which did not undergo selection, were somewhat less well correlated. Replicate 1 of the control population had an r^2 of 0.75 ($P = 0.0001$) and Replicate 2 had an r^2 of 0.85 ($P < 0.0001$). Thus, the flies harvested for RNA are very good representatives of each population as a whole.

Hierarchical Generalized Linear Model analysis reveals that selection for night sleep impacts gene expression

For each gene, the linear model analysis produced posterior distributions for the parameters as well as log-likelihood values for the full and reduced models. Point estimates (MAP) are shown in Table S5 and S6 (for females and males, respectively). For the male flies 11,778 genes passed the filtering for low expression, of which 405 were found to have a significant selection scheme effect over the generations of artificial selection (i.e., significant likelihood ratio test for the $sel \times gen$ term). Thus, the expression level shift given by the slope of the generalized linear model is different from controls and attributable to selection for long and/or short sleep. For the females 820 genes out of 9,370 with detectable expression were found to be significant. Genes with opposite trends in the short and long selection schemes were compared using the group-level parameter $\mu_{short \times gen}$ and $\mu_{long \times gen}$ (i.e. the effect that best explains both replicates): 204 genes in the males and 384 in females showed opposite trends by that criterion. Table S7 and S8 list those genes for females and males, respectively. Between males and females, 85 genes were common to both sexes. Known functions of these 85 genes from the DAVID gene ontology database are presented in Table S9. We used these 85 genes in subsequent analyses; see below. Figure 2 shows the fit for one gene.

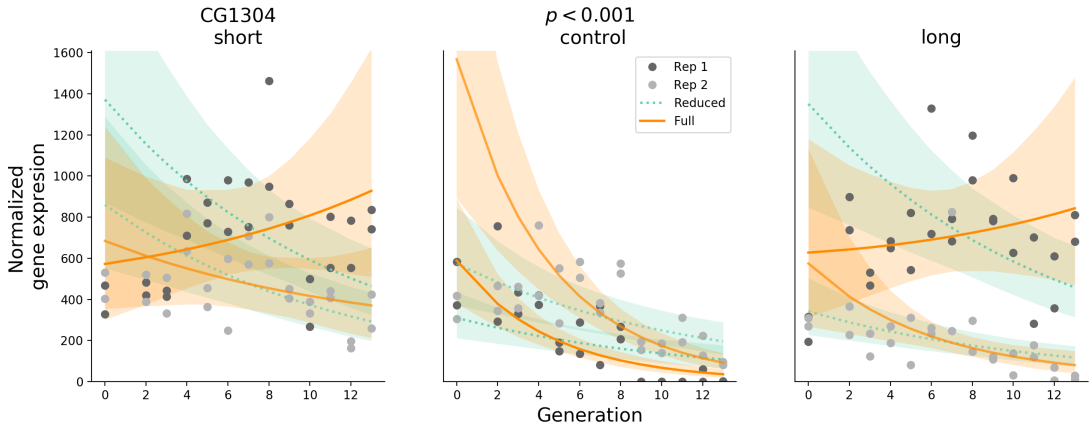


Figure 2. Fit of Hierarchical Generalized Linear Model to gene *CG1304* for flies selected for short sleep, unselected controls, and selected for long sleep. The solid lines show the expected value of full model, dashed lines for reduced model, and shaded regions show the 95% credibility interval. Replicate 1 data points are shown in dark gray, Replicate 2 in light gray.

Pairwise Spearman correlation is non-specific and significant for a large fraction of genes

We computed Spearman correlations for all pairwise combinations of the 85 genes common between sexes (Supplementary Table S10). Correlations computed using the Spearman method were found to be significant at 95% confidence for 2,999 of the 3,570 possible pairs. The confidence intervals for the correlations coefficients showed no overlap with controls for either short sleepers, long sleepers, or both populations in 1,348 of 3,570 pairs. Thus, a simple correlational analysis identifies a minimum of 38% of the possible interactions among genes as relevant.

Gaussian Process model analysis uncovers nonlinear trends and specifically identifies covariance in expression between genes

As noted above, a simple correlational analysis suggested that large numbers of genes are potentially interacting to alter sleep. Because direct computation of linear model-based correlations cannot account for non-linear effects or spurious confounding trends we fit Gaussian Process models that can account for temporal variation in multiple genes even in the absence of actual interactions between them. The 85 significant genes overlapping between males and females potentially have 3,570 pairwise interactions. To that end, the parameter of interest in the Gaussian Process model is the signal covariance between each pair of genes. This covariance is a measure of the degree of their interaction. We applied the Gaussian Process model for each of the 3,570 pairs for each selection scheme (long, short, and control). As an example, the model fit for one pair of genes from the female gene expression data is shown in Figure 3.

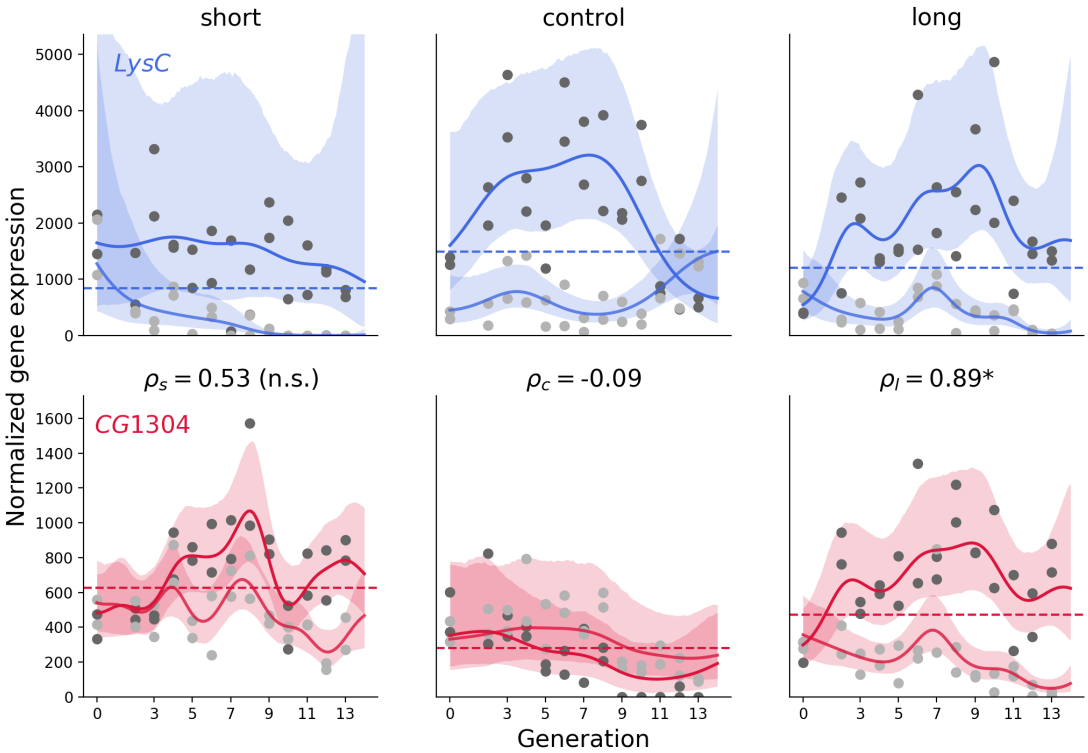


Figure 3. Fit of Gaussian Process model to pair of genes *LysC* and *CG1304*, for flies selected for short sleep, unselected controls, and selected for long sleep. The solid lines show the expected value, while the shaded regions show the 95% credibility interval. Replicate 1 data points are shown in dark gray, Replicate 2 in light gray). The expectation for correlations (ρ_{sel}) is shown for each selection scheme. An asterisk indicates significant difference from controls in selection scheme, as opposed to non-significance (n.s.).

Figure 3—figure supplement 1. Fit of Gaussian Process model to pair of genes *haf* and *CG1304*.

Figure 3—figure supplement 2. Fit of Gaussian Process model to pair of genes *CR43242* and *CG1304*.

Figure 3—figure supplement 3. Fit of single-channel Gaussian Process model to *CG1304* gene.

Figure 3—figure supplement 4. Fit of single-channel Gaussian Process model to *LysC* gene.

Convergence for all three runs was on the order of $|\hat{R} - 1| \approx 10^{-4}$, and close to the 4,000 samples expected for each run; therefore, the wide confidence intervals are likely a product of the large dispersion in the data itself. Correlation between gene expression patterns of the two genes is computed by dividing the signal covariance by the square root of the signal variance of each gene – e.g. $\rho_l = \sigma^2_{l(ij)}/\sigma_{l(i)}\sigma_{l(j)} = \sigma^2_{long(LysC,CG1304)}/\sigma_{long(LysC)}\sigma_{long(CG1304)}$ – that is, similar to computing a correlation coefficient from variances and covariances, but taken as the expectation over the posterior distribution obtained from MCMC.

Figure 3 illustrates the nonlinear trajectories of gene expression that cannot be detected by the GLM model. The two trajectories exhibited high signal covariance between the expression of the two genes in the long sleepers ($\rho_l = 0.89$) that was significantly different from controls; however, intermediate covariance in the short sleepers ($\rho_s = 0.53$) did overlap with that of controls, and therefore was not significantly different.

Figure 3 - supplement 1 shows a pair where interactions in both short and long selection schemes are different from controls, Figure 3 - supplement 2 shows another pair of genes where neither scheme is different from controls. This illustrates a range of possibilities, including a case where Spearman correlations are significant but GP correlations are not (the opposite also occurs). Figure 3 - supplements 3 and 4 fit each gene individually, and the fit does not change substantially between single to multiple channel models.

The 85 single-channel fits were good despite varying levels of dispersion and occasional outliers, indicating no issues with the Gaussian Processes ability to fit the temporal patterns of any one gene. For the two-channel inference, upwards of 90% of the chains initially converged under the criterion that $0.95 < \hat{R} < 1.05$; because the inference method is stochastic it is expected that by chance some chains may not converge and/or mix well with their replicates. Chains that initially failed were rerun up to two times. After three runs over 99% of the chains converged; the reasons for lack of convergence of the remaining were not investigated further. Figure 4 shows six heat maps (one for each sex and selection scheme combination) with the correlations for all pairs of genes calculated as described in the previous figure, summarizing the inferred interactions. Of the 3,570 correlations, 1,612 were greater than 0.5 and 98 greater than 0.9.

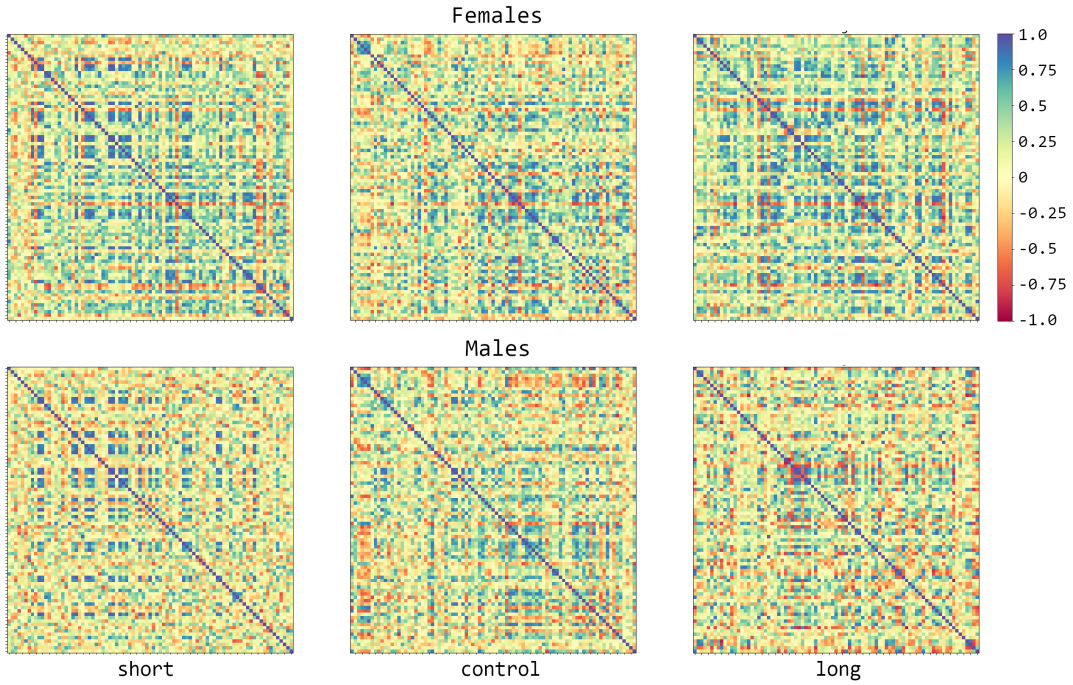


Figure 4. Signal variances and covariances normalized to range [-1,1] for females and males in each of the selection schemes: short, control, and long. Each off-diagonal square is the expected value of the interaction between two of 85 genes, for a total of 3,570 pairs.

In addition to computing expected values, the posterior distributions were used to compare the signal covariances between selection schemes and set a cutoff. Distributions of the parameter for each sex-selection scheme were assembled from the parallel MCMC runs; 145 gene pairs in the selected populations are found to be different from controls (i.e. do not overlap with them at 95% credibility for either short, long or both populations). Out of the 145, twelve gene pairs were com-

mon between males and females selected for long night sleep and one pair to males and females selected for short sleep; one gene pair was common to females in both selection schemes, and three pairs were common to males. Table S10 shows the expected values of signal covariances normalized by the variances for all two-way interactions side by side with the Spearman correlations. Table S11 shows the subset of significant Gaussian Processes correlations.

We constructed a network for each sex/selection scheme combination based on the magnitude of the correlation between genes. The network for males selected for long sleep having significant gene interactions is shown in Figure 5 (supplements 1-3 show the networks for the remaining three sex-selection scheme combinations).

For comparison, looking at significant ($\rho_{sel} \neq 0$) Spearman correlations keeps almost three thousand interactions (i.e. excludes just a bit more than a tenth of the genes), and comparing the distributions ρ_{sel} versus $\rho_{control}$ – similar to how the Gaussian Processes are compared – still has over thirteen hundred. Therefore, computing correlations between genes using covariance estimates from the Gaussian Processes greatly increases specificity over direct correlations. Furthermore, the Gaussian Processes are not only more specific but more sensitive in finding 68 gene pairs that are not found to be significant by the first Spearman approach and 18 not found by the second.

Finally, we examined known interactions between the 85 genes and any other genes using the *Drosophila* Interaction Database, DroID (***Murali et al., 2011***). We found 2,830 interactions; 8 of these were one of the 3,570 between the 85 genes, but none of them overlapped with the 145 gene pairs found to be different from controls. The gene interactions we observed may therefore be unique to extreme sleep.

Mutational analyses confirms role of candidate genes and interacting gene expression networks in sleep

We tested five genes for differences in sleep as compared to their isogenic control: *CG12560*, *CG13793*, *Cytochrome P450 6a16* (*Cyp6a16*), *highwire* (*hiw*), and *Jonah 65Aii* (*Jon65Aii*) (Table S12). All of the *Minos* insertions altered night sleep (Figure 6). Night sleep increased from 50—115 minutes beyond the w^{1118} control line (all P-values < 0.0125, the Bonferroni-corrected P-value; Table S12). Flies having a *Minos* insertion in *Jon65Aii* slept 66 minutes less than their corresponding y^1w^{67c23}control ($P < 0.0001$). Effects of the *Minos* insertions on night sleep duration were stronger in females than in males; when sexes were examined separately, only mutations in *CG12560* and *Jon65Aii* affected male night sleep duration (Table S12). Thus, all genes affected night sleep duration, but *CG12560* and *Jon65Aii* had the greatest effect on both sexes.

Gene expression decreased significantly in the *CG12560* and *Jon65Aii Minos* insertions relative to their controls (Figure 6 - supplement 1). The remaining *Minos* insertion lines had some changes in gene expression relative to the control, however, the changes were not formally significant. Potential reasons for the lack of a significant change in gene expression in the remaining lines include the position of the insertion within the targeted gene, which has variable effects on its expression, and the relatively low statistical power of the experiment.

Our baseline expectation was expression levels between knockdown and control lines should not be affected for most genes. For candidate genes, we hypothesized that the ratio of gene expression between the *Minos* insertion line and its respective control would differ considerably from 1.0; conversely, the ratio of gene expression between *Minos* insertion line and control would be approximately 1.0 for unrelated genes – that expectation is confirmed by the distributions consistently centered around unity (median $= 0.995$). We plotted the ratios of candidate genes against the distributions of the matching random sets (Figure 7). Our supposition was largely realized for *CG12560* and *Jon65Aii*, the two genes having significant knockdown in gene expression (Figure 7 and Supplemental Table S13).

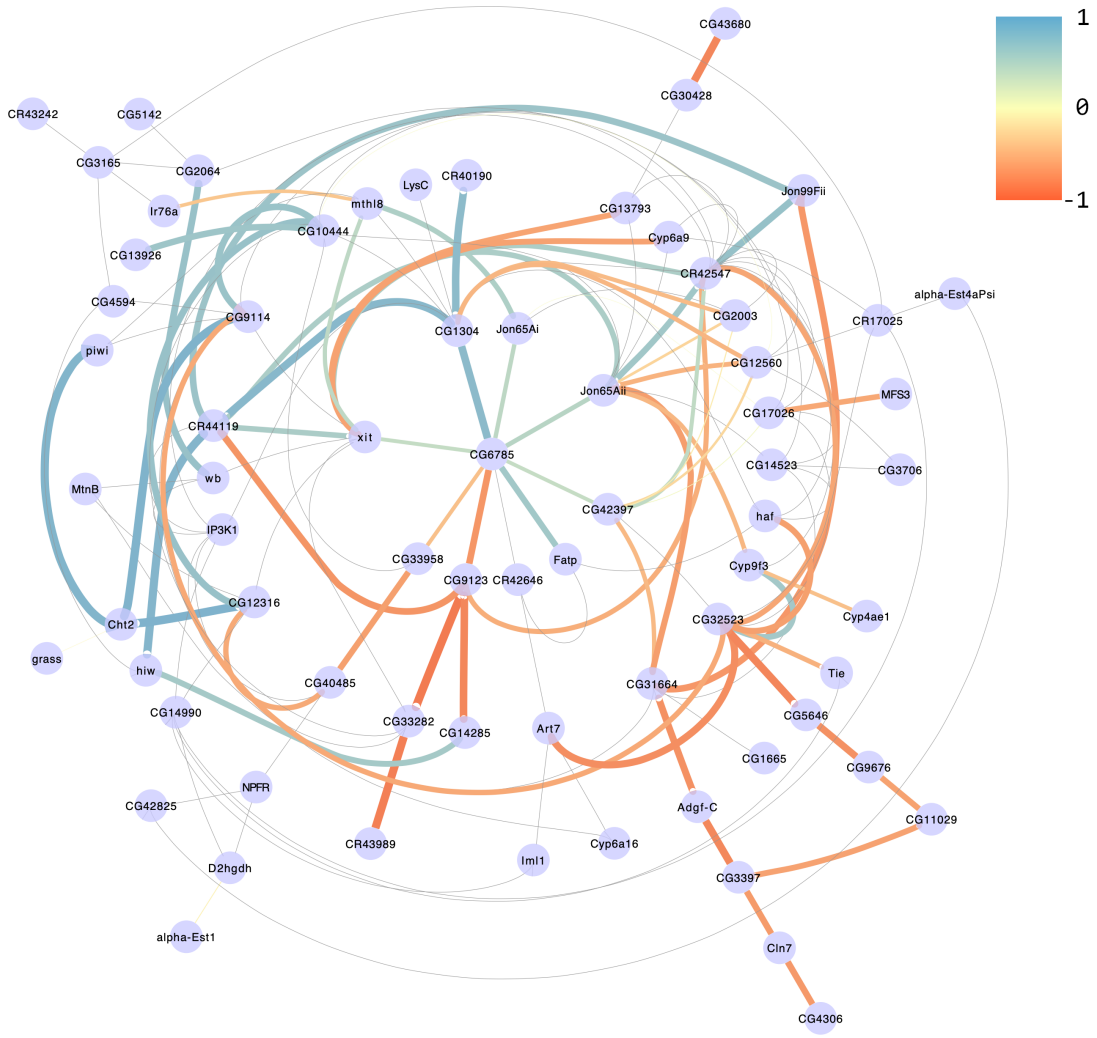


Figure 5. Gene interaction network in males selected for long sleep. Edges represent signal covariances whose posterior distributions do not overlap with that of controls at 95% credibility. Colors and line thickness indicate indicate the strength and the direction of the correlation. Thin gray lines show all 145 interactions significant for at least one of the four sex-selection scheme combinations.

Figure 5—figure supplement 1. Male, short sleepers

Figure 5—figure supplement 2. Female, long sleepers

Figure 5—figure supplement 3. Female, short sleepers

Discussion

We have shown that robust, reproducible phenotypic changes in *Drosophila melanogaster* sleep are associated with hundreds (405 in males, 820 in females) of individual shifts in gene expression – and as a consequence hundreds of thousands of potential combinations [$\binom{405}{2} > 8 \cdot 10^4$ and $\binom{820}{2} > 3 \cdot 10^5$]. Nevertheless, unique interactions important to the phenotypes are a comparatively small number (145 out of $\binom{85}{2} = 3570$ possible combinations of the 85 genes common to males and females). We have also shown that these interactions cannot be found with linear model analyses or conventional correlation calculations only, but are specifically identified using a combination of an informative experimental design with densely-sampled time points to generate a large scale data set, and a nonparametric, nonlinear model-based approach that explicitly accounts for covariance in gene expression. That complex traits can be mostly explained by additive effects of individual genes (and their expression) is a common and sometimes useful assumption. While

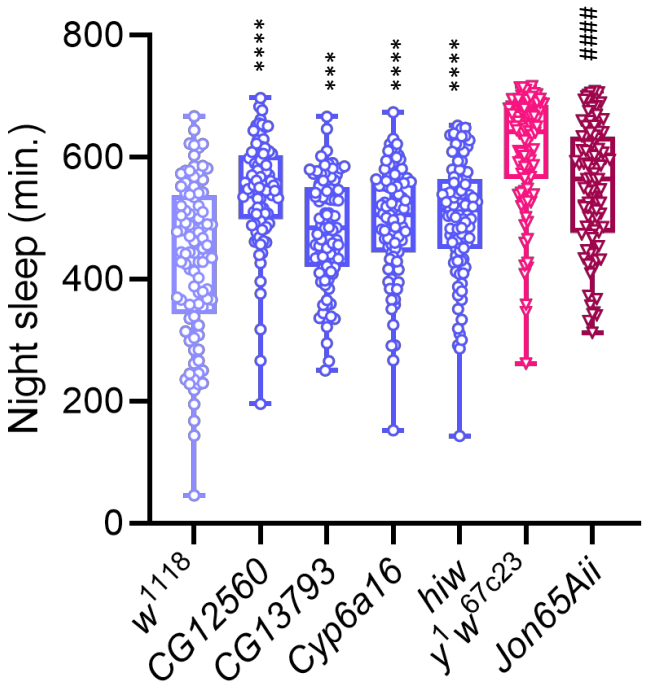


Figure 6. Night sleep in *Minos* mutations. The figure compares night sleep duration in each mutant with that of its respective control. Blue indicates *Minos* insertions with w^{1118} control strain; Red indicates *Minos* insertions with $y^1 w^{67c23}$ control strain. **** or $^{\#\#\#\#}$, $P < 0.0001$; ***, $P < 0.001$, three-way ANOVA. All P-values are less than the Bonferroni-corrected P-value of 0.0125.

Figure 6—figure supplement 1. Gene expression in *Minos* mutants. For each candidate gene, the gene expression in the *Minos* mutant and corresponding control are plotted. * or $^{\#}$, $P < 0.05$ by Kruskal-Wallis test. **(a)** *CG12560*; **(b)** *Jon65Aii*; **(c)** *CG13793*; **(d)** *Cyp6a16*; **(e)** *hiw*

it underpins preliminary analyses that allow whole-transcriptome data to be understood, it eliminates the ability to infer interactions between them from the data and stops short from identifying relevant processes. Complex traits involve multiple genes, and the actual interactions giving rise to phenotypes are likely to be highly nonlinear (***Mackay, 2014***). These nonlinearities are not a mathematical construct, but a biological reality arising from chemical kinetics. Favoring approaches that account for these features will not only increase statistical power, but understanding of actual biological mechanisms beyond simple network representations of gene expression (***DiFrisco and Jaeger, 2020***).

In most correlation and information-theory based methods the dimension (e.g. time or space) across which samples covary is only implicit (***Emmert-Streib et al., 2012***); the only possible conclusion from a significant correlation between two sets of observations is that one may have an effect on the other – i.e. the data alone does not allow the distinction between actual interactions and spurious correlation. Bioinformatic pipelines that have correlation as their starting point – in addition to carrying over its limitations – are not straightforwardly comparable to our approach (see Appendix 1). In the context of Gaussian Processes, correlation between all pairs of data points – including within the same time series, i.e. autocorrelation – is explicit in time (or other dimension), so similar trends do not necessarily imply covariance between the sets of observations. Therefore, on the one hand GPs are a nonparametric method that requires no more biological knowledge than that for computing a linear correlation; on the other hand, while not an explicit description of dynamic biological processes, it is also a model-based approach that can be used within more mechanistic formalisms like differential equations (***Äijö et al., 2013***), or potentially be used to formulate specific hypotheses and build mechanistic models.

Although somewhat self-evident, it is important to highlight the fact that to describe correla-

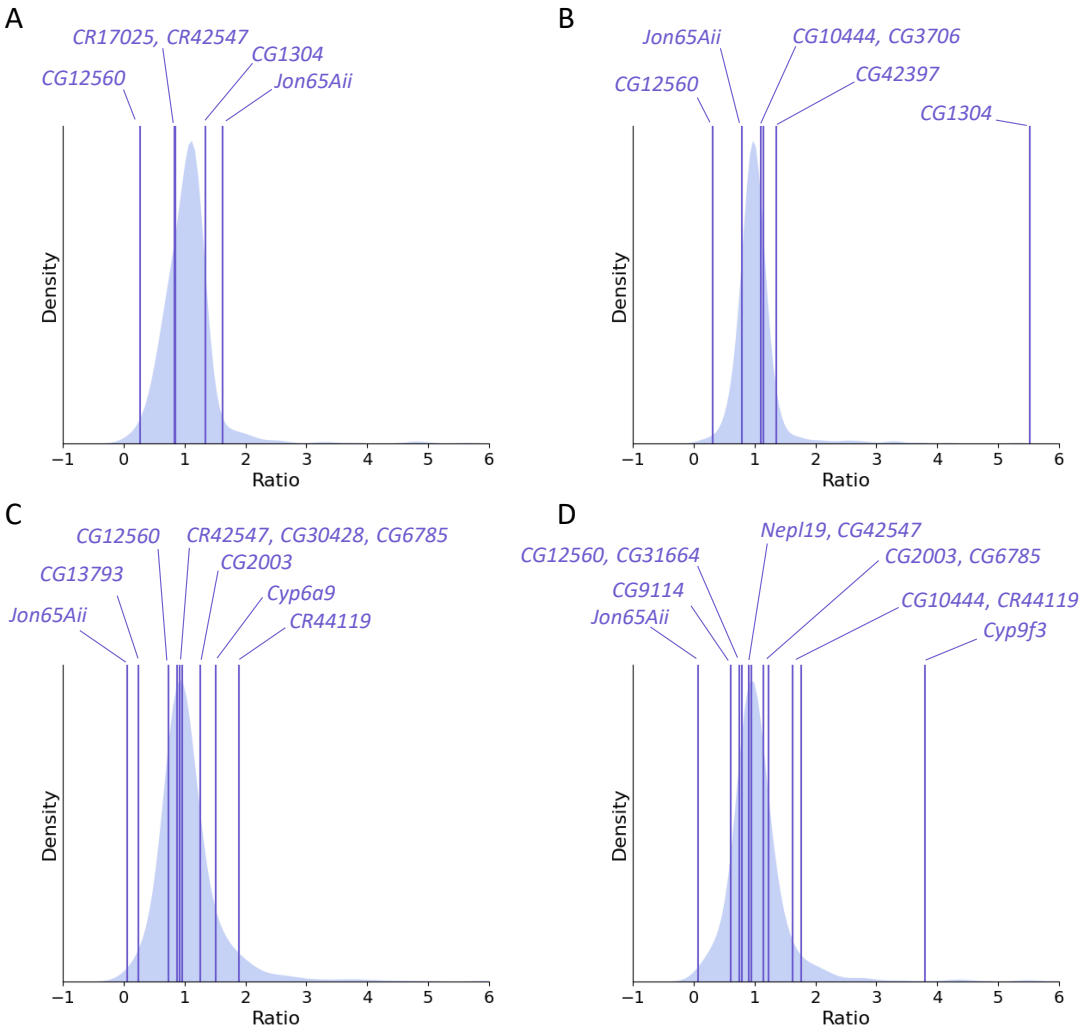


Figure 7. Comparison of ratios of gene expression between genes with significant Gaussian Process correlations and unrelated genes for *CG12560* and *Jon65Aii* mutants. Purple lines show the ratio of mutant gene expression to control for genes with significant Gaussian Process correlations. The distribution of gene ratios for $1,000$ unrelated genes is plotted in the background. Genes having the most extreme ratios are indicated; see Table S13 for the calculations. **(a)** *CG12560* females; **(b)** *CG12560* males; **(c)** *Jon65Aii* females; **(d)** *Jon65Aii* males.

Figure 7—figure supplement 1. Comparison of ratios of gene expression between genes with significant Gaussian Process correlations and unrelated genes for *CG13793*, *Cyp6a16*, and *hiw Minos* mutants. **(a)** *CG13793* females; **(b)** *CG13793* males; **(c)** *Cyp6a16* males; **(d)** *hiw* females; **(e)***hiw* males.

tions along time, multiple time points are needed – put another way, the use of a nonlinear model requires enough resolution in the data that the trajectory can be identified. To that end, a single high-resolution, large data set with a specific design, like the one generated in this work, will be more useful than several small data sets, for instance with only initial and final time points and allowing only two-sample linear comparison. Gene expression measured at the terminal generation of selection and compared among selected and control groups does identify candidate genes (***Pegoraro et al., 2020***; ***Brown et al., 2017***; ***Mackay et al., 2005***; ***Wertheim et al., 2011***; ***Sørensen et al., 2007***; ***Morozova et al., 2007***; ***Edwards et al., 2006***), but the relationship between pairs of genes is lost. Some studies evaluated gene expression during the last 2-3 generations of selection (***Telonis-Scott et al., 2009***; ***Garlapow et al., 2017***); however, the additional sampling was used to confirm consistency rather than change across time. Our approach of sampling over time enabled

us to derive interactions between genes and demonstrated that unique gene expression network profiles develop in long sleepers as compared to short sleepers.

When employing methods of increasing complexity or sophistication there is always the question of how relevant the inference is or, in other words, how "real" are the parameters or processes in the model. This pursuit of simplicity may favor the use of methods based on linear models as more palpable approaches and less prone to arbitrary assumptions about how the parameters are put together; however, it is important to realize that linear coefficients are no more real than those of any other model. On the contrary, biological processes are not restricted by our ability to comprehend them. Therefore, what may seem as an Occam's Razor-like simplicity will probably hinder accurate description of nature. Systems-level understanding of complex biology requires not only more and more detailed data, but better descriptions of the processes and methodology that captures higher-order phenomena. Equivalently, experimental validation of these phenomena will be more technically challenging to accomplish. Despite the additional difficulties, it must be recognized that methods that cannot possibly match the complexity of nature are doomed to scratch all over the surface without realizing a deeper understanding.

We observed changes in night sleep duration for all *Minos* insertions tested, and the effects were more prominent in females. Sex-specific effects of mutations on sleep are common in flies (***Harbison et al., 2009***; ***Williams et al., 2007***; ***Afonso et al., 2015***; ***Glover et al., 2019***; ***De Luca et al., 2010***; ***Álvarez-Rendón and Riesgo-Escovar, 2020***), and sex-biased effects are often noted in females (***Bushey et al., 2007***, ***2009***; ***Luo et al., 2012***; ***Huang et al., 2021***; ***Juneau et al., 2019***). Remarkably, we noted gene expression relationships among genes with predicted significant Gaussian Process correlations in the *Minos* insertion lines despite the fact that the sleep was neither extremely long or short in mutants or controls, and the genetic background of these lines (w^{1118} or y^1w^{67c23}) is completely different from the outbred Sleep Advanced Intercross Population that we used for artificial selection.

Acknowledgments

We thank the members of the NISC Consortium for sequence data and helpful discussions. This work used the computational resources of the National Institutes of Health High-Performance Computing Biowulf cluster (http://hpc.nih.gov). This research was supported by the Intramural Research Program of the National Institutes of Health, the National Heart Lung and Blood Institute. We thank N. Redekar and N. Gulzar from the NIAID Collaborative Bioinformatics Resource (NCBR) for assistance with bioinformatic processing.

Author Contributions

Conceptualization: C.S.-M., S.T.H.; Investigation: C.S.-M., Y.L.S.N., Y.L. Data curation and formal analysis: C.S.-M., Y.L., S.T.H. Writing: C.S.-M., S.T.H.

Data Availability

All RNA-Seq data from this study are available from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) under the accession number GSE202600.

Competing Interests

The authors have no competing interests to declare.

References

Aalto A, Viitasaari L, Ilmonen P, Mombaerts L, Gonçalves J. Gene regulatory network inference from sparsely sampled noisy data. Nature Communications. 2020; 11(1). doi: 10.1038/s41467-020-17217-1.

Afonso DJ, Liu D, Machado DR, Pan H, Jepson JE, Rogulja D, Koh K. TARANIS functions with cyclin A and Cdk1 in a novel arousal center to control sleep in Drosophila. Current Biology. 2015; 25(13):1717–1726.

Äijö T, Granberg K, Lähdesmäki H. Sorad: a systems biology approach to predict and modulate dynamic signaling pathway response from phosphoproteome time-course measurements. Bioinformatics. 2013 may; 29(10):1283–1291. doi: 10.1093/bioinformatics/btt130.

Álvarez-Rendón JP, Riesgo-Escovar JR. Circadian and rhythmic-related behavioral co-morbidities of the diabetic state in Drosophila melanogaster. General and Comparative Endocrinology. 2020; 295:113477.

Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015; doi: 10.1093/bioinformatics/btu638.

Arbeitman MN, Furlong EEM, Imam F, Johnson E, Null BH, Baker BS, Krasnow MA, Scott MP, Davis RW, White KP. Gene Expression During the Life Cycle of Drosophila melanogaster. Science. 2002; 297(5590):2270–2275. doi: 10.1126/science.1072152.

Arnol D, Schapiro D, Bodenmiller B, Saez-Rodriguez J, Stegle O. Modeling Cell-Cell Interactions from Spatial Molecular Data with Spatial Variance Component Analysis. Cell Reports. 2019; 29(1):202–211.e6. doi: 10.1016/j.celrep.2019.08.077.

Austin PC, Hux JE. A brief note on overlapping confidence intervals. Journal of Vascular Surgery. 2002; 36(1):194–195. doi: 10.1067/mva.2002.125015.

Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RRH, Mackay TFC. Systems genetics of complex traits in Drosophila melanogaster. Nature Genetics. 2009 mar; 41(3):299–307. doi: 10.1038/ng.332.

Bahg G, Evans DG, Galdo M, Turner BM. Gaussian process linking functions for mind, brain, and behavior. Proceedings of the National Academy of Sciences of the United States of America. 2020; 117(47):29398–29406. doi: 10.1073/pnas.1912342117.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995 jan; 57(1):289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x.

Berger RJ, Phillips NH. Energy conservation and sleep. Behav Brain Res. 1995; 69:65–73. doi: 10.1016/0166-4328(95)00002-b.

Bishop CM. Pattern recognition and machine learning. Springer; 2006.

Bonilla EV, Chai KMA, Williams CKI. Multi-task Gaussian Process prediction. In: *Advances in Neural Information Processing Systems 20* NIPS Foundation; 2008. p. 153–160.

Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017 jun; 169(7):1177–1186. doi: 10.1016/j.cell.2017.05.038.

Brown EB, Layne JE, Elchert AR, Rollmann SM. Behavioral and transcriptional response to selection for olfactory behavior in Drosophila. G3: Genes, Genomes, Genetics. 2020; 10(4):1283–1296. doi: 10.1534/g3.120.401117.

Brown EB, Patterson C, Pancoast R, Rollmann SM. Artificial selection for odor-guided behavior in Drosophila reveals changes in food consumption. BMC Genomics. 2017; 18(1):1–13. doi: 10.1186/s12864-017-4233-1.

Bushey D, Huber R, Tononi G, Cirelli C. Drosophila Hyperkinetic mutants have reduced sleep and impaired memory. Journal of Neuroscience. 2007; 27(20):5384–5393.

Bushey D, Tononi G, Cirelli C. The Drosophila fragile X mental retardation gene regulates sleep need. Journal of Neuroscience. 2009; 29(7):1948–1961.

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. Stan : A Probabilistic Programming Language. Journal of Statistical Software. 2017; 76(1). doi: 10.18637/jss.v076.i01.

Dashti HS, Jones SE, Wood AR, Lane JM, van Hees VT, Wang H, Rhodes JA, Song Y, Patel K, Anderson SG, Beaumont RN, Bechtold DA, Bowden J, Cade BE, Garaulet M, Kyle SD, Little MA, Loudon AS, Luik AI, Scheer FAJL, et al. Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. Nature Communications. 2019; 10(1). doi: 10.1038/s41467-019-08917-4.

De Luca M, Klimentidis YC, Casazza K, Moses Chambers M, Cho R, Harbison ST, Jumbo-Lucioni P, Zhang S, Leips J, Fernandez JR. A conserved role for syndecan family members in the regulation of whole-body energy metabolism. PloS one. 2010; 5(6):e11286.

Diessler S, Jan M, Emmenegger Y, Guex N, Middleton B, Skene DJ, Ibberson M, Burdet F, Götz L, Pagni M, Sankar M, Liechti R, Hor CN, Xenarios I, Franken P. A systems genetics resource and analysis of sleep regulation in the mouse. PLoS Biology. 2018; 16(8). doi: 10.1371/journal.pbio.2005750.

DiFrisco J, Jaeger J. Genetic Causation in Complex Regulatory Systems: An Integrative Dynamic Perspective. BioEssays. 2020 jun; 42(6):1900226. doi: 10.1002/bies.201900226.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013; doi: 10.1093/bioinformatics/bts635.

Edwards AC, Rollmann SM, Morgan TJ, Mackay TFC. Quantitative Genomics of Aggressive Behavior in Drosophila melanogaster. PLoS Genetics. 2006; 2(9):1386–1395. doi: 10.1371/journal.pgen.0020154.

Emmert-Streib F, Glazko GV, Altay G, Simoes RdM. Statistical inference and reverse engineering of gene regulatory networks from observational expression data. Frontiers in Genetics. 2012; 3(FEB):1–15. doi: 10.3389/fgene.2012.00008.

Falconer DS, Mackay TFC. Introduction to Quantitative Genetics (Fourth Edition); 1996.

Faria VG, Martins NE, Magalhães S, Paulo TF, Nolte V, Schlötterer C, Sucena É, Teixeira L. Drosophila Adaptation to Viral Infection through Defensive Symbiont Evolution. PLoS Genetics. 2016; 12(9):1–18. doi: 10.1371/journal.pgen.1006297.

Faria VG, Martins NE, Paulo T, Teixeira L, Sucena É, Magalhães S. Evolution of Drosophila resistance against different pathogens and infection routes entails no detectable maintenance costs. Evolution. 2015 nov; 69(11):2799–2809. doi: 10.1111/evo.12782.

Ganguly-Fitzgerald I, Donlea J, Shaw PJ. Waking Experience Affects Sleep Need in Drosophila. Science. 2006 sep; 313(5794):1775–1781. doi: 10.1126/science.1130408.

Garlapow ME, Everett LJ, Zhou S, Gearhart AW, Fay KA, Huang W, Morozova TV, Arya GH, Turlapati L, St Armour G, Hussain YN, McAdams SE, Fochler S, Mackay TFC. Genetic and Genomic Response to Selection for Food Consumption in Drosophila melanogaster. Behavior Genetics. 2017; 47(2):227–243. doi: 10.1007/s10519-016-9819-x.

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis, third edition; 2013.

Glover Z, Hodges MD, Dravecz N, Cameron J, Askwith H, Shirras A, Broughton SJ. Loss of angiotensin-converting enzyme-related (ACER) peptidase disrupts behavioural and metabolic responses to diet in Drosophila melanogaster. Journal of Experimental Biology. 2019; 222(8):jeb194332.

Hammerschlag AR, Stringer S, de Leeuw CA, Sniekers S, Taskesen E, Watanabe K, Blanken TF, Dekker K, te Lindert BHW, Wassing R, Jonsdottir I, Thorleifsson G, Stefansson H, Gislason T, Berger K, Schormair B, Wellmann J, Winkelmann J, Stefansson K, Oexle K, et al. Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. Nature Genetics. 2017 nov; 49(11):1584–1592. doi: 10.1038/ng.3888.

Harbison ST, Carbone MA, Ayroles JF, Stone EA, Lyman RF, Mackay TFC. Co-regulated transcriptional networks contribute to natural genetic variation in Drosophila sleep. Nature Genetics. 2009; doi: 10.1038/ng.330.

Harbison ST, Chang S, Kamdar KP, Mackay TFC. Quantitative genomics of starvation stress resistance in Drosophila. Genome biology. 2005; 6:R36. doi: 10.1186/gb-2005-6-4-r36.

Harbison ST, McCoy LJ, Mackay TFC. Genome-wide association study of sleep in Drosophila melanogaster. BMC Genomics. 2013; 14(1):281. doi: 10.1186/1471-2164-14-281.

Harbison ST, Sehgal A. Quantitative Genetic Analysis of Sleep in Drosophila melanogaster. Genetics. 2008 apr; 178(4):2341–2360. doi: 10.1534/genetics.107.081232.

Harbison ST, Serrano Negron YL, Hansen NF, Lobell AS. Selection for long and short sleep duration in Drosophila melanogaster reveals the complex genetic network underlying natural variation in sleep. PLOS Genetics. 2017 dec; 13(12):e1007098. doi: 10.1371/journal.pgen.1007098.

Hill VM, O'Connor RM, Shirasu-Hiza M. Tired and stressed: Examining the need for sleep. European Journal of Neuroscience. 2020; 51(1):494–508. doi: 10.1111/ejn.14197.

Honkela A, Girardot C, Gustafson EH, Liu YH, Furlong EEM, Lawrence ND, Rattray M. Model-based method for transcription factor target identification with limited data. Proceedings of the National Academy of Sciences. 2010 apr; 107(17):7793–7798. doi: 10.1073/pnas.0914285107.

Huang H, Possidente DR, Vecsey CG. Optogenetic activation of SIFamide (SIFa) neurons induces a complex sleep-promoting effect in the fruit fly Drosophila melanogaster. Physiology & Behavior. 2021; 239:113507.

Huang W, Carbone MA, Magwire MM, Peiffer JA, Lyman RF, Stone EA, Anholt RRH, Mackay TFC. Genetic basis of transcriptome diversity in Drosophila melanogaster. Proceedings of the National Academy of Sciences. 2015 nov; 112(44):E6010–E6019. doi: 10.1073/pnas.1519159112.

Huang W, Massouras A, Inoue Y, Peiffer J, Ramia M, Tarone AM, Turlapati L, Zichner T, Zhu D, Lyman RF, Magwire MM, Blankenburg K, Carbone MA, Chang K, Ellis LL, Fernandez S, Han Y, Highnam G, Hjelmen CE, Jack JR, et al. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. Genome Research. 2014 jul; 24(7):1193–1208. doi: 10.1101/gr.171546.113.

Huylmans AK, Parsch J. Population- and Sex-Biased Gene Expression in the Excretion Organs of Drosophila melanogaster. G3: Genes|Genomes|Genetics. 2014 dec; 4(12):2307–2315. doi: 10.1534/g3.114.013417.

Jansen PR, Watanabe K, Stringer S, Skene N, Bryois J, Hammerschlag AR, de Leeuw CA, Benjamins JS, Muñoz-Manchado AB, Nagel M, Savage JE, Tiemeier H, White T, Tung JY, Hinds DA, Vacic V, Wang X, Sullivan PF, van der Sluis S, Polderman TJC, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. Nature Genetics. 2019 mar; 51(3):394–403. doi: 10.1038/s41588-018-0333-3.

Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G. The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster. Nature Genetics. 2001 dec; 29(4):389–395. doi: 10.1038/ng766.

Joiner WJ. Unraveling the Evolutionary Determinants of Sleep. Current Biology. 2016; 26(20):R1073–R1087. doi: 10.1016/j.cub.2016.08.068.

Jones SE, Tyrrell J, Wood AR, Beaumont RN, Ruth KS, Tuke MA, Yaghootkar H, Hu Y, Teder-Laving M, Hayward C, Roenneberg T, Wilson JF, Del Greco F, Hicks AA, Shin C, Yun CH, Lee SK, Metspalu A, Byrne EM, Gehrman PR, et al. Genome-Wide Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration Loci. PLOS Genetics. 2016 aug; 12(8):e1006125. doi: 10.1371/journal.pgen.1006125.

Joshi SS, Sethi M, Striz M, Cole N, Denegre JM, Ryan J, Lhamon ME, Agarwal A, Murray S, Braun RE, Fardo DW, Kumar V, Donohue KD, Sunderam S, Chesler EJ, Svenson KL, O'Hara BF. Noninvasive sleep monitoring in large-scale screening of knock-out mice reveals novel sleep-related genes. bioRxiv. 2019; doi: 10.1101/517680.

Juneau BA, Stonemetz JM, Toma RF, Possidente DR, Heins RC, Vecsey CG. Optogenetic activation of short neuropeptide F (sNPF) neurons induces sleep in Drosophila melanogaster. Physiology & behavior. 2019; 206:143–156.

Kontio JAJ, Sillanpää MJ. Scalable Nonparametric Prescreening Method for Searching Higher-Order Genetic Interactions Underlying Quantitative Traits. Genetics. 2019 dec; 213(4):1209–1224. doi: 10.1534/genetics.119.302658.

Krueger JM, Obál F. A neuronal group theory of sleep function. Journal of Sleep Research. 1993 jun; 2(2):63–69. doi: 10.1111/j.1365-2869.1993.tb00064.x.

Laing EE, Möller-Levet CS, Dijk DJ, Archer SN. Identifying and validating blood mRNA biomarkers for acute and chronic insufficient sleep in humans: A machine learning approach. Sleep. 2019; 42(1):1–18. doi: 10.1093/sleep/zsy186.

Lane JM, Jones SE, Dashti HS, Wood AR, Aragam KG, van Hees VT, Strand LB, Winsvold BS, Wang H, Bowden J, Song Y, Patel K, Anderson SG, Beaumont RN, Bechtold DA, Cade BE, Haas M, Kathiresan S, Little MA, Luik AI, et al. Biological and clinical insights from genetics of insomnia symptoms. Nature Genetics. 2019 mar; 51(3):387–393. doi: 10.1038/s41588-019-0361-7.

Lin Y, Chen ZX, Oliver B, Harbison ST. Microenvironmental gene expression plasticity among individual drosophila melanogaster. G3: Genes, Genomes, Genetics. 2016; 6(12):4197–4210. doi: 10.1534/g3.116.035444.

Liu ZP. Reverse Engineering of Genome-wide Gene Regulatory Networks from Gene Expression Data. Current Genomics. 2015; 16(1):3–22. doi: 10.2174/1389202915666141110210634.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014 dec; 15(12):550. doi: 10.1186/s13059-014-0550-8.

Luo W, Chen WF, Yue Z, Chen D, Sowcik M, Sehgal A, Zheng X. Old flies have a robust central oscillator but weaker behavioral rhythms that can be improved by genetic and environmental manipulations. Aging cell. 2012; 11(3):428–438.

Ly S, Pack AI, Naidoo N. The neurobiological basis of sleep: Insights from Drosophila. Neuroscience and Biobehavioral Reviews. 2018; 87:67–86. doi: 10.1016/j.neubiorev.2018.01.015.

MacKay DJ. Information theory, inference and learning algorithms. Cambridge university press; 2003.

Mackay TFC, Heinsohn SL, Lyman RF, Moehring AJ, Morgan TJ, Rollmann SM. Genetics and genomics of Drosophila mating behavior. Proceedings of the National Academy of Sciences. 2005 may; 102(Supplement 1):6622–6629. doi: 10.1073/pnas.0501986102.

Mackay TFC, Richards S, Stone EA, Barbadilla A, Ayroles JF, Zhu D, Casillas S, Han Y, Magwire MM, Cridland JM, Richardson MF, Anholt RRH, Barrón M, Bess C, Blankenburg KP, Carbone MA, Castellano D, Chaboub L, Duncan L, Harris Z, et al. The Drosophila melanogaster Genetic Reference Panel. Nature. 2012 feb; 482(7384):173–178. doi: 10.1038/nature10811.

Mackay TFC. Epistasis and quantitative traits: Using model organisms to study gene-gene interactions. Nature Reviews Genetics. 2014; 15(1):22–33. doi: 10.1038/nrg3627.

Mackay TF, Lyman RF. Drosophila bristles and the nature of quantitative genetic variation. Philosophical Transactions of the Royal Society B: Biological Sciences. 2005; 360(1459):1513–1527.

McDowell IC, Manandhar D, Vockley CM, Schmid AK, Reddy TE, Engelhardt BE. Clustering gene expression time series data using an infinite Gaussian process mixture model. PLoS Computational Biology. 2018; 14(1):1–27. doi: 10.1371/journal.pcbi.1005896.

Melkumyan A, Ramos F. Multi-kernel Gaussian processes. In: *IJCAI International Joint Conference on Artificial Intelligence*; 2011. p. 1408–1413. doi: 10.5591/978-1-57735-516-8/IJCAI11-238.

Morozova TV, Anholt RRH, Mackay TFC. Phenotypic and transcriptional response to selection for alcohol sensitivity in Drosophila melanogaster. Genome Biology. 2007; 8(10):1–15. doi: 10.1186/gb-2007-8-10-r231.

Murali T, Pacifico S, Yu J, Guest S, Roberts GG, Finley RL. DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. Nucleic Acids Research. 2011 jan; 39(suppl_1):D736–D743. doi: 10.1093/nar/gkq1092.

Parisi M, Nuttall R, Naiman D, Bouffard G, Malley J, Andrews J, Eastman S, Oliver B. Paucity of Genes on the Drosophila X Chromosome Showing Male-Biased Expression. Science. 2003; 299(5607):697–700. doi: 10.1126/science.1079190.

Pegoraro M, Flavell LMM, Menegazzi P, Colombi P, Dao P, Helfrich-Forster C, Tauber E. The genetic basis of diurnal preference in Drosophila melanogaster. BMC Genomics. 2020; 21(1). doi: 10.1186/s12864-020-07020-z.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. Sex-dependent gene expression and evolution of the Drosophila transcriptome. Science. 2003; 300(5626):1742–1745.

Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. The MIT Press; 2006. doi: 10.7551/mitpress/3206.001.0001.

Ruscio J. Constructing confidence intervals for Spearman's rank correlation with ordinal data: A simulation study comparing analytic and bootstrap methods. Journal of Modern Applied Statistical Methods. 2008; 7(2):416–434. doi: 10.22237/jmasm/1225512360.

Scharf MT, Naidoo N, Zimmerman JE, Pack AI. The energy hypothesis of sleep revisited. Progress in Neurobiology. 2008; 86(3):264–280. doi: 10.1016/j.pneurobio.2008.08.003.

Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. Heredity. 2015 may; 114(5):431–440. doi: 10.1038/hdy.2014.86.

Schmidt MH. The energy allocation function of sleep: A unifying theory of sleep, torpor, and continuous wakefulness. Neuroscience & Biobehavioral Reviews. 2014 nov; 47:122–153. doi: 10.1016/j.neubiorev.2014.08.001.

Schulz E, Speekenbrink M, Krause A. A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions. Journal of Mathematical Psychology. 2018; doi: 10.1016/j.jmp.2018.03.001.

Serrano Negron YL, Hansen NF, Harbison ST. The Sleep Inbred Panel, a Collection of Inbred Drosophila melanogaster with Extreme Long and Short Sleep Duration. G3: Genes|Genomes|Genetics. 2018 sep; 8(9):2865–2873. doi: 10.1534/g3.118.200503.

Sørensen JG, Nielsen MM, Loeschcke V. Gene expression profile analysis of Drosophila melanogaster selected for resistance to environmental stressors. Journal of Evolutionary Biology. 2007; 20(4):1624–1636. doi: 10.1111/j.1420-9101.2007.01326.x.

Telonis-Scott M, Hallas R, McKechnie SW, Wee CW, Hoffmann AA. Selection for cold resistance alters gene transcript levels in Drosophila melanogaster. Journal of Insect Physiology. 2009; 55(6):549–555. doi: 10.1016/j.jinsphys.2009.01.010.

Tononi G, Cirelli C. Sleep and the Price of Plasticity: From Synaptic and Cellular Homeostasis to Memory Consolidation and Integration. Neuron. 2014; 81(1):12–34. doi: 10.1016/j.neuron.2013.12.025.

Velten B, Braunger JM, Arnol D, Argelaguet R, Stegle O. Identifying temporal and spatial patterns of variation from multi- modal data using MEFISTO. bioRxiv. 2020 dec; p. 2020.11.03.366674. doi: 10.1101/2020.11.03.366674.

Villaverde AF, Banga JR. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. Journal of The Royal Society Interface. 2014 feb; 11(91):20130505. doi: 10.1098/rsif.2013.0505.

Wayne ML, Telonis-Scott M, Bono LM, Harshman L, Kopp A, Nuzhdin SV, McIntyre LM. Simpler mode of inheritance of transcriptional variation in male Drosophila melanogaster. Proceedings of the National Academy of Sciences. 2007 nov; 104(47):18577–18582. doi: 10.1073/pnas.0705441104.

Wertheim B, Kraaijeveld AR, Hopkins MG, Walther Boer M, Godfray HCJ. Functional genomics of the evolution of increased resistance to parasitism in Drosophila. Molecular Ecology. 2011; 20(5):932–949. doi: 10.1111/j.1365-294X.2010.04911.x.

Williams JA, Sathyanarayanan S, Hendricks JC, Sehgal A. Interaction between sleep and the immune response in Drosophila: a role for the NFκB relish. Sleep. 2007; 30(4):389–400.

Xie L, Kang H, Xu Q, Chen MJ, Liao Y, Thiyagarajan M, O'Donnell J, Christensen DJ, Nicholson C, Iliff JJ, Takano T, Deane R, Nedergaard M. Sleep drives metabolite clearance from the adult brain. Science. 2013; 342(6156):373–377. doi: 10.1126/science.1241224.

Zhang Y, Malone JH, Powell SK, Periwal V, Spana E, MacAlpine DM, Oliver B. Expression in Aneuploid Drosophila S2 Cells. PLoS Biology. 2010 feb; 8(2):e1000320. doi: 10.1371/journal.pbio.1000320.

Zhang Y, Sturgill D, Parisi M, Kumar S, Oliver B. Constraint and turnover in sex-biased gene expression in the genus Drosophila. Nature. 2007 nov; 450(7167):233–237. doi: 10.1038/nature06323.

Supplemental Information

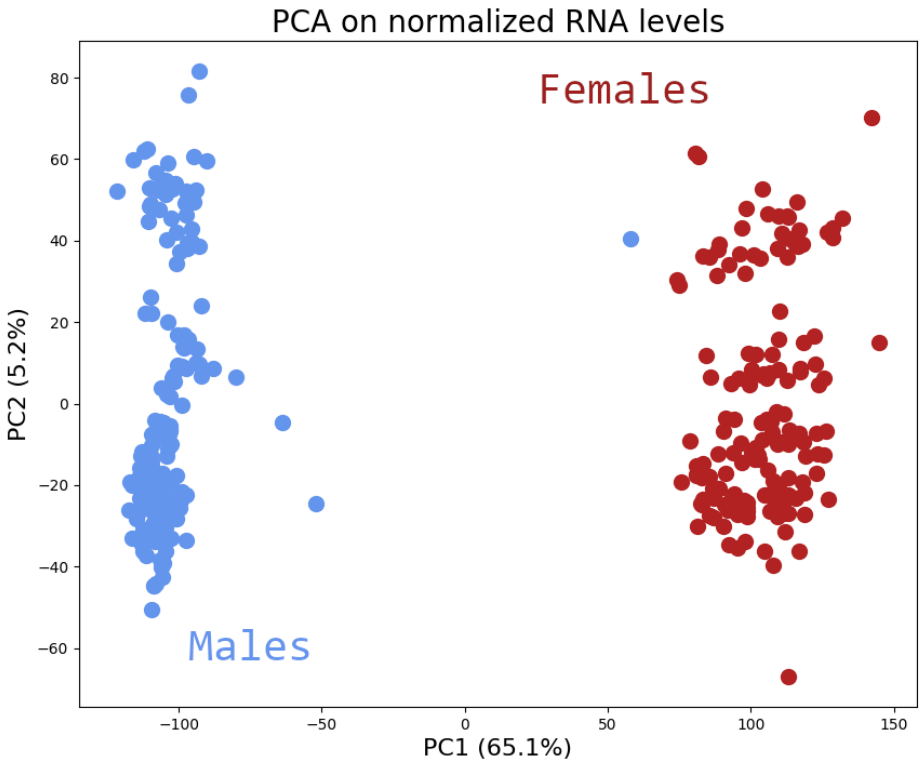


Figure S1. Principal Component Analysis on matrix of normalized expression data shows complete separation of sexes along the first component, which explains 65% of the variance in the data.

Table S1. Quantitative genetics of the response to selection for long or short night sleep and related sleep parameters. For each trait, the ANOVA analysis results are presented. Source indicates each factor in the model. gen, generation; rep, replicate; sel, selection; $d.f.$, degrees of freedom; M.S., Type III mean squares; F, F ratio statistic; P, P–value.

Table S2. Quantitative genetics of the response to selection for long or short night sleep per generation. For each sleep trait, the ANOVA analysis results are presented for each generation. Source indicates each factor in the model. rep, replicate; sel, selection; $d.f.$, degrees of freedom; M.S., Type III mean squares; F, F ratio statistic; P, P-value.

Table S3. Quantitative genetics of control populations. For each sleep trait, the ANOVA analysis results are presented. gen, generation; rep, replicate; sel, selection; $d.f.$, degrees of freedom; MS, Type III mean squares; F, F ratio statistic; P, P-value.

Table S4. Correlated response of sleep trait coefficient of environmental variance (CV_E) to selection for long or short night sleep duration. For each sleep trait listed, the ANOVA results are presented. $d.f.$, degrees of freedom; M.S., Type III mean squares; F, F ratio statistic; P, P-value.

Table S5. GLM analysis results for each gene in females are shown as a row; the Maximum a Posteriori (MAP) parameter estimates and log-likelihoods are shown as well as p-values computed from the likelihood ratio test. Significance statistics corrected for multiple testing are also included, as well as the normalized counts for all samples.

Table S6. GLM analysis results for each gene in males are shown as a row; the Maximum a Posteriori (MAP) parameter estimates and log-likelihoods are shown as well as p-values computed from the likelihood ratio test. Significance statistics corrected for multiple testing are also included, as well as the normalized counts for all samples.

Table S7. Genes with opposite slopes for the short and long interaction terms of generation in females

Table S8. Genes with opposite slopes for the short and long interaction terms of generation in males

Table S9. Gene Ontology analysis results for 85 significant genes common to males and females.

Table S10. Correlations obtained from normalizing Gaussian Process signal covariances (GP correlation) and from Spearman Correlation for each of the six sex, selection scheme combinations

Table S11. Expected values for the correlations obtained from normalizing Gaussian Process signal covariances (GP correlation) not overlapping with controls for each of the six sex, selection scheme combinations (value missing if overlapping in that condition)

Table S12. Effects of *Minos* insertions on sleep. For each gene the table lists the Flybase ID, Bloomington *Drosophila* Stock Center (BDSC) number, *Minos* genotype, and isogenic control line. For each sleep trait, the number of flies tested and mean sleep phenotype is given for sexes combined and females and males separately. P-values are listed for each term in the ANOVA model for sexes combined and for males and females separately. Significance is indicated by bold P-values.

Table S13. Gene expression ratios calculated from *Minos* insertion line and control RNA-Seq data. For each target gene the table lists the corresponding control line, the Flybase ID and gene symbol of the gene predicted to interact with the target gene, the normalized expression of the interacting gene for the *Minos* target gene line and isogenic control line, and the ratio of normalized expression (*Minos*/control)

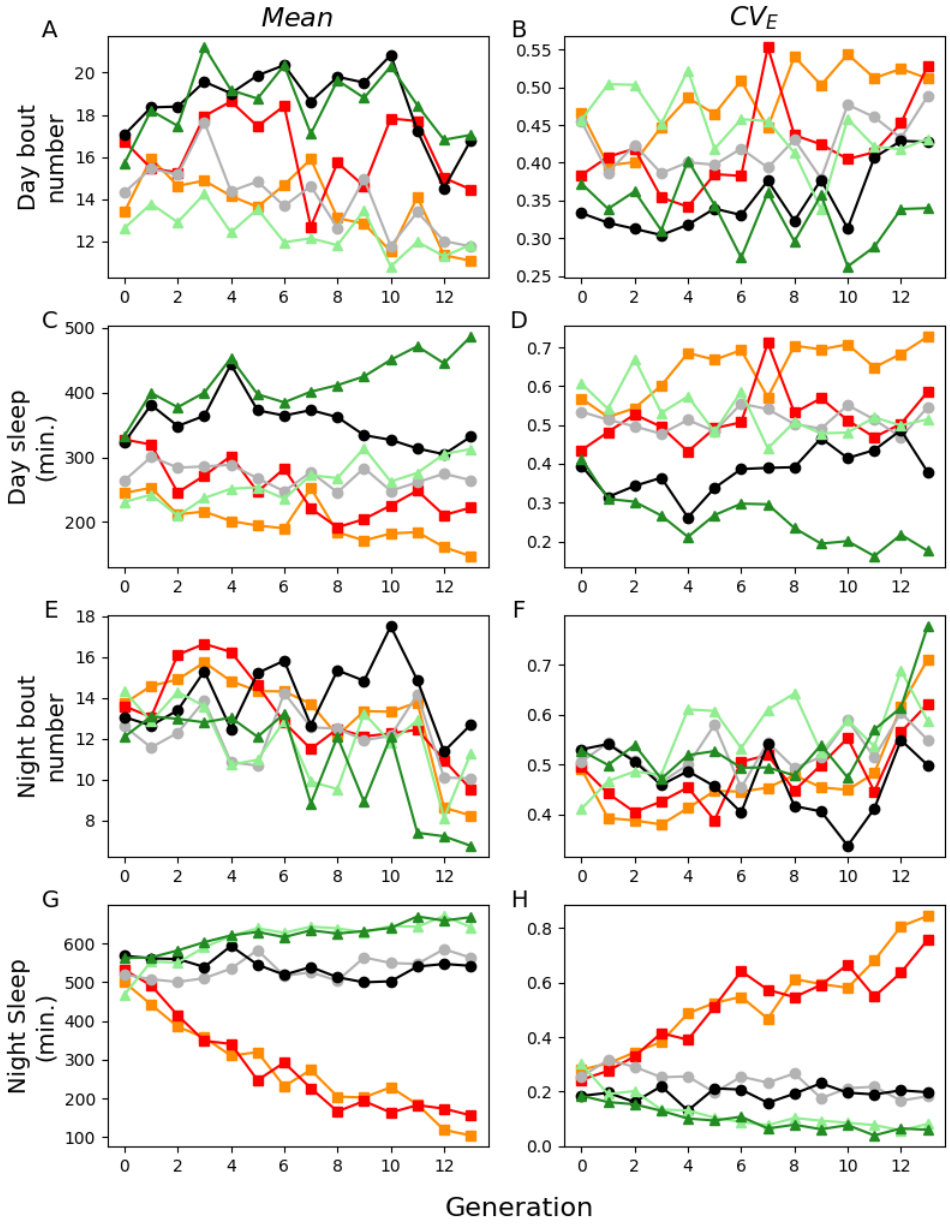


Figure S2. Correlated response to selection for long/short night sleep and associated coefficient of environmental variation. A, day bout number; B, day bout number coefficient of environmental variation (CV_E); C, day sleep; D, day sleep CV_E; E, night bout number; F, night bout number CV_E; G, night sleep; H, night sleep CV_E; I, waking activity; J, waking activity CV_E; K, sleep latency; L, sleep latency CV_E; M, day average bout length; N, day average bout length CV_E; O, night average bout length; P, night average bout length CV_E. Light green, Replicate 1 long-sleeper population; Dark green, Replicate 2 long-sleeper population; Orange, Replicate 1 short-sleeper population; Red, Replicate 2 short-sleeper population; Gray, Replicate 1 control population; Black, Replicate 2 control population. CV_E, phenotypic variation.

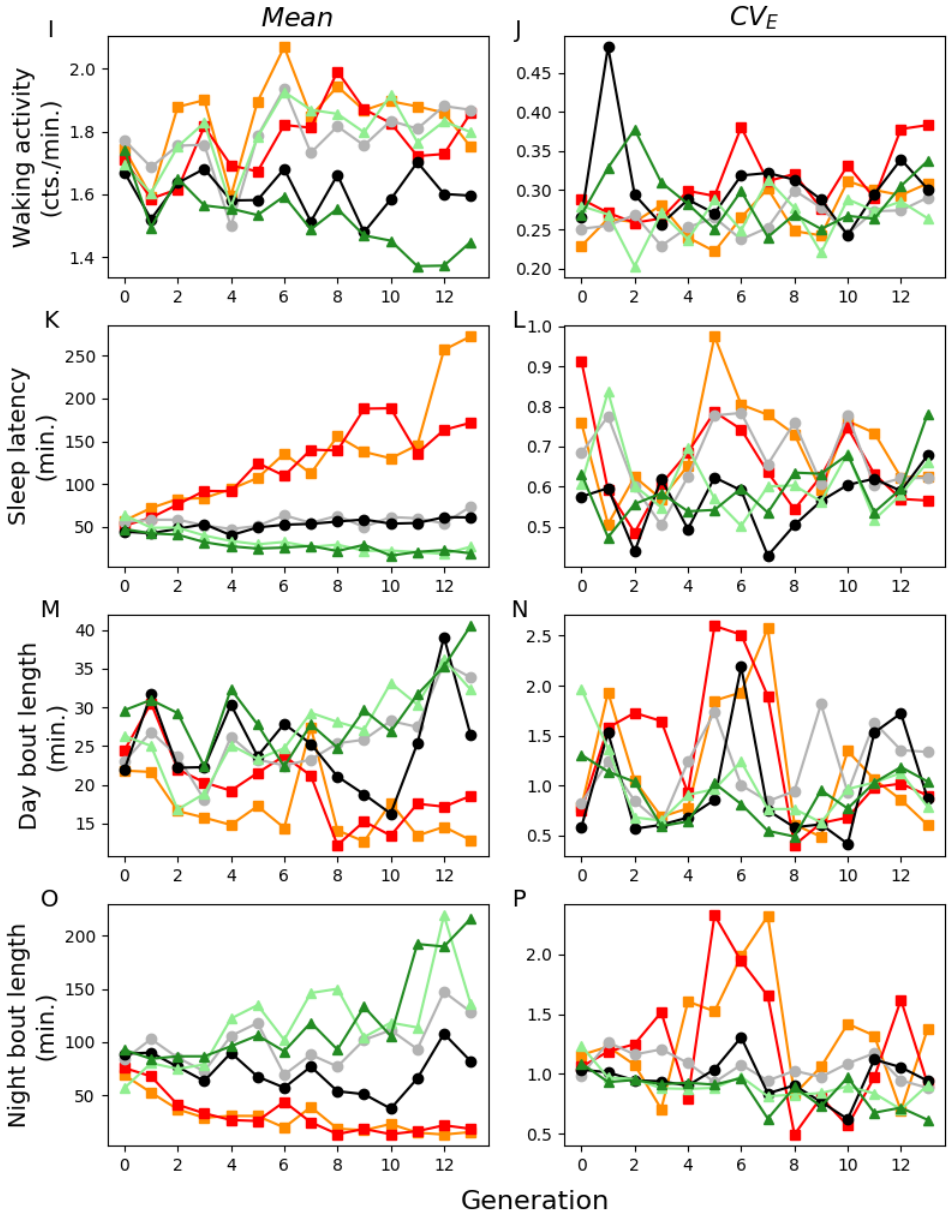


Figure S2. (Continued).

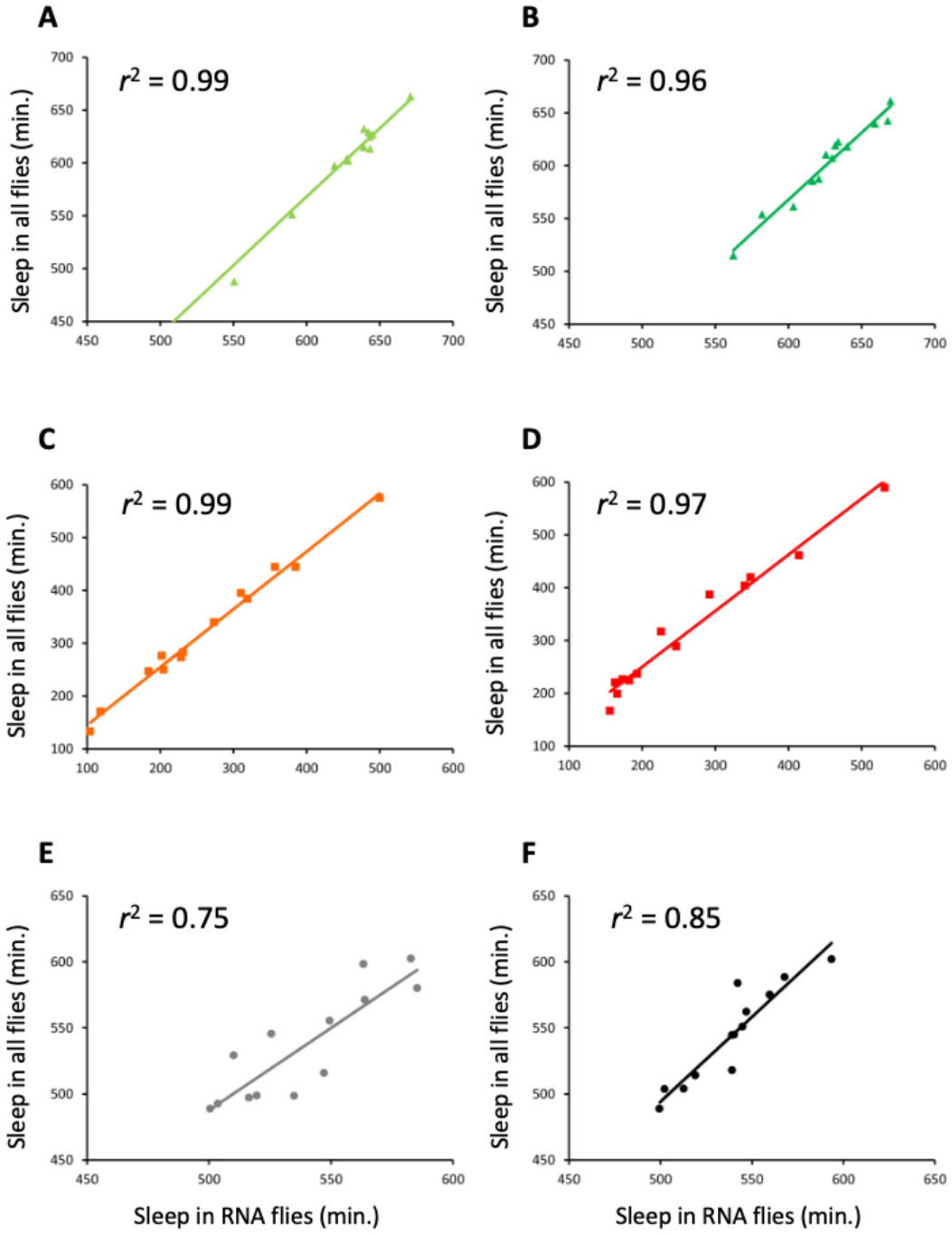


Figure S3. Correlation of night sleep between flies harvested for RNA and all flies in the population. A, long-sleeping Replicate 1; B, long-sleeping Replicate 2; C, short-sleeping Replicate 1; D, short-sleeping Replicate 2; E, control Replicate 1; F, control Replicate 2

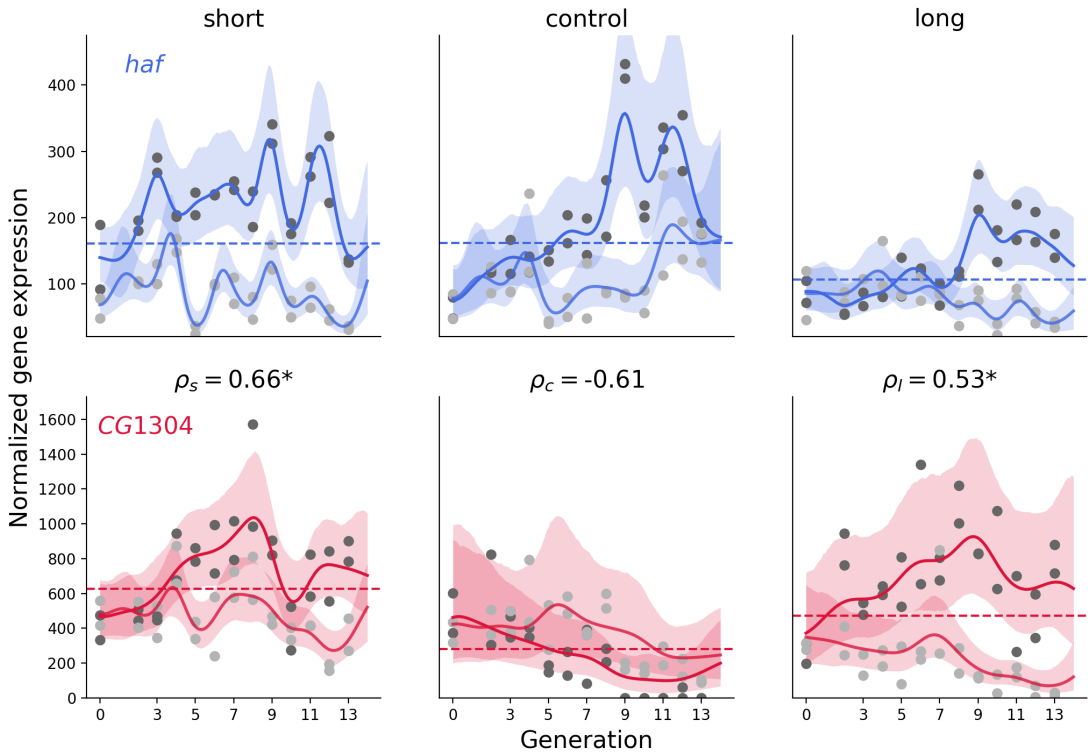


Figure 3—figure supplement 1. Fit of Gaussian Process model to pair of genes *haf* and *CG1304*.

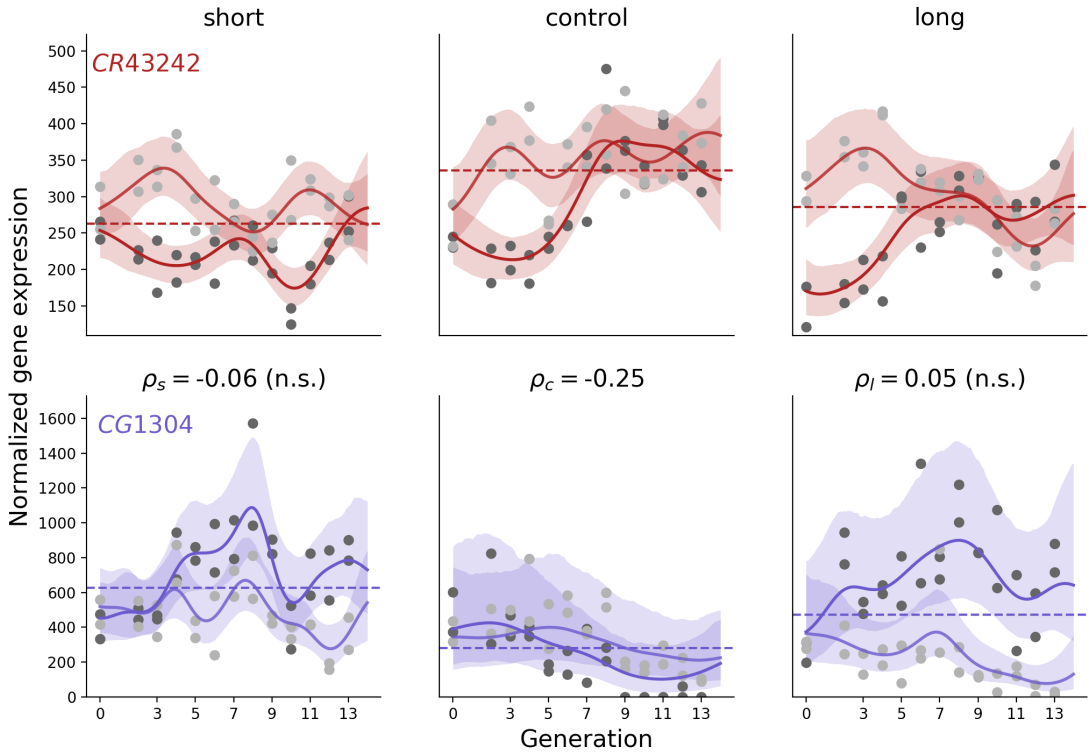


Figure 3—figure supplement 2. Fit of Gaussian Process model to pair of genes *CR43242* and *CG1304*.

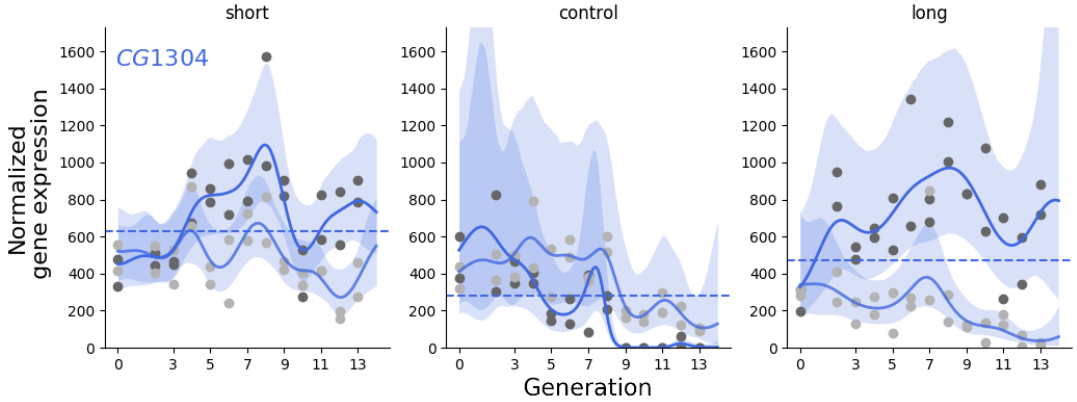


Figure 3—figure supplement 3. Fit of single-channel Gaussian Process model to *CG1304* gene.

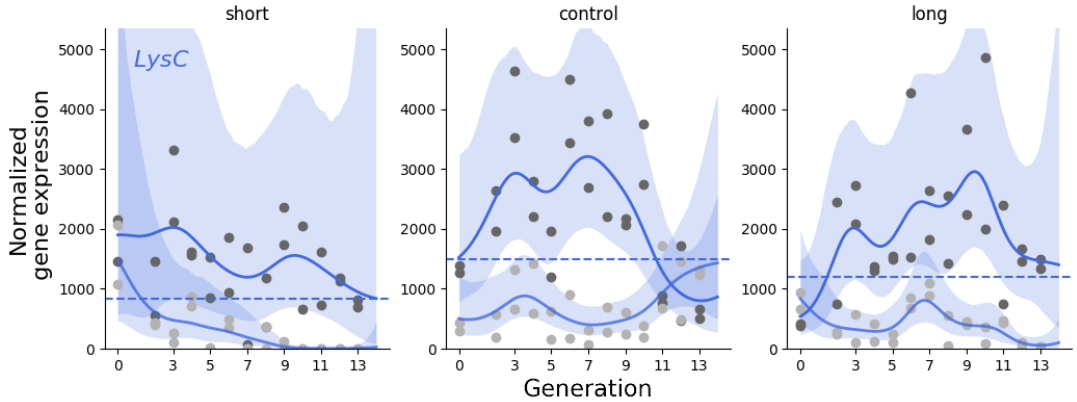


Figure 3—figure supplement 4. Fit of single-channel Gaussian Process model to *LysC* gene.

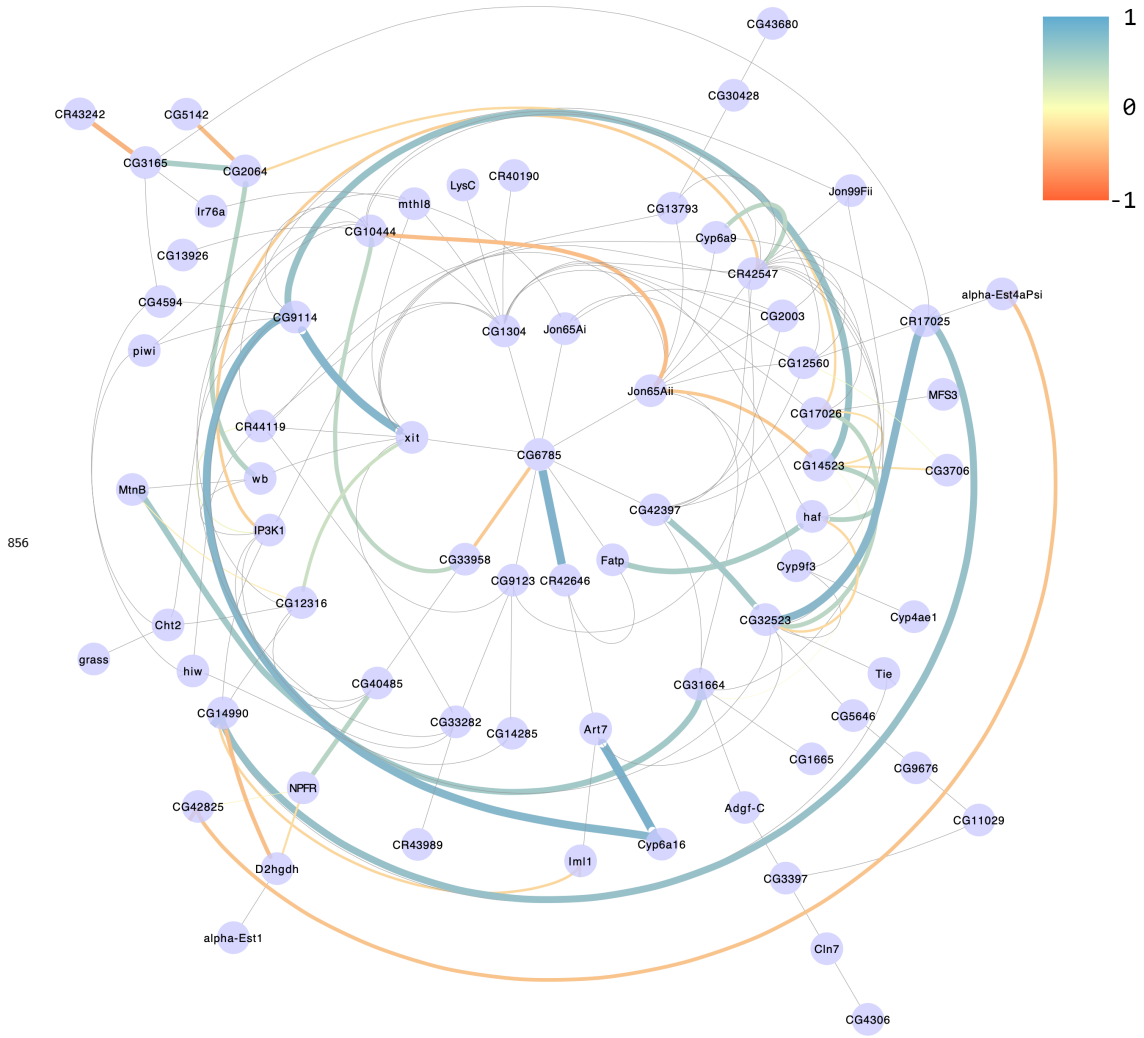


Figure 5—figure supplement 1. Male, short sleepers

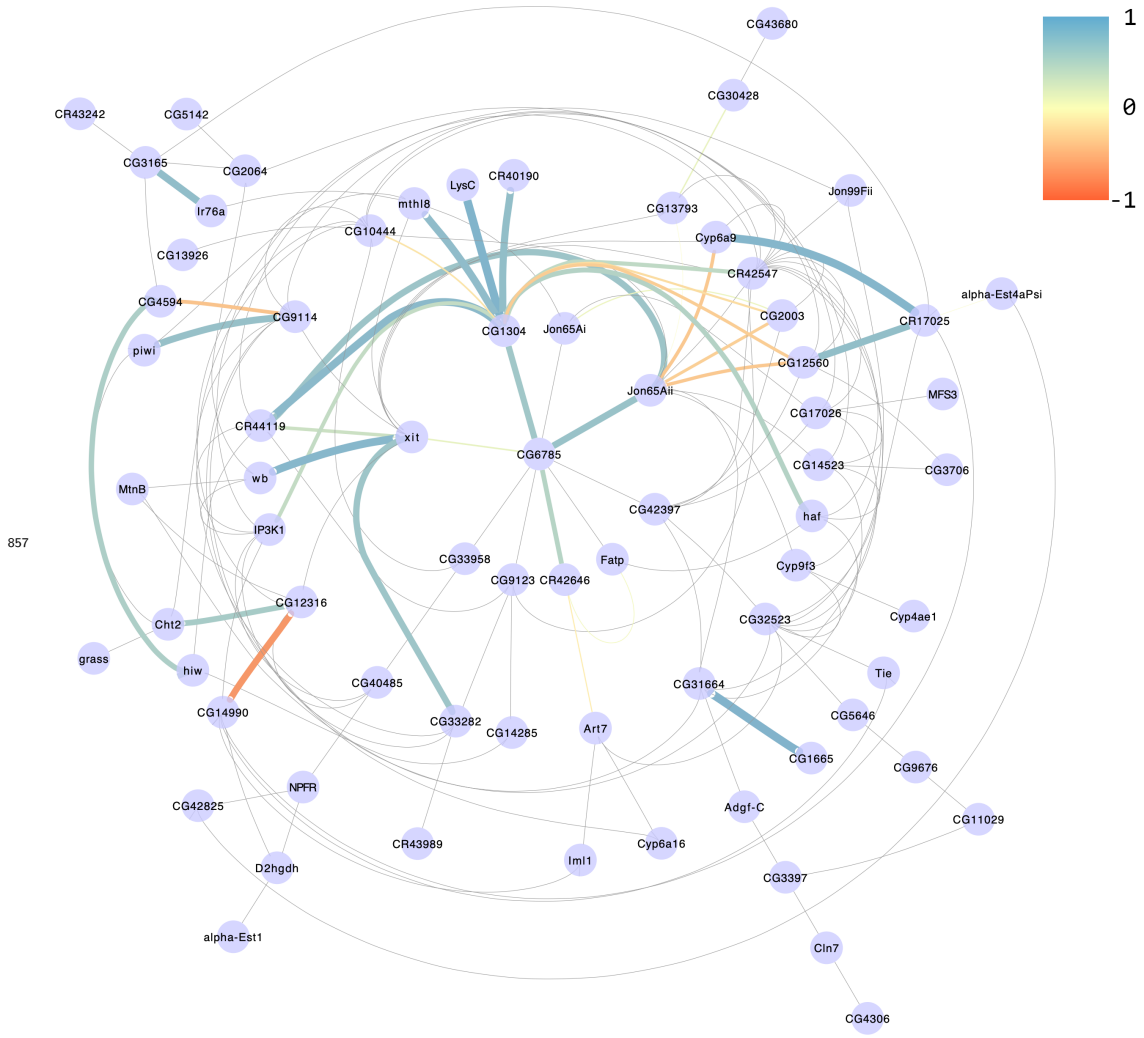
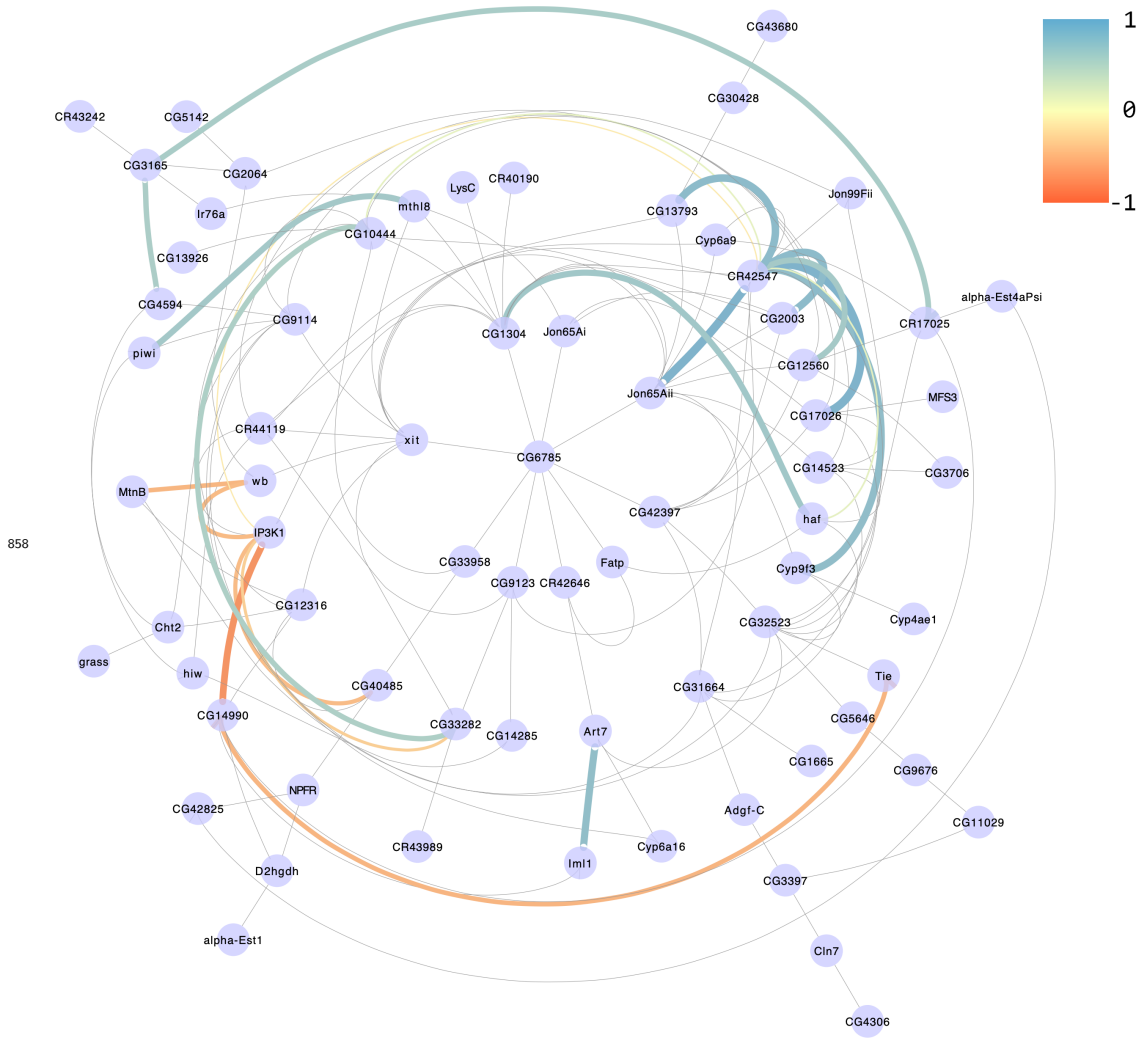


Figure 5—figure supplement 2. Female, long sleepers



858

Figure 5—figure supplement 3. Female, short sleepers

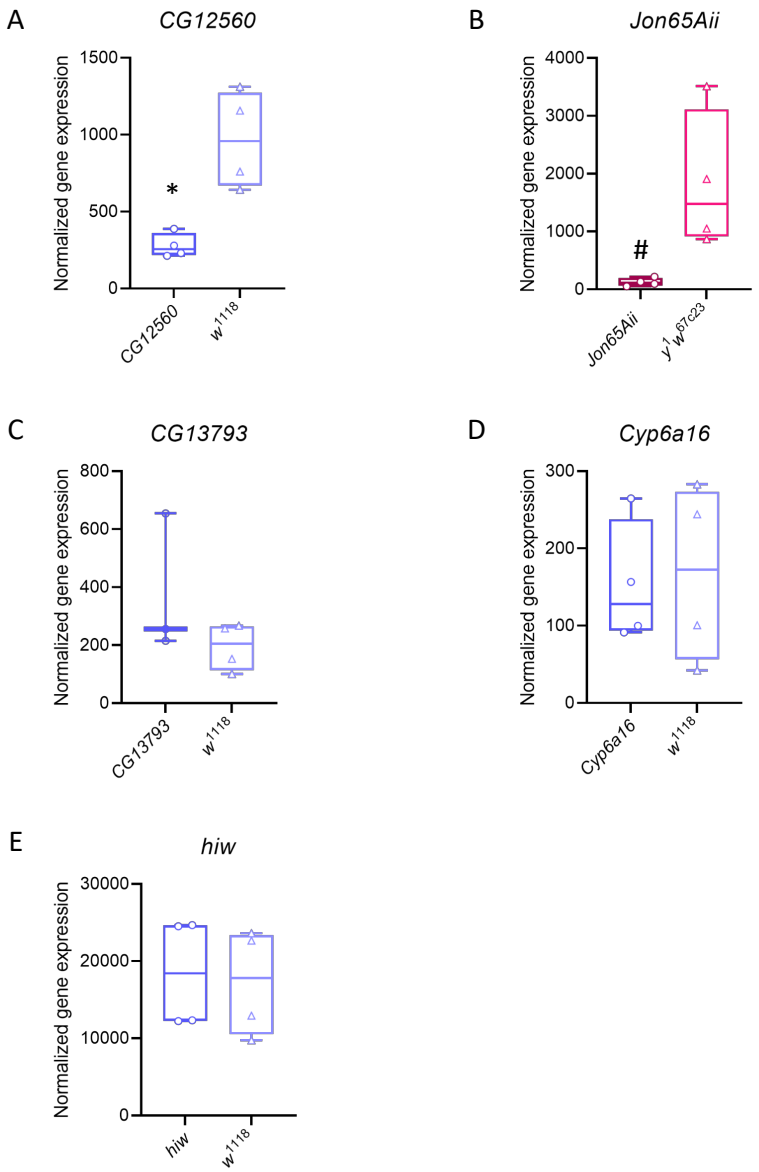


Figure 6—figure supplement 1. Gene expression in *Minos* mutants. For each candidate gene, the gene expression in the *Minos* mutant and corresponding control are plotted. * or #, $P < 0.05$ by Kruskal-Wallis test. **(a)** *CG12560*; **(b)** *Jon65Aii*; **(c)** *CG13793*; **(d)** *Cyp6a16*; **(e)** *hiw*

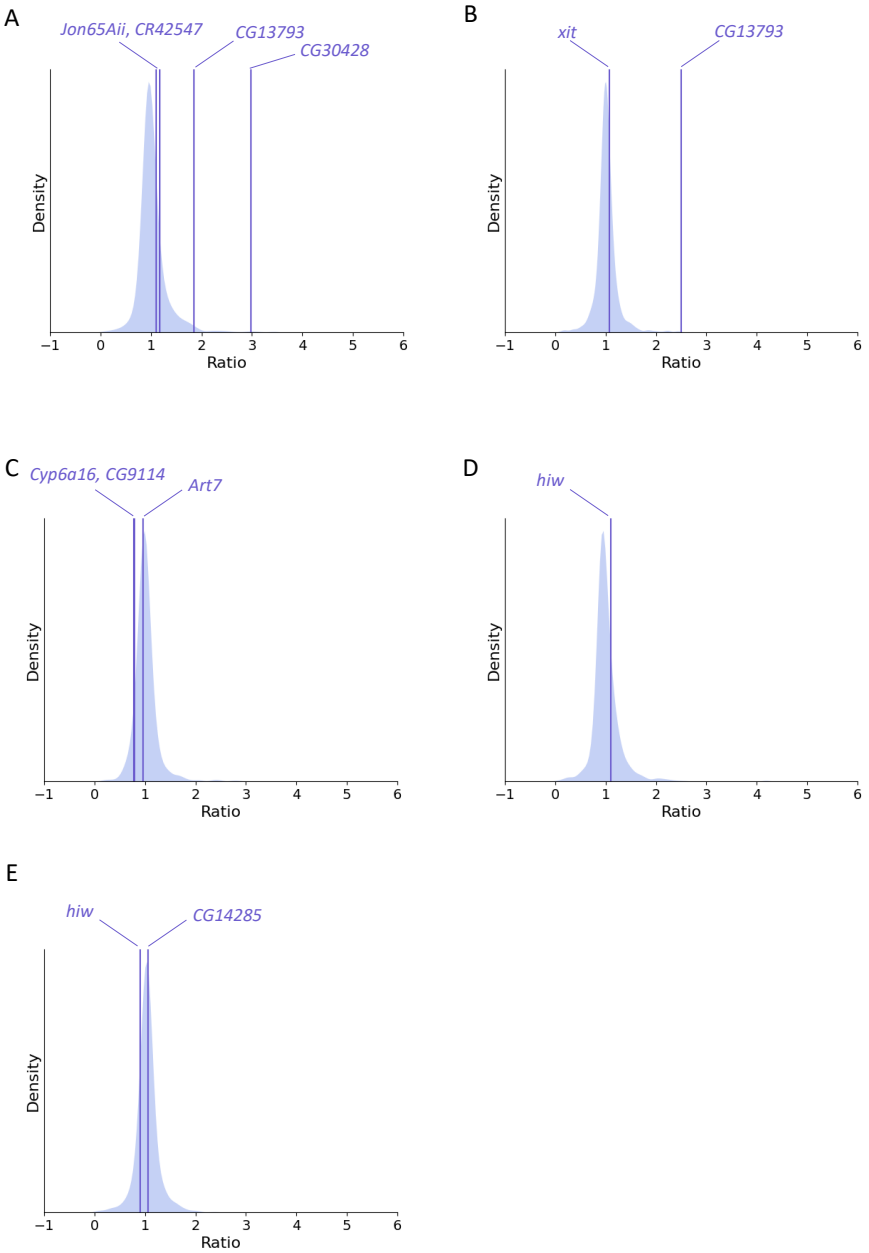


Figure 7—figure supplement 1. Comparison of ratios of gene expression between genes with significant Gaussian Process correlations and unrelated genes for *CG13793*, *Cyp6a16*, and *hiw Minos* mutants. **(a)** *CG13793* females; **(b)** *CG13793* males; **(c)** *Cyp6a16* males; **(d)** *hiw* females; **(e)***hiw* males.