Selection against expression noise explains the origin of the hourglass pattern of Evo-Devo

Jialin Liu[1,2]*, Michael Frochaux[2,3], Vincent Gardeux[2,3], Bart Deplancke[2,3], Marc Robinson-Rechavi[1,2]*

1 Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

2 Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

3 Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL)

* corresponding authors: jialin.liu@unil.ch, marc.robinson-rechavi@unil.ch

The evolution of embryological development has long been characterized by deep conservation. Both morphological and transcriptomic surveys have proposed a "hourglass" model of Evo-Devo [1,2]. A stage in mid-embryonic development, the phylotypic stage, is highly conserved among species within the same phylum [3–7]. However, the reason for this phylotypic stage is still elusive. Here we hypothesize that the phylotypic stage might be characterized by selection for robustness to noise and environmental perturbations. This could lead to mutational robustness, thus evolutionary conservation of expression and the hourglass pattern. To test this, we quantified expression variability of single embryo transcriptomes throughout fly *Drosophila melanogaster* embryogenesis. We found that indeed expression variability is lower at extended germband, the phylotypic stage. We explain this pattern by stronger histone modification mediated transcriptional noise control at this stage. In addition, we find evidence that histone modifications can also contribute to mutational robustness in regulatory elements. Thus, the robustness to noise does indeed contributes to robustness of gene expression to genetic variations, and to the conserved phylotypic stage.

Phenotypes can vary even among isogenic individuals in homogenous environments, suggesting that stochastic effects contribute to phenotypic diversity [8,9]. Gene expression variability among genetically identical individuals under uniform conditions, hereafter "variability", is one of the most important stochastic processes in the mapping of genotype to phenotype. It is caused by a combination of molecular noise (stochastic biochemical effects, e.g., transcriptional burst process based transcriptional noise) and other effects (variation in cells and their environment, e.g., distribution of molecules at cell division) [10–13]. Precise regulation of gene expression is notably important during development [14], however, this process inevitably has to deal with stochasticity [15]. This tension between precision and stochasticity in development raises questions, such as whether some stages are more robust to gene expression stochasticity. And whether natural selection against expression variability can transfer to mutational robustness, causing the evolutionary conservation of the phylotypic stage. To answer these questions, we investigated expression variability across fly embryonic development.

We generated 288 single embryo 3' end transcriptomes using BRB-seq [16], at eight developmental stages covering the whole fly embryogenesis, with 3h intervals (Figure 1A). After quality control, 239 samples were kept (Figure S1, S2). On average, we obtained over 5 million uniquely mapped reads of protein coding genes per embryo. Based on multidimensional scaling analysis (MDS), 150 embryos follow the developmental trajectory, while there is a small cluster of 89 embryos collected at different time points mixed together (Figure 1B). The samples in this cluster appear to be unfertilized eggs (Methods and Figure S3). All further analysis was performed only on the 150 fertilized embryos.

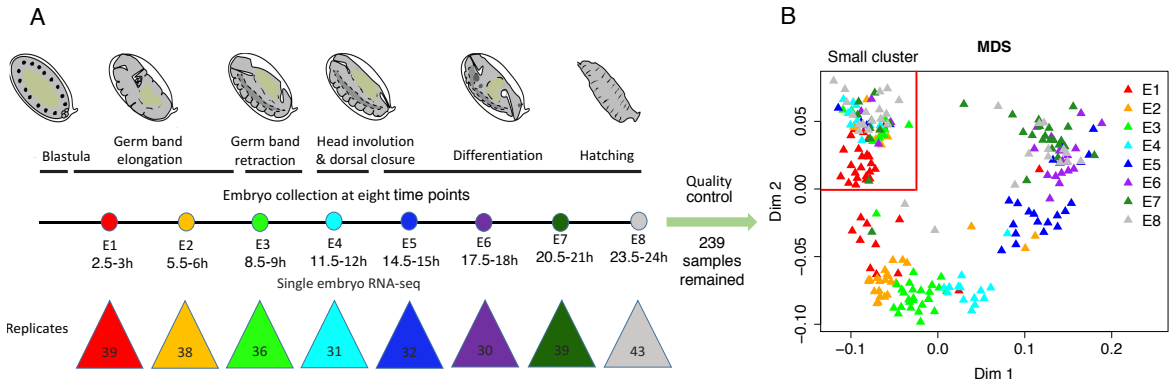

Figure 1: Studying expression variability throughout embryogenesis

A. Method outline. We performed single embryo BRB-seq[16] at eight developmental stages, indicated by different colored dots. The number of samples collected at each stage is indicated in the colored triangles. Embryo images adapted from [17] and used with permission from Springer Nature (License Number: 4547630238607) and from the authors.

B. Multidimensional scaling analysis (MDS) of 239 high quality samples. Different colors indicate different stages. The samples can be split into two groups: a small cluster in the top-left delimited by two red lines; and the remaining samples, which are organized according to embryonic stage. Only the 150 samples which follow embryonic stages were used for further analysis.

We measured expression variability as Adjusted SD, standard deviation (SD) of expression between replicates corrected for expression level (Methods and Figures S4-6). This expression variability follows an hourglass pattern overall, with a global minimum at E3 (Figure 2A), corresponding to the phylotypic stage of fly [7]. There is also a local minimum at E6. This is consistent with the pattern of transcriptome divergence between fly and

mosquito *Anopheles gambiae*, with the global minimum at E3, and a local minimum at E6 [18]. Our observations are robust to the use of different variability metrics (Figure S7), and to sampling (bootstrap analysis, Figure S8). Bootstrap results also suggest that the minimum of variability extends over E3 to E4. The embryo transcriptome is dominated by zygotic transcripts 2.5h after egg laying[19], so the high variability in E1 and E2 is not directly caused by maternal transcripts. We didn't find any significant functional enrichment for genes which follow the hourglass variability pattern. Overall, expression variability is not equally distributed throughout embryogenesis, and gene expression at the phylotypic stage appears more robust to stochastic factors than at other stages.

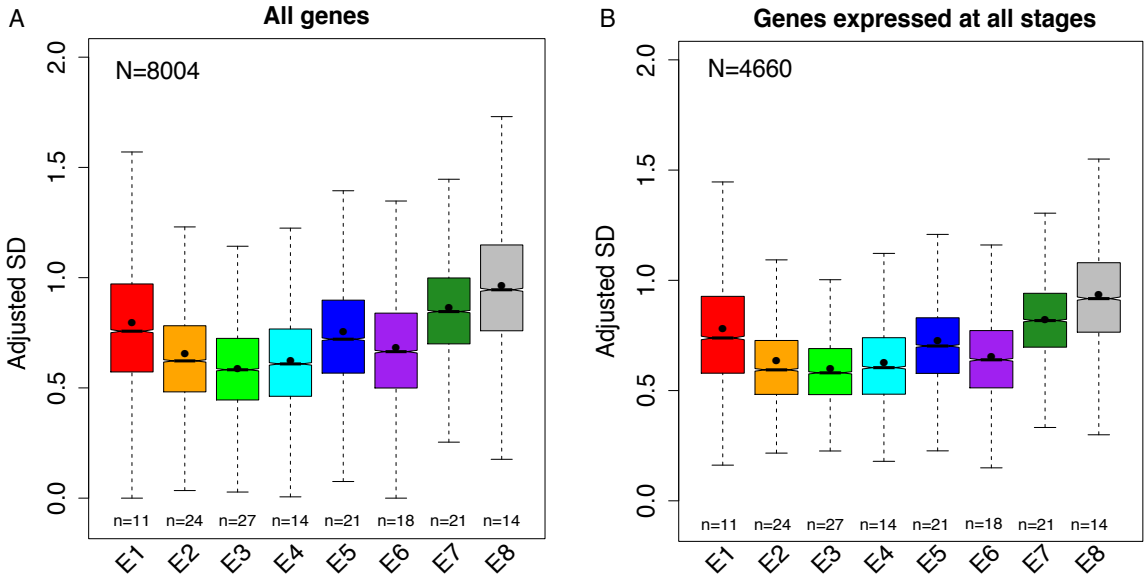

Figure 2: The phylotypic stage (E3) has lower expression variability

The number of individual samples used in each development stage is indicated below each box. The number of genes analyzed is indicated in the top-left corner of each plot. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range (IQR), and the box shows the lower and upper intervals of IQR together with the median. The black dot in each box indicates the mean.

A. Expression variability pattern of all genes which passed quality control. We performed pairwise Wilcoxon tests between any two stages to test the significance. The multiple test corrected p-values (Benjamini–Hochberg method) are shown in Table S1; they are all $< 10^{-7}$.

B. Expression variability pattern of genes expressed at all stages. We performed pairwise Wilcoxon tests between any two stages to test the significance. The multiple test corrected

p-values (Benjamini–Hochberg method) are shown in Table S2; they are all $< 10^{-5}$ *except E2 vs. E4, for which p-values = 0.24.*

The variation in expression variability could either be due to changes in the set of active genes, with genes differing in their intrinsic variability levels, or to genome-wide changes in the regulation of variability. To test this, we first reproduced our results restricted to the subset of genes which are expressed at all stages. Under the first explanation, we would expect to lose the hourglass variability pattern, but the pattern is maintained (Figure 2B). We performed additional tests: restricting to genes with constant expression level over development (Figure S9A); restricting to transcription factors (Figure S9B); and contrasting genes with dispersed or precise promoters (Figure S10), following Schor et al [20]. Dispersed promoters seem to be more robust to mutations, which might also translate into robustness to noise. Despite a loss of power with fewer genes, there remains an hourglass pattern of expression variability in all cases. Interestingly, the precise promoter genes have higher variability than the dispersed promoter genes except at E3, thus a strongest hourglass pattern. Overall, these results suggest that the lower variability at E3 is due to genome-wide regulation mechanisms more than to changes in the gene set.

Histone modifications can regulate transcriptional noise [21–25], notably through the modulation of transcriptional burst frequency [22–24]. For example, high levels of histone modifications can increase chromatin accessibility, leading to an increase in transcriptional burst frequency, which leads to minimizing noise. To check this role of histone modifications, we analyzed four available euchromatin histone modifications at six developmental stages [26]. For each gene, we calculated the mean modification signal (background-subtracted tag density) separately for proximal promoters and for gene bodies [23]. Higher modification signal genes tend to have lower variability for all histone modifications (Figure 3A). This supports a role in minimizing transcriptional noise, and is consistent with previous studies in yeast and mammals [22,23].

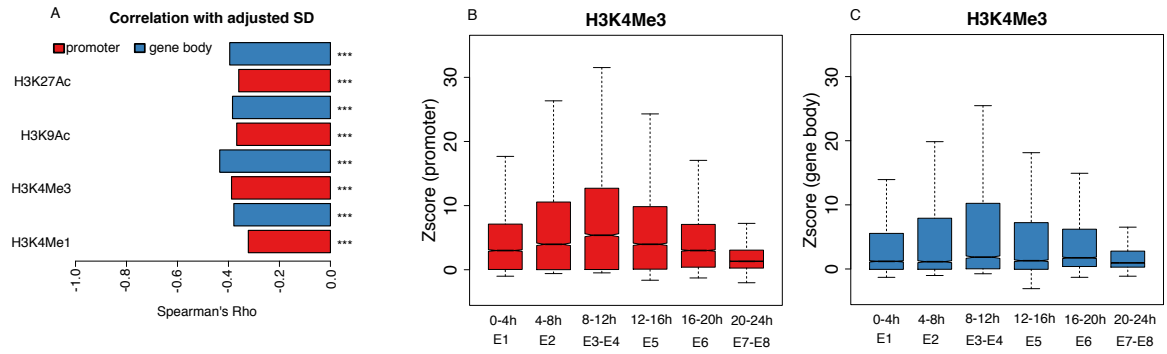

Figure 3: Histone modification signal and expression variability

Red and blue represent histone modification signals calculated on the proximal promoter (4kb around the transcription start site – TSS) and the gene body, respectively.

*A. Spearman's correlation coefficient between histone modification signal (background-subtracted tag density) and expression variability. Here, for each gene, both its variability and its histone modification signal are the mean value across stages. ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$, NS, $P \geq 0.05$.*

B. Proximal promoter H3K4Me3 signal (Z score relative to intergenic signal) in different stages. Corresponding stages of our expression variability data are indicated below. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range (IQR), and the box shows the lower and upper intervals of IQR together with the median. We performed pairwise Wilcoxon tests between any two stages to test the significance. The multiple test corrected p-values (Benjamini–Hochberg method) are shown in Table S3; they are all $<10^{-7}$ except 4-8h vs. 12-16h, for which p-value = 0.68.

C. Gene body H3K4Me3 signal (Z score relative to intergenic signal) in different stages. Corresponding stages of our single embryo BRB-seq data are indicated below. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range (IQR), and the box shows the lower and upper intervals of IQR together with the median. We performed pairwise Wilcoxon tests between any two stages to test the significance. The multiple test corrected p-values (Benjamini–Hochberg method) are shown in Table S4; they are all $< 10^{-5}$ except 0-4h vs. 20-24h, for which p-value = 0.26; and 8-12h vs. 16-20h, for which p-value = 0.26.

The gene-level relation between histone modifications and expression variability raises the possibility that the pattern of expression variability across development could be driven by changes in histone modification signal. To compare histone modification signal between stages,

we normalized gene and promoter signal by that on intergenic regions (Methods), which are not expected to change histone modification signal between stages. All histone marks present an hourglass-like pattern, with the highest signal at 8-12h (except for H3K4Me1 on gene body, where it is a local but not global maximum), corresponding to E3 and E4, i.e. the lowest expression variability, for both promoters and gene body (Figures 3B-C, S11). Moreover, for all histone marks on gene body, as well as H3K4Me1 on promoters, there is another local maximum at 16-20h, corresponding to E6. Generally, these results support changes in histone modification signal over development, with a correspondence between stronger histone modification signal and lower expression variability.

Several studies have suggested that mechanisms which confer robustness to stochastic variation can also buffer the effects of genetic variation [14,27,28]. If histone modifications can buffer the effect of genetic variation on gene expression, we should observe that genes with higher histone modification signal are less sensitive to mutations in their regulatory regions, and are thus less conserved. Indeed, genes with higher histone modification signal tend to have less conserved core promoter sequences [29] (49 bp upstream TSS and 10 bp downstream from the TSS) between species (phastCons score; Figure 4A). They are also less conserved within a population (promoter nucleotide diversity π; Figure S12). The phastCons pattern remains using 200 bp or 400 bp regions, but disappears using 1 kb regions (Figure S13), indicating a relatively narrow region around the TSS under this balance of selection and robustness.

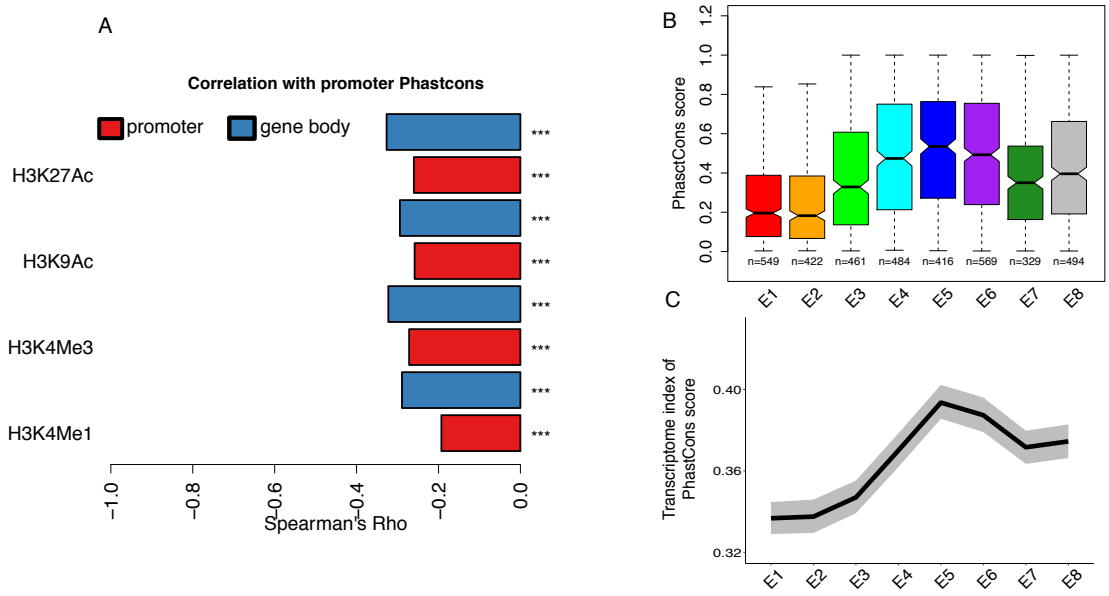

Figure 4: Histone modification signal and promoter sequence conservation

The promoter sequence conservation is the mean of the phastCons score over experimentally identified core promoter regions (49 bp upstream TSS to 10 bp downstream of the TSS) [29].

*A. Spearman's correlation coefficient between histone modification signal (background-subtracted tag density) and promoter sequence conservation. Red and blue represent histone modification signals calculated from the proximal promoter (4 kb around the TSS) and gene body respectively. Here, for each gene, the histone modification signal is the mean value across stages. ***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$, NS, $P \geq 0.05$.*

B. Variation of promoter sequence conservation for stage specific genes. The number of genes in each development stage is indicated below each box. We performed pairwise Wilcoxon test between any two stages to test the significance. The multiple test corrected p-values (Benjamini–Hochberg method) are shown in Table S5.

C. Transcriptome index of promoter phastCons score across development. The grey area indicates 95% confidence interval estimated from bootstrap analysis.

Since histone modifications appear to buffer genetic variation in gene expression, and since the E3 stage has stronger modification signals, the lower expression divergence in E3 between species [7] might be caused either by stronger purifying selection on mutations in regulatory regions, or by histone modifications buffering the consequences of mutations in these regions. In the first case, we expect genes specifically expressed at E3 to have higher sequence conservation on promoters. In the second case, we expect the opposite pattern, since mutations that are buffered would behave nearly neutrally. To test this, we identified

196 genes specifically expressed in each stage and compared their promoter sequence
197 conservation. We found that genes specific of E3 have a relatively weak promoter sequence
198 conservation (Figure 4B), supporting a stronger buffering mechanism rather than stronger
199 purifying selection on sequences. The transcriptome indexes of conservation and of diversity
200 (mean promoter sequence conservation and mean π, respectively, weighted by expression)
201 extend this observation to the full transcriptome (Figure 4C; Figure S14). These results
202 support a role of buffering effects on regulatory mutations in the hourglass pattern of
203 expression divergence in fly embryogenesis. Essential genes, and highly connected genes,
204 have lower variability (Figure S17), which supports that variability is detrimental, and that
205 mechanisms which reduce it are adaptive. Thus natural selection on robustness against
206 expression variability could contribute to the phylotypic stage conservation at
207 macroevolutionary scale.
208
209 We have found an uneven distribution of variability, and thus of robustness of the process
210 of gene expression, across development, which mirrors the hourglass Evo-Devo model [1,2].
211 Stage E3 is the most robust to stochastic variation on gene expression, with lower expression
212 variability, and is the phylotypic stage of fly, with conservation between species [7]. Although
213 mutational robustness can evolve under natural selection theoretically [30], the conditions are
214 too restrictive to be relevant in practice. We propose that the mutational buffering effect of
215 histone modifications is a by-product of selection for minimizing transcriptional noise. Thus,
216 our model is that selection for robustness to noise and environmental perturbations in a key
217 embryonic stage has led to the evolutionary conservation over large time scales which
218 characterizes the phylotypic stage.
219
220
221

Methods

Availability of code

Data files and analysis scripts are available on GitHub:
https://github.com/ljljolinq1010/expression-noise-across-fly-embryogenesis.

Availability of data

Expression datasets have been deposited to the Gene Expression Omnibus with accession number GSE128370.

Embryo collection and RNA extraction

Fly lines (w^{1118}) were obtained from the Bloomington stock center and reared at room temperature on a standard fly medium with 12 hours light dark cycle. The fly medium we used is composed of: 6.2 g Agar powder (ACROS N. 400400050), 58.8 g Farigel wheat (Westhove N. FMZH1), 58.8 g yeast (Springaline BA10), 100 mL grape juice; 4.9 mL Propionic acid (Sigma N. P1386), 26.5 mL of Methyl 4-hydroxybenzoate (VWR N. ALFAA14289.0) solution (400 g/L) in 95% ethanol, 1 L Water. 100 to 150 flies were transferred to cages, which were sealed to a grape agar plate (1:1 mixture of 6% agar and grape juice). We used 4 separate cages to collect the embryos. The adults were kept overnight on this plate before being transferred to a new plate supplemented with yeast paste. Synchronization of eggs on this plate lasted for 2 hours before being swapped with a new plate supplemented with yeast paste. We let the adults lay eggs for 30 min before removing the plate and letting the eggs incubate for the desired time. Eggs were harvested using the following protocol. First a 1:1 bleach (Reactolab 99412) 1x PBS mix was poured on the plate and incubated for 2 min. During this incubation, we used a brush to lightly scrape the surface to mobilize the embryos. We then poured the PBS-bleach mixture through a sieve, washed the plate with 1x PBS, and poured the wash on the same sieve. We washed the sieve several time with 1x PBS until the smell of bleach disappeared. Single embryos were then manually transferred to Eppendorf containing 50 μL beads and 350 μL Trizol (lifetechnologies AM9738). The tubes were homogenized in a Precellys 24 Tissue Homogenizer at 6000 rpm for 30 seconds. Samples were then transferred to liquid nitrogen for flash freezing and stored at –80°C. For RNA extraction, tubes were thawed on ice, supplemented with 350 μL of 100% Ethanol before homogenizing again with the same parameters. We then used the Direct-zol™ RNA Miniprep R2056 Kit, with the following modifications: we did not perform DNAse I treatment, we added another 2 min centrifugation

into an empty column after the RNA Wash step, finally elution was performed by adding 8 μL of RNAse-free water to the column, incubation at room temperature for 2 min and then centrifugation for 2 min. RNA was transferred to a low-binding 96 well plate and stored at -80°C.

Bulk RNA Barcoding and sequencing (BRB-seq)

The BRB-seq is a technique for multiplexed RNA-seq [16] which is able to provide high-quality 3' transcriptomic data at a low cost (e.g. 10-fold lower than Illumina Truseq Stranded mRNA-seq). The data (fastq files) generated from BRB-seq are multiplexed and asymmetrical paired reads. Read R1 contains a 6 bp sample barcode, while read R2 contains the fragment sequence to align to the reference genome.

1. Library preparation

RNA quantity was assessed using picogreen (Invitrogen P11496). Samples were then grouped according to their concentration in 96-well plates and diluted to a final concentration determined by the lowest sample concentration on the plate. RNA was then used for gene expression profiling using BRB-seq. In short, the BRB-seq protocol starts with oligo-dT barcoding, without TSO for the first-strand synthesis (reverse transcription), performed on each sample separately. Then all samples are pooled together, after which the second-strand is synthesized using DNA PolII Nick translation. The sequencing library is then prepared using cDNA augmented by an in-house produced Tn5 transposase preloaded with the same adapters (Tn5-B/B), and further enriched by limited-cycle PCR with Illumina compatible adapters. Libraries are then size-selected (200 - 1000 bp), profiled using High Sensitivity NGS Fragment Analysis Kit (Advanced Analytical, #DNF-474), and measured using Qubit dsDNA HS Assay Kit (Invitrogen, #Q32851). In total, we generated four libraries. For details of library information, please check Table S20.

2. Sequencing

Libraries were mixed in equimolar quantities and were then sequenced on an Illumina Hi-Seq 2500 with pair-end protocol (read R2 with 101 bp) at the Lausanne Genomic Technologies Facility.

RNA-seq analysis

1. Generating expression matrix

The fastq files were first demultiplexed by using the "Demultiplex" tool from BRB-seqTools suite (available at https://github.com/DeplanckeLab/BRB-seqTools). Then, we trimmed the polyA sequences of the demultiplexed files by using the "Trim" tool. Next, the STAR aligner [31] was used to map the trimmed reads to the reference genome of fly *Drosophila melanogaster* (BDGP6, Ensembl release 91 [32]). Finally, the read count of each gene was obtained with HTSeq [33].

2. Filtering samples and genes

Low-quality samples need to be filtered out, since they might bias results of downstream analyses. In order to assess sample quality, we calculated the number of uniquely mapped reads and of expressed genes for each sample [34]. We removed samples with <0.3 million uniquely mapped reads or with <4500 expressed genes (Figure S1). We confirmed that these filtered samples are indeed outliers in a multidimensional scaling analysis (MDS) (Figure S15). Since lowly expressed genes have larger technical error, to minimize the technical noise, we need to remove lowly expressed genes as well. We first calculated counts per million (cpm) with the edgeR package [35] for each gene. Then we removed genes with mean cpm across samples ≤1, as suggested by Lun et al.[34]. Finally, for the remaining genes, we re-transformed their cpm values to the original count values for the downstream normalization analysis. After filtering, we obtained an expression count matrix with 239 samples (Figure S2) and 8004 protein coding genes.

3. Normalization and batch effect correction

Because BRB-Seq retains only the 3′ end of the transcript, we performed sample normalization by using quantile normalization with log transformation in the voom package [36], but without transcript length normalization. To remove potential batch effects across the four libraries, we applied the ComBat function in the sva package [37] to the normalized and log2 transformed expression data. For genes with expression values less than 0 after Combat, or with original expression values equal to 0, we change its values to 0 after Combat correction as suggested by Kolodziejczyk et al [38].

Multidimensional scaling analysis (MDS)

A number of factors could be invoked to explain the two groups observed in our multidimensional scaling analysis (MDS) (Figure 1B), but they should also explain that only one group is structured according to developmental time. The obvious hypothesis that they correspond to male and female embryos does not explain that structure, and is also not

supported by examining X/autosome gene expression ratios (Figure S16). An alternative hypothesis is that samples in the small cluster are unfertilized eggs. If an egg is not fertilized, after completion of meiosis, development will be arrested [39], but they are visually indistinguishable. This hypothesis is confirmed by at least two lines of evidence, in addition to the lack of developmental time structure. First, for expression correlation, all samples in the small cluster are highly correlated with unfertilized egg, while the correlations in the other samples gradually decrease with development (Figure S3A). Second, all the samples from the small cluster are enriched with meiosis related genes (Figure S3B). Thus we excluded the small cluster for downstream analyses, i.e. we used 150 embryos with an average of 18 individuals per developmental stage.

Metrics of expression variability

Expression variability is generally measured by the coefficient of variation (CV) [40]. However, a gene's CV is strongly dependent on its RNA abundance (Figure S4). While this is an inherent property of time-interval counting processes (such as a Poisson process), it makes the comparison of variability between different conditions difficult [38,41]. Distance to median (DM, the distance between the squared CV of a gene and its running median) has been proposed as a variability metric that is independent of expression level [38,41,42]. However, the DM is still strongly negatively correlated with the mean expression level in our data (Figure S5). To avoid this dependency, we developed another variability measure, the adjusted standard deviation (adjusted SD), by calculating the ratio between observed SD and expected SD. Following the same approach as Barroso et al. [43], we performed polynomial regressions to predict the expected SD from mean expression. Since the adjusted SD metric works much better than DM in terms of accounting for the confounding effects of mean expression (Figure S6), we adopted it as a measure of expression variability in our study. As observed in yeast [42,44], we found that essential genes and hubs (proteins in the center of protein-protein interaction network) have lower expression variability than other genes (Figure S17), indicating selection to reduce it. This observation provides a control that we are indeed measuring biologically relevant expression variability.

Detailed calculation of expression variability:

1. Adjusted SD.

For each gene, we computed standard deviation (SD) in each stage and over all stages. Then we fitted a polynomial model to predict the global (across all stages) SD from the global mean

353 expression. We increased the degrees of the model until there was no more significant
354 improvement (tested with ANOVA, $p<0.05$ as a significant improvement). Then, based on this
355 best fitting model, for each gene, we computed its predicted global SD based on its global mean
356 expression. Finally, the adjusted SD of a gene in one stage is this gene's SD in its corresponding
357 stage divided by its predicted global SD. This method is derived from Barroso et al. [43], but
358 computing adjusted SD rather than adjusted variance.
359 2. Distance to median: the distance between the squared coefficient of variation (CV) of a
360 gene and its running median.
361 For each gene, we computed its squared CV in each stage and over all stages. Then, we ordered
362 genes based on their global (across all stages) mean expression. Next, we defined series of
363 sliding windows of 50 genes with 25 genes overlap, starting from lowest global mean
364 expression. Finally, the distance to median of a gene in one stage is the stage specific log10
365 squared CV minus the median of global log10 squared CV in this gene's corresponding
366 window. R code was modified from the DM function of the scran package [34].
367
368 **Bootstrap analysis**
369 For each stage, we randomly sampled the same number of samples. Then, we computed the
370 adjusted SD based on these random samples. We repeated the first two steps 500 times. Each
371 time, we only kept the median of the adjusted SD for each stage. Thus in each stage we obtained
372 500 medians. Finally, we performed a Wilcoxon test to test the significance of the difference
373 between the bootstrapped values of different stages.
374
375 **ChIP-Seq data analysis**
376 1. Histone modification signal datasets
377 The signal data files of four euchromatin histone modification marks (H3K4me1, H3K4me3,
378 H3K9ac, and H3K27ac) at six developmental stages (0-4h, 4-8h, 8-12h, 12-16h. 16-20h, 20-
379 24h) were downloaded from modENCODE [26] (NCBI GEO: GSE16013) (March, 2018). The
380 signal is smoothed, background-subtracted tag density. The signal was precomputed along
381 the genome in 35-bp windows.
382 2. Histone modification signal for promoter and gene body
383 For each gene, as suggested by Nicolas et al. [23], we separately calculated the mean signal
384 of its proximal promoter (2 kb upstream to 2 kb downstream for transcription start site (TSS))
385 and of its gene body (TSS to transcription end site (TES)) by using the bedtools "map"

386 command [45]. The TSS and TES information was retrieved from Ensembl release 91 [32]. For
387 a gene with several TSS and TES, we use its mean coordinates.
388 3. Histone modification signal Z score transformation
389 For each modification mark in each stage, the signal value was transformed into a Z score by
390 subtracting the mean signal across intergenic regions and dividing by the standard deviation
391 signal of the intergenic regions. The intergenic region were defined by removing all proximal
392 promoter regions and gene body regions with the bedtools "subtract" command [45]. Our
393 assumption is that on average such intergenic regions are not the target of active histone
394 modification signal, and thus allow to normalize between libraries. Then, for each gene, we
395 re-calculated the mean signal (Z score) of its proximal promoter (2 kb upstream to 2 kb
396 downstream for transcription start site (TSS)) and of its gene body (TSS to transcription end
397 site (TES)) by using the bedtools "map" command [45].
398
399 **Identification of stage specifically expressed genes**
400 Following the same approach as previously [46], we first defined 8 stage specific expressed
401 artificial expression profile (Figure S18A). Then, for each gene, we performed Pearson's
402 correlation between its real expression and this artificial expression. Finally, for each
403 artificial expression profile, we kept genes with top 10% correlation coefficient as the
404 corresponding stage specifically expressed genes (Figure S18B).
405
406 **Identification of hourglass expression variability genes**
407 Similar to the stage specifically expressed gene identification approach, we correlated each
408 gene's variability profile with the median across all genes. Then, we kept genes with the top
409 10% correlation coefficient as the ones following the global hourglass variability profile.
410
411 **Identification of genes expressed at all stages**
412 For each gene, we calculated the average expression across replicates in each stage. Then,
413 we defined the average expression > 1 as expressed.
414
415 **Identification of genes with constant expression across all stages**
416 For each gene, we first preformed one-way analysis of variance (ANOVA) to compare the
417 means of expression in different stage. Then, we calculated the *q-values* for multiple test

418 correction. Finally, the constantly expressed genes were defined as genes with *q-values* >
419 0.05.
420
421 **Gene ontology (GO) enrichment analysis**
422 We performed GO enrichment analysis for hourglass expression variability genes by using
423 the topGO [47] R package with the "elim" algorithm.
424
425 **Single Nucleotide Polymorphism (SNP) data**
426 The SNP data for 205 *D. melanogaster* inbred lines were downloaded from the Drosophila
427 Genetic Reference Panel (DGRP [48]) (December, 2018).
428
429 **Nucleotide diversity (π) calculation**
430 We calculated nucleotide diversity of promoters with vcftools [49].
431
432 **Transcriptome index analysis**
433 A "transcriptome index" [50,51] is a weighted mean of a feature over all genes, where the
434 weights are the expression levels of the genes at each condition (e.g., developmental stage).
435 The transcriptome index of phastCons was calculated as:

$$TPI_s = \frac{\sum_{i=1}^{n} phastCons_i * e_{is}}{\sum_{i}^{n} e_{is}},$$

436
437 where s is the developmental stage, $phastCons_i$ is the promoter sequence conservation score
438 of gene i, n is the total number of genes, and e_{is} is the expression level (log transformed) of
439 gene i in developmental stage s. For the transcriptome index of nucleotide diversity (π) the
440 same formula is used, replacing $phastCons_i$ by π_i.
441
442 **Meiosis related genes and transcription factors**
443 The Meiosis related genes and transcription factors were downloaded from AmiGO [52] (May,
444 2018).
445
446 **Individual unfertilized eggs RNA-seq data**
447 The normalized and log transformed expression matrix of individual unfertilized eggs was
448 downloaded from NCBI GEO: GSE68062 [53] (May, 2018).

449

450 **Dispersed and precise promoters**

451 The annotation of genes with dispersed or precise promoters was downloaded from Schor et
452 al [20] (June, 2019). Dispersed promoters are often associated with ubiquitously expressed
453 genes, have more dispersed patterns of transcriptional initiation, and do not contain a TATA
454 box. On the contrary, precise promoters are typically associated with restricted tissue-specific
455 expression and with a TATA box, and have a single predominant TSS.

456

457 **Essential gene annotation and protein connectivity datasets**

458 The gene essentiality and protein connectivity datasets were downloaded from OGEE v2 [54]
459 (March, 2018).

460

461 **PhastCons score**

462 The pre-computed sequence conservation score phastCons [55] of fly genome was downloaded
463 from http://hgdownload.soe.ucsc.edu/goldenPath/dm3/phastCons15way/ (February, 2018).
464 Higher value means higher conservation.

465

466 **Experimentally validated core promoters**

467 Experimentally validated transcription start sites (TSSs) were downloaded from the
468 Eukaryotic Promoter Database (EPD) [29] (May, 2018). For a gene with several TSSs, we
469 selected the most representative one (the TSS that has been validated by the largest number
470 of samples). The core promoter region was defined as 49 bp upstream TSS to 10 bp
471 downstream of the TSS [29]. We used EPD defined TSSs here because they are more accurate
472 for defining core promoters, whose function is expected to be related to sequence
473 conservation. Whereas for histone modification signal for promoter and gene body we used
474 Ensembl defined TSSs to be consistent with the source of TES information, and precision
475 was less important in defining broader proximal promoters.

476

Acknowledgements

We thank Virginie Braman for help with embryo collection and library preparation. We thank Daniel Alpern for help with the BRB-seq technology. We thank Richard Benton, David Garfield and Laurent Keller for critically reading and commenting on the manuscript. We thank Andrea Komljenovic and other members of the Robinson-Rechavi lab for helpful discussions. Part of the computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the SIB Swiss Institute of Bioinformatics. JL and MRR are supported by Swiss National Science Foundation grants 31003A_153341 / 1 and 31003A_173048. MRR and BD are supported by SystemsX.ch grant AgingX.

Author contributions

JL designed the work with input from MRR, MF and BD. MF led all experiments. JL performed all computational analyses. VG and BD contributed expertise in the BRB-seq experiments. JL and MRR interpreted the results with input from all the other authors. JL wrote the first draft of the paper. JL and MRR finalized the paper with input from all the other authors.

Supplementary figure legends

Figure S1: Relationship between uniquely mapped reads and expressed genes

Each dot represents one sample. The black dots indicate low quality samples with <4500 expressed genes or with <0.3 million uniquely mapped reads. The 239 orange colored samples were retained for downstream analysis ("high quality samples").

Figure S2: Proportion of retained samples in each development stage

The number of retained samples and of total samples in each stage is indicated in the bottom of each bar.

Figure S3: Evidence that the samples from the small cluster are unfertilized eggs

A. Boxplot of Spearman's correlation coefficients (rho) of expression between individual unfertilized eggs and each sample from the small cluster or from the large cluster, showing that the small cluster has an expression profile of unfertilized eggs. The lower and upper intervals indicated by the dashed lines ("whiskers") represent 1.5 times the interquartile range (IQR), and the box shows the lower and upper intervals of IQR together with the median.

B. Expression heat map of meiosis related genes across all samples, showing that their expression decreases over development for the large cluster, but is high in all samples of the small cluster, consistent with unfertilized eggs.

For testing of an alternative explanation of the two clusters as being males and females, see Figure S16.

Figure S4: Relationship between average expression and coefficient of variation at each stage

Pearson's correlation between average expression and coefficient of variation in each development stage is indicated in the top left of each subfigure.

Figure S5: Relationship between average expression and distance to median at each stage

Pearson's correlation between average expression and distance to median in each development stage is indicated in the top left of each subfigure.

Figure S6: Relationship between average expression and adjusted SD at each stage

Pearson's correlation between average expression and adjusted SD in each development stage is indicated in the top left of each subfigure.

Figure S7: Variation of expression variability across development using alternate measures of variability

A. Variability measured by adjusted SD; unlike in Figure 2, the variability in E1 was calculated using all samples from both small and large clusters.
B. Variability measured by coefficient of variation (CV).
C. Variability measured by distance to median (DM).

The legend is the same as for Figure 2. We performed pairwise Wilcoxon test between any two stages to test the significance. The multiple test corrected *p*-values (Benjamini–Hochberg method) are shown in Tables S6, S7 and S8.

Figure S8: Bootstrap analysis of the variability calculation

We performed pairwise Wilcoxon test between any two stages to test the significance. The multiple test corrected *p*-values (Benjamini–Hochberg method) are shown in Table S9.

Figure S9: Variation of expression variability across development for different categories of genes

A. Genes with constant expression level over development**.**
B. Transcription factor.

The legend is the same as for Figure 2. We performed pairwise Wilcoxon test between any two stages to test the significance. The multiple test corrected *p*-values (Benjamini–Hochberg method) are shown in Tables S10 and S11.

Figure S10: Variation of expression variability across development for dispersed promoter genes and for precise promoter genes

For each stage, the first and the second box represents dispersed promoter genes and precise promoter genes respectively. The legend is the same as for Figure 2. We performed pairwise Wilcoxon test between any two stages to test the significance separately for dispersed promoter genes and for precise promoter genes. The multiple test corrected *p*-values (Benjamini–Hochberg method) are shown in Tables S12 and S13.

Figure S11: Histone modification signal across development

The legend is the same as for Figure 3B and 3C. The median signal value in each development stage is indicated above each box. We performed pairwise Wilcoxon test between any two stages to test the significance. The multiple test corrected *p*-values (Benjamini–Hochberg method) for H3K4Me1, H3K27Ac and H3K9Ac are shown in Tables S14-S19.

Figure S12: Spearman's correlation coefficient between histone modification signal and promoter nucleotide diversity (π).

The legend is the same as for Figure 4A.

Figure S13: Spearman's correlation coefficient between histone modification signal and promoter sequence conservation for different definitions of promoter width

The figure legend is the same as in Figure 4A.

A. Promoter defined as 200 bp around TSS

B. Promoter defined as 400 bp around TSS

C. Promoter defined as 1000 bp around TSS

Figure S14: transcriptome index of π across development.

The legend is the same as for Figure 4C.

Figure S15: Multidimensional scaling analysis for all samples

Different colors indicate different stages. The solid triangles represent high quality samples according to Figure S1; the hollow triangles represent low quality samples which were discarded.

Figure S16: Mapping of X/autosome gene expression ratios to the multidimensional scaling analysis plot

We calculated the ratio of mean expression between genes from the X chromosome and from the autosomes for each sample. Red represents high ratio, blue represents low ratio. For Drosophila, dosage compensation is achieved by increasing expression of X chromosome genes in males. Since the dosage compensation is still incomplete during development, females should have a higher ratio of mean expression between genes from the X chromosome

and from the autosomes. Here, we found both high ratio samples and low ratio samples are well mixed in both the cluster and large clusters. Thus, we reject the hypothesis that the two different clusters are due to sex.

Figure S17: Relationship between expression variability and protein importance

We used the average variability across all development stages.

A. We split genes into 10 equally sized bins based on expression variability. The proportion of essential genes was fit by regression (the first degree of polynomial), whose R^2 and *p*-value are indicated in the top-left corner of each graph. The median expression variability of each bin was plotted on the x-axis.

B. Spearman's correlation between connectivity in a protein-protein interaction network and expression variability. The coefficient and *p*-value are indicated in the top-right. Loess regression lines are plotted in red.

Figure S18: Detection of stage specific genes

A. The artificial expression profile.

B. The expression of identified stage specific genes. The bold black line represents the median expression, the two gray lines represent 25th and 75th quantiles of expression, respectively.

References

1. Duboule, D. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. *Development* **1994**, 135–142 (1994).
2. Raff, R. A. *The shape of life : genes, development, and the evolution of animal form*. (University of Chicago Press., 1996).
3. Irie, N. & Kuratani, S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* **2**, 248 (2011).
4. Hu, H. *et al*. Constrained vertebrate evolution by pleiotropic genes. *Nat. Ecol. Evol.* **1**, 1722–1730 (2017).
5. Levin, M., Hashimshony, T., Wagner, F. & Yanai, I. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. *Dev. Cell* **22**, 1101–8 (2012).
6. Zalts, H. & Yanai, I. Developmental constraints shape the evolution of the nematode mid-developmental transition. *Nat. Ecol. Evol.* **1**, 0113 (2017).
7. Kalinka, A. T. *et al*. Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–4 (2010).
8. Raj, A., Rifkin, S. A., Andersen, E. & van Oudenaarden, A. Variability in gene expression underlies incomplete penetrance. *Nature* **463**, 913–8 (2010).
9. Li, X. *et al*. Systems Properties and Spatiotemporal Regulation of Cell Position Variability during Embryogenesis. *Cell Rep.* **26**, 313-321.e7 (2019).
10. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–26 (2008).
11. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464 (2005).
12. Eling, N., Morgan, M. D. & Marioni, J. C. Challenges in measuring and understanding biological noise. *Nat. Rev. Genet.* 1 (2019).
13. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic Gene Expression in a Single Cell. *Science* **297**, 1183–1186 (2002).
14. Waddington, C. H. Canalization of development and genetic assimilation of acquired characters. *Nature* **150**, 563–565 (1942).
15. Oates, A. C. What's all the noise about developmental stochasticity? *Development* **138**, 601–7 (2011).
16. Alpern, D. *et al*. BRB-seq: ultra-affordable high-throughput transcriptomics enabled

644 by bulk RNA barcoding and sequencing. *Genome Biol*. **20**, 71 (2019).
645 17. Levin, M. *et al*. The mid-developmental transition and the evolution of animal body
646 plans. *Nature* **531**, 637–41 (2016).
647 18. Schep, A. N. & Adryan, B. A Comparative Analysis of Transcription Factor
648 Expression during Metazoan Embryonic Development. *PLoS One* **8**, e66826 (2013).
649 19. Tadros, W. & Lipshitz, H. D. The maternal-to-zygotic transition: a play in two acts.
650 *Development* **136**, 3033–3042 (2009).
651 20. Schor, I. E. *et al*. Promoter shape varies across populations and affects promoter
652 evolution and expression noise. *Nat. Genet*. **49**, 550–558 (2017).
653 21. Benayoun, B. A. *et al*. H3K4me3 breadth is linked to cell identity and transcriptional
654 consistency. *Cell* **158**, 673–88 (2014).
655 22. Wu, S. *et al*. Independent regulation of gene expression level and noise by histone
656 modifications. *PLOS Comput. Biol*. **13**, e1005585 (2017).
657 23. Nicolas, D., Zoller, B., Suter, D. M. & Naef, F. Modulation of transcriptional burst
658 frequency by histone acetylation. *Proc. Natl. Acad. Sci. U. S. A*. **115**, 7153–7158
659 (2018).
660 24. Weinberger, L. *et al*. Expression Noise and Acetylation Profiles Distinguish HDAC
661 Functions. *Mol. Cell* **47**, 193–202 (2012).
662 25. Faure, A. J., Schmiedel, J. M. & Lehner, B. Systematic Analysis of the Determinants
663 of Gene Expression Noise in Embryonic Stem Cells. *Cell Syst*. **5**, 471-484.e4 (2017).
664 26. Nègre, N. *et al*. A cis-regulatory map of the Drosophila genome. *Nature* **471**, 527–531
665 (2011).
666 27. Lehner, B. Genes Confer Similar Robustness to Environmental, Stochastic, and
667 Genetic Perturbations in Yeast. *PLoS One* **5**, e9035 (2010).
668 28. Meiklejohn, C. D. & Hartl, D. L. A single mode of canalization. *Trends Ecol. Evol*. **17**,
669 468–473 (2002).
670 29. Dreos, R., Ambrosini, G., Périer, R. C. & Bucher, P. The Eukaryotic Promoter
671 Database: expansion of EPDnew and new promoter analysis tools. *Nucleic Acids Res*.
672 **43**, D92–D96 (2015).
673 30. Wagner, G. P., Booth, G. & Bagheri-Chaichian, H. A Population Genetic Theory of
674 Canalization. *Evolution* **51**, 329 (1997).
675 31. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
676 (2013).

677 32. Zerbino, D. R. *et al*. Ensembl 2018. *Nucleic Acids Res*. **46**, D754–D761 (2018).
678 33. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-
679 throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
680 34. Lun, A. T. L., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level
681 analysis of single-cell RNA-seq data. *F1000Research* **5**, 2122 (2016).
682 35. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
683 differential expression analysis of digital gene expression data. *Bioinformatics* **26**,
684 139–40 (2010).
685 36. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear
686 model analysis tools for RNA-seq read counts. *Genome Biol*. **15**, R29 (2014).
687 37. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray
688 expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
689 38. Kolodziejczyk, A. A. *et al*. Single Cell RNA-Sequencing of Pluripotent States Unlocks
690 Modular Transcriptional Variation. *Cell Stem Cell* **17**, 471–85 (2015).
691 39. Avilés-Pagán, E. E. & Orr-Weaver, T. L. Activating embryonic development in
692 Drosophila. *Semin. Cell Dev. Biol* **84**, 100-110 (2018).
693 40. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and
694 control. *Science* **309**, 2010–3 (2005).
695 41. Tung, P.-Y. *et al*. Batch effects and the effective design of single-cell gene expression
696 studies. *Sci. Rep*. **7**, 39921 (2017).
697 42. Newman, J. R. S. *et al*. Single-cell proteomic analysis of S. cerevisiae reveals the
698 architecture of biological noise. *Nature* **441**, 840–846 (2006).
699 43. Barroso, G. V., Puzovic, N. & Dutheil, J. Y. The Evolution of Gene-Specific
700 Transcriptional Noise Is Driven by Selection at the Pathway Level. *Genetics* **208**, 173–
701 189 (2018).
702 44. Lehner, B. Selection to minimise noise in living systems and its implications for the
703 evolution of gene expression. *Mol. Syst. Biol*. **4**, 170 (2008).
704 45. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
705 genomic features. *Bioinformatics* **26**, 841–2 (2010).
706 46. Liu, J. & Robinson-Rechavi, M. Adaptive Evolution of Animal Proteins over
707 Development: Support for the Darwin Selection Opportunity Hypothesis of Evo-Devo.
708 *Mol. Biol. Evol*. **35**, 2862–2872 (2018).
709 47. Alexa, A., Rahnenfuhrer, J. & Lengauer, T. Improved scoring of functional groups

from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).

48. Huang, W. *et al*. Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Res*. **24**, 1193–208 (2014).

49. Danecek, P. *et al*. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–8 (2011).

50. Liu, J. & Robinson-Rechavi, M. Developmental Constraints on Genome Evolution in Four Bilaterian Model Species. *Genome Biol. Evol*. **10**, 2266–2277 (2018).

51. Domazet-Loso, T. & Tautz, D. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010).

52. Carbon, S. *et al*. AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288–9 (2009).

53. Paris, M., Villalta, J. E., Eisen, M. B. & Lott, S. E. Sex Bias and Maternal Contribution to Gene Expression Divergence in Drosophila Blastoderm Embryos. *PLoS Genet*. **11**, e1005592 (2015).

54. Chen, W.-H., Lu, G., Chen, X., Zhao, X.-M. & Bork, P. OGEE v2: an update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res*. **45**, D940–D944 (2017).

55. Siepel, A. *et al*. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. **15**, 1034–50 (2005).