


# Comparative studies of genomic and epigenetic factors influencing transcriptional variation in two insect species

Xin Wu , Neharika Bhatia ,<sup>1</sup> Christina M. Grozinger ,<sup>2</sup> Soojin V. Yi <sup>1,3,\*</sup>

<sup>1</sup>School of Biological Sciences, Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA 30332, USA

<sup>2</sup>Department of Entomology, Center for Pollinator Research, Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16801, USA

<sup>3</sup>Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, CA 93106, USA

\*Corresponding author: Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, Santa Barbara, CA 93106, USA.

Email: soojinyi@ucsb.edu

## Abstract

Different genes show different levels of expression variability. For example, highly expressed genes tend to exhibit less expression variability. Genes whose promoters have TATA box and initiator motifs tend to have increased expression variability. On the other hand, DNA methylation of transcriptional units, or gene body DNA methylation, is associated with reduced gene expression variability in many species. Interestingly, some insect lineages, most notably Diptera including the canonical model insect *Drosophila melanogaster*, have lost DNA methylation. Therefore, it is of interest to determine whether genomic features similarly influence gene expression variability in lineages with and without DNA methylation. We analyzed recently generated large-scale data sets in *D. melanogaster* and honey bee (*Apis mellifera*) to investigate these questions. Our analysis shows that increased gene expression levels are consistently associated with reduced expression variability in both species, while the presence of TATA box is consistently associated with increased gene expression variability. In contrast, initiator motifs and gene lengths have weak effects limited to some data sets. Importantly, we show that a sequence characteristics indicative of gene body DNA methylation is strongly and negatively associated with gene expression variability in honey bees, while it shows no such association in *D. melanogaster*. These results suggest the evolutionary loss of DNA methylation in some insect lineages has reshaped the molecular mechanisms concerning the regulation of gene expression variability.

**Keywords:** gene expression variability; DNA methylation; insects; core promoter elements

## Introduction

Expression levels of the same genes show variability at different levels of biological units, from cells to populations. The widely observed variability of gene expression at different levels is thought to share some common underlying molecular mechanisms (Ho et al. 2008; Tirosch et al. 2009; Lehner 2010; Li et al. 2010; Alemu et al. 2014). This variation, which we will refer to as gene expression variability, is implicated in several phenotypic traits such as aging, development, disease, and immunity (Ho et al. 2008; Li et al. 2010; Hagai et al. 2018; Bashkeel et al. 2019).

Previous studies of gene expression variability have discovered that highly expressed genes tend to have reduced variability between individuals in diverse taxa (Bird 1995; Choi and Kim 2008; Huh et al. 2013; Wu, Lindsey et al. 2020). It is hypothesized that natural selection has reduced the expression variability of highly expressed genes due to the advantages associated with the improved control of the inherent stochasticity of transcription and subsequent protein synthesis (Fraser et al. 2004; Newman et al. 2006; Wang and Zhang 2011; Barroso et al. 2018). Genes that are constitutively highly expressed are typically essential housekeeping genes, where noise is therefore minimized by natural selection (Fraser et al. 2004; Wang and Zhang 2011; Barroso et al. 2018).

Other traits that significantly associated with gene expression variability include genomic features such as gene length, presence of a TATA box, and initiator motifs (Park et al. 2012; Huh et al. 2013; Ravarani et al. 2016; Faure et al. 2017; Wu, Lindsey et al. 2020). The presence of a TATA box is strongly associated with high gene expression variability, with other core promoter elements such as initiator motifs and GC motifs being associated with higher gene expression variability to a much lesser degree (Faure et al. 2017). Therefore, genomic features can play significant roles in shaping gene expression variability.

Gene body DNA methylation, which is an ancestral form of epigenetic regulation in animal genomes, is negatively associated with gene expression variability in humans (Huh et al. 2013). Studies in insects, in particular from hymenopteran species including fire ants and wasps, also reported similar observations (Zeng and Yi 2010; Hunt et al. 2013; Wang et al. 2016; Wu, Lindsey et al. 2020). However, the relative contribution of gene body DNA methylation compared to the aforementioned genomic features has not been examined in insects. One of the reasons for this lack of knowledge is due to the fact that the most extensively studied model insect *Drosophila melanogaster*, and other species in the Diptera, lacks DNA methylation which is best explained by lineage-specific loss (Sarda et al. 2012; Yi 2012). To address the

Received: July 19, 2022. Accepted: August 05, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

gap of knowledge, here we examined relative impacts of different genomic features on gene expression variability through a comparative analysis of honey bees (*Apis mellifera*), a hymenopteran lineage possessing ancestral gene body methylation, and *D. melanogaster*. We integrated data from 15 *D. melanogaster* studies and 8 *A. mellifera* studies to comprehensively address these questions (Table 1). Our study confirms impacts of several genomic features on gene expression variability in both species. Intriguingly, our analyses indicate pervasive and significant effects of DNA methylation on gene expression variability in *A. mellifera*. These results provide new insights into regulation of gene expression variability in insects.

## Materials and methods

### Gene expression data

We analyzed a total of 20 published RNA-seq studies for this study, 12 of which are from fly (*D. melanogaster*) and 8 from honey bee (*A. mellifera*; Table 1). Our *D. melanogaster* datasets were chosen from a diverse set of laboratories as well as recently published with at least 10 samples (no more than 2 years old). The *A. mellifera* studies were all of the RNA-seq datasets we could access, as well as being fairly recent and a minimum of 10 samples (one from 2012, the rest were from 2016 to 2020).

### Data processing

Reads for each study were trimmed to remove low-quality reads and adaptors using default Trim\_galore! (Martin 2011) settings. Trimmed reads were then aligned to their respective genomes, *amel* 4.5 and *dmel* r6.33 for *A. mellifera* and *D. melanogaster*, respectively, using HISAT2 with soft clipping disabled (parameter setting: `-sp 1,000, 1,000`). Following alignment, gene counts were generated with HTSeq (Anders et al. 2015) default parameters and imported into R (R Core Team 2014) for further downstream analyses. Gene expression for each study was quantified and normalized using the “estimateSizeFactors” function in the DESeq2

package (Love et al. 2014). To remove lowly expressed genes, we removed genes with counts less than 5 and also required a gene to be expressed in at least 10% of all samples in the study. Gene expression variation was measured as the % coefficient of variation (CV) of gene expression (Huh et al. 2013) and CpG O/E values for the *A. mellifera* genome were calculated as previously described (Lindsey et al. 2018).

### Data processing

Core promoter element designations for TATA boxes and initiator motifs were obtained from the Eukaryotic Promoter Database (Cavin Perier et al. 1998; Dreos et al. 2017). Briefly, promoter classifications for each organism were downloaded from the database using the “EPDnew selection tool” as performed in a previous study (Faure et al. 2017).

### Statistics

For our full linear model, gene expression variation was used as the response variable for the following quadratic model:  $\log_{10}(\text{CV}) \sim \log_2(\text{expression}) + \log_2(\text{expression})^2 + \log_{10}(\text{gene length}) + \text{TATA box} + \text{Initiator motif} + X$ , where X are additional covariates from each experiment based on its metadata file. In our second set of linear models, we first regressed out the effect of gene expression with  $\log_{10}(\text{CV}) \sim \log_2(\text{expression}) + \log_2(\text{expression})^2$  and then using the residuals as the response variable mirroring the full linear model:  $\text{residuals} \sim \log_{10}(\text{gene length}) + \text{TATA box} + \text{Initiator motif} + X$ . Partial correlation was performed using the “pcorr” function in R with gene expression as the variable that was controlled for and gene length and CpG O/E (*A. mellifera* studies only) as the response variables.

## Results

### Core promoter elements are significant contributors to gene expression variation

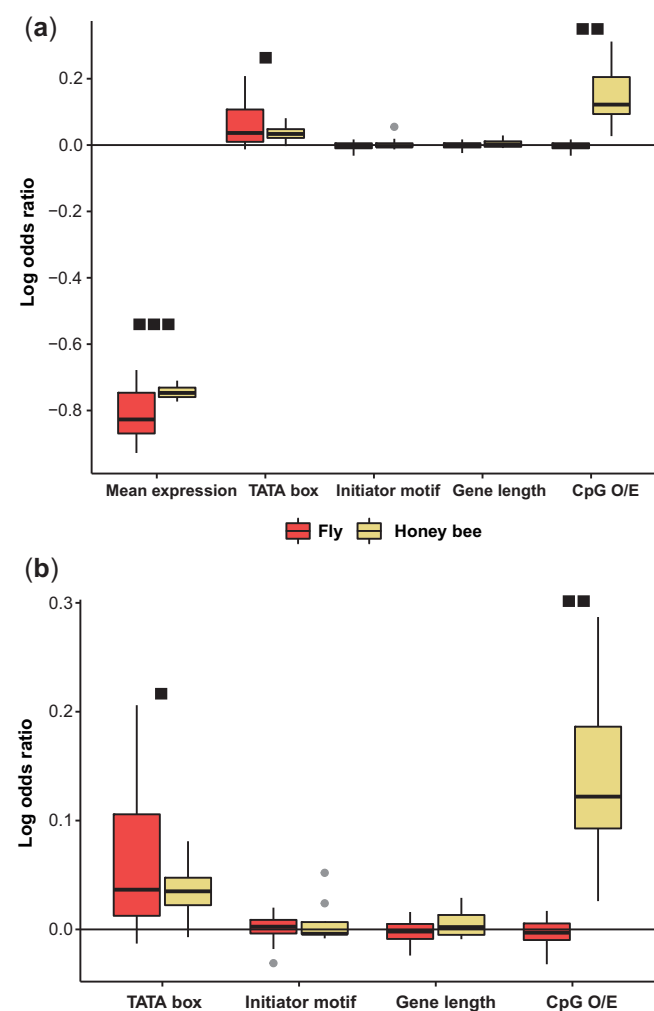
For all the data sets (Table 1), we first quantified gene expression variability as the CV of gene expression, as in previous studies (Huh et al. 2013; Islam et al. 2014; Fan et al. 2016). We then evaluated the contributing factors using a linear model using the following covariates: mean gene expression, gene length, presence of a TATA box, and presence of an initiator motif (see Methods). Our main motivation was to examine the impact of DNA methylation on gene expression variability. However, for data sets in *A. mellifera*, matching data on DNA methylation and gene expression are sparse, and the existing data sets tend to have small number of biological replicates. Therefore, for *A. mellifera* data sets, we included CpG O/E as an additional covariate which is an approximate measure of DNA methylation (Elango et al. 2009). Notably, CpG O/E strongly negatively correlates with the measured DNA methylation levels in honey bee and other invertebrates in literature (e.g. Sarda et al. 2012), and also in a compiled data set from *A. mellifera* (Supplementary Fig. 2).

As expected, average gene expression was strongly negatively correlated with gene expression variability and was by far the most significant term with the largest coefficient in the linear model in all datasets (Huh et al. 2013; Islam et al. 2014; Fan et al. 2016; Wu, Lindsey et al. 2020) (Fig. 1a and Supplementary Table 1). Following mean expression, the presence of a TATA box in the gene promoter region was a significant term in 12 of the 15 *D. melanogaster* datasets (Supplementary Table 1) and in 6 of the 8 *A. mellifera* datasets (Supplementary Table 1). The presence of a TATA box was positively and significantly correlated with gene expression variability in all but one study (Hood et al. 2020). The

**Table 1.** Source datasets for gene expression variability.

Source publication	Sample size	Sample type	GEO accession
<i>Drosophila melanogaster</i>			
Genç et al. (2020)	23	Head	GSE153225
Seong et al. (2020)	42	Head, fat body, testes	GSE140950
Wang et al. (2020)	24	Head	GSE109489
Shah et al. (2021)	24	Brain	GSE140663
Lindsey et al. (2021)	48	Whole body	GSE162666
Hood et al. (2020)	24	Whole body	GSE158189
Agren et al. (2020)	72	Whole body	GSE155395
Miozzo and Nagoshi, unpublished results	18	Brain	GSE156890
Thackray et al. (2020)	36	Brain	GSE144028
Brown et al. (2020)	30	Head	GSE144433
Weigelt et al. (2020)	78	Brain, Thorax, Fat Body	GSE130158
<i>Apis mellifera</i>			
Shpigler et al. (2017)	180	Mushroom body	GSE85876
Harris et al. (2019)	29	Head and pupa	GSE116629
Liberti et al. (2019)	18	Brain	GSE127185
Rutter and Cook (2020)	24	Whole body	GSE121885
Naeger and Robinson (2016)	48	Mushroom body	GSE85433
Jasper et al. (2020)	38	Fat body and brain	GSE145395
Traniello et al. (2020)	96	Brain	GSE130700
Doublet et al. (2016)	11	Brain	GSE81664

other core promoter element, presence of initiator motif, was only significant in approximately half of the studies (6 out of 12 *D. melanogaster* studies; 4 out of 8 *A. mellifera* studies; [Supplementary Table 1](#)). The direction of correlation for the initiator motif was also less consistent than the previous 2 discussed factors, as the coefficient was positive in 3 of the 6 *D. melanogaster* datasets it was significant in and in 2 of the 4 *A. mellifera* datasets it was significant in ([Supplementary Table 1](#)). Lastly, gene length, while a significant term in the majority of datasets, also failed to display a consistent direction of correlation in either *D. melanogaster* or *A. mellifera* datasets. In conclusion, in the linear models, we observed a strong and significant negative correlation between mean expression and expression variability along with consistent, though not always significant, correlation between



**Fig. 1.** Linear model covariate coefficients summary. We used the coefficient of variation of gene expression as a proxy for gene expression variability and modeled it using a set of key factors hypothesized to affect transcriptional noise. The log odds ratio was calculated for each covariate coefficient (*Methods*) in each study. a) Box plot of log ratio of covariate coefficients for the full linear model including gene expression, presence of a TATA box, presence of an initiator motif, gene length, and CpG O/E. b) Log ratio of covariate coefficients with mean expression regressed out using a quadratic model (*Methods*). The number of boxes above the boxplots represents the percentage of studies the covariate is significant in. Three boxes mean the covariate is statistically significant ( $P < 0.05$ ) in all studies (100%), 2 boxes represent significance in 50–70% of all studies, and 1 box represents significance in at most 30% of the studies. Boxes for CpG O/E only apply for the *A. mellifera* datasets and not the *D. melanogaster* data sets.

the presence of a TATA box and expression variability ([Fig. 1a](#)). The other promoter element, initiator motifs, and gene length did not show a consistent relationship with expression variability.

Interestingly, CpG O/E, which we used as a proxy for the level of gene body methylation, was highly significantly and positively correlated with gene expression variability in *A. mellifera* data sets ([Fig. 1](#) and [Supplementary Tables 1 and 2](#)). Thus, in addition to gene expression, CpG O/E displayed strong and stable correlation with gene expression variability across all *A. mellifera* datasets. Since CpG O/E itself is negatively correlated with DNA methylation, these results imply that gene body DNA methylation is associated with reduced gene expression variability.

Because of the strong effects of mean gene expression on gene expression variability, we applied another strategy to control for this effect. We first regressed out mean expression using a quadratic model (see *Methods*). We used the quadratic model as it was shown to have fairly unbiased residual distributions and previously applied to model the relationship between gene expression and expression variability ([Alemu et al. 2014](#)). The residual from this regression would reflect the remaining variability independent of the mean gene expression levels, which then can be interrogated for other genomic factors. This analysis yielded almost identical results as our initial linear models ([Supplementary Table 2](#)). For the TATA box term, the significance at the  $P < 0.05$  threshold and the direction of correlation remained the same for all *A. mellifera* studies. Similarly, the  $P$ -value for the TATA box term was nearly the same for the *D. melanogaster* datasets, with only one study ([Thackray et al. 2020](#)), having a small change going from  $P = 0.055$  in the full model to  $P = 0.048$  ([Supplementary Table 2](#)). For the initiator motif term, the direction changed for one study ([Thackray et al. 2020](#)) while significance remained the same for all *D. melanogaster* studies ([Supplementary Table 2](#)). Gene length, as with the other covariates, was the same across all studies with the exception of one study ([Brown et al. 2020](#)), which was no longer statistically significant after regressing out the effects of gene expression ([Supplementary Table 2](#)). Due to the expected strong effects of mean expression on expression variability, there was a sharp drop off in  $R^2$  values across these analyses. Specifically, after regressing out gene expression, only 3 *D. melanogaster* and 2 *A. mellifera* studies had models explaining more than 10% of the variance in expression variability. Nevertheless, the results of both linear model approaches indicate that the presence of a TATA box in the gene promoter region is consistently correlated with higher expression variability ([Fig. 1b](#)).

When mean expression is regressed out, the CpG O/E term was the most impactful term in the *A. mellifera* datasets ([Fig. 1b](#) and [Supplementary Table 2](#)). It was highly significant in all 8 studies ( $P < 0.001$  in all cases) with an average coefficient of 0.14 ([Fig. 1b](#) and [Supplementary Table 2](#)). None of the *D. melanogaster* datasets contained significant CpG O/E terms which was an expected result given the lack of genomic DNA methylation ([Fig. 1b](#); [Sarda et al., 2012](#); [Yi, 2012](#)). Therefore, controlling for gene expression yielded similar results suggesting that outside of gene expression itself, DNA methylation had the biggest impact on gene expression variability followed by the core promoter elements ([Fig. 1b](#) and [Supplementary Table 2](#)).

We also used a partial correlations approach ([Kim and Yi 2006; 2007](#)) to examine effects of covariates free from the effects of gene expression. Specifically, we separately applied partial correlations for each numerical variable (gene length for both organisms in addition to CpG O/E for *A. mellifera*) while controlling for mean expression. Using this method, gene length was a

significant term in 10 *D. melanogaster* and 6 *A. mellifera* datasets (Supplementary Table 3). CpG O/E was once again highly significant in all 8 *A. mellifera* datasets with an average partial Pearson correlation of 0.157 (Supplementary Table 3).

These results remained (Supplementary Table 4) when we used only a subset of gene sets that are likely to be methylated, using the CpG O/E cutoff of 0.8, following Sarda et al. (2012). Furthermore, to test if the results were affected by the tissue heterogeneity, we used a subset of data sets from brain (3 from the fly studies, 3 from honey bee studies, Table 1). Analyses of this subset of data recapitulate the same patterns as seen in the whole data set (Supplementary Fig. 2). Even though we need to further consider additional tissues, this analysis indicates that the significance of gene body DNA methylation on gene expression variability is apparent in the brain data sets from the 2 species. We also examined experimentally determined DNA methylation data sets (Wu, Galbraith et al. 2020) and gene expression (Galbraith et al. 2015). These data sets are from bees in the same genetic background although the exact individuals are not matched. Using DNA methylation levels instead of CpG O/E, we performed the same statistical analyses. We found that in the full model, DNA methylation level is a highly significant predictor of gene expression variability, followed by the mean expression level ( $P < 10^{-15}$  for both terms, Supplementary Table 5). When the expression level is regressed out, DNA methylation was the best predictor of gene expression variability (Supplementary Table 5).

## Discussion

By utilizing the emerging richness of large-scale gene expression data sets, we initiated a study of the genomic and epigenetic factors underlying gene expression variability in a canonical model insect fruit fly (*D. melanogaster*) and the honey bee (*A. mellifera*). Since *D. melanogaster* does not exhibit widespread DNA methylation and *A. mellifera* does, this comparison is useful to ask whether DNA methylation is associated with gene expression variability in insects. While a negative association between DNA methylation and gene expression variability was previously found in 3 studies (Hunt et al. 2013; Wang et al. 2016; Wu, Lindsey et al. 2020), these were limited in sample size and did not control for other genomic features.

We show that, as in other taxa, highly expressed genes tend to have reduced expression variability in both flies and honey bees. In addition, the presence of TATA boxes was often associated with increased noise, consistent with previous studies (Blake et al. 2003; Lehner 2008; Ravarani et al. 2016; Faure et al. 2017). Initiator motifs and gene lengths had some effects in some data sets, but their effects were not consistent across the data sets (Fig. 1).

While the effect of gene expression levels and TATA box and initiator motifs were consistent between *D. melanogaster* and *A. mellifera*, these 2 species differ by the presence (*A. mellifera*) and absence (*D. melanogaster*) of DNA methylation. To contrast the effect of DNA methylation on these 2 species, we utilized CpG O/E as a proxy measurement for gene body DNA methylation (Elango et al. 2009). As expected from the lack of DNA methylation in *D. melanogaster* genome, CpG O/E in *D. melanogaster* is unimodally distributed and shows no relationship with gene expression (Fig. 1b). In contrast, in *A. mellifera*, for all of our statistical methods (full linear model, linear model with mean expression regressed out, and partial correlations), the CpG O/E term was highly significantly and positively correlated with gene expression variability (Fig. 1 and Supplementary Tables 1–3). Thus, both

mean expression, which was by far the most significant and impactful covariate, and CpG O/E displayed strong and stable correlation with gene expression variation across all *A. mellifera* datasets. Since CpG O/E itself is negatively correlated with DNA methylation (Sarda et al. 2012 and Supplementary Fig. 1), these results align with previous findings in both mammals and insects that DNA methylation is associated with reduced gene expression variation (Huh et al. 2013; Wu, Lindsey et al. 2020). These results strongly imply that gene body DNA methylation is associated with gene expression variability in *A. mellifera*, and its effect is stronger and more consistent than other genomic features such as TATA box, initiator motif and gene lengths.

The molecular mechanism underlying this association between DNA methylation and gene expression variation is not resolved in *A. mellifera*, but studies in other taxa have revealed several potential pathways (Choi and Kim 2009; Coleman-Derr and Zilberman 2012; Huh et al. 2013). For example, gene body methylation may directly or indirectly reduce spurious intragenic transcription, by avoiding erroneous intron retention (Horvath et al. 2019) or interacting with other epigenetic modifications (Coleman-Derr and Zilberman 2012). Our results beg the question of how *D. melanogaster* regulate expression variability in the absence of DNA methylation. There is no indication of *D. melanogaster* having greater variability of gene expression in our data, and there is very little difference in terms of  $R^2$  of any of the statistical models. It is likely that alternative mechanisms have evolved in *D. melanogaster* to compensate the lack of DNA methylation. For example, gene body DNA methylation in *A. mellifera* and several histone markers in *D. melanogaster* are highly associated (Hunt et al. 2013). Comparing taxa where the loss of DNA methylation has occurred recently could provide some insight into the evolution of regulatory mechanisms underlying gene expression variability.

## Data availability

All data used in this study are found in public domain and the accession numbers are indicated in Table 1.

Supplemental material is available at G3 online.

## Acknowledgments

The authors thank the members of the Yi lab and anonymous reviewers for their helpful comments.

## Funding

This work was supported by the National Science Foundation [grant number MCB-0950896] to SVY and CMG.

## Conflicts of interest

None declared.

## Literature cited

- Agren JA, Munasinghe M, Clark AG. Mitochondrial-Y chromosome epistasis in *Drosophila melanogaster*. *Proc Biol Sci*. 2020;287(1937):20200469.
- Alemu EY, Carl JW, Jr, Bravo HC, Hannenhalli S. Determinants of expression variability. *Nucleic Acids Res*. 2014;42(6):3503–3514.

- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31(2):166–169.
- Barroso GV, Puzovic N, Duthel JY. The evolution of gene-specific transcriptional noise is driven by selection at the pathway level. *Genetics*. 2018;208(1):173–189.
- Bashkeel N, Perkins TJ, Kærn M, Lee JM. Human gene expression variability and its dependence on methylation and aging. *BMC Genomics*. 2019;20(1):941.
- Bird AP. Gene number, noise reduction and biological complexity. *Trends Genet*. 1995;11(3):94–100.
- Blake WJ, Kaern M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. *Nature*. 2003;422(6932):633–637.
- Brown EB, Layne JE, Elchert AR, Rollmann SM. Behavioral and transcriptional response to selection for olfactory behavior in *Drosophila*. G3 (Bethesda). 2020;10(4):1283–1296.
- Cavin Perier R, Junier T, Bucher P. The eukaryotic promoter database EPD. *Nucleic Acids Res*. 1998;26(1):353–357.
- Choi JK, Kim YJ. Epigenetic regulation and the variability of gene expression. *Nat Genet*. 2008;40(2):141–147.
- Choi JK, Kim YJ. Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nat Genet*. 2009;41(4):498–503.
- Coleman-Derr D, Zilberman D. Deposition of histone variant H2A.Z within gene bodies regulates responsive genes. *PLoS Genet*. 2012; 8(10):e1002988.
- Doublet V, Paxton RJ, McDonnell CM, Dubois E, Nidelet S, Moritz RFA, Alaux C, Le Conte Y. Brain transcriptomes of honey bees (*Apis mellifera*) experimentally infected by two pathogens: black queen cell virus and *Nosema ceranae*. *Genom Data*. 2016;10:79–82.
- Dreos R, Ambrosini G, Groux R, Perier RC, Bucher P. The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Res*. 2017;45(D1):D51–D55.
- Elango N, Hunt BG, Goodisman MA, Yi SV. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proc Natl Acad Sci U S A*. 2009; 106(27):11206–11211.
- Fan J, Salathia N, Liu R, Kaeser GE, Yung YC, Herman JL, Kaper F, Fan J-B, Zhang K, Chun J, et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat Methods*. 2016;13(3):241–244.
- Faure AJ, Schmiedel JM, Lehner B. Systematic analysis of the determinants of gene expression noise in embryonic stem cells. *Cell Syst*. 2017;5(5):471–484.e4.
- Fraser HB, Hirsh AE, Giaever G, Kumm J, Eisen MB. Noise minimization in eukaryotic gene expression. *PLoS Biol*. 2004;2(6):e137.
- Galbraith DA, Yang X, Nino EL, Yi S, Grozinger C. Parallel epigenomic and transcriptomic responses to viral infection in honey bees (*Apis mellifera*). *PLoS Pathog*. 2015;11(3):e1004713.
- Genç Ö, An J-Y, Fetter RD, Kulik Y, Zunino G, Sanders SJ, Davis GW. Homeostatic plasticity fails at the intersection of autism-gene mutations and a novel class of common genetic modifiers. *Elife*. 2020;9:e55775.
- Hagai T, Chen X, Miragaia RJ, Rostom R, Gomes T, Kunowska N, Henriksson J, Park J-E, Proserpio V, Donati G, et al. Gene expression variability across cells and species shapes innate immunity. *Nature*. 2018;563(7730):197–202.
- Harris KD, Lloyd JPB, Domb K, Zilberman D, Zemach A. DNA methylation is maintained with high fidelity in the honey bee germline and exhibits global non-functional fluctuations during somatic development. *Epigenetics Chromatin*. 2019;12(1):62.
- Ho JWK, Stefani M, dos Remedios CG, Charleston MA. Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics*. 2008;24(13):i390–i398.
- Hood SE, Kofler XV, Chen Q, Scott J, Ortega J, Lehmann M. Nuclear translocation ability of Lipin differentially affects gene expression and survival in fed and fasting *Drosophila*. *J Lipid Res*. 2020; 61(12):1720–1732.
- Horvath R, Laenen B, Takuno S, Slotte T. Single-cell expression noise and gene-body methylation in *Arabidopsis thaliana*. *Heredity (Edinb)*. 2019;123(2):81–91.
- Huh I, Zeng J, Park T, Yi SV. DNA methylation and transcriptional noise. *Epigenetics Chromatin*. 2013;6(1):9.
- Hunt BG, Glastad KM, Yi SV, Goodisman MA. Patterning and regulatory associations of DNA methylation are mirrored by histone modifications in insects. *Genome Biol Evol*. 2013;5(3):591–598.
- Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, Lönnerberg P, Linnarsson S. Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods*. 2014;11(2):163–166.
- Jasper WC, Brutscher LM, Grozinger CM, Nino EL. Injection of seminal fluid into the hemocoel of honey bee queens (*Apis mellifera*) can stimulate post-mating changes. *Sci Rep*. 2020;10(1):11990.
- Kim S-H, Yi S. Correlated asymmetry between sequence and functional divergence of duplicate proteins in *Saccharomyces cerevisiae*. *Mol Biol Evol*. 2006;23(5):1068–1075.
- Kim S-H, Yi S. Understanding relationship between sequence and functional evolution in yeast proteins. *Genetica*. 2007;131(2):151–156.
- Lehner B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol Syst Biol*. 2008;4: 170.
- Lehner B. Genes confer similar robustness to environmental, stochastic, and genetic perturbations in yeast. *PLoS One*. 2010;5(2): e9035.
- Li J, Liu Y, Kim T, Min R, Zhang Z. Gene expression variability within and between human populations and implications toward disease susceptibility. *PLoS Comput Biol*. 2010;6(8):e1000910.
- Liberti J, Görner J, Welch M, Dosselli R, Schiøtt M, Ogawa Y, Castleden I, Hemmi JM, Baer-Imhoof B, Boomsma JJ, et al. Seminal fluid compromises visual perception in honeybee queens reducing their survival during additional mating flights. *Elife*. 2019;8:e45009.
- Lindsey ARI, Bhattacharya T, Hardy RW, Newton ILG. Wolbachia and virus alter the host transcriptome at the interface of nucleotide metabolism pathways. *mBio*. 2021;12(1):1–17.
- Lindsey ARI, Kelkar YD, Wu X, Sun D, Martinson EO, Yan Z, Rugman-Jones PF, Hughes DST, Murali SC, Qu J, et al. Comparative genomics of the miniature wasp and pest control agent *Trichogramma pretiosum*. *BMC Biol*. 2018;16(1):54.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011;17(1):10–12.
- Naeger NL, Robinson GE. Transcriptomic analysis of instinctive and learned reward-related behaviors in honey bees. *J Exp Biol*. 2016; 219(Pt 22):3554–3561.
- Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*. 2006; 441(7095):840–846.
- Park J, Xu K, Park T, Yi SV. What are the determinants of gene expression levels and breadths in the human genome? *Hum Mol Genet*. 2012;21(1):46–56.
- Ravarani CN, Chalancon G, Breker M, de Groot NS, Babu MM. Affinity and competition for TBP are molecular determinants of gene expression noise. *Nat Commun*. 2016;7:10417.

- Rutter L, Cook D. bigPint: a Bioconductor visualization package that makes big data pint-sized. *PLoS Comput Biol.* 2020;16(6): e1007912.
- Sarda S, Zeng J, Hunt BG, Yi SV. The evolution of invertebrate gene body methylation. *Mol Biol Evol.* 2012;29(8):1907–1916.
- Seong K-H, Ly NH, Katou Y, Yokota N, Nakato R, Murakami S, Hirayama A, Fukuda S, Kang S, Soga T, et al. Paternal restraint stress affects offspring metabolism via ATF-2 dependent mechanisms in *Drosophila melanogaster* germ cells. *Commun Biol.* 2020; 3(1):208.
- Shah EJ, Huttemann M, Sanderson TH, Gurdziel K, Ruden DM. Inhibiting mitochondrial cytochrome c oxidase downregulates gene transcription after traumatic brain injury in *Drosophila*. *Front Physiol.* 2021;12:628777.
- Shpigler HY, Saul MC, Murdoch EE, Cash-Ahmed AC, Seward CH, Sloofman L, Chandrasekaran S, Sinha S, Stubbs LJ, Robinson GE, et al. Behavioral, transcriptomic and epigenetic responses to social challenge in honey bees. *Genes Brain Behav.* 2017;16(6): 579–591.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2014. <http://www.R-project.org/>.
- Thackray AM, Lam B, Shahira Binti Ab Razak A, Yeo G, Bujdoso R. Transcriptional signature of prion-induced neurotoxicity in a *Drosophila* model of transmissible mammalian prion disease. *Biochem J.* 2020;477(4):833–852.
- Tirosh I, Reikhav S, Avraham AL, Barkai N. A yeast hybrid provides insight into the evolution of gene expression regulation. *Science.* 2009;324(5927):659–662.
- Traniello IM, Bukhari SA, Kevill J, Ahmed AC, Hamilton AR, Naeger NL, Schroeder DC, Robinson GE. Meta-analysis of honey bee neurogenomic response links Deformed wing virus type A to precocious behavioral maturation. *Sci Rep.* 2020;10(1):3101.
- Wang M, Peng IF, Li S, Hu X. Dysregulation of antimicrobial peptide expression distinguishes Alzheimer's disease from normal aging. *Aging (Albany NY).* 2020;12(1):690–706.
- Wang X, Werren JH, Clark AG. Allele-specific transcriptome and methylome analysis reveals stable inheritance and Cis-regulation of DNA methylation in *Nasonia*. *PLoS Biol.* 2016;14(7): e1002500.
- Wang Z, Zhang J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc Natl Acad Sci U S A.* 2011;108(16):E67–E76.
- Weigelt CM, Sehgal R, Tain LS, Cheng J, Eßer J, Pahl A, Dieterich C, Grönke S, Partridge L. An insulin-sensitive circular RNA that regulates lifespan in *Drosophila*. *Mol Cell.* 2020;79(2):268–279.e5. e265.
- Wu X, Galbraith DA, Chatterjee P, Jeong H, Grozinger CM, Yi SV. Lineage and parent-of-origin effects in DNA methylation of honey bees (*Apis mellifera*) revealed by reciprocal crosses and whole-genome bisulfite sequencing. *Genome Biol Evol.* 2020; 12(8):1482–1492.
- Wu X, Lindsey ARI, Chatterjee P, Werren JH, Stouthamer R, Yi SV. Distinct epigenomic and transcriptomic modifications associated with Wolbachia-mediated asexuality. *PLoS Pathog.* 2020;16(3): e1008397.
- Yi SV. Birds do it, bees do it, worms and ciliates do it too: DNA methylation from unexpected corners of the tree of life. *Genome Biol.* 2012;13(10):174.
- Zeng J, Yi S. DNA methylation and genome evolution in honeybee: gene length, expression, functional enrichment co-vary with the evolutionary signature of DNA methylation. *Genome Biol Evol.* 2010;2:770–780.

Communicating editor: J. Comeron