ISSN 0006-3509, Biophysics, 2018, Vol. 63, No. 1, pp. 43–51. © Pleiades Publishing, Inc., 2018.
Published in Biofizika, 2018, Vol. 63, No. 1, pp. 54–77.

Inference of Transcription Factor Regulation Patterns Using Gene Expression Covariation in Natural Populations of *Drosophila melanogaster*[1, 2]

N. M. Osman[a, c], T. H. Kitapci[a], S. Vlaho[a], Z. Wunderlich[b], and S. V. Nuzhdin[a, d, *]

[a] *University of Southern California, Los Angeles, California, USA*
[b] *University of California, Irvine, California, 92697 USA*
[c] *National Research Centre, Dokki, Giza, 12622 Egypt*
[d] *Saint Petersburg Polytechnical University, St. Petersburg, 195251 Russia*
**e-mail: snuzhdin@usc.edu*

Received December 2, 2016

Abstract—Gene regulatory networks control the complex programs that drive development. Deciphering the connections between transcription factors (TFs) and target genes is challenging, in part because TFs bind to thousands of places in the genome but control expression through a subset of these binding events. We hypothesize that we can combine natural variation of expression levels and predictions of TF binding sites to identify TF targets. We gather RNA-seq data from 71 genetically distinct F1 *Drosophila melanogaster* embryos and calculate the correlations between TF and potential target genes' expression levels, which we call "regulatory strength." To separate direct and indirect TF targets, we hypothesize that direct TF targets will have a preponderance of binding sites in their upstream regions. Using 14 TFs active during embryogenesis, we find that 12 TFs showed a significant correlation between their binding strength and regulatory strength on downstream targets, and 10 TFs showed a significant correlation between the number of binding sites and the regulatory effect on target genes. The general roles, e.g. *bicoid*'*s* role as an activator, and the particular interactions we observed between our TFs, e.g. *twist*'*s* role as a repressor of *sloppy paired* and *odd paired,* generally coincide with the literature.

Keywords: Drosophila melanogaster, gene regulatory networks, transcription factors

DOI: 10.1134/S0006350918010128

1. INTRODUCTION

Intricate gene regulatory networks are responsible for the patterning events during development that generate a complex adult animal from a single fertilized egg. These networks are composed of transcription factors (TFs), chromatin remodelers, co-activators, and signaling pathway components and are largely encoded in the genome in pieces of regulatory DNA, e.g. enhancers. Enhancers are located in the non-coding portions of the genome and are composed of TF binding sites (Davidson 2010; Peter and Davidson, 2011). The number, strength, and arrangement of TF binding sites within a enhancer help to determine the expression pattern driven by the enhancer (Li et al., 2008; Vaquerizas et al., 2009; Nuzhdin et al., 2010; Yang et al., 2011; Levo and Segal, 2014).

The networks that specify the anterior-posterior and dorsal-ventral axes during early *Drosophila* development are established model systems for studying gene regulatory networks. The anterior-posterior patterning network is a transcriptional cascade, in which genes in upstream layers of the cascade regulate the expression of genes downstream. Maternally deposited genes control the expression patterns of gap genes. These gap genes, which constitute the most upstream zygotic layer of the cascade, control pair-rule gene patterns, which then specify the expression patterns of segment polarity genes (Nasiadka et al., 2002; Bonn and Furlong, 2008). The dorsal-ventral patterning system begins with the ventral activation of the Toll signalling pathway, which activates Dorsal, an NF-kappaB family TF that activates several other TFs important for dorsal-ventral axis specification (Levine and Davidson, 2005; Bonn and Furlong, 2008).

One challenge of studying gene regulatory networks is finding all the connections between the TFs active in the network. Traditionally, this is accomplished through a combination of mutant and overex-

[1] The article is published in the original.

[2] Supplementary materials are available for this article at 10.1134/S0006350918010128 and are accessible for authorized users.

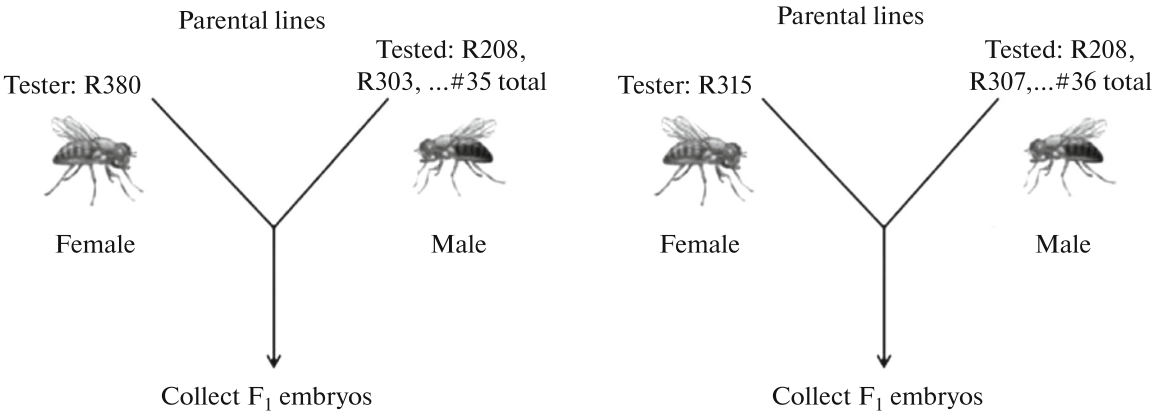

Fig. 1. Crossing scheme for F_1 samples. We generated F_1 embryos using two tester female strains, R380 and R315 from the DGRP. In total, 71 distinct crosses were performed using these two tester strains and various male strains.

pression experiments for TFs, enhancer bashing, biochemical footprinting of TF binding sites, and more recently, functional genomics approaches like ChIP-seq and RNA-seq have been combined with statistical inference to elucidate gene regulatory networks (Howard and Ingham, 1986; Bergman et al., 2005; Bonneau, 2008; Busser et al., 2008; Bumgarner and Yeung, 2009; Park, 2009; Pepke et al., 2009; Wang and Huang, 2014).

Here, we combine genome-wide transcriptional profiling of individual *D. melanogaster* lines with TF binding site analysis to find the connections in the early embryonic patterning networks. We hypothesize that the levels of a TF and its target genes will co-vary and use bioinformatic analysis of predicted TF binding sites to separate direct and indirect effects, in a similar fashion to previous studies (Kliebenstein, 2009; Lewis et al., 2010; Mostafavi et al., 2014). To measure covariation between TFs and their target genes, we used naturally occurring variation in TF and target gene expression levels, measured in four to five hour old embryos (Nuzhdin et al., 2008). We expected to observe significant expression variation between genotypes because earlier studies have found significant transcript level differences among *D. melanogaster* genotypes for roughly 10% of the whole body transcriptome (Jin et al., 2001; Nuzhdin et al., 2004). We mated males with different genotypes to two tester female lines, resulting in F1 embryos with distinct genetic backgrounds that drive variation in TF and target gene expression levels (Fig. 1).

Since we used linear regressions for the analysis, we confined the analysis to mostly additive genetic variation, which is best exposed in Fl heterozygous individuals (Nuzhdin et al., 2012). Our aim was to determine how a TF's control of target genes correlated with the strength and number of binding sites for the TF in the target gene's presumptive regulatory DNA and to see if our results would coincide with current literature, indicating a strong correlation between a TF's binding strength or number of binding sites and its regulatory strength on a target gene (Li et al., 2008; Nuzhdin et al., 2010).

2. MATERIALS AND METHODS

D. melanogaster Lines and Embryo Collections

Experiments were conducted using Drosophila Genetic Reference Panel (DGRP) lines (Mackay et al., 2012) and Winters lines, which were collected from an organic orchard in Winters, California and inbred to achieve homozygosity (Campo et al., 2013). All stocks were maintained on normal cornmeal-based food and kept at an approximate temperature of 25°C in a 12:12 light to dark cycle. Females from two DGRP lines, Raleigh 315 and Raleigh 380, were crossed to males from various Raleigh and Winters lines (Fig. 1). Crosses were performed in 6 oz square bottom bottles that were capped with a Petri dish containing grape juice agar. The next morning at 8 h bottles were recapped with Petri dishes containing yeast to collect embryos laid overnight. At 9 h, these Petri dishes were then discarded and the bottles were recapped with new Petri dishes for exactly one hour. At 10 h, the new Petri dishes containing the embryos were collected and incubated at 25°C for 4 h, yielding 4–5 h old embryos. After incubation, the embryos were dechorionated by submerging them in a 50% bleach solution for one and a half minute. The embryos were then washed with deionized water and stored in Ambion TRIzol reagent (Life Technologies no. 15596-026) at –80°C. 71 crosses were analyzed in this study. 25 crosses were sequenced as two biological replicates and the remaining 46 crosses were sequenced without replicates (Table S1). In total, 96 samples were sequenced.

Construction of transcriptome Libraries and RNA-Sequencing

For each sample, RNA was extracted using the Direct-Zol RNA-prep Kit following the protocol from Zymo Research. mRNA was purified with the Ambion Dynabeads mRNA Purification Kit (Product no. 61006) and fragmented using the Fragmentation Kit (Product no. AM8740), followed by cDNA synthesis with random hexamer primers. Blunt ends were generated with the help of the Quick Blunting Kit (NEB Product no. E1201L) and a single A base was added with the Klenow Fragment 3'–5' exo-nuclease (NEB Product no. M0212L). Illumina adaptors were ligated onto the cDNA fragments with the Quick Ligation Kit (NEB Product no. M2200L). Size selection of fragments were done using Agencourt AMPure XP beads (Beckman Coulter Product no. A63880) with a ratio of 0.7 beads to total volume. Finally 96 samples were tagged by 12 Illumina indexes and 8 custom built barcodes and enriched before being sequenced in a 96-well platform (Dunham and Friesen, 2013) on an Illumina HiSeq 2500 in paired end 100 base-pair mode. Raw Illumina reads will be deposited on NCBI (SRA ID will be inserted here) after acceptance and reads are currently available here: http://rri-nuzhdin-2.cts.usc.edu/thkitapci/data/.

RNA Analysis Including Mapping and Normalization of Sequencing Reads

RNA-Seq reads were mapped to *D. melanogaster* reference genome sequence (dm3/BDGP5.75) using STAR (2.4.0k) with default parameters (Dobin et al., 2013). Only uniquely and concordantly mapped reads were used for the further analysis. Raw read counts were generated using HTSeq with default parameters (Anders et al., 2014) and using the annotation file (dm3/BDGP5.75.gtf). Read counts from samples corresponding to the same genotype were merged together. Raw read counts were normalized using RPKM (reads per kilobase per million reads) (Mortazavi et al., 2008). Analysis scripts are available at (https://github.com/thkitapci/Inference_of_TF_regulatory_networks.git).

Gene Expression Covariation Analysis

We were interested in segmentation genes (maternally-expressed TFs (*bicoid* and *caudal*), gap genes (*giant*, *Kruppel*, *knirps*, *hunchback*, and *tailless*), primary pair-rule genes (*even skipped, hairy, runt* and *fushi tarazu*), and segment polarity gene (*engrailed*)), in addition to genes in dorsal-ventral patterning (*snail* and *twist*), all of which play an important role in patterning during *D. melanogaster* embryogenesis (Sandmann et al., 2007; Campos-Ortega and Hartenstein, 2013). Furthermore, the segmentation genes have shown an abundance of expression variation during *D. melanogaster* embryogenesis (Nuzhdin et al., 2010).

To characterize the regulatory strength between TFs of interest and their target genes we calculated the Spearman's correlation coefficient between the expression levels of each TF and other mapped genes (7805 genes) that had a high expression levels across 59 samples. Positive regulatory strengths are those with correlation coefficients > 0, negative strengths are those with correlation coefficients < 0.

To find the presumptive regulatory DNA for each target gene, we found the DNase accessible regions within a 5000 bp window upstream of each target gene, using earlier measurements of DNase accessible regions of stage 10 (4 h) and stage 11 (5 h and 40 min) embryos of *D. melanogaster* (Thomas et al., 2011). To find binding sites in each accessible region, we used the PATSER program (Hertz and Stormo, 1999), assuming 47% GC content, and the position weight matrices (PWMs) for each TF of interest from Fly Factor Survey Database (Zhu et al., 2011) (Table S2). We used the PATSER option -li to calculate a binding score cutoff based on each PWM's information content. We further refined the sites by selecting the strongest 5, 10 or 15% of binding sites, based on the PATSER binding scores, which correlate with binding strength. We used the 10% threshold in the main text, because the 5% threshold seemed too stringent and significantly reduced our samples size and the 15% threshold seemed too permissive and abolished correlations between regulatory strength and binding site content (Fig. S5).

Using these sites, we calculated the average strength or number of binding sites in each target gene's accessible upstream DNA, discarding any target genes with fewer than three binding sites. Restricting the analysis to target genes with more than two binding sites increased the correlation between binding site content and regulatory strength. PATSER's calculated scores, which roughly correspond to the log-likelihood of a k-mer being a TF binding site, were used as a proxy for binding site strength in this analysis.

To test whether there was a significant correlation between a TF's binding site content and its regulatory strength, we calculated the Spearman's correlation coefficient between positive and negative regulatory strengths and average strengths or numbers of binding sites, using R. P-values were corrected using the Benjamini and Hochberg method (Benjamini and Hochberg, 1995). All figures were generated using the R statistical program and all scripts were written in PERL language. Raw and normalized count files and scripts were used for analysis can be retrieved from (https://github.com/thkitapci/Inference_of_TF_regulatory_networks.git).

Regulatory Interaction Analysis between TFs

For this analysis, we used the potential target genes for each TF of interest to find the subset of potential targets that were TFs themselves. Spearman's correlation coefficients between these TFs were calculated and tested for significance ($p < 0.001$, after multiple testing correction by Benjamini and Hochberg method) using *R*.

Replication Analysis

To confirm that our results on the regulatory interactions between TFs are reproducible and have biological significance, we divided our dataset into two datasets based on the tester genotype used, namely crosses with R315 and crosses with R380. The Spearman's correlation coefficients of these TFs were calculated independently in these datasets and tested for significance ($p < 0.05$, after multiple testing correction by Benjamini and Hochberg method) using *R*.

3. RESULTS

Measuring the Transcriptome of D. melanogaster Lines

To measure the covariation between TFs and presumptive target genes, we characterized the transcriptomes embryos of individual *D. melanogaster* F1 lines (Fig. 1) four to five hours after fertilization. The transcriptomes of 96 *D. melanogaster* samples were sequenced and reads were mapped to 15682 genes that were annotated in the *D. melanogaster* reference genome. To characterize the sequencing quality of the data, the number of mapped genes and the average number of reads in each sample were analyzed. We define a gene as being mapped in a sample when it has a non-zero read count. There was variation in the average read count in each sample (Fig. 2a). A higher average read count allowed more genes to be identified, and as expected, there was a positive relationship between average read count and mapped gene count in a sample. However, this relationship is not linear, since increasing the sequencing depth will identify more genes only to a certain point. We see that as the mean read count is increased, the number of genes sequenced approaches 11000, suggesting an upper limit to the number of genes detectable in each sample (Fig. 2a).

To account for differences in sequencing depth, due to either poor sample preparation, lower concentrations of RNA, or sequencing bias that results from mRNA fragmentation (Wang et al., 2009), we normalized the data by the lengths of each mapped gene and millions of reads to calculate the Reads Per Kilobase of transcript per Million mapped reads (RPKM), a quantitative measure of gene expression (Mortazavi et al., 2008). The average RPKM value across all mapped genes is consistent across all the samples that were included in downstream analysis, with mean = 37.9 RPKM and standard deviation = 2.8 RPKM. Twelve samples with fewer than 7500 mapped genes displayed a wide range in average and standard deviation RPKM values and were removed from further analysis (Fig. 2b). Among the remaining 59 samples, 7805 genes were detected in at least 50 samples (Fig. 2c). We considered these 7805 genes in the 59 samples in the downstream analysis. To measure "regulatory strength," we calculated the Spearman's correlation coefficient between the expression levels of 14 TFs active in the anterior-posterior and dorsal-ventral patterning networks with their potential target genes.

Gene Expression Covariation Is Related to TF Binding Site Content

To test the hypothesis that gene expression covariation (regulatory strength) between a TF and a potential target gene is related to TF binding site content in the target gene's regulatory DNA, we generated two data sets that describe either the average strength or number of TF binding sites in each target gene's presumptive regulatory DNA region. To start, we assumed that a gene's regulatory DNA was in the 5 kb upstream region of its transcription start site and further narrowed down this region to include only the regions identified as DNAse hypersensitive during stage 10 and stage 11 of development, which roughly corresponds to the time of our embryo collection window. Because we do not have comprehensive annotations of the enhancers in the genome, we assume that these proximal, accessible regions of the genome will contain the enhancers of the potential target genes and consider the implications of this assumption in the Discussion. Binding sites for each TF of interest were detected in these regions using each TF's position weight matrices (PWMs) and the PATSER program (see Methods and Materials). To determine which binding sites to consider, for each TF, PATSER was used to calculate a binding score cutoff based on PWM's information content, and then we considered sites that fell into the top 5 and 10% of the binding site score distribution as potential binding sites for each TF. We then calculated the average strength or number of binding sites located in each potential target gene's presumptive regulatory DNA for each TF. In the main text, Figs. S1 and S3, we show the results for the 10% cutoff, and results for the 5% cutoff are in Figs. S2 and S4. Using a 15% cutoff abolished correlations between regulatory strength and binding site content, and therefore was considered to be too permissive as a threshold (Fig. S5).

To compare each TF's regulatory strengths to each potential target's TF binding site content, we separated the regulatory strength data into positive and negative correlations. We found a strong relationship between the positive regulatory strengths and average binding site strengths of *bicoid* and *engrailed*, while *giant* and *hunchback* have a strong relationship between negative regulatory strengths and average binding site strength in this stage of *D. melanogaster* development (Fig. 3). We also found a significant correlation between the regulatory strength and the average strength of TF binding sites for *even skipped, fushi tarazu, hairy, Kruppel, runt, snail, tailless* and *twist* (Figs. S1, S2, Tables S3, S4). *Bicoid, engrailed, fushi tarazu, runt* and *twist* show significant correlations between positive regulatory strength and the average strength of TF binding sites, suggesting that these TFs act as activators in this stage of development. *Even-skipped, giant, hairy, hunchback, Kruppel, snail* and *tailless* had a significant correlation between negative regulatory strength and the average strength of TF binding sites, suggesting that they act as repressors in this stage of the development (Figs. S1, S2, Tables S3, S4).We also found a significant correlation between the number of binding sites and the activating strength of *bicoid, engrailed, fushi tarazu, Kruppel, runt* and *snail. Engrailed, even skipped, giant, hunchback, Kruppel* and *tailless* had a significant correlation between the number of binding sites and their repressing strength (Figs. S3, S4, Tables S3, S4).

These observations are somewhat consistent with the known biology of these TFs. *Bicoid* and *twist* (Struhl et al., 1989; Leptin, 1991; Cripps et al., 1998; Stathopoulos et al., 2002; Schroeder et al., 2004; Zeitlinger et al., 2007; Ochoa-Espinosa et al., 2009; Porcher and Dostatni, 2010) are known to act as transcriptional activators. *Giant, tailless* (Wu et al., 1998; Hewitt et al., 1999; Schroeder et al., 2004; Morán and Jiménez, 2006; Yáñez et al., 2013), *even-skipped*, and *hairy* (Manoukian and Krause, 1992; Barolo and Levine, 1997; Jiménez et al., 1997; Kobayashi et al., 2001; Fujioka et al., 2002; Bianchi-Frias et al., 2004) are known as transcriptional repressors. *Runt* and *fushi tarazu* can act as activators or repressors, but we only detect their roles as activators in our study (Hiromi and Gehring 1987; Kramer et al., 1999; Yu et al., 1999; Nasiadka et al., 2000; Wheeler et al., 2000). Hunchback can act as transcriptional activator or a repressor, though we only detect its repressive role in our study (Zuo et al., 1991; Schroeder et al., 2004; Staller et al., 2015). There is some evidence that *Kruppel* (Sauer and Jäckle, 1991; Zuo et al., 1991; Sauer et al., 1995; La Rosée-Borggreve et al., 1999), *engrailed* (Heemskerk et al., 1991; Tabata et al., 1992; Alexandre and Vincent, 2003) and *snail* (Rembold et al., 2014) can act as bifunctional TFs.

Uncovering Regulatory Interaction between TFs

Because the anterior-posterior and dorsal-ventral patterning networks involve interactions between TFs; we analyzed our data to see if we could uncover these interactions. We found that the maternal gene *bicoid* had a positive correlation with most TFs, which is consistent with *bicoid*'s role as an activator. Meanwhile, tailless showed a repressive interaction with *giant* and *knirps,* again consistent with the known biology (Sánchez and Thieffry, 2001; Jaeger et al., 2004; Jaeger, 2011; Liu et al., 2013; Gula and Samsonov,

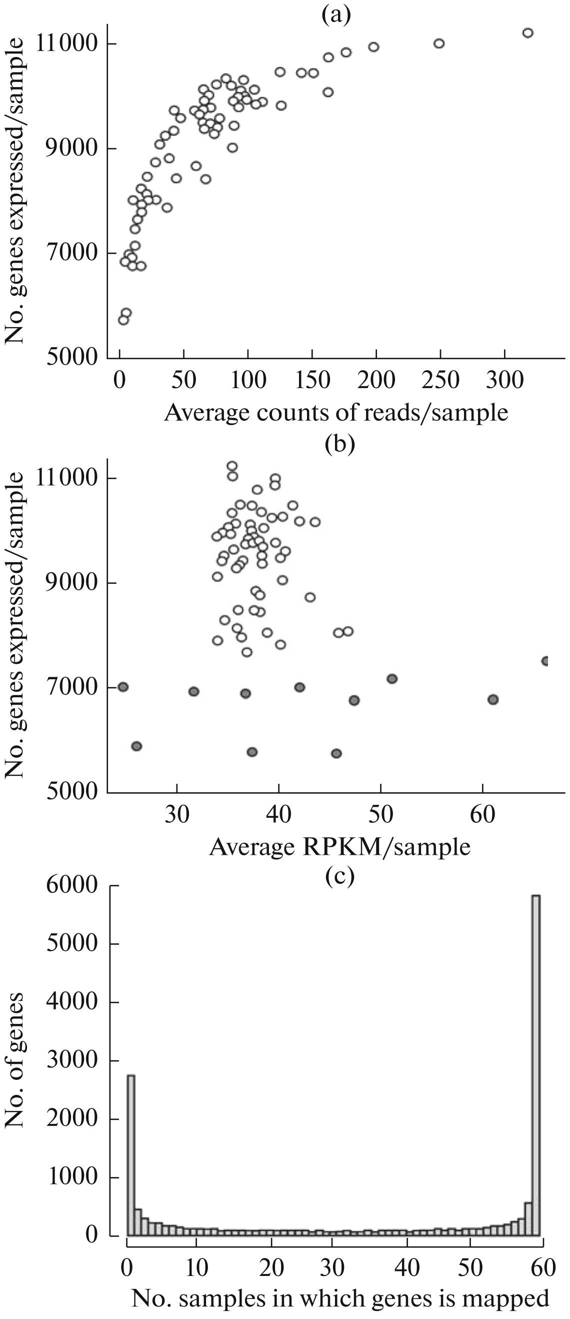

Fig. 2. Analysis of *D. melanogaster* line transcriptome data. (a) Here we plot average read count as a function versus the number of genes expressed before RPKM normalization. As read count increases, so does the number of expressed (mapped) genes reaching until a plateau of ~11000 expressed genes. (b) Here we plot average RPKM versus the number of genes expressed. For samples with more than 7500 mapped (expressed) genes, the average RPKM value is consistent. For samples with less than 7500 expressed genes (red dots), the average RPKM values vary greatly. These samples were discarded from downstream analysis. (c) Here we plot the number of genes mapped across the samples. Only genes that are expressed in at least 50 samples were used in downstream analysis.

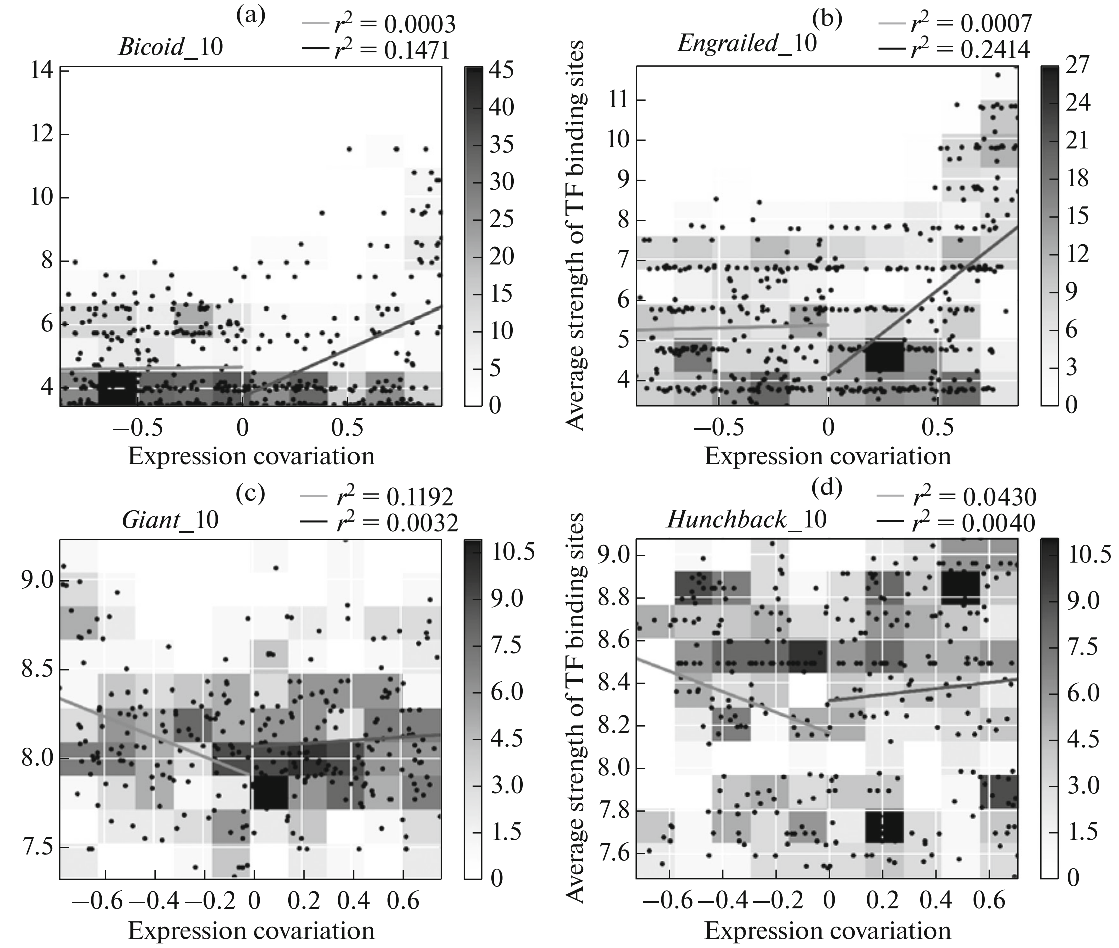

Fig. 3. The relationship between regulatory strength and average TF binding site strength. Here we plot the relationship between regulatory strength and average TF binding strength for *bicoid, engrailed, giant* and *hunchback* with their target genes using the top 10% of binding scores for each TF. The *x*-axis shows the correlations between the expression level of each TF and its target genes across all our samples. The *y*-axis shows the average TF binding site strength for binding sites located in the assumed region of regulatory DNA for each target gene. Each black dot represents a target for the TF in the panel with at least three binding motifs. The blue and red lines show the linear regression for the positively and negatively correlated target genes, respectively, and r^2 values are displayed for these best fit lines. (a, b) Activating strengths (positive expression covariation) of *bicoid* and *engrailed* are correlated with the average binding site strength, suggesting *bicoid* and *engrailed* act as activators at this stage of the development. (c, d) Repressing strengths (negative expression covariation) of *giant* and *hunchback* are correlated with average binding site strength.

2015). Twist seems to repress sloppy paired and *odd paired*, as suggested in a previous study (Sandmann et al., 2007) (Table 1, Fig. S6).

To make sure our findings do not depend on genetic background, we have used two tester female strains (R315 and R380), similar to (Nuzhdin et al., 2010). When we analyzed the data for each of these tester strains separately, the estimates of regulatory strength appear nearly identical between two testers (Fig. S7, Table S5). This high degree of replication establishes the robustness of the technique.

4. DISCUSSION

In this study, we used transcriptional profiling of *D. melanogaster* lines to analyze the relationship between the TF binding sites and the regulatory control of their target genes using two metrics of binding site content: average strength and number of binding sites. We found significant correlations for several TFs of interest. This suggests that the strength and number of binding sites for a particular TF in regulatory DNA regions are correlated with the regulatory control of its target genes. Our results are consistent with previous studies, which suggested that the numbers and strengths of TF binding sites are correlated, albeit imperfectly, with a particular TF's regulatory strength on its target genes (Li et al., 2008; MacArthur et al., 2009; Franco-Zorrilla et al., 2014). A previous study of five TFs (*bicoid, caudal, giant, hunchback*, and *Kruppel*) during the first 5–8 h of *D. melanogaster* development showed that there was an association between the strength of TF binding and the regulatory control of their target genes (Nuzhdin et al., 2010).

Engrailed exhibited the highest correlation between the regulatory control of its target genes and its strength and number of binding sites (Tables S3, S4). We hypothesize that this strong correlation is due to *engrailed* having a generally high level of expression at this time point in development, especially compared with the previous developmental stages, while other TFs we examined do not show the same distinct difference in expression (Graveley et al., 2011; Hammonds et al., 2013).

Overall, our results show that the number and strength of the significant associations of covariation in expression with binding strength and number of binding sites is modest. There are many reasons for this observation. Each of these TFs is expressed in a complex spatial and temporal pattern, so whole-embryo expression data will obscure regulatory events in a small number of cells. Predicted TF binding sites may not correspond to in vivo binding (Biggin and Tjian, 2001; Levine and Tjian, 2003). Our assumption that the DNAse accessible regions 5 kb upstream of a target gene will regulate its expression will cause us to miss enhancers located in other parts of the genome and may include regions that do not act as enhancers. Therefore, our proxy for TF binding is only approximate. TFs can interact with co-factors that modify the TF's ability to activate or repress transcription of their target genes (Björklund et al., 1999; Tanay, 2006). Furthermore, some instances where we found high covariance between TFs and genes with low affinity binding sites could be due to the presence of intermediate proteins that bind to a given site and increase the binding potential of the TF (Björklund et al., 1999; Mannervik et al., 1999). Despite all these caveats, we are able to use predicted TF binding sites and expression covariation to confirm some known regulatory interactions in this gene regulatory network (Li et al., 2008; MacArthur et al., 2009; Nuzhdin et al., 2010).

We can envision how the expression of TFs and regulated genes could co-vary due to indirect regulation. However, the observed relationship between strength of regulation and binding site strength indicates a direct component in this covariance. Further insight comes from comparing activation and repression. Consider the activator *bicoid*, for instance. If the covariances are due to indirect effects, their signs might be both positive and negative. One would then hypothesize that negative regulation must be indirect, while positive regulation might be either direct or indirect. Then, the strength of binding shall not be associated with the magnitude of negative covariances, while the positive regulation may, exactly as observed on the Fig. 3. The strength of such a covariance likely reflects the proportion of direct influences in our dataset.

Generally speaking, the number of the significant associations of covariation in expression with binding strength and number of binding sites is small. Taking into consideration that transcriptional machinery is complex and the potential influence of intermediate proteins, we are not alarmed that the magnitude of significant associations of covariation is low. The correlations we observed with multiple TFs between the binding strength with a particular target gene and the regulatory strength supports the idea that the strength of TF binding is a mechanistically sound predictor of the strength of the regulatory effect. Moreover, our results on the regulatory interactions between TFs validate findings in previous studies, such as *bicoid*'s role as an activator.

Table 1. Regulatory interaction between TFs. We consider corrected p-values < 0.001 to be significant

TF	Target	Spearman's Correlation Coefficient	Corrected p-value
Bicoid	Caudal	0.62	1.56E-05
Bicoid	Knirps	0.66	3.00E-06
Bicoid	Kruppel	0.69	5.33E-07
Bicoid	Giant	0.67	1.72E-06
Bicoid	Hunchback	0.61	1.56E-05
Bicoid	Tailless	0.64	7.46E-06
Bicoid	Runt	0.68	1.43E-06
Bicoid	Hairy	0.60	4.10E-05
Bicoid	Tailless	−0.53	6.54E-05
Bicoid	Twist	−0.59	4.25E-05
Bicoid	Sloppy paired	−0.62	7.46E-06
Bicoid	Odd paired	−0.68	5.33E-07

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Institutes of Health (NIH) Grants R00HD073191 (to ZW) and U01GM103804 and R01GM102227 (to SVN). SVN work was also supported by RSF grants 16-16-00007 (gene expression covariance analysis and analysis of regulatory interactions) and 14-14-00302 (construction of transcriptome libraries and RNA sequencing). We would like to thank Peter Chang and Yerbol Kurmangaliyev for their help and useful discussions.

REFERENCES

1. E. H. Davidson, Nature **468**, 911 (2010).
2. I. S. Peter and E. H. Davidson, Cell **144**, 970 (2011).
3. X. Li, S. MacArthur, R. Bourgon, et al., PLoS Biol. **6**, e27 (2008).
4. J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe, Nat. Rev. Genet. **10**, 252 (2009).

5. S. V. Nuzhdin, A. Rychkova, and M. W. Hahn, Trends Genet. **26**, 51 (2010).

6. S. Yang, H. K. Yalamanchili, X. Li, et al., Bioinformatics **27** (21), 2972 (2011).

7. M. Levo and E. Segal, Nat. Rev. Genet. **15**, 453 (2014).

8. A. Nasiadka, B. H. Dietrich, and H. M. Krause, Adv. Dev. Biol. Biochem. **12**, 155 (2002).

9. S. Bonn and E. E. Furlong, Curr. Opin. Genet. Dev. **18**, 513 (2008).

10. M. Levine and E. H. Davidson, Proc. Natl. Acad. Sci. U. S. A. **102**, 4936 (2005).

11. K. Howard and P. Ingham, Cell **44**, 949 (1986).

12. C. M. Bergman, J. W. Carlson, and S. E. Celniker, Bioinformatics **21**, 1747 (2005).

13. R. Bonneau, Nat. Chem. Biol. **4**, 658 (2008).

14. B. W. Busser, M. L. Bulyk, and A. M. Michelson, Curr. Opin. Genet. Dev. **18**, 521 (2008).

15. R. E. Bumgarner and K. Y. Yeung, Comput. Syst. Biol. **541**, 225 (2009).

16. P. J. Park, Nat. Rev. Genet. **10**, 669 (2009).

17. S. Pepke, B. Wold, and A. Mortazavi, Nat. Methods **6**, S22 (2009).

18. Y. R. Wang and H. Huang, J. Theor. Biol. **362**, 53 (2014).

19. D. J. Kliebenstein, Plant Syst. Biol. **553**, 227 (2009).

20. J. A. Lewis, I. M. Elkon, M. A. McGee, et al., Genetics **186**, 1197 (2010).

21. S. Mostafavi, A. Ortiz-Lopez, M. A. Bogue, et al., J. Immunol. **193**, 4485 (2014).

22. S. V. Nuzhdin, D. M. Tufts, and M. W. Hahn, Evol. Dev. **10**, 683 (2008).

23. W. Jin, R. M. Riley, R. D. Wolfinger, et al., Nat. Genet. **29**, 389 (2001).

24. S. V. Nuzhdin, M. L. Wayne, K. L. Harmon, and L. M. McIntyre, Mol. Biol. Evol. **21**, 1308 (2004).

25. S. V. Nuzhdin, M. L. Friesen, and L. M. McIntyre, Trends Genet. **28**, 421 (2012).

26. T. F. Mackay, S. Richards, E. A. Stone, et al., Nature **482**, 173 (2012).

27. D. Campo, K. Lehmann, C. Fjeldsted, et al., Mol. Ecol. **22**, 5084 (2013).

28. J. P. Dunham and M. L. Friesen, Cold Spring Harb. Protoc. **9**, 820 (2013).

29. A. Dobin, C. A. Davis, F. Schlesinger, et al., Bioinformatics **29**, 15 (2013).

30. S. Anders, P. T. Pyl, and W. Huber, Bioinformatics **31** (2), 166 (2015).

31. A. Mortazavi, B. A. Williams, K. McCue, et al., Nat. Methods **5**, 621 (2008).

32. T. Sandmann, C. Girardot, M. Brehme, et al., Genes Dev. **21**, 436 (2007).

33. J. A. Campos-Ortega and V. Hartenstein, *The Embryonic Development of Drosophila melanogaster* (Springer Sci. Business Media, 2013).

34. S. Thomas, X.-Y. Li, P. J. Sabo, et al., Genome Biol. **12**, R43 (2011).

35. G. Z. Hertz and G. D. Stormo, Bioinformatics **15**, 563 (1999).

36. L. J. Zhu, R. G. Christensen, M. Kazemian, et al., Nucleic Acids Res. **39**, D111 (2011).

37. Y. Benjamini and Y. Hochberg, J. R. Stat. Soc. Series B **57** (1), 289 (1995).

38. Z. Wang, M. Gerstein, and M. Snyder, Nat. Rev. Genet. **10**, 57 (2009).

39. G. Struhl, K. Struhl, and P. M. Macdonald, Cell **57**, 1259 (1989).

40. M. Leptin, Genes Dev. **5**, 1568 (1991).

41. R. M. Cripps, B. L. Black, B. Zhao, et al., Genes Dev. **12**, 422 (1998).

42. A. Stathopoulos, M. Van Drenth, A. Erives, et al., Cell **111**, 687 (2002).

43. M. D. Schroeder, M. Pierce, J. Fak, et al., PLoS Biol. **2**, e271 (2004).

44. J. Zeitlinger, R. P. Zinzen, A. Stark, et al., Genes Dev. **21**, 385 (2007).

45. A. Ochoa-Espinosa, D. Yu, A. Tsirigos, et al., Proc. Natl. Acad. Sci. U. S. A. **106**, 3823 (2009).

46. A. Porcher and N. Dostatni, Curr. Biol. **20**, R249 (2010).

47. X. Wu, R. Vakani, and S. Small, Development **125**, 3765 (1998).

48. G. F. Hewitt, et al., Development **126**, 1201 (1999).

49. E. Morán and G. Jiménez, Mol. Cell. Biol. **26**, 3446 (2006).

50. J. O. Yáñez-Cuna, E. Z. Kvon, and A. Stark, Trends Genet. **29**, 11 (2013).

51. A. S. Manoukian and H. M. Krause, Genes Dev. **6**, 1740 (1992).

52. S. Barolo and M. Levine, EMBO J. **16**, 2883 (1997).

53. G. Jiménez, Z. E. Paroush, and D. Ish-Horowicz, Genes Dev. **11**, 3072 (1997).

54. M. Kobayashi, R. E. Goldstein, M. Fujioka et al., Development **128**, 1805 (2001).

55. M. Fujioka, G. L. Yusibova, N. H. Patel, et al., Development **129**, 4411 (2002).

56. D. Bianchi-Frias, A. Orian, J. J. Delrow, et al., PLoS Biol. **2**, e178. (2004).

57. Y. Hiromi and W. J. Gehring, Cell **50**, 963 (1987).

58. S. G. Kramer, T. M. Jinks, P. Schedl, and J. P. Gergen, Development **126**, 191 (1999).

59. Y. Yu, M. Yussa, J. Song, et al., Mech. Dev. **83**, 95 (1999).

60. A. Nasiadka, A. Grill, and H. M. Krause, Development **127**, 2965 (2000).

61. J. C. Wheeler, K. Shigesada, J. P. Gergen, and Y. Ito, Semin. Cell Dev. Biol. **11**, 369 (2000).

62. P. I. Zuo, D. Stanojević, J. Colgan, et al., Genes Dev. **5**, 254 (1991).

63. M. V. Staller, B. J. Vincent, M. D. J. Bragdon, et al., Proc. Natl. Acad. Sci. U. S. A. **112**, 785 (2015).

64. F. Sauer and H. Jäckle, Nature **353**, 563 (1991).

65. F. Sauer, J. D. Fondell, Y. Ohkuma, et al., Nature **375**, 162 (1995).

66. A. La Rosée-Borggreve, T. Häder, D. Wainwright, et al., Mech. Dev. **89**, 133 (1999).

67. J. Heemskerk, S. DiNardo, R. Kostriken, and P. H. O'Farrell, Nature **352**, 404 (1991).

68. T. Tabata, S. Eaton, and T. B. Kornberg, Genes Dev. **6**, 2635 (1992).

69. C. Alexandre and J. P. Vincent, Development **130**, 729 (2003).

70. M. Rembold, L. Ciglar, J. O. Yáñez-Cuna, et al., Genes Dev. **28**, 167 (2014).

71. L. Sánchez and D. Thieffry, J. Theor. Biol. **211**, 115 (2001).

72. J. Jaeger, M. Blagov, D. Kosman, et al., Genetics **167** (4), 1721 (2004).

73. J. Jaeger, Cell. Mol. Life Sci. **68**, 243 (2011).

74. F. Liu, A. H. Morrison, and T. Gregor, Proc. Natl. Acad. Sci. U. S. A. **110**, 6724 (2013).

75. I. A. Gula and A. M. Samsonov, Bioinformatics **31**, 714 (2015).

76. S. MacArthur, X.-Y. Li, J. Li, et al., Genome Biol. **10**, R80 (2009).

77. J. M. Franco-Zorrilla, I. López-Vidriero, J. L. Carrasco, et al., Proc. Natl. Acad. Sci. U. S. A. **111**, 2367 (2014).

78. B. R. Graveley, A. N. Brooks, J. W. Carlson, et al., Nature **471** (7339), 473 (2011).

79. A. S. Hammonds, C. A. Bristow, W. W. Fisher, et al., Genome Biol. **14**, R140 (2013).

80. M. D. Biggin and R. Tjian, Funct. Integr. Genomics **1**, 223 (2001).

81. M. Levine and R. Tjian, Nature **424**, 147 (2003).

82. S. Björklund, G. Almouzni, I. Davidson, et al., Cell **96**, 759 (1999).

83. A. Tanay, Genome Res. **16**, 962 (2006).

84. M. Mannervik, Y. Nibu, H. Zhang, and M. Levine, Science **284**, 606 (1999).