EVOLUTION & DEVELOPMENT 10:6, 683–689 (2008)

Abundant genetic variation in transcript level during early *Drosophila* development

Sergey V. Nuzhdin,[a] Danielle M. Tufts,[a] and Matthew W. Hahn[b,*]

[a]Department of Evolution and Ecology, University of California, Davis, CA 95616, USA
[b]Department of Biology and School of Informatics, Indiana University, Bloomington, IN 47405, USA
*Author for correspondence (email: mwh@indiana.edu)

SUMMARY Variation in gene expression may underlie many important evolutionary traits. However, it is not known at what stage in organismal development changes in gene expression are most likely to result in changes in phenotype. One widely held belief is that changes in early development are more likely to result in changes in downstream phenotypes. In order to discover how much genetic variation for transcript level is present in natural populations, we studied zygotic gene expression in nine inbred lines of *Drosophila melanogaster* at two time points in their development. We find abundant variation for transcript level both between lines and over time; close to half of all expressed genes show a significant line effect at either time point. We examine the contribution of maternally loaded genes to this variation, as well as the contribution of variation in upstream genes to variation in their downstream targets in two well-studied gene regulatory networks. Finally, we estimate the dimensionality of gene expression in these two networks and find that—despite large numbers of varying genes—there appear to be only two factors controlling this variation.

INTRODUCTION

It has become increasingly evident that differences in gene expression underlie many phenotypic differences within and between species (reviewed in Raff 1996; Carroll et al. 2001; Davidson 2001; Wray et al. 2003). Microarray studies in mice (Karp et al. 2000; Schadt et al. 2003), humans (Schadt et al. 2003), fish (Oleksiak et al. 2002), flies (Jin et al. 2001; Wayne et al. 2004), corn (Schadt et al. 2003), and yeast (Cavalieri et al. 2000; Brem et al. 2002) all indicate that genetic variation in transcript abundance is pervasive within populations. In addition, studies examining between-species variation in gene expression have also found abundant differences (e.g., Enard et al. 2002; Oleksiak et al. 2002; Rifkin et al. 2003; Nuzhdin et al. 2004). How this variation in transcript levels translates into phenotypes, however, still remains to be elucidated in a vast majority of cases (Wray et al. 2003).

Two main questions on the relationship between genotype and phenotype at the level of gene expression stand out. First, at what stage in organismal development are changes in gene expression most likely to result in changes in phenotype? Although it has been thought for many years that changes in early development might play a large role in morphological changes (e.g., Gould 1977), no studies that we know of have examined genome-wide variation among individuals in transcript level during early embryonic/zygotic stages. Second, we wish to know how similar phenotypes are maintained when such a large number of genes vary in expression. For instance, surveys on adult *Drosophila melanogaster* have found that 10–25% of all genes show variation in transcript level among individuals within the species (Jin et al. 2001; Wayne et al. 2004), and that approximately 30% of all genes differ in transcript level from the almost morphologically indistiguishable sister species, *Drosophila simulans* (Michalak and Noor 2003; Meiklejohn et al. 2003; Nuzhdin et al. 2004). One possibility for this apparent stasis of observable phenotypes is that most differences in adult gene expression are simply noise, with no functional consequences; it may be that all the evolutionarily important differences are expressed before adult life stages.

Another, perhaps more interesting, possibility is that the variation at any stage is structured such that those genes that vary act in concert rather than orthogonally to one another. If there are only a few dimensions along which all variations act, then the apparent glut of diversity may only translate at the phenotypic level into very few observable differences. The mechanisms by which early development proceeds—in which transcription factors bind to their targets in an orchestrated set of connections—may underlie such a structured output. The gene regulatory networks (GRNs) for development specify the logic maps that control the connections between transcription factors and their targets (Levine and Davidson 2005). Whereas previous studies of gene expression have considered variation at each gene individually, or in small

pathways of interacting genes (e.g., Tarone et al. 2005), our aim here is to show that GRNs can be used to reliably predict variation in a large set of interacting genes.

In this article, we examine gene expression among nine inbred lines of *D. melanogaster*, at two time points in development. We show that there is abundant variation in transcript levels both among lines and over time across development. We examine the contribution that maternal and zygotic gene expression make to these differences, and compare our results with previous work on gene expression during *D. melanogaster* development. We also show that the connections in two GRNs—for segmentation and dorsal–ventral patterning—can be used to predict the relationship in gene expression between upstream and downstream genes. Finally, we show that this variation appears to act in two dimensions in both networks: one working to activate target genes and the other to repress them.

MATERIALS AND METHODS

The flies originated from the Wolfskill Orchard in Winters, CA. They were established by mating a single pair of progeny for each of nine gravid females sampled in nature. Each line was made inbred by at least 20 generations of full-sib mating. Flies were raised on standard cornmeal medium with yeast, with an excess amount of yeast added to the top of the vials to increase the body size of the females. The flies were kept at room temperature with normal day and night light cycles. Approximately 200 young, nonvirgin females were collected from each line and allowed to lay on 25% grape/3% agar plates supplied with yeast paste overnight, or approximately 18 h. They were then transferred onto fresh plates containing the same medium and allowed to lay for 1 h in a quiet, dark place. Consecutively, they were transferred to fresh plates three times more at 1-h intervals. Plates were washed with deionized water to remove the embryos and filtered using coffee filters. After allowing 30–50 embryos per line develop for 5 h from laying (time point 1) or for 8 h (time point 2, one of the lines—127—was sampled 7 h after laying), the tubes were flash frozen using liquid nitrogen and stored at −20°C in 50 µl of RNAlater (Ambion, Austin, TX, USA). Two RNA samples per line per time point (5 and 8 h) were extracted using the manufacturer's TRIzol reagent protocol (Invitrogen, Carlsbad, CA, USA). The concentrations of these samples were tested using a spectrophotometer and Na_2HPO_4 spec solution. The 36 samples (9 lines × 2 time points × 2 samples) of extracted RNA were labeled using the one-cycle cDNA Synthesis protocol from Affymetrix. cDNA was made from the extracted RNA by first making a T7-Oligo(dT) Primer Master Mix and allowing it incubate with the samples for 10 min at 70°C and cooling for 2 min at 4°C. A First-Strand Master Mix was added, and the samples were incubated for 2 min at 42°C. After 200U/µl SuperScript II was added, the samples were incubated for an additional hour at 42°C and then cooled at 4°C for 2 min. A Second-Strand Master Mix was added to the samples and then they were incubated at 16°C for 2 h and then cooled for 2 min at 4°C. The samples were incubated for another 5 min at 16°C after T4 DNA polymerase was added, the samples were cooled at 4°C for 2 min, and 0.5 M EDTA was added. The now double-stranded cDNA was cleaned up using spin columns and 100% ethanol. An IVT reaction mix was added to the samples to synthesize biotin-labeled cRNA, and they were allowed to incubate at 37°C overnight (approximately 16 h). The newly biotin-labeled cRNA was cleaned up and quantified using a spectrophotometer. Sample purity was between 1.96 and 1.7 (A260/A280). Fragmented samples were then stored at −20°C until hybridizations. Hybridizations to the Affymetrix Drosophila 2.0 GeneChip microarray took place in the Microarray Core Facility at UC Davis. All raw data from the experiment were deposited in the GEO database (http://www.ncbi.nlm.nih.gov/geo/) under the series record GSE9982.

The transcript levels were reconstructed from feature hybridizations using ArrayAssist, with subsequent log and variance normalization using the PLIER procedure (Therneau and Ballman 2005). There was no evidence for spot saturation in any of the arrays and all analyses were conducted on unadjusted data. Intensity values are weighted averages of the set of oligonucleotide probes for each gene. The data are available in supporting information, Table S1. As a minority of the genes are expected to be expressed in 5–8 h-old embryos, the transcript-level data were purged of genes called Absent by the Affymetrix MAS5 procedure in more than half the samples, which left 5065 genes (shown in supporting information, Table S2).

Downstream analyses of gene-level hybridization intensities were performed in SAS (SAS Institute, Cary, NC, USA) using proc GLM. Normalized transcript levels from the microarray hybridizations were fit to the following model: $Y_{ijk} = \mu + l_i + t_j + \varepsilon_{ijk}$, where the parameter µ is the overall mean transcript abundance for each gene, and the terms l and t stand for line (random) and time (fixed) effect ("k" represents the replicate arrays). We did not include the line-by-time interaction term as only four observations are available in each of nine lines for a total of 36 microarrays.

Factor analysis and factor loadings were estimated on mean-centered data using an oblique rotation in FACTOR Proc (SAS Institute). Following FACTOR Proc guidelines, we used variables (including factors themselves) standardized to have a unit variance (in this case, standardized regression coefficients are equivalent to correlation coefficients). Specifically, we used the options method = prin, priors = sms, rotate = promax. The resulting set of eigenvalues was plotted in a SCREE plot, and the number of factors was chosen such that a sharp drop-off between eigenvalues was apparent, and a reasonable proportion of the variation was still explained (Stevens 1996). Once the number of factors was identified, the analysis was repeated for that fixed number of factors to estimate loading values (the correlation between individual genes and the estimated network). Coffman et al. (2005) implemented extensive simulations to establish a sensible way to apply factor analyses to microarray data, which typically have rather few samples but many observations; we followed the general recommendations of this work.

RESULTS AND DISCUSSION

Variation in early embryo transcript levels

We first examined the total number of genes showing significant variation among lines of *D. melanogaster* in early zygotic

gene expression. Of the 5065 genes expressed in a majority of lines (see "Materials and methods"), 3754 showed significant variation among lines across both time points at $P < 0.05$ (supporting information, Table S3). The expected number of significant genes at this threshold is 253, giving a false discovery rate (FDR) of 0.07 (Benjamini and Hochberg 1995).

If we examine variation in gene expression at each time point separately, we find 3084 genes with significant line effects at 5 h ($P < 0.05$; FDR = 0.08) and 2794 genes with significant line effects at 8 h ($P < 0.05$; FDR = 0.09). There are 1948 genes that show significant effects at both time points in development, which implies that the remaining genes only have genotypic variation in gene expression during some fraction of Drosophila development. It is also possible that we have less power to detect line effects in these 1948 genes because the variance in expression changes over time; unfortunately, we have too few samples to detect such changes. In addition, 1780 genes show significant time effects (FDR = 0.14), indicating that they differ in expression across the two time points of development. These changes involve both increases and decreases in transcript level (see next section).

Comparing our results with previous studies of gene expression in adult flies, we find evidence that there is more genetic variation in embryos than adults, and a larger effect of age on gene expression in embryos than adults. The earliest studies of genetic variation in *D. melanogaster* adults found evidence for an effect of genotype on 10–25% of all genes (Jin et al. 2001; Meiklejohn et al. 2003; Rifkin et al. 2003; Gibson et al. 2004), and an effect of age for only 1% of genes (Jin et al. 2001). We find that 74% (3754/5065) of all expressed genes show a significant genotypic effect. Likewise, 35% (1780/5065) of expressed genes show a significant age effect across the two time points sampled here. We believe that our results are consistent with previous studies of embryonic gene expression in *D. melanogaster*. In a comprehensive study of gene expression in a single genotype across Drosophila development, Arbeitman et al. (2002) found that 86% (3483/4028) of genes showed significant variation over time; of these, 60% (2089/4028) appeared to vary across the first 20 h of development. Together with the results presented here and those of Jin et al. (2001), this appears to indicate that fluctuating levels of gene expression are typical of early development and are relatively rare in adults.

As our results show relatively more variation in gene expression than previous studies, it is unlikely that our results are due to an idiosyncratic experimental design or microarray platform: many of these studies in *D. melanogaster* have also used Affymetrix GeneChip arrays (e.g., Michalak and Noor 2003; Nuzhdin et al. 2004; Wayne et al. 2004). Because only two replicates per line were used, it is possible that the variance has been misestimated. Likewise, our small number of replicates means that we cannot determine whether the variance is heteroskedastic over time. Either of these situations could lead to an overestimation of the number of significant results. However, our experimental design does not differ considerably from similar microarray experiments, and we therefore do not believe that our results are due solely to statistical error. Our results do indicate that there is abundant genetic variation for early gene expression. Extensive variation in early zygotic transcript levels implies that there is a large source of genetic variation on which evolution can act in early development. This variation may underlie many changes in morphology and behavior (e.g., Kim et al. 2000), and is predicted to be a major source of evolutionary novelty (Raff and Kaufman 1983).

One question raised by our results is the source of variation in levels of transcription: that is, is it zygotic or maternal? Many transcripts are maternally loaded into the egg during oogenesis, with either no direct zygotic expression or a gradual increase in zygotic expression as maternal transcripts turn over (Davidson 1986). Estimates put the proportion of maternally deposited genes as high as 30% (Arbeitman et al. 2002). Maternal variation in either the production of these transcripts or in the loading of these transcripts can result in apparent variation in expression in the embryo; likewise, variation among embryos in rates of maternal mRNA turnover can result in varying transcript levels. Although all of this does not mean that the observed variation is not genetic, it does indicate the possibility that control of expression lies in adult flies. Even when genetic control of an individual gene's expression lies with the embryo, however, it still does not indicate that the important source of variation is in the embryo. For instance, if nutrient deposition is genetically controlled by the mother, then it may be that the embryos we have collected have varying developmental rates and are therefore not developmentally synchronized. If developmental rates vary in a way that is somehow proportional to the amount of nutrient supplied, there will appear to be genetic variation in embryonic gene expression, even though the source of this genetic variation lies with the maternal parent. Such "molecular heterochrony" (cf. Kim et al. 2000) may underlie some significant fraction of variation in gene expression. With these caveats in mind, we still conclude that abundant genetic variation—at least partially of zygotic origin—is present for transcript levels during early development.

Maternal and zygotic gene expression

To further examine the interplay between maternal and zygotic genes, as well as variation in expression between lines and over time, we directly compared our results with those of Arbeitman et al. (2002). These researchers defined five classes of genes: (1) changing during development; (2) strictly maternal; (3) maternal; (4) strictly zygotic; and (5) zygotic. Those genes classified as "changing during development" showed variation over time in the first 20 h in their experiment

Table 1. Variation in gene expression

Gene category	Arbeitman et al. (2002)	Found on Affymetrix microarray	Expressed	Varying among lines	Varying over time	Increasing	Declining
Changes during development	2089 (AST7[2])	2065[1] (ST4)	1215 (ST5)	950 (ST6)	698 (ST7)	224 (ST8)	491 (ST9)
Strictly maternal	27 (AST12)	27 (ST10)	21 (ST11)	21 (ST12)	19 (ST13)	0	19 (ST14)
Maternal	49 (AST13)	50 (ST15)	41 (ST16)	35 (ST17)	22 (ST18)	2 (ST19)	20 (ST20)
Strictly zygotic	53 (AST19)	57 (ST27)	29 (ST28)	25 (ST29)	22 (ST30)	19 (ST31)	3 (ST32)
Zygotic	532 (AST18)	514 (ST21)	176 (ST22)	141 (ST23)	127 (ST24)	89 (ST25)	38 (ST26)

[1]In a few instances, two different measurements per gene were obtained from Affymetrix slides, causing slight mismatches of numbers in the table (for instance, Arbeitman et al. [2002] detected 53 strictly zygotic genes, which on Affymetrix microarray are represented by 57 probes). We retained all of them in this table (all the data can be downloaded from supporting information tables).
[2]AST stands for supporting information table in Arbeitman et al. [2002] and ST is for the supporting information table in this article.

($n = 2089$); "strictly maternal" were strongly degraded after fertilization and did not reappear until female oogenesis ($n = 27$); "maternal" genes showed gradual declines in transcript levels during development ($n = 49$); "strictly zygotic" genes increased expression by at least 10-fold in the first 6.5 h of development ($n = 53$); and "zygotic" showed gradual increases in transcript levels ($n = 532$; Arbeitman et al. 2002).

Overall, our data on expression variation over time were highly consistent with these previous findings (Table 1). For instance, we detected expression in 21 of the 27 "strictly maternal" genes on the Affymetrix array. As expected, 19 of the 21 significantly differ in transcript level between the 5- and 8 h-old embryos, and every one of these genes declines in expression. For the "strictly zygotic" genes, 19 of the 22 that showed differences in expression over time showed the expected increase in expression. However, three genes showed an unexpected pattern of decreased expression (Table 1), although this pattern is not significantly different from the expected ($\chi^2 = 1.4$, $P = 0.23$). There are even more previously defined "zygotic" genes that show decreases in expression over development in our experiment (38 of 176 expressed), a significant excess relative to the expectation that zygotic genes are all increasing in expression ($\chi^2 = 57.5$, $P = 3.4 \times 10^{-14}$). We attribute these and other similar deviations from expected patterns to genetic differences between our lines and those of Arbeitman and colleagues (however, inconsistencies among platforms are a persistent feature of microarray analyses [Yauk et al. 2004] that might also contribute to the differences we observe). Although genes that showed the most extreme changes in transcript levels in the previous experiment showed similar changes here, there was much more lability in those genes that showed only gradual, modest changes. These results imply that annotating the function of genes based on the transcriptional profile of a single genotype may often result in misannotation and incorrect functional assignment. As this is done quite often in many organisms and throughout "systems biology" (e.g., Spellman et al. 1998), results from single-genotype experiments should be viewed with the appropriate amount of caution.

Our data show a huge amount of genetic variation in expression for all five classes of genes (Table 1), although not all of this variation results in a reversal in the direction of transcript-level changes. Between 78% and 100% of all genes that have been classified as maternal or zygotic appear to differ in transcript level among our lines. There are minor differences in the proportion of genes with varying transcript levels between maternal and zygotic, with a higher proportion for "maternal" than "zygotic" as well as for "strictly maternal" compared with "strictly zygotic" (Table 1). This difference in variation between maternal and zygotic genes has been predicted by some models because maternal effects are only expressed in a single sex (Demuth and Wade 2007), although the differences observed here are not close to the expected 2:1 ratio (Barker et al. 2005). It is important to note that, overall, fewer genes are significantly different between time points than among lines. Accordingly, imperfect synchronization during egg collections cannot completely account for the among-line variation we observe (see Table 1).

GRNs

One major goal of studies into GRNs (Levine and Davidson 2005) is to be able to describe variation in transcript levels in terms of the interactions between genes in the network (e.g., Tarone et al. 2005). In order to achieve this goal, however, we need data on both the structure of the network and on the genetic variation in gene transcript levels. Detailed knowledge of the GRNs for both *Drosophila* segmentation (Schroeder

et al. 2004) and dorsal–ventral patterning (Levine and Davidson 2005) provides us with much of the information needed to mechanistically describe variance–covariance patterns in transcription. These patterns can then be used to validate individual protein–DNA interactions that represent "*cis*-regulatory transactions" inferred to be present in the GRN via other means (Levine and Davidson 2005).

Embryonic segmentation is an outcome of the maternal gradient genes—*bicoid* (*bcd*), *hunchback* (*hb*), *caudal* (*cad*), *Torso* (*Tor*), and *Stat92E* (*D-Stat*)—affecting downstream gap factors—*Kruppel* (*Kr*), *knirps* (*kni*), *giant* (*gt*), and *tailless* (*tll*) (Carroll 1990; Rivera-Pomar and Jackle 1996). Crossregulation among these genes establishes their patterns of spatial and temporal expression, as well as those of their downstream targets. Both computational and experimental results have linked these upstream genes to target genes at later stages in the segmentation GRN (Schroeder et al. 2004). We hypothesized that variation in the transcript abundance of upstream genes should result in variation in the transcript abundance of downstream targets of these genes. For instance, the *even skipped* (*eve*) stripe 1 *cis*-regulatory module is strongly bound by *bcd* and *Kr* in one of the regions of the embryo, and is also more weakly bound by *hb* and *gt* in another region. This implies transcriptional control of *eve* by *bcd* and *Kr*. To test this relationship, we regressed the transcript level of *eve* (predicted variable) on the transcript levels of *bcd* and *Kr* (predictor variables). We developed this and other regression models for the segmentation GRN in exact accordance with fig. 4 from Schroeder et al. (2004). We omitted genes for which evidence of among-line variation is missing.

We were able to construct 34 separate regression models describing the relationships between upstream effector genes and their downstream targets. A description of the full set of models and their overall fit to the data are summarized in supporting information, Table S33. Out of 34 models in total, 30 were significant at $P<0.05$, four of them at $P<0.0001$. The effects of upstream genes were both positive and negative; they therefore appear to act as both activators and repressors (see next section). We conclude that this analysis represents a compelling case of a high overall fit of the variance–covariance structure of transcript levels to the pattern expected from previous molecular genetic experiments.

The dorsal–ventral GRN is composed of nearly 60 genes (see fig. 2 in Levine and Davidson 2005). As we used a rather stringent cut-off for calling a gene "Expressed" (called present in more than half the samples), many of these genes did not meet this criterion (supporting information, Table S2). We therefore limited our analysis to a consecutive stretch of the GRN consisting of *cactus* (*cact*), *dorsal* (*dl*), *easter* (*ea*), *pelle* (*pll*), *spatzle* (*spz*), *tube* (*tub*), *snail* (*sna*), *stumps* (also called *hbr*), and *twist* (*twi*)—each expressed and varying (see supporting information, Table S3). From the relationships among these genes, we were able to construct seven regression models (Fig. 1). Of the seven total models, five were significant at $P<0.05$ and three of these were significant at $P<0.01$ (Fig. 1). Random pairing of 1000 genes in our experiment shows that six of the seven models have correlations higher than expected ($P<0.05$). Although we are able to make fewer comparisons than in the segmentation GRN, our results also show a good concordance between the variance–covariance structure of the transcript level and the previous experimental evidence on the structure of the dorsal–ventral GRN.

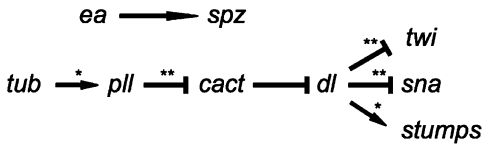

Fig. 1. Relationships between genes in the dorsal–ventral gene regulatory network. Each arrow represents a regression model, with upstream genes affecting their downstream targets based on the directionality of each arrow. $^{*}P<0.05$, $^{**}P<0.01$.

The dimensionality of variation

Although encouraging, the above analyses perhaps do not fully capture the biological mechanisms underlying transcriptional variation. Imagine, for example, that what really varies between samples is a gradient of *hb*. Downstream, *hb* variation generates variation in *hb* targets; these in turn generate variation in their own downstream targets, ad infinitum. Accordingly, a single factor might cascade down the GRN, resulting in numerous variance–covariance profiles all fitting seemingly different regression models. Described in more intuitive terms, we can say that transcriptional variation might have low dimensionality. To search for the number of dimensions potentially accounting for variation in multiple expression profiles, we used factor analysis (cf. Coffman et al. 2005).

As applied to array data, factor analysis is an analytic approach that can describe the covariation among a set of genes through the estimation of factors (Coffman et al. 2005). Individual factors represent putative biological mechanisms by which genes are coregulated; in this case, they likely represent individual transcription factors. The factor model that results from such analyses represents a set of coordinately expressed genes (genes may participate in multiple factor models). Factor analysis represents the relationship between each gene and the factor as a load between -1 and 1, where the value indicates the strength and direction of each factor's influence on transcript levels. Following the recommendations of Coffman et al. (2005) for the analysis of microarray data, we initially limited our analyses to the upstream genes of the segmentation (*Kr*, *D-Stat*, *bcd*, *cad*, *gt*, *hb*, *kni*, *tll*, and *Tor*) and dorsal–ventral (*cact*, *dl*, *ea*, *pll*, *spz*, *tub*, *sna*, *stumps*, and *twi*) GRNs.

For the segmentation GRN, we retained two factors because their eigenvalues exceeded one, the appearance of the SCREE plot, and at least three different genes significantly loaded (>0.4) on each factor (see "Materials and methods" and supporting information, Table S34). The two selected factors account for 85% of the total variance in transcript level. The loadings of genes on rotated factors are shown in Table 2. Generally, maternal genes load positively on Factor 2 and negatively on Factor 1. Gap genes, in contrast, load negatively on Factor 2 but positively on Factor 1. That this split so closely resembles the previously defined roles of maternal and gap genes is an esthetically and scientifically pleasing outcome of the analysis. It also authenticates the relationships between genes that have previously only been defined by computational means, or by empirical means in only a small number of genotypes.

More specifically, maternal genes vary in two dimensions, with *D-Stat*, *cad*, and *Tor* positively and highly significantly loading on Factor 2, and *bcd* and *hb* loading negatively and highly significantly on Factor 1. Gap genes appear to vary in a single dimension with high positive loading on Factor 1. It is tempting to speculate that genetic variation in these genes causes downstream variation in gap genes. We should point out, however, that correlation is not equivalent to causation. Two factors largely account for variation of the other segmentation genes as well. Every single gene in the segmentation GRN (*D, Btd, cnc, ems, eve, fkh, ftz, h, hkb, knrl, nub, oc, odd, pdm2, run, and slp2*: see fig. 4 in Schroeder et al. 2004) is positively correlated with Factor 1, resembling the pattern for maternal genes and contrasting with the loadings of gap genes. For most of these relationships, the correlations are close to 1 and highly significant (with the exception of *cnc*: $r = 0.14$, $P = 0.716$, and *nub*: $r = 0.66$, $P = 0.055$). Similar to maternal genes, negative correlations are typical with Factor 2 (supporting information, Table S35). These patterns remain robust if we analyze either of the time points alone, and the same general structure of factors is recovered when the difference in expression between time points is analyzed rather than the individual transcript levels at each (results not shown). We conclude that nearly all variation in transcript levels of segmentation genes is explained by two factors.

Based on arguments identical to those for the segmentation GRN, two factors were retained for the dorsal–ventral GRN (supporting information, Table S36). These two factors account for 90% of among-line variation. The second factor nearly perfectly covaries with the *ea* gene, which is far upstream in the modeled portion of the network. The first factor is marked, again due to nearly perfect correlation, by the *cact* gene (Table 2). Most of the genes significantly load on this factor, but show negative correlations with *cact*. This is expected as *cact* is suppressed by *pll*, and is itself a suppressor of *dl*, although the true cause of these relationships is unknown. Overall, variation in the transcript levels of downstream genes—*twi*, *sna*, and *stumps*—appears to be jointly accounted for by the two factors, with different strengths of effects for each. When the genes in the remaining dorsal–ventral GRN are correlated with the identified factors, most of the variation for most of genes is accounted for by the two factors (supporting information, Table S37).

Table 2. Factor analysis of expression variation

Gene	Loading on factor 1	Loading on factor 2	Gene	Loading on factor 1	Loading on factor 2
Kr	0.84*	−0.19	*cact*	1.04*	0.16
D-Stat	0.29	1.04*	*dl*	−0.88*	0.15
bcd	−0.74*	0.35	*ea*	0.30	1.06*
cad	−0.38	0.72*	*pll*	−0.92*	−0.01
gt	0.68*	−0.25	*spz*	−0.70*	0.38
hb	−0.94*	−0.40	*tub*	−0.44*	0.67*
kni	0.79*	−0.18	*sna*	0.64*	−0.42*
tll	0.84*	−0.19	*stumps*	0.24	−0.81*
Tor	−0.28	0.76*	*twi*	0.99*	0.12

The genes from the segmentation network are on the left, those from the dorsal–ventral network on the right.
*Significant correlations at $P<0.05$ are starred.

Overall, we conclude that the variance–covariance structure of transcriptional variation fits reasonably well with the known hierarchical structure of GRNs. In addition, few dimensions of variation appear to account for most of the variation in transcript level. This result reaffirms many previous studies that have shown that maternal genes lie upstream of gap genes; it also shows that these genes may act as either activators or repressors. We do not think this confirmation should be surprising, although we have shown for the first time that the known mechanistic relationships are recapitulated by patterns of genetic variation. One standard interpretation of our results finding a small number of factors for both the segmentation and the dorsal–ventral GRNs is that there may be just a few mutations controlling all of the downstream variation. This would imply that most of the variation in gene expression that we observe lies in *trans*-acting factors, rather than many *cis*-acting changes in the varying genes; this contradicts some previous results in Drosophila (e.g., Wittkopp et al. 2004). An alternative interpretation, however, is that the structure of the GRNs is such that even multiple *cis*-acting mutations acting throughout genes in the network would result in only a small number of factors. This would come about because the network strongly constrains the effects of each member gene; if each can only act locally as either an activator or a repressor, the emergent behavior of the network may resemble the action of only two factors. As we cannot distinguish between these two possibilities at the moment, an answer will have to await further linkage studies (e.g., Wayne et al. 2004).

Acknowledgments

We thank L. McIntyre and two anonymous reviewers for comments that helped to improve the manuscript. We acknowledge support from National Institutes of Health grant R01-GM076643A to S. V. N. and M. W. H.

REFERENCES

Arbeitman, M. N., et al. 2002. Gene expression during the life cycle of *Drosophila melanogaster*. *Science* 297: 2270–2275.

Barker, M. S., Demuth, J. P., and Wade, M. J. 2005. Maternal expression relaxes constraint on innovation of the anterior determinant, bicoid. *Public Libr. Sci. Genet*. 1: 527–530.

Benjamini, Y., and Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B-Methodological* 57: 289–300.

Brem, R. B., Yvert, G., Clinton, R., and Kruglyak, L. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* 296: 752–755.

Carroll, S. B. 1990. Zebra patterns in fly embryos: activation of stripes or repression of interstripes? *Cell* 60: 9–16.

Carroll, S. B., Grenier, J. K., and Weatherbee, S. D. 2001. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Science Inc., Malden, MA.

Cavalieri, D., Townsend, J. P., and Hartl, D. L. 2000. Manifold anomalies in gene expression in a vineyard isolate of *Saccharomyces cerevisiae* revealed by DNA microarray analysis. *Proc. Natl. Acad. Sci. USA* 97: 12369–12374.

Coffman, C. J., Wayne, M. L., Nuzhdin, S. V., Higgins, L. A., and McIntyre, L. M. 2005. Identification of co-regulated transcripts affecting male body size in *Drosophila*. *Genome Biol*. 6: R53.

Davidson, E. H. 1986. *Gene Activity in Early Development*. Academic Press, Orlando, FL.

Davidson, E. H. 2001. *Genomic Regulatory Systems: Development and Evolution*. Academic Press, San Diego.

Demuth, J. P., and Wade, M. J. 2007. Maternal expression increases the rate of *bicoid* evolution by relaxing selective constraint. *Genetica* 129: 37–43.

Enard, W., et al. 2002. Intra- and interspecific variation in primate gene expression patterns. *Science* 296: 340–343.

Gibson, G., et al. 2004. Extensive sex-specific nonadditivity of gene expression in *Drosophila melanogaster*. *Genetics* 167: 1791–1799.

Gould, S. J. 1977. *Ontogeny and Phylogeny*. Belknap, Cambridge, MA.

Jin, W., Riley, R. M., Wolfinger, R. D., White, K. P., Passador-Gurgel, G., and Gibson, G. 2001. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet*. 29: 389–395.

Karp, C. L., et al. 2000. Identification of complement factor 5 as a susceptibility locus for experimental allergic asthma. *Nat. Immunol*. 1: 221–226.

Kim, J., Kerr, J. Q., and Min, G.-S. 2000. Molecular heterochrony in the early development of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 97: 212–216.

Levine, M., and Davidson, E. H. 2005. Gene regulatory networks for development. *Proc. Natl. Acad. Sci. USA* 102: 4936–4942.

Meiklejohn, C. D., Parsch, J., Ranz, J. M., and Hartl, D. L. 2003. Rapid evolution of male-biased gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 100: 9894–9899.

Michalak, P., and Noor, M. A. F. 2003. Genome-wide patterns of expression in *Drosophila* pure species and hybrid males. *Mol. Biol. Evol*. 20: 1070–1076.

Nuzhdin, S. V., Wayne, M. L., Harmon, K. L., and McIntyre, L. M. 2004. Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Mol. Biol. Evol*. 21: 1308–1317.

Oleksiak, M. F., Churchill, G. A., and Crawford, D. L. 2002. Variation in gene expression within and among natural populations. *Nat. Genet*. 32: 261–266.

Raff, R. A. 1996. *The Shape of Life: Genes, Development, and the Evolution of Animal Form*. The University of Chicago Press, Chicago.

Raff, R. A., and Kaufman, T. C. 1983. *Embryos, Genes, and Evolution: The Developmental-Genetic Basis of Evolutionary Change*. Macmillan Publishing Co. Inc., New York.

Rifkin, S. A., Kim, J., and White, K. P. 2003. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet*. 33: 138–144.

Rivera-Pomar, R., and Jackle, H. 1996. From gradients to stripes in *Drosophila* embryogenesis: filling in the gaps. *Trends Genet*. 12: 478–483.

Schadt, E. E., et al. 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature* 422: 297–302.

Schroeder, M. D., et al. 2004. Transcriptional control in the segmentation gene network of *Drosophila*. *Public Libr. Sci. Biol*. 2: 1396–1410.

Spellman, P. T., et al. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* 9: 3273–3297.

Stevens, J. 1996. *Applied Multivariate Statistics for the Social Sciences*. Lawrence Erlbaum Associates Inc., Mahwah, NJ.

Tarone, A. M., Nasser, Y. M., and Nuzhdin, S. V. 2005. Genetic variation for expression of the sex determination pathway genes in *Drosophila melanogaster*. *Genet. Res*. 86: 31–40.

Therneau, T. M., and Ballman, K. V. 2005. What does PLIER really do? Technical Report Series No. 75, Department of Health Science Research, Mayo Clinic, Rochester, MN.

Wayne, M. L., Pan, Y. J., Nuzhdin, S. V., and McIntyre, L. M. 2004. Additivity and trans-acting effects on gene expression in male *Drosophila simulans*. *Genetics* 168: 1413–1420.

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. 2004. Evolutionary changes in *cis* and *trans* gene regulation. *Nature* 430: 85–88.

Wray, G. A., et al. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol*. 20: 1377–1419.

Yauk, C. L., Berndt, M. L., Williams, A., and Douglas, G. R. 2004. Comprehensive comparison of six microarray technologies. *Nucleic Acids Res*. 32: e124.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Table S1–Table S37

Please note: Wiley-Periodicals, Inc. are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.