

# Widespread Transcriptional Autosomal Dosage Compensation in *Drosophila* Correlates with Gene Expression Level

Ashley A. McAnally and Lev Y. Yampolsky\*

Department of Biological Sciences, East Tennessee State University

\*Corresponding author: E-mail: yampolsk@etsu.edu.

**Accepted:** 15 December 2009 **Associate editor:** William Martin

## Abstract

Little is known about dosage compensation in autosomal genes. Transcription-level compensation of deletions and other loss-of-function mutations may be a mechanism of dominance of wild-type alleles, a ubiquitous phenomenon whose nature has been a subject of a long debate. We measured gene expression in two isogenic *Drosophila* lines heterozygous for long deletions and compared our results with previously published gene expression data in a line heterozygous for a long duplication. We find that a majority of genes are at least partially compensated at transcription, both for ½-fold dosage (in heterozygotes for deletions) and for 1.5-fold dosage (in heterozygotes for a duplication). The degree of compensation does not vary among functional classes of genes. Compensation for deletions is stronger for highly expressed genes. In contrast, the degree of compensation for duplications is stronger for weakly expressed genes. Thus, partial transcriptional compensation appears to be based on regulatory mechanisms that insure high transcription levels of some genes and low transcription levels of other genes, instead of precise maintenance of a particular homeostatic expression level. Given the ubiquity of transcriptional compensation, dominance of wild-type alleles may be at least partially caused by the regulation at transcription level.

**Key words:** dosage compensation, dominance, regulation of transcription, deletions, duplications, *Drosophila*.

## Introduction

In bisexual organisms with chromosomal sex determination, sex-linked genes occur in different dosages in the two sexes, and this dosage difference in genes not directly related to sex determination is compensated by a variety of mechanisms, such as deactivation of one of X chromosomes in mammalian females (Lyon 1988; Straub and Becker 2007) or higher expression of X-linked genes in *Drosophila* males (Baker et al. 1994; Stuckenholz et al. 1999; Gupta et al. 2006). Much less is known about autosomal dosage compensation. Gene-specific data indicate that autosomal dosage compensation may be a common phenomenon in *Drosophila* and that it is likely to occur on the transcriptional level (Devlin et al. 1982; Birchler et al. 1990). Broader evidence that a broad spectrum of autosomal genes are under transcriptional regulation compensating for aneuploidy comes from transcriptome analysis of trisomies (FitzPatrick et al. 2002) and cancer-associated aneuploidies (Tsafrir et al. 2006; Williams et al. 2008) in human and mice and a variety of

aneuploid genotypes in yeast (Torres et al. 2007). Although transcriptional level, as detected by microarrays, generally followed the DNA dosage trend in these studies, two important trends became apparent: at least some genes in aneuploid regions are, in fact, transcribed at a nearly normal diploid level and many misregulated genes are not located in aneuploid regions (FitzPatrick et al. 2002; Tsafrir et al. 2006). In contrast to mammalian and yeast data, two recent studies (Gupta et al. 2006; Stenberg et al. 2009; Zhang et al. 2009) indicated that autosomal dosage compensation is a rule rather than exception in *Drosophila* (see below).

If autosomal transcriptional dosage compensation is indeed common in at least some organisms, three important questions can be asked. First, do these mechanisms act on gene-specific level or on the level of larger chromosomal segments? Second, are these mechanisms capable of fine-tuned regulation of transcription level compensating for both deficiencies and duplications or do they operate on a more coarse scale, that is, assuring that genes with

high-demand products are expressed at a sufficiently high level? Finally, can autosomal dosage compensation be the fundamental basis of dominance?

Dominance is a pervasive although not a universal property of genes. There are two aspects of dominance that require explanation: 1) why do most genes exhibit complete dominance of one of the alleles rather than additivity of the action of two alleles (i.e., why is codominance a relatively rare phenomenon), and 2) why is it the wild-type allele that is usually dominant. Possible explanation of these factors has been the subject of perhaps the fiercest debate between the founding fathers of modern synthesis. Fisher (1928) proposed that dominance of the wild-type alleles is the result of selection on modifier genes, which epistatically mask the action of the mutant allele in a heterozygote. By necessity such evolution can only occur in the heterozygous subpopulation. This idea was met with criticism by cofounders of the modern synthesis (Wright 1934; Haldane 1939). Instead, Wright suggested that dominance is an inherent property of the physiological systems, perhaps evolved through selection to provide a safety margin in the action of a single functional copy of a gene, which would allow to accommodate for environmental fluctuations and for the lack of activity of the other gene. Such selection would act on the entire population, not on heterozygotes only, and is therefore much more powerful.

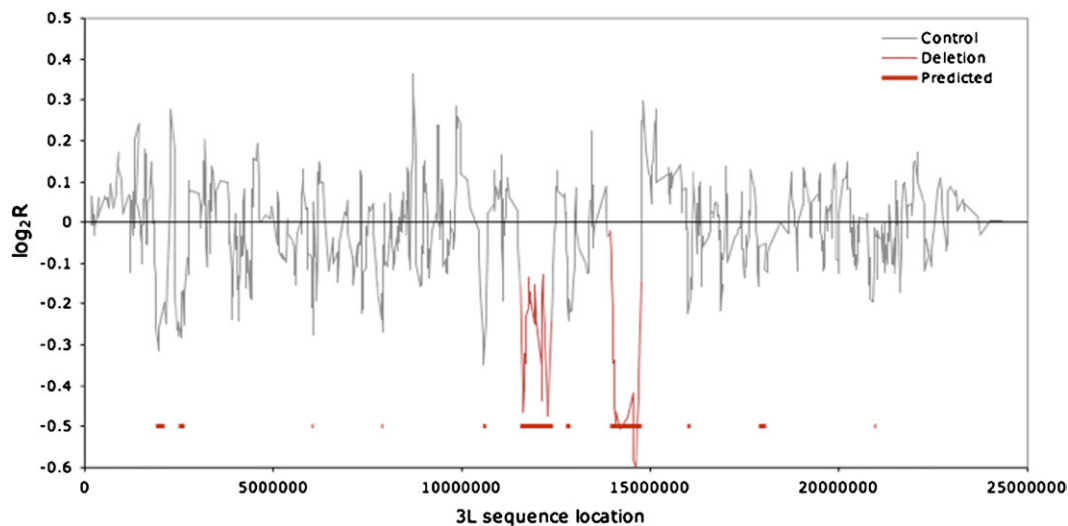
A variety of studies during last 30 years provided strong evidence in favor of the physiological theory of dominance. In particular, the metabolic control theory (Kacser and Burns 1981) implies that enzymes functioning in metabolic pathways are bound to exhibit a diminishing return relationship between activity of an individual enzyme and the flux of metabolites through the whole pathway, which ultimately determines the phenotype. Thus, 2-fold change in protein titer (as in a heterozygote for a loss-of-function mutation) is usually negligible in terms of the resulting phenotype. This idea was supported by negative correlation between the strength of a mutant allele and its dominance (Charlesworth 1979; Crow 1979; Crow and Simmons 1983; Phadnis and Fry 2005). Further support to the physiological dominance theory came from the observation that dominance is prevalent in organisms, which spend much of their life cycle in haploid phase, such as in *Chlamydomonas* (Orr 1991) or fission yeast (Baek et al. 2008). Finally, loss-of-function mutations are more likely to be fully recessive in enzyme-coding genes than in genes coding for structural or regulatory proteins (Fisher and Scambler 1994; Veitia 2002; Kondrashov and Koonin 2004) and for proteins that are less likely to form protein complexes (Papp et al. 2003). Due to these findings, the physiological mechanisms behind dosage compensations are thought to be acting primarily on the protein level.

On the other hand, dosage compensation may as well occur at the level of transcription, resulting in a mechanism of dominance independent from the protein function. Eu-

karyotic transcriptional machinery is equipped with a stunning variety of mechanisms enabling fine-tuned regulation of transcription (Lee et al. 2002). Such mechanisms often incorporate negative feedbacks allowing adjustment of transcription level to match the environmental fluctuations or tissue-specific developmental needs; these feedbacks may as well provide compensation for a loss-of-function mutation in one of the alleles. On the other hand, one can hypothesize that, in many genes, the regulation of gene expression may be a lot less sophisticated and lacking fine-tuned gene-specific homeostasis mechanisms. In fact, the number of transcription factors present in the genome is constrained by the limits set by coding theory: unlimited increase of the number of transcription factors capable of recognizing specific nucleotide sequences would lead to the increase of misrecognition errors (Itzkovitz et al. 2006). Instead, there may be genes coding for high demand proteins, which are constitutively transcribed at a high rate, and genes coding for low demand proteins, whose transcription is maintained at low level. If this is true, one would expect that transcriptional dosage compensation to be widespread and to correlate with the overall gene expression level. Indeed, the pervasive nature of autosomal transcriptional compensation has been recently demonstrated in *Drosophila* (Gupta et al. 2006; Stenberg et al. 2009). In particular, Stenberg et al. (2009) report that the degree of compensation is different in genes with different degree of tissue specificity of expression: ubiquitously expressed genes are stronger compensated for deficiencies and less effectively for a duplication than tissue-specific ones. Stenberg et al. (2009) also suggest that there is no correlation between the degree of compensation and overall expression level. Here, we test these results using three sets of *Drosophila* genes. We measured the level of gene expression in two isogenic deletional lines using oligonucleotide microarrays and compared our results with published data on transcriptional compensation in heterozygotes for a duplication (Gupta et al. 2006).

## Materials and Methods

Two *DrosDel* (Ryder et al. 2007) *Drosophila melanogaster* isogenic lines, Df(3L)ED4475 and Df(3L)ED4543, heterozygous for long deletions on 3L chromosomal branch both maintained against the TM6C balancer, were used for microarray experiment. Twenty-five adult flies 2–5 days after eclosion were frozen in liquid nitrogen and used for RNA extraction by Trizol method (Invitrogen) in 12 replicates from each line. Samples were reverse transcribed, labeled, and hybridized to two-color 14k oligonucleotide microarrays by Canadian *Drosophila* Microarray Centre (Mississauga, Ontario, Canada; Neal et al. 2003). The two lines, which serve as controls for each other, were alternated between Cy3 and Cy5 dyes to minimize possible channel bias. Fluorescence intensity data (background intensity



**FIG. 1.**—Sliding average ( $\pm 5$  genes) of expression ratio along 3L chromosome branch. Gray line: diploid genes (wild type:wild type); red line: genes in deletions (Deletion:wild type). Thick red bars at  $\log_2 R = -0.5$  indicate genes identified as hemizygous by a neural network prediction.

subtracted) were analyzed by JMP Genomics 3.0 (SAS Institute 2007). The data are available at Gene Expression Omnibus (GEO) (Edgar et al. 2002; <http://www.ncbi.nlm.nih.gov/gds>; accession number GSE14799).

Data were ANOVA-normalized to eliminate the effects of arrays and channels, mean intensity of fluorescence for hemizygous and diploid line calculated for each gene and the ratio of the intensity in the line with the deletion to that in the control line ( $R$ ) analyzed. For a comparison with Gupta et al. (2006), in which raw data are available and to avoid possible expression level-dependent bias in expression ratios, the correlation between expression ratio and mean expression level was also analyzed without any normalization; the results are virtually identical. Line ED4475 had slightly but statistically significantly higher fluorescence across the genome (ED4475: ED4543 = 1.017 for diploid genes), and therefore, the ratio was adjusted by this factor. This adjustment had no effect on any of the findings reported. Presence or absence of expression for each gene was evaluated by comparison of average fluorescence across all arrays with the internal blank (AutoBlank) intensities and genes with intensities significantly ( $P < 0.05$ ) above the AutoBlank level were considered detectable. The distribution of  $R$  in hemizygous genes was analyzed by maximum likelihood as a superposition of two normal distributions one with the mean of 0.5 (uncompensated genes), the other (partially compensated genes) with unknown mean; this unknown mean, standard deviations of both distributions and the frequency of uncompensated genes were simultaneously optimized.

In order to verify that the observed expression ratios in deletion lines are not an artifact of inherent microarray noise, we performed a neural network-based analysis of the distribution of expression ratios along the chromosome.

A neural network consisting of three nodes in one hidden layer and implemented within JMP Genomics package (SAS Institute 2007) was provided the input data in the form of expression ratio for each gene with detectable expression along with the expression ratios of this gene's five nearest neighbors on each side (again, using only gene with detectable expression). The network was then trained to recognize each deletion using the data on the other deletion as the training set. The network recognized genes within deletions with 4.4% of false negatives (almost all on the deletion breakpoints) and 9.75% false positives (fig. 1). More than half of the false positives were present in runs of more than three adjacent genes and may represent actual unknown short deletions present in the DrosDel lines we used. We conclude, therefore, that our data contain a sufficiently strong signal to allow us the comparison of expression ratios between hemizygous and diploid genes.

Following Stenberg et al. (2009) approach, to compare ubiquitously expressed genes to tissue-specific genes, genes were classified as "ubiquitous" if they had expression value of at least 6 in all 26 adult and larval tissues present in the FlyAtlas database (Chintapalli et al. 2007).

Gupta et al. (2006) data were obtained from GEO (Edgar et al. 2002), accession numbers GSM37444, GSM37445, GSM37447, GSM37448, GSM77751, and GSM77752. These data contain expression levels in heterozygotes for a long duplication Dp(2;2)Cam3 relative to heterozygotes for a deletion Df(2L)JH located within the duplication. Thus, genes located within the duplication but outside of the deletion are compared with the diploid wild type (1.5-fold dosage), whereas genes located within the deletion are compared in lines with 3-fold gene dosage. There are 289 genes (2.1% of the genome) in the Dp(2;2)Cam3 duplication, 59 of which (0.4% of the genome) lay within

the deletion Df(2L)JH. See Gupta et al. (2006) for RNA extraction, microarray hybridization and scanning, and data handling details.

## Results

Deletions ED4475 and ED4543 contain 119 and 65 genes, respectively, comprising about 1.3% of the genome. Of these 184 genes all 184 demonstrated transcription level resulting in higher than background fluorescence, however, in only 69 (0.5% of the genome) the average fluorescence across all arrays was significantly ( $P < 0.05$ ) above the AutoBlank level. The distribution of expression ratios in genes with detectable expression along the chromosomal branch 3L, along with the neural network identification of hemizygous regions is presented on figure 1. Further statistical analysis, except the distribution shown on fig. 2A, is done only on genes with detectable expression. The distribution of ratio of mean fluorescence intensities ( $\log_2 R$ ) of hemizygous genes (i.e., genes within the deletions) is significantly different from that of the genes outside of the deletions ( $P < 10^{-12}$ , fig. 2A). It is shifted to the left (lower expression of hemizygous genes) with the modal class at  $\log_2 R = -0.1$  ( $R = 0.93$ ) and with a distinct shoulder at  $\log_2 R = -1$  (the value expected in case of no transcriptional compensation). The frequency of observations within  $-1.32$  to  $-0.74$  range (corresponding to the range of ratios 0.4–0.6) among hemizygous is significantly higher than among diploid genes ( $\chi^2 = 20.9$ ,  $P < 0.0001$ ). The distribution of  $\log_2 R$  differed somewhat between the two deletion lines ( $P < 0.05$ ) with the secondary mode around  $-1$  more pronounced in line ED4543 (fig. 1), indicating that the frequency of noncompensated genes may differ across chromosomal regions. The distribution of  $\log_2 R$  significantly deviated from normal in both deletion lines (Shapiro–Wilk test,  $P < 0.004$ ). Assuming a contaminated normal distribution (i.e., a distribution, in which the majority of observations is from a specified normal distribution, but a small proportion is from a normal distribution with a different mean and/or variance) and optimizing all parameters (both means, both variances, and the degree of contamination), the maximal likelihood estimate of the frequency of uncompensated genes  $F_{0.5}$  was 0.05. Values of  $F_{0.5}$  exceeding 0.25 and 0.35 can be excluded with significance level of 0.05 and 0.01, respectively. Thus, majority of genes within the two deletions exhibit at least partial compensation at the stage of transcription.

A similar result is apparent from the data from Gupta et al. (2006; fig. 2B). The distribution of the ratio of expression intensity in heterozygotes for duplication relative to wild type and heterozygotes for duplication relative to heterozygotes for a deletion are significantly higher than the control (expression ratios of genes outside of the aberrations;  $P < 10^{-12}$ ) but with modes close to 0, indicating

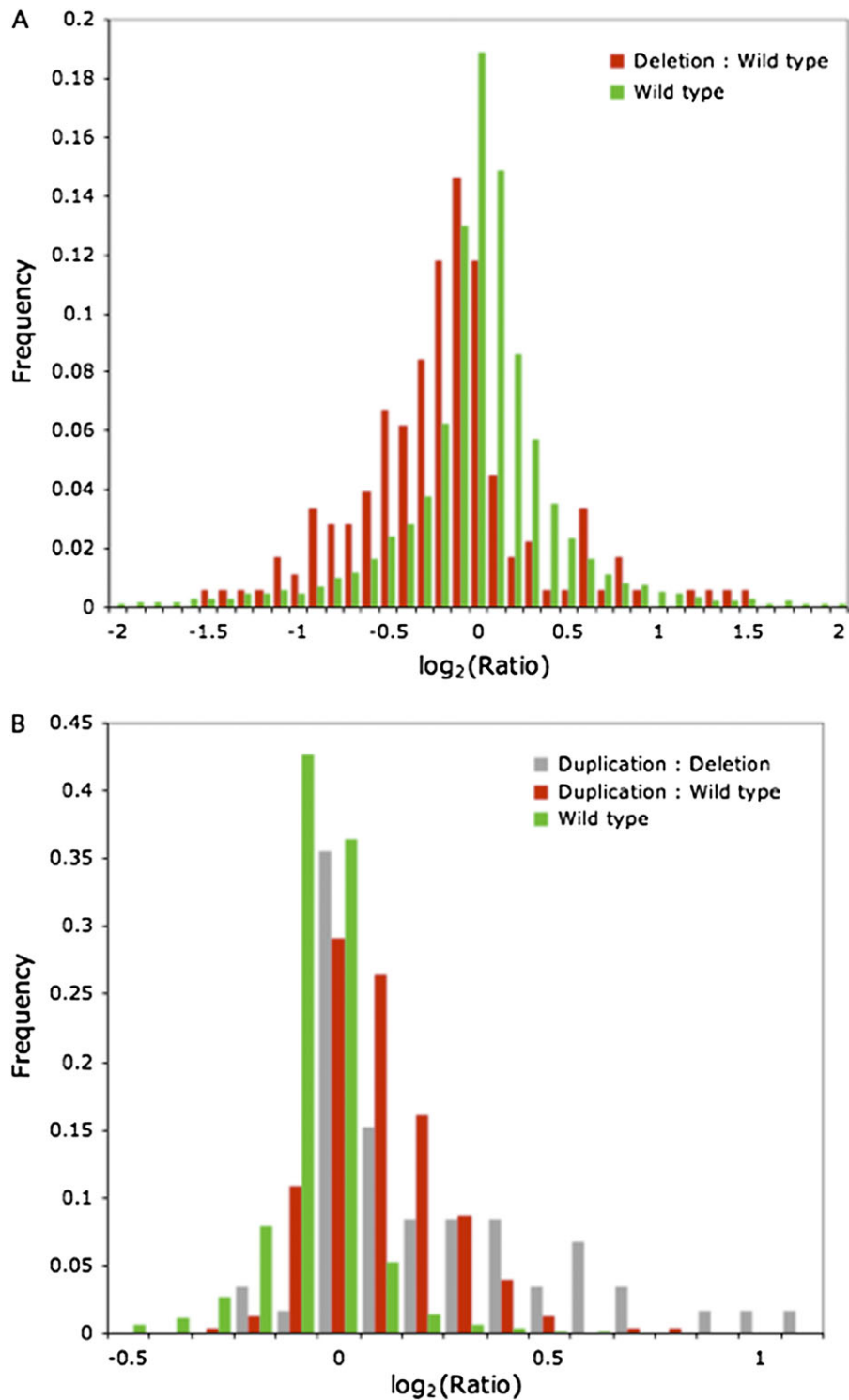
nearly complete dosage compensation in many genes and relatively few genes with no dosage compensation at all. In both studies, the degree of dosage compensation appears to be independent from the molecular function category to which genes belong (fig. 3).

There was, as predicted, a positive correlation between expression ratio ( $\log_2 R$ ) and overall expression level ( $\log_2 M$ ). For heterozygotes, for deletions, the regression (fig. 4A;  $\log_2 R = -0.77 + 0.077 \times \log_2 M$ ;  $P < 0.007$ ;  $P < 0.02$  with the uppermost outlier point removed) suggests that genes with barely detectable expression are nearly noncompensated ( $\log_2 R = -1$ ), whereas genes with highest observed expression level approach full compensation ( $\log_2 R = 0$ ). For heterozygotes, for a duplication versus diploid wild type, the regression (fig. 4B;  $\log_2 R = -0.77 + 0.14 \times \log_2 M$ ;  $P < 0.0001$ ) suggests that there is a nearly complete compensation of low expression genes ( $\log_2 R = 0$ ), whereas highly expressed genes demonstrate little dosage compensation ( $\log_2 R = \log_2(1.5) = 0.58$ ). The regression of expression ratio in heterozygotes for a duplication to that is heterozygotes for a deletion (3-fold dosage difference) over overall expression level is also positive (fig. 4B;  $\log_2 R = -1.41 + 0.25 \times \log_2 M$ ;  $P < 0.0008$ ). The corresponding regression for genes outside of the aberrations is nearly horizontal for both data sets, suggesting that the observed correlations are not a product of a bias in the expression data. It should be noted that this result is identical for both normalized and nonnormalized data. The observed correlations are also not artifacts of data fanning. For example, for deletions, expression ratios for genes with lower expression are expected tend to be located below the  $\log_2 R = 0$  line simply for the reason of being underexpressed in hemizygous state, whereas expression ratios for highly expressed genes, if not affected by mean expression, would be expected to cluster around  $\log_2 R = -1$ , not  $\log_2 R = 0$  as are points corresponding to diploid genes.

Figure 5 shows the relationship between the degree of tissue specificity of expression and degree of compensation in two deletions in this study (A) and in a duplication (Gupta et al. 2006; B). In one of the deletions, ED4475 (partly overlapping with the deletion used in Stenberg et al. 2009), ubiquitously expressed genes are less strongly compensated than tissue-specific genes ( $P < 0.003$ ), but in the other deletion, ED4542, the difference is not significant. Likewise, for the duplication, transcriptional compensation of ubiquitously expressed genes is weaker than that of tissue-specific genes, in a reversal of the pattern observed on figure 3B in Stenberg et al. (2009).

## Discussion

Our data confirm the recently observed phenomenon of widespread transcriptional compensation of genes located in haploid and triploid areas in *Drosophila* lines with

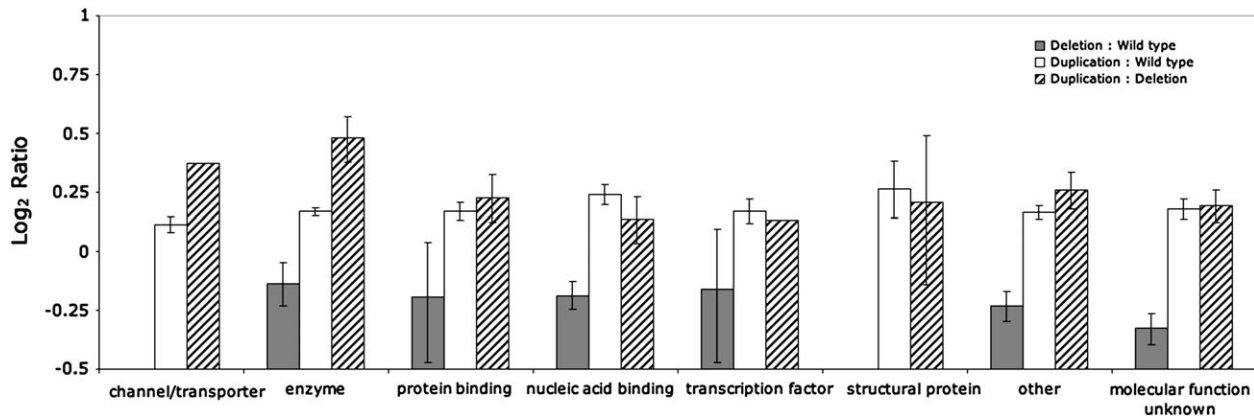


**FIG. 2.**—Distribution of logarithms of expression ratios. (A): this study. Deletion:wild type (red) and wild type:wild type (green). (B): Gupta et al. (2006) data. Duplication:wild type (red), Duplication:Deletion (gray), and wild type:wild type (green).

chromosomal aberrations, in a striking contrast to data recently reported in mammals (Williams et al. 2008) and yeast (Torres et al. 2007). This conclusion is based on the compar-

ison of Aberration:wild type expression ratios to diploid expression ratios and is, therefore, dependent on the assumption that chromosomal aberrations as such have no





**Fig. 3.**—Expression ratio in heterozygotes for deficiencies relative to wild type (dark bars, this study), heterozygotes for a duplication relative to wild type (white bars, Gupta et al. 2006), and in heterozygotes for duplication relative to heterozygotes for a deficiency (gray bars, Gupta et al. 2006), by molecular function. Error bars are standard errors reflecting variation among genes within molecular function category. One-way analysis of variance results for the three data sets were:  $F = 0.64$ ;  $P > 0.69$ ;  $F = 0.48$ ,  $P > 0.87$ , and  $F = 1.06$ ,  $P > 0.4034$  for this study, Gupta et al. (2006), duplication versus wild type and Gupta et al. (2006), duplication versus deletion, respectively.

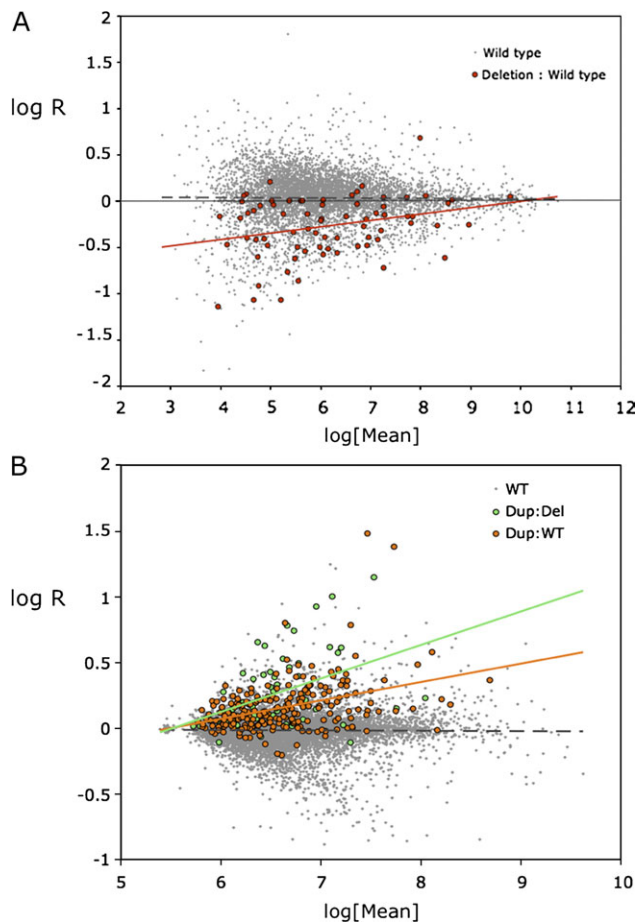
effect on the expression of diploid genes. With that caveat in mind, we also demonstrate that the degree of such compensation does not depend on the molecular function of the coded protein. Stenberg et al. (2009) reported that Aberration:wild type expression ratios are distributed normally, indicating a universal (or at least chromosome wide or segment wide) rather than gene-specific compensatory mechanism. Here we demonstrate that for in both deletions studied the distribution of Deletion:wild type ratios was significantly deviant from normal, with a strong suggestion of a shoulder around  $\log_2 R = -1$ , which suggest that some genes are stronger compensated than others, perhaps through the action of at least some gene-specific regulatory mechanisms. On the other hand, higher level of transcriptional compensation in tissue-specific genes than in ubiquitously expressed genes (see below) was observed for one deletion but not for the other, suggesting that some chromosomal segment-specific regulation may be taking place.

Furthermore, in contrast to Stenberg et al. (2009), we observe a correlation between the degree of compensation and the overall expression level of a gene. Genes with higher expression level are more likely to be compensated for a deletion but less likely to be compensated for a duplication. A positive correlation between transcriptional compensation of a deletion and overall expression level (fig. 4A) can be explained by the existence of genes, which are constantly expressed at the highest possible level, limited by gene-specific factor such as the availability of RNA polymerase complexes or individual transcription factors. However, this explanation is incompatible with the observed positive correlation between compensation for duplication and expression level (fig. 4B): transcriptional limitation of highly expressed genes would result in this correlation having a negative not positive sign. We hypothesize that this find-

ing indicates that highly expressed genes are equipped with a regulatory feedback mechanism more efficient in preventing underexpression than in preventing overexpression. In contrast, genes with low overall expression are more efficiently regulated to prevent overexpression in case of an overdose than to prevent underexpression in case of half the normal dosage.

We hypothesize that the relationship between the degree of transcriptional dosage compensation and overall gene expression may be widespread. Although Stenberg et al. (2009) reported such relationship for only 1 comparison out of 4, a careful examination of their figure S4A, suggests that at least one other comparison may be approaching significance, demonstrating the same pattern we observed in this study: highly expressed genes are strongly compensated for the deficiency, perhaps with a hint of a nonmonotonic relationship (see also fig. S2 in Stenberg et al. 2009). It might be suggested that the reason why in Stenberg et al. (2009) did not observe this correlation, despite using very similar deletion lines, is that in this study we used a more stringent criterion to exclude genes with expression rate near or below detection level (and, therefore, meaningless expression ratios) by excluding genes with transcript signal not statistically different from the AutoBlank signal across arrays. Meanwhile, in Stenberg et al. (2009) study, an arbitrary cut-off was used, possibly leaving in some genes with low signal-to-noise ratio. This would account both for the lack of a significant correlation with the expression level and for the apparent nonmonotonic relationship between expression ratio in deficiencies and expression level (fig. 4A and C in Stenberg et al. 2009).

We also partly confirm the finding of Stenberg et al. (2009) that ubiquitous genes are less effectively compensated for the dosage deficiency. The fact that this



**FIG. 4.**—Correlation between logarithm of mean expression level and logarithm of Aberration: wild type expression ratio. (A): this study (red circles: Deletion:wild type). (B): Gupta et al. (2006) data (orange circles: Duplication: wild type; green circles: Duplication:Deletion). Small circles are control (diploid) genes on both A and B.  $\log_2 R = 0$  corresponds to full dosage compensation.

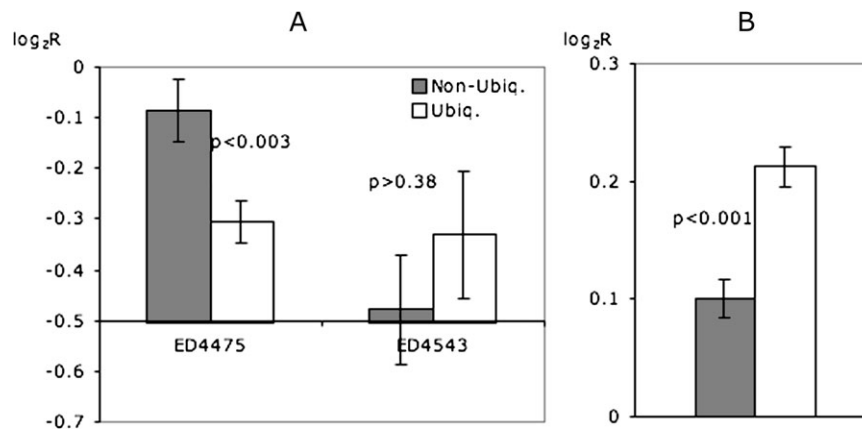
relationship is observed in one deletion but not the other, corroborates Stenberg et al. (2009) idea that such compensatory mechanisms may be chromosome region specific or chromosomal region specific. Observe that the correlation with gene expression level is not confounded with this result: ubiquitously expressed genes tend to have a higher expression level (simply due to the way they are defined for this analysis), so if the correlations with expression level were artifacts of confounding with the ubiquitousness of gene expression, it would have been expected to the opposite of what is actually observed. Stenberg et al. (2009) suggest that transcription of ubiquitously expressed genes tend to be limited by copy number. Although consistent with the patterns observed for deletions in this study and by Stenberg et al. (2009), this explanation is inconsistent with higher degree of compensation of ubiquitously expressed genes: if these genes were universally copy number limited, one

would expect the transcription level to be directly proportional to the number of copies of the gene.

Rather, we hypothesize that the observed patterns can be explained by the existence of a continuum of genes with respect to most general types of regulatory mechanisms, genes on one end of the distribution being compensated more efficiently for underexpression than for overexpression and with the pattern reversed on the opposite end of the continuum. The former types of genes tend to: 1) have high overall expression or be expressed in a tissue-specific manner, 2) demonstrate a stronger dosage compensation for a deficiency, and 3) demonstrate little dosage compensation for duplication. Genes of the second type tend to have low overall expression or be expressed ubiquitously, demonstrate little compensation for deficiency and greater compensation for duplication. At present it is impossible to differentiate between two possibilities: 1) highly expressed genes are equipped with a regulatory mechanism more efficiently preventing underexpression in order to maintain the required high level of expression and 2) a regulatory mechanism more efficient in preventing underexpression than overexpression causes the cognate gene to have a higher than average baseline expression level. One consideration speaks in favor of the former hypothesis: highly expressed genes are known to evolve slower (Pál et al. 2001; Drummond and Wilke 2008), which indicates that essential or household genes are overrepresented among genes with high expression level. One might speculate that household genes are more likely than nonessential genes to evolve a regulatory mechanism maintaining a necessary minimal level of transcription but more permitting of overexpression.

Likewise, it remains unknown whether the observed compensation patterns are a manifestation of some general regulatory mechanisms (whether gene-specific or chromosome-wide) capable of detecting copy number imbalance and predating the individual aberrations, or the result of a recent and fast adaptation to compensate for the aberration, which occurred in these particular lines since the introduction of the chromosomal abnormality. Some of the lines used in these studies (such as Df(2L)JH in Gupta et al. 2006; Stenberg et al. 2009) are fairly old and had an ample opportunity to evolve; others, as DrosDel lines in Stenberg et al. (2009) and this study are younger but have also been maintained in a balanced hemizygous state for some time.

One observation, namely the comparison of expression rate in heterozygotes for a duplication to that in heterozygotes for a deletion in Gupta et al. (2006) experiment, is puzzling. If transcriptional compensation of genes in deletion Df(2L)JH is characterized by the same patterns observed deletions in our experiment, then the expected correlation between expression ratio Duplication:Deletion and overall gene expression should be negative. Yet, this regression has a positive slope, in fact, higher than the regression of



**Fig. 5.**—Strength of transcriptional compensation in tissue-specific and ubiquitously expressed genes in two deletions (A, this study) and in a duplication (B, Gupta et al. 2006).

Duplication: Wild type ratios. Either the pattern observed in our data is region or chromosome specific, or the deletion Df(2L)JH (which contains only 59 genes) happens to be enriched with genes fully compensated despite their low expression level.

High degree of transcriptional compensation in heterozygotes for a deletion suggests that recessivity of most loss-of-function mutations in *Drosophila* can be explained by transcriptional compensation. This implies that relatively rare dominant mutant alleles either are not compensated at transcription or are gain-of-function mutations. There are only three genes with known dominant mutant alleles located within studied deletions (*Dichaete*, *frizzled*, and *breathless*), and none of them show detectable expression, so we cannot test the hypothesis that these genes are less likely to be compensated than genes with fully recessive mutations. It might be noted, however, that the line ED4543, which is hemizygous for *Dichaete*, does not exhibit the dominant phenotype of classic *Dichaete* alleles (extended and elevated wings), indicating that the classic alleles may be of gain-of-function type.

We hypothesize that dominance of the wild-type allele caused by transcription-level compensation is a by-product of the regulatory mechanisms whose purpose is to maintain the expression level to meet changing environmental or developmental conditions rather than a direct result of selection to compensate for mutant alleles. This hypothesis is consistent with the theoretical prediction that selection to compensate for mutations is weak, whereas selection to maintain the optimal gene expression is strong (Hurst and Randerson 2000) and with the observation that mammalian genes possess abundant variation for such optimization (Rockman and Wray 2002). It is also consistent with the increased frequency of codominance of deleterious alleles observed in genes whose products are involved in protein–protein interactions (Papp et al. 2003). Such interactions require a balance between expression levels of all genes in

a group of interacting genes, which imposes constraints on the evolution of regulation of individual genes, resulting in lower opportunity for transcriptional compensation.

Because highly expressed genes demonstrate a more complete compensation for deletions, we predict that transcriptional compensation-based dominance of the wild-type alleles should be more common in highly expressed genes, whereas dominant mutations are more likely in genes with low overall expression. Moreover, we can hypothesize that transcriptional-level dominance can be of two types: in genes with high expression loss-of-function mutations are compensated at transcription, whereas in genes with low expression high levels of gene products are simply not necessary, that is, haploinsufficiency is unlikely.

We also found that, unlike compensation on the protein level, transcriptional compensation appears to be independent of protein function. We do not see any evidence of greater transcriptional compensation of enzyme-coding genes than regulatory genes coding for transcription factors and nucleic acid-binding proteins. It is hard to imagine that all genes for transcription factors are regulated at transcription, because it implies an endless pyramid of transcription factors for transcription factors and leads to low fidelity of regulation (Itzkovitz et al. 2006). In addition, not every transcription regulation mechanism will automatically compensate for mutant alleles. For this to occur, the regulatory mechanism must be based on a negative feedback detecting abundance or activity of the gene product. Positive regulatory mechanisms, for example those, which detect a particular environmental variable, independent from the gene product will not result in transcriptional compensation of mutations. It should be therefore possible to test the hypothesis that transcriptional compensation is a by-product of evolution of negative feedback regulation of transcription by measuring transcription level in genes known to respond to environmental cues and in genes known to respond to abundance or activity of their own products. We predict that



the first group will demonstrate lower transcriptional compensation of mutations than the second one.

## Acknowledgments

We are grateful to Kaiguo Mo and Tim Westwood for the help with microarray methodology and to Eugene Koonin, Alex Kondrashov, Brian Oliver, Jan Larson, Mike Whitlock, and Jim Fry for helpful discussions of the manuscript. This study was supported by East Tennessee State University Student–Faculty Collaboration grant and NSF-0525447.

## Literature Cited

- Baek ST, et al. 2008. Genome-wide identification of haploinsufficiency in fission yeast. *J Microbiol Biotechnol.* 18:1059–1063.
- Baker BS, Gorman M, Marin I. 1994. Dosage compensation in *Drosophila*. *Annu Rev Genet.* 28:491–521.
- Birchler JA, Hiebert JC, Paigen K. 1990. Analysis of autosomal dosage compensation involving the alcohol dehydrogenase locus in *Drosophila melanogaster*. *Genetics.* 124:679–686.
- Charlesworth B. 1979. Evidence against Fisher's theory of dominance. *Nature.* 278:848–849.
- Chintapalli VR, Wang J, Dow JA. 2007. Using FlyAtlas to identify better *Drosophila melanogaster* models of human disease. *Nat Genet.* 39:715–720.
- Crow JF. 1979. Minor viability mutants in *Drosophila*. *Genetics.* 92:s165–s172.
- Crow JF, Simmons MJ. 1983. The mutation load in *Drosophila*. In: Ashburner M, Carson HL, Thompson JN Jr., editors. *The genetics and biology of Drosophila*. London: Academic Press. Vol. 3c, pp. 1–35.
- Devlin RH, Holm DG, Grigliatti TA. 1982. Autosomal dosage compensation *Drosophila melanogaster* strains trisomic for the left arm of chromosome 2. *Proc Natl Acad Sci U S A.* 79:1200–1204.
- Drummond DA, Wilke CO. 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell.* 134:341–352.
- Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30:207–210.
- Fisher E, Scambler P. 1994. Human haploinsufficiency—one for sorrow, two for joy. *Nat Genet.* 7:5–7.
- Fisher RA. 1928. The possible modification of the response of the wild type to recurrent mutations. *Am Nat.* 62:115–126.
- FitzPatrick DR, et al. 2002. Transcriptome analysis of human autosomal trisomy. *Hum Mol Genet.* 11:3249–3256.
- Gupta V, et al. 2006. Global analysis of X-chromosome dosage compensation. *J Biol.* 5:3.
- Haldane JBS. 1939. The theory of the evolution of dominance. *J Genet.* 37:365–374.
- Hurst LD, Randerson JP. 2000. Dosage, deletions and dominance: simple models of the evolution of gene expression. *J Theor Biol.* 205:641–647.
- Iitzkovitz S, Tlusty T, Alon U. 2006. Coding limits on the number of transcription factors. *BMC Genomics.* 7:239.
- Kacser H, Burns JA. 1981. The molecular basis of dominance. *Genetics.* 97:639–666.
- Kondrashov FA, Koonin EV. 2004. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. *Trends Genet.* 20:287–291.
- Lee TI, et al. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science.* 298:799–804.
- Lyon MF. 1988. X-chromosome inactivation and the location and expression of X-linked genes. *Am J Human Genet.* 42:8–16.
- Neal SJ, Gibson ML, So AK, Westwood JT. 2003. Construction of a cDNA-based microarray for *Drosophila melanogaster*: a comparison of gene transcription profiles from SL2 and Kc167 cells. *Genome.* 46:879–892.
- Orr AH. 1991. A test of Fisher's theory of dominance. *Proc Natl Acad Sci U S A.* 88:11413–11415.
- Pál C, Papp B, Hurst LD. 2001. Highly expressed genes in yeast evolve slowly. *Genetics.* 158:927–931.
- Papp B, Pál C, Hurst LD. 2003. Dosage sensitivity and the evolution of gene families in yeast. *Nature.* 424:194–197.
- Phadnis N, Fry JD. 2005. Widespread correlations between dominance and homozygous effects of mutations: implications for theories of dominance. *Genetics.* 171:385–392.
- Rockman MV, Wray GA. 2002. Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol.* 19:1991–2004.
- Ryder E, et al. 2007. The DrosDel deletion collection: a *Drosophila* genomewide chromosomal deficiency resource. *Genetics.* 177:615–629.
- SAS Institute. 2007. JMP Genomics 3.0 for SAS. Cary (NC): SAS Institute.
- Stenberg P, et al. 2009. Buffering of segmental and chromosomal aneuploidies in *Drosophila melanogaster*. *PLoS Genet.* 5(5):e1000465.
- Straub T, Becker PB. 2007. Dosage compensation: the beginning and end of generalization. *Nat Rev Genet.* 8:47–57.
- Stuckenholz C, Kageyama Y, Kuroda MI. 1999. Guilt by association: non-coding RNAs, chromosome-specific proteins and dosage compensation in *Drosophila*. *Trends Genet.* 15:454–458.
- Torres EM, et al. 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science.* 317(5840):916–924.
- Tsafir D, et al. 2006. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res.* 66:2129–2137.
- Veitia RA. 2002. Exploring the etiology of haploinsufficiency. *BioEssays.* 24:175–184.
- Williams BR, et al. 2008. Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science.* 322(5902):703–709.
- Wright S. 1934. Physiological and evolutionary theories of dominance. *Am Nat.* 68:24–53.
- Zhang Y, Malone JH, Powell SK, et al. Forthcoming. 2010. Expression in aneuploid *Drosophila* S2 cells. *PLoS Biology.*