Genetical toxicogenomics in Drosophila identifies master-modulatory loci that are regulated by developmental exposure to lead

Douglas M. Ruden [a,*], Lang Chen [b], Debra Possidente [c], Bernard Possidente [d], Parsa Rasouli [a], Luan Wang [a], Xiangyi Lu [a], Mark D. Garfinkel [e], Helmut V.B. Hirsch [d], Grier P. Page [f]

[a] Institute of Environmental Health Sciences, Wayne State University, Detroit, MI, United States
[b] Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, United States
[c] Department of Biological Sciences, University at Albany, SUNY, Albany, NY, United States
[d] Department of Biology, Skidmore College, Saratoga Springs, NY, United States
[e] Department of Environmental Health Sciences, University of Alabama at Birmingham, Birmingham, AL, United States
[f] Statistics and Epidemiology Unit, RTI International, Atlanta, GA, United States

ARTICLE INFO

ABSTRACT

The genetics of gene expression in recombinant inbred lines (RILs) can be mapped as expression quantitative trait loci (eQTLs). So-called "genetical genomics" studies have identified locally acting eQTLs (cis-eQTLs) for genes that show differences in steady-state RNA levels. These studies have also identified distantly acting master-modulatory trans-eQTLs that regulate tens or hundreds of transcripts (hotspots or transbands). We expand on these studies by performing genetical genomics experiments in two environments in order to identify trans-eQTL that might be regulated by developmental exposure to the neurotoxin lead. Flies from each of 75 RIL were raised from eggs to adults on either control food (made with 250 μM sodium acetate), or lead-treated food (made with 250 μM lead acetate, PbAc). RNA expression analyses of whole adult male flies (5–10 days old) were performed with Affymetrix DrosII whole genome arrays (18,952 probesets). Among the 1389 genes with cis-eQTL, there were 405 genes unique to control flies and 544 genes unique to lead-treated ones (440 genes had the same cis-eQTLs in both samples). There are 2396 genes with trans-eQTL which mapped to 12 major transbands with greater than 95 genes. Permutation analyses of the strain labels but not the expression data suggests that the total number of eQTL and the number of transbands are more important criteria for validation than the size of the transband. Two transbands, one located on the 2nd chromosome and one on the 3rd chromosome, co-regulate 33 lead-induced genes, many of which are involved in neurodevelopmental processes. For these 33 genes, rather than allelic variation at one locus exerting differential effects in two environments, we found that variation at two different loci are required for optimal effects on lead-induced expression.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Removal of lead from paint and gasoline has substantially reduced blood lead levels in the United States, but lead exposure still remains a serious public health problem (White et al., 2007). Even if blood lead levels in children are below the current community action level (10 μg/dl), chronic lead exposure can impair cognitive function (e.g., Counter et al., 2000; Jusko et al., 2008; Kordas et al., 2006; Min et al., 2007; Needham et al., 2005; Stewart et al., 2006; Surkan et al., 2007; White et al., 2007) and increase the likelihood of delinquency (Dietrich et al., 2001).

Chronic lead exposure has developmental and behavioral effects in both vertebrate and invertebrate animals. Among its effects are changes in motor activity levels in rhesus monkeys (Lasky and Laughlin, 2001), rats (Tang et al., 1994) and Drosophila (Hirsch et al., 2003) and impairments in learning and memory in tadpoles (Strickler-Shaw and Taylor, 1991), mice (Sun et al., 2005) and rats (Alber and Strupp, 1996; Garavan et al., 2000; Jett et al., 1997; Moreira et al., 2001; Morgan et al., 2000). Despite considerable progress in understanding the physiological mechanisms of how lead affects neuronal plasticity during development

* Corresponding author at: Institute for Environmental Health Sciences, Wayne State University, 2727 Second Avenue, Room 4000, Detroit, MI 48201-2654, United States.
   E-mail addresses: douglasr@wayne.edu, douglas.ruden@gmail.com (D.M. Ruden).

(Chang et al., 2008; Marchetti, 2003; Toscano and Guilarte, 2005) and in the adult (Gilbert et al., 2005; Schneider and Decamp, 2007; Verina et al., 2007; White et al., 2007), the genetic mechanisms lead disturbs are not well understood.

We thus decided to develop *Drosophila melanogaster* as a model for study of the behavioral, physiological and genetic effects of chronic lead exposure during development. *Drosophila* take up lead from the nutrient medium they are reared on (Cohn et al., 1992) and the dose response curve for the lead body burden has a significant linear component (Hirsch et al., 2003). As is the case for vertebrates, lead is toxic in *Drosophila* (larval LC50 ($\pm$standard error): 6.60 ± 0.64 mM PbAc added to medium) and delays pre-adult development (lowest concentration extending larval development, 1.2 mM PbAc) without changing adult body weight (Akins et al., 1992). Lead-dependent behavioral changes have been harder to demonstrate in *Drosophila*. An early study reported that, even at a high dose (3.07 mM PbAc), phototaxis, locomotion and learning are unaffected (Akins et al., 1992). In sharp contrast, at very much lower doses courtship behavior (which is influenced by experience during the first few days of adult life; Barth et al., 1997b; Hirsch et al., 1995; Hirsch and Ghiradella, 2004) shows biphasic lead-dependent changes: facilitation at low (10 and 40 μM PbAc) doses and inhibition at higher (100 and 250 μM PbAc) ones (Hirsch et al., 2003). Locomotor activity is also affected, but only at higher (250 μM PbAc) doses, and then only for males (Hirsch et al., 2003). Furthermore, as is the case for mammals (Chen et al., 2007; Cooper and Manalis, 1983; Cooper et al., 1984; Kiraly and Jones, 1982; Ming and Song, 2005; Petit and LeBoutillier, 1979; Toscano and Guilarte, 2005; White et al., 2007), lead affects synaptic development in *Drosophila* (100 μM PbAc; Morley et al., 2003). Recent evidence suggests that these synaptic effects may involve lead-dependent changes in calcium regulation (He et al., 2009). Finally, QTL analysis using the recombinant inbred lines (RILs) identified a site for lead-dependent changes in locomotor activity (Hirsch et al., 2009).

Toxicogenomic or pharmacogenomic studies generally involve administering various doses of a toxin or drug to an organism or cell line, and then measuring changes in expression of all or nearly all genes using microarrays at either a single end point or as a time series (reviewed in Foster et al., 2007). If two compounds produce similar patterns of changed gene expression, they are said to have similar "fingerprints" suggesting that they function in a similar manner, for instance in a cell line (Coe et al., 2007) or in yeast (Perlstein et al., 2007). As the cost of whole genome microarrays and other technologies to measure global gene expression levels has dropped, use of these approaches has become widespread. However, because they often identify toxin-dependent changes in expression for hundreds or even thousands of genes, it is difficult if not impossible to take the next step and validate all the results through follow-up genetic or molecular studies.

What is needed is an approach that reduces to a more manageable level the number of candidate toxin-regulated genes. Fortunately a new multi-dimensional strategy, "genetical genomics," has recently been developed to identify master-modulatory loci that regulate the expression of dozens or hundreds of other genes (Broman, 2005; de Koning and Haley, 2005; Jansen and Nap, 2001; Li and Burmeister, 2005; Rockman and Kruglyak, 2006). Genetical genomics combines two methodologies: microarray-based whole transcriptome analyses and extracting quantitative trait loci (QTLs) using recombinant inbred lines (RILs) (see WebQTL at http://www.genenetwork.org for several examples of eQTL studies). Global gene expression levels are determined for each recombinant inbred line, and then QTL software (*e.g.*, R/QTL, Broman et al., 2003) is used to look for statistical association between specific chromosomal loci and transcript levels for all the genes. With this approach regulation of many genes in an organism is linked by the QTL analysis to specific chromosomal loci, termed expression QTLs (eQTLs). The eQTLs affecting a gene's steady-state mRNA levels can either be closely linked (*cis*-eQTLs) or unlinked, even on a different chromosome (*trans*-eQTLs) (Broman, 2005).

Drosophila has a number of features that are desirable in genetical genomic studies using multicellular organisms. First, the *Drosophila* genome ($\sim$125 Mbp; $\sim$662 cM) is about one tenth the size of the rat or mouse genome, which would simplify re-sequencing large numbers of strains to identify quantitative trait nucleotides, and there is a wide array of genetic tools that allow one to quickly map and knock out the genes identified as eQTL peaks in flies (Page and Ruden, 2005). Also *Drosophila* recombinant inbred lines (RILs) are available that were made by mapping roo transposons in recombinants of two highly divergent parental strains, *Oregon R* (*ORE*) and *Russian 2b* (*2B*) (Nuzhdin et al., 1993). While the *C. elegans* genome is even smaller, in *C. elegans* organs are often present as multi-nucleated syncytia. In addition, they lack some epithelial properties and signaling pathways found in *Drosophila* and vertebrates, including humans (Chisholm and Hardin, 2005; Podbilewicz, 2006). Also, the entire *C. elegans* nervous system consists of just 302 identifiable neurons with all neuronal connections known (*e.g.*, Ye et al., 2008). *Drosophila*, on the other hand, shares important features with mammals including epithelial properties and a central nervous system that more closely resembles those of mammals. For example, the *Drosophila* brain consists of $\sim$500,000 neurons (Bate and Martinez Arias, 1993) whose synaptic connections involve proteins that are highly conserved with those found in vertebrates (Koh et al., 2000) and, as in humans, the developing nervous system is affected by early experience (Barth et al., 1997a; Urbach and Technau, 2003).

We found that by using the new genetical toxicogenomic approach we could identify candidate regulatory loci underlying lead toxicity. Surprisingly, we identified a small number of *trans*-eQTLs, each of which apparently influences the expression of a large number of other genes in a lead-dependent manner. The eventual identification of lead-dependent master-modulatory genes will be profoundly important in deepening our understanding of the developmental and physiological consequences of lead exposure.

2. Results

2.1. Lead alters the regulation of cis-eQTLs and trans-eQTLs

We performed gene expression analyses of 18,952 probesets (using Affymetrix Dros2 gene expression arrays) in adult male *Drosophila* from 75 recombinant inbred lines (RILs) that were mapped with 82 polymorphic markers between the two parental strains, *Oregon R* (*ORE*) and *Russian 2B* (*2B*). The 18,952 probesets correspond to $\sim$14,000 genes that have been validated in the latest genome browser and $\sim$5000 candidate regions that have not been validated as expressed genes. To avoid confusion, unless we are discussing a specific gene or genes, we will refer to the expression of probesets rather than genes because not all probesets correspond to validated genes. Expression analyses were performed from each RIL in the presence or absence of developmental and adult exposure to 250 μM lead acetate (see Section 4; Supplementary Table S1 has the expression data for all for all 18,952 probesets for all 150 conditions).

We identified a total of 1389 different probesets with *cis*-eQTL (*i.e.*, within a 5 cM sliding window). There are 405 *cis*-eQTL in control flies only, 544 in treated flies only, and 440 in both conditions (Table 1 and Fig. 1a; Supplementary Table S2 has the complete *cis*-eQTL gene lists and data). The criteria for significant *cis*-eQTL are that the permutation LOD scores have a *P*-value of less than 0.0001 (this is the nominal *P*-value of the LOD threshold based on 1000 permutations of the strain labels), which corresponds to a

Table 1

Number of genes and GO categories in control and lead-treated *trans*-eQTL transbands. The 5 control (con) and 7 lead-treated (trt) *trans*-eQTL transbands and their approximate cytological locations are indicated in the first column. The second column shows the location of a transband by linkage group and centimorgans (cM) (1 = X chromosome (bands 1–20), 2 = 2L and part of 2R (bands 21–60), 3 = distal 2R (bands 61–80), and 4 (chromosome 3 (bands 81–100). GO, is the most significant gene ontological (GO) category for the genes in the transband. The false discovery rate (FDR) is shown in parentheses. # permutations with this GO category, is the number of permutations (out of 200) that have a transband with significant GO (FDR ≤ 0.01).

Transbands	cM	Number of genes	GO (FDR)	# permutations with this GO category
con_27B	2_46	117	Cell Cycle (0.01)	0/200
con_50DF	2_161	152	Alternative splicing (4.41E−06)	19/200
con_70C	4_136	103	None	
con_72A	4_141	134	Hydrolase (3.87E−04)	17/200
con_73D	4_146	278	Hydrolase (1.43E−26)	17/200
trt_3E	1_11	245	Developmental Protein (5.68E−03)	15/200
trt_30AB	2_71	194	Actin Binding (4.80E−04)	0/200
trt_57F	3_6	96	Golgi apparatus (0.01)	0/200
trt_63A	4_21	140	Mitochondrial part (0.0022)	14/200
trt_65A	4_31	98	None	
trt_73D	4_146	222	Hydrolase (7.65E−12)	17/200
trt_77E	4_156	103	Carbohydrate metabolic process (3.45E−15)	2/200

false discovery rate (FDR) of about 4% using the Benjamini-Hochberg method. Permutation LOD scores are calculated for each of the 18,952 probesets independently. The strain labels for the microarray data are permuted 1000 times and LOD scores are calculated for each marker interval for each probeset. Some QTL studies use a LOD cutoff of 3–3.5 to indicate a "significant" QTL peak, but the permutation method for calculating the LOD scores is more robust because it depends on the *P*-values which are calculated for each probeset independently. Consequently, for one probeset, a LOD score of 2.8 at one interval could have a *P*-value of less than 0.0001, whereas for a second probeset a LOD score of 3.8 at some other interval is needed to reach this threshold. The idea behind permutation LOD score methodology is that the 40 highest LOD scores (out of 1000 permutations) for a QTL at a particular interval represents the critical value that provides an FDR = 0.04 (4%), which is a measure of the experiment-wide type I errors (*i.e.*, false positives). In other words, if there is a real relationship between a marker and a QTL, such as an eQTL, then only 4% of the permutations will have a higher LOD score at a particular interval than the experimental data. Permutation analyses have to be conducted to determine the LOD scores with this method because, since each probeset has a different result after 1000 permutations,

there is no mathematical method to determine permutation LOD scores directly for each probeset.

To further improve the data, we also eliminated potential "false eQTLs" that might be caused by single nucleotide polymorphisms at Affymetrix probe sites (Fig. 1c; Supplementary Table S2). Our method removes false eQTLs in genes that have one or more outlier probes in a probeset, presumably because the outliers correspond to probes that contain a SNP (as described in Section 4.10). As expected, a much larger percentage of the *cis*-eQTLs are false by this criteria compared with *trans*-eQTLs (Fig. 1c). Since the Affymetrix DrosII gene array that we used includes 18,952 probesets, the 1389 *cis*-eQTL represent ~7% of the probesets.

We identified a total of 2396 probesets with *trans*-eQTLs (~13% of the probesets), 948 in control flies only, 1191 in lead-treated flies only, and 357 in both control and lead-treated flies (Fig. 1b; Table 1). As with *cis*-eQTLs, the criteria for significant *trans*-eQTLs are *P*-values of <0.0001 which corresponds to a FDR of about 4% (see Supplementary Table S3 for complete *trans*-eQTL probeset lists and data). This is consistent with published studies in other organisms. The mean heritability for steady-state mRNA levels is 38.7% without treatment effects, and heritability increased to 40.6% when treatment effects are considered (Fig. 2a and b). Heritability is the proportion of phenotypic variation (in our case, steady-state mRNA levels) in a population that is attributable to genetic variation among individuals, and "mean heritability" is the average heritability for all of the probesets. Heritability was calculated as described in Section 4.8. When all of the *cis*- and *trans*-eQTLs were analyzed, there are more eQTL with "significant" *P*-values ($P < 0.05$) than expected at random (Fig. 2c). A random event would have equal frequencies across the *P*-value range (*i.e.*, the bars would all be the same height), or possibly a dip in the number of *P*-values in the significant range ($P < 0.05$).
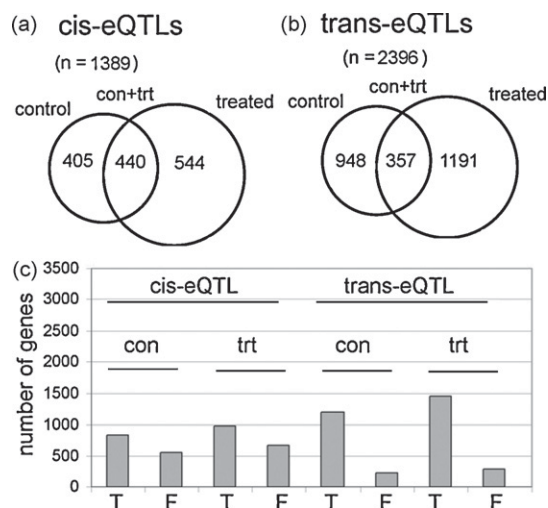
2.2. Trans-eQTLs identify transbands (hotspots) for lead-induced changes in gene expression

Our genetical toxicogenomic study identified 12 significant *trans*-eQTL "transbands" (5 in control samples and 7 in treated samples) at 11 different loci (the transband at 4_146 cM (73D) is located in both samples), each containing between 96 and 278 probesets (Table 1). A "transband" or "hotspot" is a group of *trans*-eQTLs that lie in a nearly vertical line (*i.e.*, within a 5 cM window) in a "*cis-trans*-plot" that plots eQTL location on the *x*-axis and gene location on the *y*-axis (Fig. 3a). The nearly diagonal line corresponds to *cis*-eQTLs because these eQTLs coincide (*i.e.*, within a 5 cM window) with the gene locations. Transbands are often interpreted as being a group of genes that are co-regulated by a

Fig. 1. Control and lead-treated *cis*-eQTLs and *trans*-eQTLs. (a) Venn diagrams of numbers of genes with *cis*-eQTLs and (b) *trans*-eQTLs in each category (con only, con + trt only, and treated only). Detailed information is in Supplementary Tables S2 and S3. (c) Graph of the number of true (T) and false (F) *cis*-eQTL based on outlier analyses of the 14 probes in each of the 18,952 probesets. Notice that *cis*-eQTL have a higher percentage of false eQTL than do *trans*-eQTL.
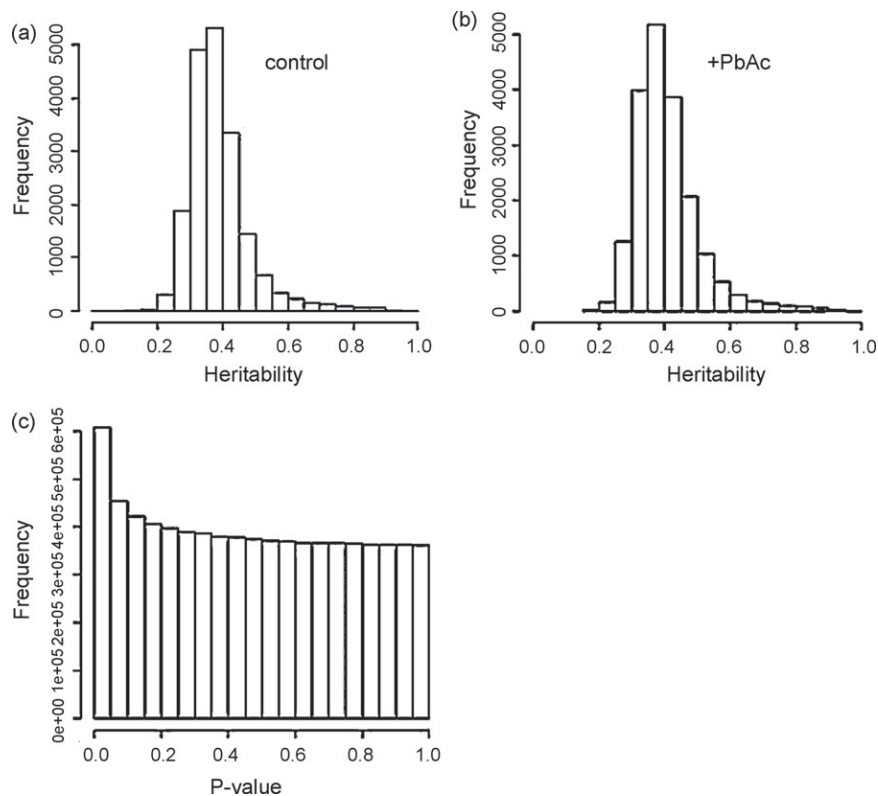
Fig. 2. Heritability and *P*-values for eQTL data. (a) Heritability estimates without factoring in treatment (environment) effects. (b) Heritability estimates with treatment effects factored in. (c) *P*-values for all markers (82 markers), and every 5 cM between markers by all expression traits. As can be seen there are more 'significant' ($P < 0.05$) than expected at random.

master-modulatory gene that is encoded at the genomic location of the transband. The concept is that a DNA sequence polymorphism in a master-modulatory gene (*e.g.*, a deleterious SNP that decreases the activity of a transcriptional activator, such as by a missense mutation in the activating domain or a promoter mutation that decreases the expression of the master-modulatory gene) would cause all of the genes in a transband to be coordinately expressed at a low level when the master-modulatory protein has one genotype (*i.e.*, it has the 'bad' SNP), but coordinately expressed at a high level when the master-modulatory protein has the other genotype (*i.e.*, it has the 'good' SNP).

The size of the transbands was determined based on the number of probesets with significant *trans*-eQTL linkage in a 5 cM window in the experimental data. To find the minimum size of a transband, we determined that 96 probesets in a 5 cM window corresponds to $P < 0.05$, chi-square test, so this was used as the minimum size. As shown in a "*cis-trans*-plot" of the eQTL data, untreated males have significant control *trans*-eQTL transbands at chromosomal regions 27B, 50DF, 70C, 72A and 73D (Fig. 3a (red); Supplementary Table S3). In contrast, males reared in lead-treated food have significant *trans*-eQTL transbands at chromosomal regions 3E, 30AB, 57F, 63A, 65A, 73D and 77E (Fig. 3a (black); Supplementary Table S3). This result suggests that the *trans*-regulation of most probesets in the *trans*-eQTL transbands (with the possible exception of 73D) differ between control and lead-treated flies. As is standard in *Drosophila*, unless otherwise indicated, we will refer to the transbands by their approximate cytological location rather than by cM linkage (*e.g.*, 4_146, where 4 is the linkage group and 146 is in cM, corresponds roughly to 73D). As a technical note, chromosome 2 is split into two linkage groups (2 and 3) because of unlinked roo elements on the right arm of chromosome 2. The other chromosomes (X, 3, and 4) are single linkage groups.

A possible explanation for the different genetic locations of most of the transbands in the two environments is that the transbands are correlation artifacts amongst co-expressed genes and are therefore false transbands (Breitling et al., 2008). *In silico* simulation studies have shown that false *trans*-eQTLs often cluster randomly to a small number of transbands because of artifacts in the architecture of the RILs (Perez-Enciso, 2004). In order to determine whether the transbands in Table 1 are true or false, we did 100 permutations of the strain labels of the 75 RILs of the data from the control-exposed flies (Fig. 4a and b) and another 100 permutations with the data from the lead-treated flies (Supplementary Fig. 4C, D and Table S4) (Breitling et al., 2008). For each permutation, we analyzed: (1) the number of transbands, (2) the number of eQTLs, (3) the number of probesets that are the same (*i.e.*, they match) in the control and lead-treated transbands, and (4) the size (total number of probesets with eQTLs) in all of the transbands. To be consistent with the experimental data, significant *in silico* permutation transbands have at least 96 probesets with an eQTL in a 5 cM window (*i.e.*, 1–4.99 cM, 5–9.99 cM, etc.). The results are as follows:

(1) The number of transbands was significantly larger for the experimental data (which we sometimes call real data) compared with the *in silico* permutation data (which we call "permutation data") ($P < 0.05$, chi-squared test). Only 6 of the 200 permutations have greater than 4 transbands (3%), whereas there are 5 control transbands and 7 treated transbands in the experimental data. The largest number of transbands in the permutation data was 7, and this occurred once (Fig. 4e). When the numbers of transbands in the control and treated datasets were added, there were 12 transbands in the real data but the largest number in the combined 200 permutations was 8 transbands (Supplementary Table S4).
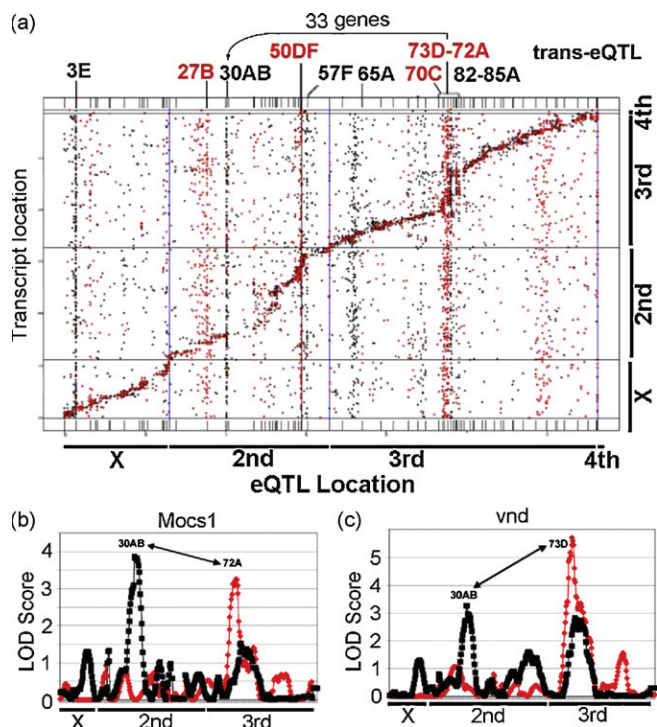
Fig. 3. Results of a genetical genomics study of control and lead-treated from 75 RIL *Drosophila* lines. (a) Locations of the eQTLs are represented on the *x*-axis and locations of transcripts (based on their chromosomal locations) are represented on the *y*-axis. eQTL locations for *cis* regulated genes every 5 cM are present along the diagonal. Notice the control *trans*-eQTL transbands at 27B, 50DF, 70C, and 73D–72A. Also, notice the lead-induced transbands at 3E, 30AB, 57F, 65A, and 82D–85A. The genetic and physical maps of the lines are not perfectly aligned along the diagonal, due to regions with greatly reduced recombination in certain regions (bands 50–57 and 75–85) in the RILs compared to traditional crosses. It is not clear why these regions had reduced recombination. (b) Graphs of eQTL analysis of *Mocs1*, which is co-regulated by *trans*-eQTL at 30AB and 72A, and (c) *vnd*, which is co-regulated by *trans*-eQTL at 30AB and 73D. The LOD Score is the log of the probability distribution, as described in Section 4. *Red*, eQTL of genes from control flies. *Black*, eQTL of genes from lead-treated flies.

(2) The number of eQTLs was significantly higher for the real data compared with the permutation data ($P < 0.05$, chi-square test). Only 2 of the 200 permutations have greater than 2780 eQTLs, whereas there are 2780 control eQTLs and 3350 treated eQTLs in the real data. The false eQTLs were included in these analyses (*i.e.*, the eQTLs are for probesets that have one or more outlier probes), but the real data and the permutation data have the same frequency of false eQTLs, so this probably will not affect the results. The largest two permutation transbands contain 4040 and 2921 eQTLs (Fig. 4f). When adding the number of eQTL in the control and treated datasets, there were 6130 eQTL in the real data, but the largest number in the combined 200 permutations was 4658 (Supplementary Table S4).

(3) The total number of matching probesets with eQTLs in control and treated transbands was significantly higher in the real data compared with the permutation data ($P < 0.01$, chi-square test). There were a total of 225 matching probesets with eQTLs in the real data transbands (control versus treated), but the largest number of matching probesets with eQTLs in the permutation data transbands was 58 (control permutation 1 was compared with treated permutation 1, etc.). In fact, only 22 of the 100 permutation datasets had greater than one matching probeset with eQTLs in the transbands, and the total number of matching probesets with eQTLs in the combined 200 permutations datasets was only 223 (Supplementary Table S5). We

analyzed the matching probesets with eQTLs in the real data more carefully and we identified 9 pairs of transbands that had 3 or more matching probesets with eQTLs in the transbands (Table 2). The largest match in the real data had 83 probesets with eQTLs which are regulated by the same transband at chromosome region 73D (4-146_73D) in both control and treated samples (Table 3). The *same transband* regulating one or more genes only occurred one time in the permutation dataset—one permutation had 11 genes regulated by the same transband in both control and lead-treated datasets (Supplementary Table S5).

The number of matching probesets with eQTLs in *two different transbands* (in control versus treated) was also significantly higher in the real data compared with the permutation data ($P < 0.05$, chi-square test). The largest category of matching probesets with eQTLs in the permutation data was in transbands that were separated by greater than 5 cM on the same chromosome or on different chromosomes. In this category, one permutation had 49 probesets with eQTLs, a second permutation had 26 probesets with eQTLs, and a third permutation had 18 probesets with eQTLs in two different transbands in the control and lead-treated permutation datasets. All of the remaining permutations had 13 or fewer probesets in two different transbands (Supplementary Table S5). This is in comparison to the 33 probests with eQTLs (corresponding to 33 genes) in the real data that were in the 30AB transband on chromosome 2 in the lead-treated dataset and in 2 adjacent transbands (72A and 73D) on chromosome 3 in the control dataset (Table 3). These 33 "co-expressed" genes will be discussed further in a later section.

(4) The size of the transbands was not significantly higher for the real data compared with the permutation data (Fig. 4g and h). There were over 20 transbands in the 200 permutations that were larger than the cutoff of 96 probesets with eQTLs. There would have to be over 800 genes with eQTLs in a transband to be a significant excess, but the largest transband in the real data was only 278 genes with eQTLs (Fig. 4g and h; Table 1).

The fourth conclusion, that the transband size in the real data is not significantly higher than the permutation data, was not unexpected because false *trans*-eQTLs often cluster randomly to a small number of transbands (Perez-Enciso, 2004). Nevertheless, the significance of the first three analyses (the total number of transbands, the total number of probesets with eQTLs, and the total number of matching probesets with eQTLs) suggest that our other approaches (1–3) might be valid for identifying eQTLs, transbands, and co-regulated genes in the experimental data (see Section 3).

2.3. Some transband probesets are polarized for expression based on genotype and/or treatment

In the *cis*-eQTL analysis, we eliminated false *cis*-eQTL artifacts that are caused by probe(s) in a probeset that poorly hybridize to the sample because of SNPs or other sequence mismatches. If the false *cis*-eQTLs are not eliminated, there would be an apparent polarization of expression (steady-state mRNA levels) based on genotype if one of the two parental strains was used to make the expression array (Chesler et al., 2005). However in studies such as ours where neither parental strain has been sequenced, there should be no polarity for expression based on genotype because there are likely an approximately equal number of SNPs in *ORE* and *2B* compared with the reference strain (y^1; cn^1 bw^1 sp^1) (Adams et al., 2000). Fortunately *trans*-eQTLs are less susceptible to this SNP-based artifact because, by definition, there are no SNPs that map to the probesets in a transband (Chesler et al., 2005).
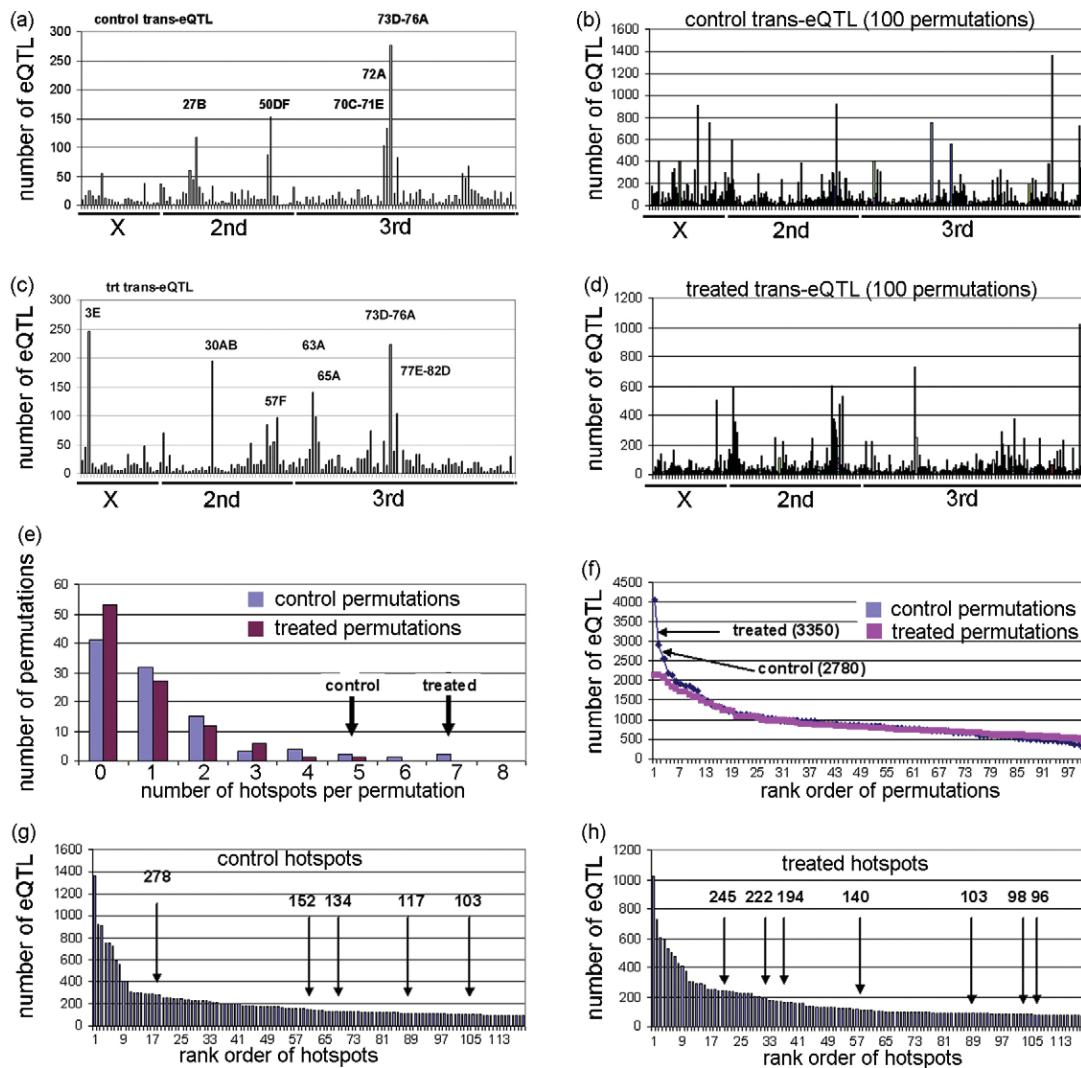
Fig. 4. Permutation analysis of control and treated eQTL data. (a and c) Plot of control and treated eQTL data showing genome location (x-axis) and the number of eQTL (y-axis). The 5 control and 7 treated eQTL transbands are indicated. (b and d) Plot of 100 control and 100 treated permutation analyses in which the RIL labels were randomized. (e) Plot of the number of transbands per permutation (x-axis) and the number of permutations with the corresponding number of transbands (y-axis). The experimental control and treated transband numbers are indicated (arrows). (f) Plot of the rank order of permutations (x-axis) versus the number of eQTL in the transband (y-axis). The number of eQTL in the control and treated transbands from the experimental data are indicated (arrows). (g and h) Plots of rank order of transbands (x-axis) versus the number of eQTL in the corresponding transband (y-axis). The control and treated transbands are indicated (arrows).

Therefore polarization of expression with the genotype of genes in a transband as a dependent variable (which we call "polarization of expression based on genotype") might indicate a common regulator for a majority of genes in a transband.

We calculated the polarity of expression based on both genotype and environment (±Pb). In these analyses, we found that expression of probesets with trans-eQTL overall show no polarity of expression based on genotype (data not shown). In other words, of the 2396 total trans-eQTLs (see Fig. 1), the number of probesets that have higher steady-state mRNA levels when the genotype is ORE is approximately the same as when the genotype is 2B. However the probesets in individual transbands often show such polarity, suggesting that steady-state mRNA levels are regulated by genes that are encoded at the transband loci. For example, if a transcription activator at a transband locus is expressed at a low level in 2B but at a high level in ORE, then there should be a polarity of expression where ORE relative expression is higher than 2B relative expression for all of the genes regulated by the activator in the transband.

We analyzed polarity of expression based on genotype in the 12 transbands by determining the ratio of the expression level of a probeset with one genotype over the expression level of the same probeset with the other genotype with the same environmental treatment. Specifically we determined the ratio (ORE − Pb)/(2B − Pb) and (ORE + Pb)/(2B + Pb) for all of the probesets in the 12 transbands (Table 2). In the experimental data from flies with no lead treatment, 5 transbands are enriched in probesets that have higher expression in ORE (con_50DF, con_72A, con_73D, trt_30AB, and trt_73D) and 2 transbands have a significant enrichment in probesets that have higher expression in 2B (trt_3E and trt_65A) ($P < 0.001$, chi-square test; Table 2; Fig. 5c; Supplementary Table S3). In the lead-treatment group, 5 transbands have a significant enrichment in probesets that have higher expression in ORE (con_72A, con_73D, trt_30AB, trt_57F, and trt_73D) and 4 transbands have a significant enrichment in probesets that have higher expression in 2B (con_27B, trt_3E, trt_63A, and trt_65A) ($P < 0.001$, chi-square test; Table 2; Fig. 5d; Supplementary Table S3). Of the 12 transbands, only two did not show a polarity of expression based on genotype in at least one of the two environments (con_70C and trt_77E).

Since we analyzed relative expression levels in two environments, and only 1 of the transbands was present at the same location in both environments (con_73D and trt_73D), it seemed

Table 2

Polarity of 5 control and 7 treated *trans*-eQTL transbands. The transband locations are in the first row (see Table 1). Polarity is determined in four manners (first row). For example, (B + Pb/B − Pb) is the ratio of the level of expression of a gene with the B genotype in the presence of lead over the level of expression of the same gene in the absence of lead. The ratio ($n > 1$:$n < 1$), where $n > 1$ is the number of genes in a transband that have a relative increase in expression and $n < 1$ is the number of genes in the same transband that have a relative decrease in expression.

Transband	B + Pb/B − Pb ($n > 1$:$n < 1$)	O + Pb/O − Pb ($n > 1$:$n < 1$)	O − Pb/B − Pb ($n > 1$:$n < 1$)	O + Pb/B + Pb ($n > 1$:$n < 1$)
con_27B	(72:45)	**(83:34)**[*]	(71:46)	*(34:83)*[*]
con_50DF	**(105:47)**[*]	*(38:114)*[*]	**(118:34)**[*]	(85:67)
con_70C	(53:50)	(52:51)	(57:46)	(54:49)
con_72A	**(105:29)**[*]	**(104:30)**[*]	**(108:26)**[*]	**(110:24)**[*]
con_73D	**(170:108)**[*]	**(170:108)**[*]	**(178:100)**[*]	**(176:102)**[*]
trt_3E	*(89:156)*[*]	*(61:184)*[*]	*(82:163)*[*]	*(60:185)*[*]
trt_30AB	**(124:70)**[*]	**(156:38)**[*]	**(127:67)**[*]	**(162:32)**[*]
trt_57F	*(21:75)*[*]	(51:45)	(44:52)	**(71:25)**[*]
trt_63A	(75:65)	*(32:108)*[*]	(75:65)	*(32:108)*[*]
trt_65A	(33:65)	*(17:81)*[*]	*(10:88)*[*]	*(11:87)*[*]
trt_73D	**(128:94)**[*]	**(125:97)**[*]	**(136:86)**[*]	**(138:84)**[*]
trt_77E	(63:40)	(62:41)	(63:40)	(61:42)
Total	**(1038:844)**[*]	(951:931)	**(1069:813)**[*]	**(994:888)**[*]

Total, the total number of genes in all 12 transbands that either increase or decrease expression in a column.

[*] $P < 0.001$ (chi-squared test). Bold text, indicates a significant enrichment in genes in a transband with a relative increase in expression. Italicized text, indicates a significant enrichment in genes in a transband with a relative decrease in expression.

likely that some of the transbands might also be polarized for expression based on environment. For example, if a transband encodes a transcription activator that is induced by lead, then all of the genes in the transband would have higher relative expression levels in the lead-treatment group. To determine whether there is polarity for expression of probesets in a transband based on environment, we determined the ratios $(2B + Pb)/(2B − Pb)$ and $(ORE + Pb)/(ORE\text{-}Pb)$ for all of the probesets in the 12 transbands (Table 2). We determined that 5 transbands with the *2B* genotype are significantly enriched in probesets that are up-regulated by lead (con_50DF, con_72A, con_73D, trt_30AB, and trt_73D) and 2 transbands are significantly enriched in probesets that are down-regulated by lead (trt_3E and trt_57F) ($P < 0.001$, chi-square test; Table 2; Fig. 5b; Supplementary Table S3). We also determined that 5 transbands with the *ORE* genotype are significantly enriched in probesets that are up-regulated by lead (con_27B, con_72A, con_73D, trt_30AB, and trt_73D) and 4 transbands are significantly enriched in probesets that are down-regulated by lead (con_50DF, trt_3E, trt_63A, and trt_65A) ($P < 0.001$, chi-square test; Table 2; Fig. 5a; Supplementary Table S3).

We do not fully understand the reason for the polarization of gene expression in the transbands based on the genotype and/or the environment, but it suggests that trans-regulators with differing sensitivities to lead might be involved in co-regulating the genes in a transband (see Section 3). In the next section, we analyze this further by investigating specific probesets that are co-regulated by two different transbands in the two environments.

2.4. Co-regulation of genes by two different trans-eQTLs

As discussed in the previous section, further analysis of probesets from the real data identified a group of 33 probesets that are in two adjacent transbands on the 3rd chromosome in the control dataset (19 probesets at 4_141 cM (72A) and 14 probesets at 4_146 cM (73D)) and in a single transband at a different location
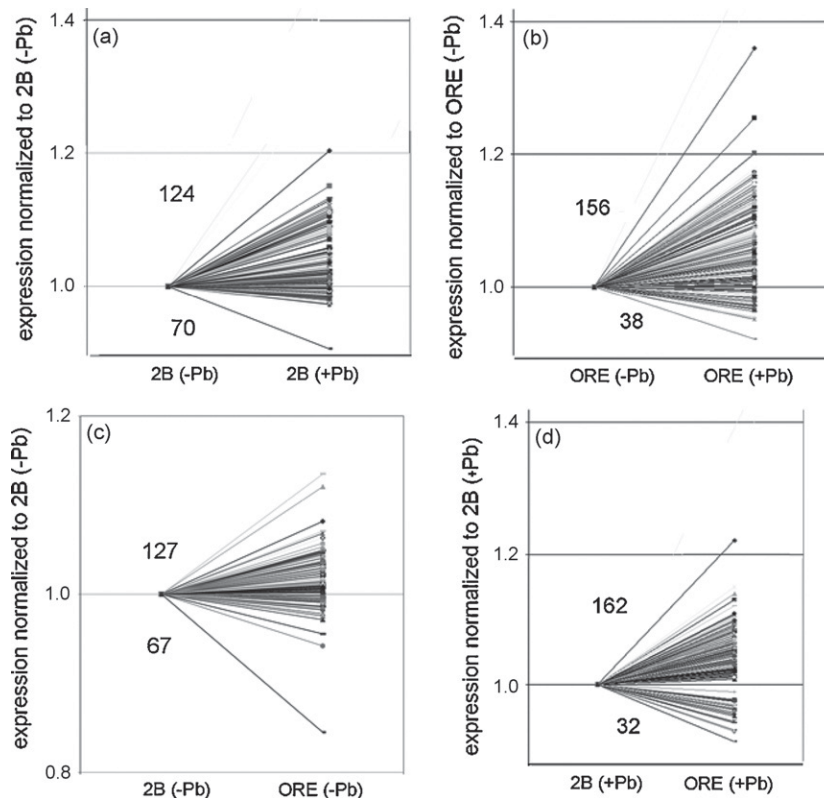

Fig. 5. The transband at 30AB is polarized for genes that are induced by lead. (a and b) More genes have an increase in relative expression when the transband has either the *2B* (a) or the *ORE* (b) genotype in the presence of lead. This indicates that there is a significant polarity of increased expression for the probesets in the 30AB transband after lead exposure regardless of genotype. (c and d) When the 30AB transband has the ORE genotype, there is a polarization of increased expression in either the presence (a) or absence (d) of lead. This indicates that there is a significant polarization of increased expression for the probesets in the 30AB transband when the genotype is *ORE* regardless of environment. The polarization level of all of the hotspots is in Supplementary Table 3.

Table 3

Genes regulated by transbands in both control and treated samples. Shown are matches between *trans*-eQTL transbands in control and treated data. Only 3 control and 3 treated transbands had four or more matches in genes amongst the transbands. Control (con), treated trt, chromosome and map positions in cM (4-136, 4-141, 4-146, 2-71, and 4-156), and approximate cytological positions (70C–71E, 72A, 73D, 30AB, and 77E) are indicated.

	trt_2-71_30AB	trt_4-146_73D	trt_4-156_77E
con_4-136_70C	4	14	4
con_4-141_72A	19	25	3
con_4-146_73D	14	83	30

on the 2nd chromosome in the lead-treated dataset (2_71 cM (30AB); Table 3; Supplementary Table S6). We refer to these 33 probesets as being "co-regulated" by two different *trans*-eQTLs. The proteins encoded by the 33 probesets (*i.e.*, 33 genes) are involved in several potentially lead-related processes, such as heavy metal sequestration or storage (*MtnA* and *Mocs1*) (Balamurugan et al., 2007; Balamurugan and Schaffner, 2006; Egli et al., 2006a,b), nervous system development and function (*tsp2a*, *tsp42ec*, *tsp42er*, CG4301, CG31720, CG5096, inx7, vnd) (Fradkin et al., 2002; Kopczynski et al., 1996; Sinenko and Mathey-Prevot, 2004; Todres et al., 2000), detoxification (CG4562, CG10505, rtet) (Mehta et al., 2005), and proteolysis (pcl, CG33127, κTry) (Schubert, 2008) (Table 4).

Comparison of the microarray signal intensities of these 33 genes co-regulated by the 2nd and 3rd chromosome transbands indicates that all 33 of them are significantly up-regulated by lead (Fig. 6a and b; Supplementary Table S6). Interestingly there is approximately the same (or lower) expression when the 30AB transband has the ORE or the 2B genotype in the absence of lead

(*i.e.*, $(ORE - Pb)/(2B - Pb) \leq 1$), but an increase in expression in ORE versus 2B in the presence of lead (*i.e.*, $(ORE + Pb)/(2B + Pb) > 1$) (Fig. 6c and d; Supplementary Table S4). There is a similar pattern for the 72A or 73D transbands (Fig. 6e and f; Supplementary Table S6). Such "crossing of the lines" in this type of GxE plot is suggestive of a genotype-by-environment (GxE) interaction (Gibson, 2008). More rigorous GxE-eQTL analyses are described in the next two paragraphs.

Log of distribution (LOD) plots of the 33 co-regulated genes shows a significant peak at 72A-73D in the control data and two significant peaks at 30AB and 72A–73D in the treated data (Fig. 7a and b). To determine whether there is a significant GxE interaction among the 33 genes co-regulated by two different trans-eQTLs, we performed analysis of variance (ANOVA) with the model: intensity = G + E + GxE + error, where G is the genotype (ORE or 2B), E is the environment (control or lead), and "error" is the root mean square error. In a plot of the GxE term, this analysis indicates that there is no significant GxE peak at 30AB and a peak for 3 probesets at 72A–73D. "Significance" lines at LOD = 3 were drawn in Fig. 7 for reference because this LOD is a common threshold used in QTL studies. However, as discussed earlier, the "significance" was determined for each probeset by permutation analyses and differed for each probeset (see Section 4). All of the peaks at 72A–73D in Fig. 7a are significant, and all of the peaks at 30AB and 72A–73D are significant in Fig. 7b (Supplementary Table S6).

Next using this "simple" model (intensity = G + E + GxE + error), we calculated the GxE term for all 18,952 probesets. Interestingly, we identified a major GxE hotspot located at 50DF that contains 819 probesets (Fig. 8a). The most significant GO category that is enriched for the genes that correspond to the 819 probesets is

Table 4

Description of genes co-regulated by two trans-eQTL loci. There are 19 probesets co-regulated by 72A and 30AB (CG14872 is represented by two probesets) and 14 genes co-regulated by 73D and 30AB.

|	Affy_ID	GO:Biological process	Gene location
Co-regulated by 72A and 30AB			
1. CG14872	1623028_at	Transporter activity	88F7
2. CG17544	1623069_s_at	Fatty acid metabolism, Pristenoyl-CoA oxidase	37D4,5
3. CG15347	1623761_at	None	7E11
4. CG31087	1624569_at	Nuclear gene	96D1
5. CG4781	1625816_at	Protein binding	60D10
6. tsp2a	1626767_at	Synapse formation, tetraspanin 2a	1F4
7. CG8907	1626881_at	None	89E8
8. tsp42ec	1630051_at	Synapse formation, Tetraspanin	42E5
9. CG4301	1630471_at	Ion transport, Ca-transporting ATPase	14C4
10. Mocs1	1631994_a_at	Mo utilization, Mo-cofactor binding protein (CG33048)	68A4
11. CG15422	1632575_at	Pathogenesis	24D3
12. pcl	1633955_at	Proteolysis, Pepsinogen-like	1B2
13. pes	1634913_s_at	Defense response, Scavenger receptor	28D3
14. tsp42er	1634977_at	Synapse formation, Tetraspanin	42E7
15. CG11961	1637042_a_at	Peptidase activity	56D1,2
16. CG31720	1637306_at	Signal transduction, G-protein couple receptor	31B1
17. myo61F	1638278_s_at	Vesicle-mediated transport, Cytoplasmic myosin	61F6
18. CG5096	1639478_at	Neuron development, Receptor in neurons	31D11
19. CG14872	1640720_a_at	Transporter activity	88F7
Co-regulated by 73D and 30AB			
20. CG8062	1624181_at	Fatty acid biosynthesis Mono-carboxylic acid transporter	18C2
21. CG4562	1625778_at	Detoxification Toxin transporter	92B4
22. CG33127	1626718_at	Proteolysis Trypsin	21B8
23. CG6484	1626734_at	Glucose homeostasis, Glucose transporter	54C10
24. CG3987	1627047_at	Muscle development, Mesoderm development	88E2
25. RpA-70	1627380_at	Protein biosynthesis, Ribosomal protein A-70	84F6
26. CG14299	1628826_at	Nuclear gene	91C6
27. kappaTry	1632055_at	Proteolysis, Kappa trypsin, CG12388	47E3
28. MtnA	1632873_at	Metal transport, Metallothionine	85E9
29. vnd	1635383_at	Neuron development, Ventral nervous system defective	1B10
30. CG1246	1638174_at	None	62E8
31. inx7	1638225_a_at	Gap junction protein, Innexin 7	6E4
32. CG3884	1638816_at	None	49E4
33. CG10505	1639368_at	Detoxification, Toxin transporter	57D7,8
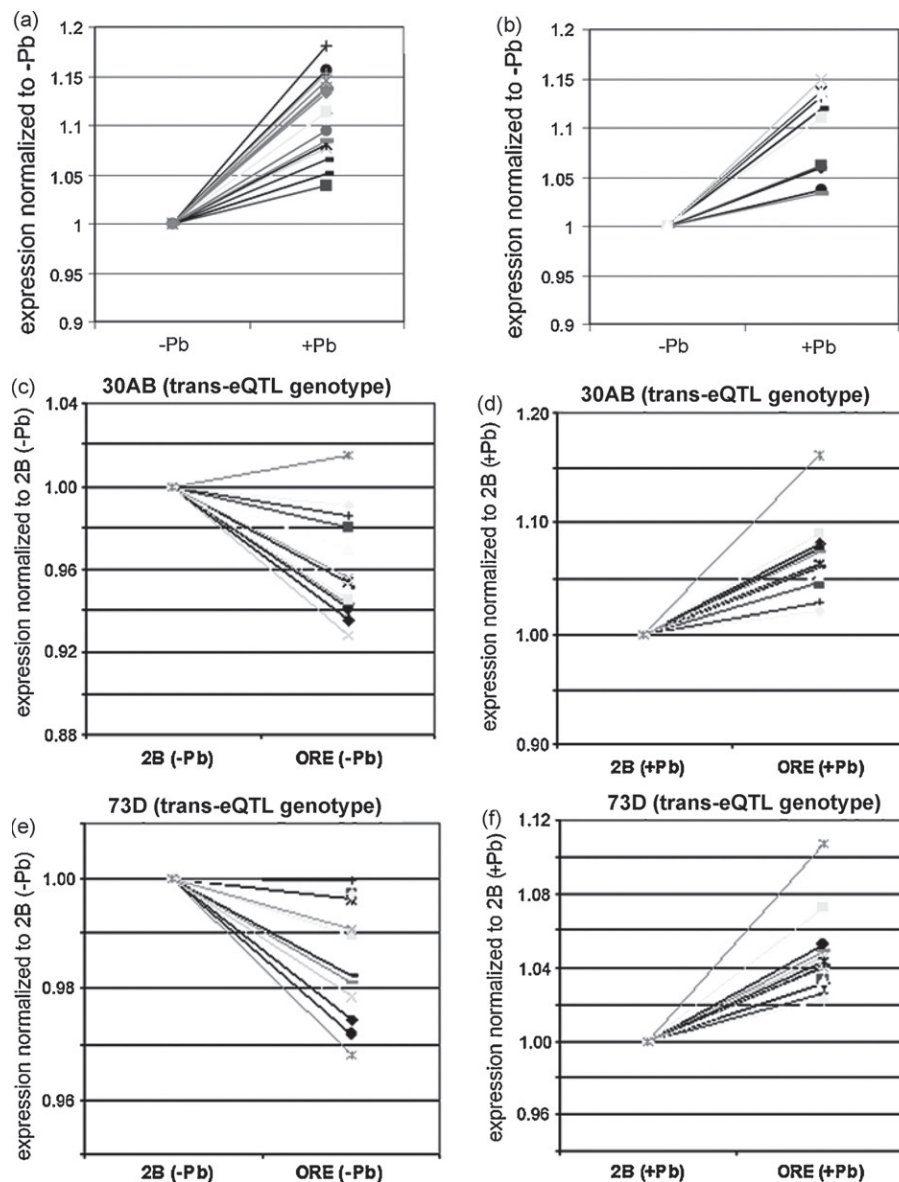
Fig. 6. Effect of lead on the expression of genes co-regulated by 30AB and 73D. (a) All 19 genes that are co-regulated by 72A and 30AB show an increase in expression after lead exposure. (b) All 14 genes co-regulated by 73D and 30AB also show an increase in expression after lead exposure. The data in a and b is from the mean expression levels from the 75 RILs in the absence of lead and the mean levels from the 75 RILs in the presence of lead. (c–f) Effects of genotype and environment on the 14 genes co-regulated by the transbands located 30AB and 73D. The relative expression of the 14 genes with the indicated transband genotype was normalized to 1 and the effect of changing the hotpot genotype is shown. Notice that relative expression decreases slightly (except for one gene) in the absence of lead but increases in the presence of lead when the transband genotype changes from *2B* to *ORE*. A similar pattern is seen for all 19 genes that are co-regulated by the transbands located at 72A and 30AB (Supplementary Table S4).

"nervous system development" (FDR < E−10). Table 5 shows the 59 genes in this GO category, and all 59 of these genes are down-regulated by lead. However, the meaning of this result is not clear because very few of the 819 probesets (53) have a main effect eQTL located at 50DF, and the robustness of a GxE effect in the absence of a main effect is uncertain (see Section 3). The significance of the 819 probesets with GxE interactions in the "simple" model is further questioned by analyzing GxE interactions in the 200 permutations of the strain labels. We found that 34 permutations have GxE hotspots with greater than 819 probesets (Fig. 8b and c) and that 94 of the permutations have greater total numbers of GxE-eQTL (Fig. 8d).

Although the finding that the GxE-eQTL hotspot is enriched in neurodevelopmental genes is intriguing, the fact that few of the 33 genes regulated by *trans*-eQTL at 30AB or 72A–73D have GxE-eQTL LOD scores greater than 3 (Fig. 7c) suggests that the original

ANOVA model is incomplete. Therefore, we added a new term to the model, L (line, the RIL designation), as a non-dependent covariate. The new model (which we will refer to as the "full" model) is: intensity = G + E + L + GxE + error (Fig. 7d). This full model indicates that a greater number of probesets have significant GXE peaks at 30AB and 72A–73D (compare Fig. 7c and d). Interestingly, the most significant peak is at 61A, which does not correspond to a main effect eQTL peak. As with the 50DF peak discussed above, the meaning of a GxE peak in the absence of a main effect peak is uncertain. Because it is computationally extensive, we have not yet completed the permutation analysis of the GxE interactions with Line as a non-dependent covariate for all 18,952 probesets. However, the permutation LOD scores (with L as an independent covariate) for the 59 "nervous system development" genes are shown in Table 5 and all but 8 of them are significant (*i.e.*, permutation LOD > 3).
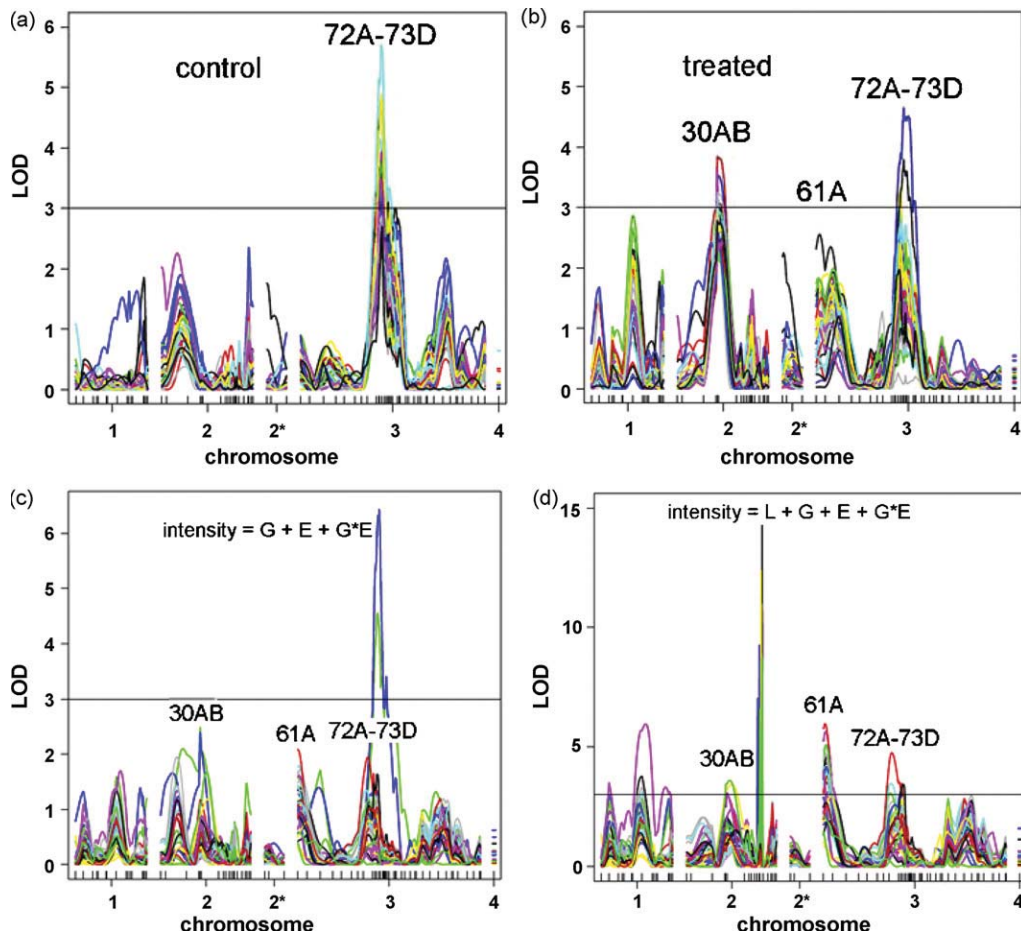
Fig. 7. QTL analyses of 33 genes co-regulated by two loci. (a) eQTL graphs of 33 genes in control flies. (b) eQTL graphs of 33 genes in treated flies. (c) GxE graphs of 33 genes where line (L) is not an independent covariate (intensity = G + E + GxE + error). (d) GxE graphs of 33 genes with the full model (intensity = L + G + E + GxE + error). More genes cross the LOD = 3 cutoff at 30AB, 61A, and 72A–73D in the "full" model as compared with the model lacking line as a covariate (c).
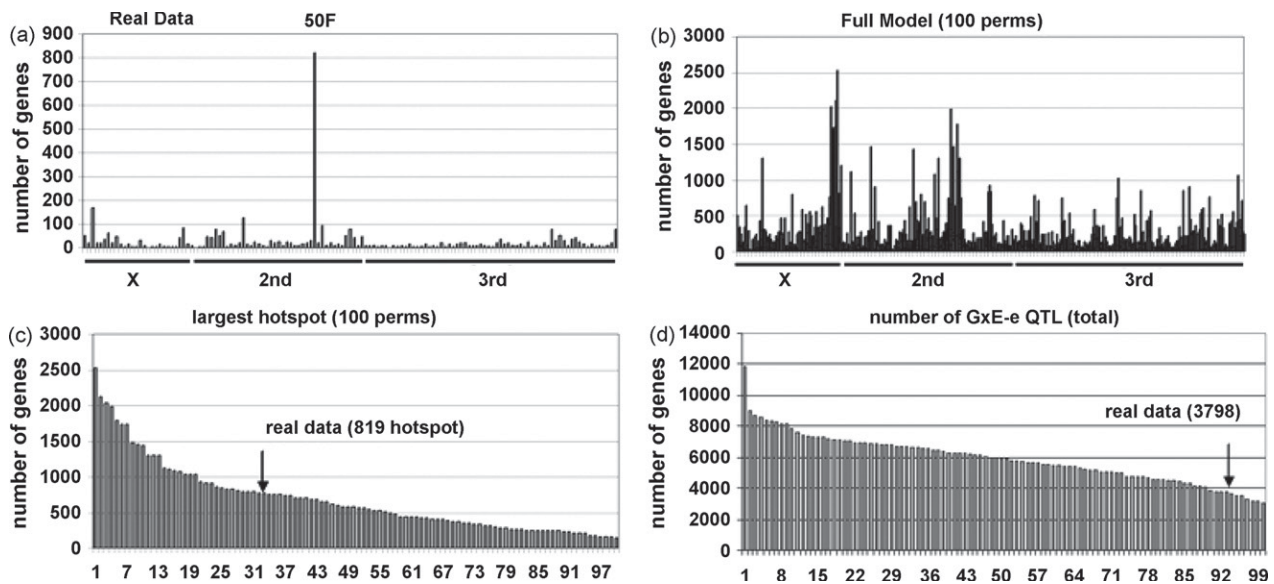

Fig. 8. Permutation analysis of the GxE-eQTL data. (a) The experimental GxE-eQTL data are plotted with eQTL location (*x*-axis) and number of probesets with eQTL at the corresponding location (*y*-axis) (intensity = G + E + GxE + error). The hotspot with 812 probesets at 50DF is indicated (50F). (b) The GxE-eQTL results of 100 permutations is plotted (intensity = G + E + GxE + error). (c) The rank order of GxE-eQTL hotspots (*x*-axis) in the 100 permutations is plotted versus the number of genes with eQTL in the largest transband (*y*-axis). The experimental data (real data) has 819 probesets in the largest transband. (d) The rank order of the number of GxE-eQTL in the 100 permutations is plotted. Notice that most (94%) of the permutations have more GxE-eQTL than the experimental data (arrow).

Table 5

Nervous system development genes down-regulated by lead. All 59 genes are down-regulated by the GxE-eQTL hotspot located at 50DF. The permutation LOD scores are shown. <3, NS (not significant). For 8 of the genes, the LOD score is less than 3 and not significant.

Gene name (GO: nervous system development)	LOD score	Gene name (GO: nervous system development)	LOD score
ABNORMAL CHEMOSENSORY JUMP 6	4.4	KUZBANIAN	3.9
ALK	<3, NS	LADYBIRD LATE	<3, NS
AMYLOID PROTEIN PRECURSOR-LIKE	8.2	LAMIN	<3, NS
ANK2	<3, NS	LEONARDO PROTEIN	5.1
BEATEN PATH IA	5.8	LIM1	3.4
BEATEN PATH IIA	5.3	LONGITUDINALS LACKING	3.4
BONUS	4.3	MASTERMIND	8.3
BROAD	5.0	MINIBRAIN	4.4
CACOPHONY	6.0	MUSCLEBLIND	6.2
CADHERIN-N	7.9	NETRIN-B	6.4
CHARLATAN	5.8	NEUROGLIAN	6.9
CHOLINE ACETYLTRANSFERASE	7.4	OCELLILESS	5.3
CONNECTIN	9.6	PHYLLOPOD	6.4
COUCH POTATO	5.1	POINTED	5.3
CUT	4.9	PROTEIN SPLIT ENDS	4.3
DACHSHUND	3.9	PURITY OF ESSENCE	4.8
DERAILED	4.9	RETAINED	3.6
DICHAETE	4.2	RHOGAP-93B	<3, NS
DISCS LARGE 1	7.4	RHOGAP-P190	<3, NS
DLIC2	<3, NS	ROUNDABOUT	5.2
DUMPY	5.0	SCABROUS	8.2
EMBRYONIC LETHAL, ABNORMAL VISION	7.2	SCALLOPED	5.9
FASCICLIN 2	6.0	SCRATCH	6.7
FRUITLESS	4.2	SCRIBBLER	4.8
G PROTEIN O 47A	7.1	STARRY NIGHT	10.6
G PROTEIN-SUBUNIT 13F	5.1	STILL LIFE	5.0
HOMOTHORAX	6.0	TRIO	<3, NS
INSULIN-LIKE RECEPTOR	4.2	TURTLE	4.9
KIN OF IRRE	3.5	VEIN	6.3
KRUPPEL	3.8		

3. Discussion

In the first genetical genomic study with RILs performed in insects, we show evidence that adult *D. melanogaster* have 12 major master-modulatory *trans*-eQTLs that contain from 96 to 278 genes each, which places this organism somewhere between *C. elegans* and mice in terms of the number and complexity of the *trans*-eQTL targets (transbands) (Broman, 2005; Li et al., 2006). We extended the genetical genomics approach by comparing control flies with flies exposed to the developmental neurotoxin lead and identified transbands that differ in the two environments. The most interesting finding is that 33 genes are co-regulated by two transbands, one on the second chromosome at 30AB and one on the 3rd chromosome at either 72A or 73D which is located in the adjacent 5 cM window that was used to define transbands. Because the resolution of QTLs is often larger than 5 cM, it is likely that the 3rd chromosome transbands are the same locus, which for simplicity we will call the "73D transband" in Fig. 9 and in the following paragraphs.

These 33 co-regulated genes show significant genotype-by-environment (GxE) interactions for both the 30AB transband and the 73D transband. We have recently shown by traditional phenotypic QTL analysis that the 30AB locus is involved in lead-dependent changes in locomotor activity, which we believe is an important validation of the 30AB eQTL transband (Hirsch et al., 2009). We propose a model based on our data that a *trans*-regulator located at 73D increases expression of its target genes when it binds lead, regardless of the genotype of the 73D transband (Fig. 9a). In contrast, a second *trans*-regulator located at 30AB only increases expression (or otherwise increases steady-state levels of mRNAs via transcriptional or post-transcriptional mechanisms) of the co-regulated target genes in the presence of lead when it has the *ORE* genotype (Fig. 9b). This model can explain much of the data but other possibilities exist. For example, the putative *trans*-regulator could increase the stability of the target gene mRNAs or

there could be an indirect effect on steady-state mRNA levels. Fine-mapping the genes that underlie the QTLs, which will be aided by sequencing both of the *ORE* and *2B* genomes, should help in determining the precise mechanism involved.

The first genetical genomics study that identified genes with significant GxE interactions was a study in *C. elegans* in which the
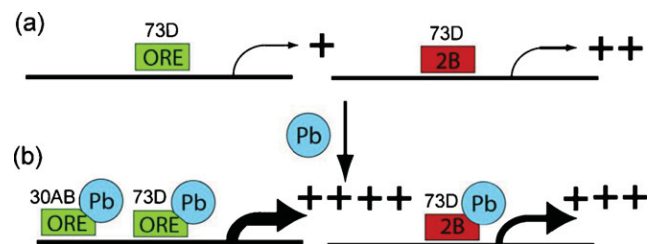

Fig. 9. Model for co-regulation by *trans*-eQTL at 30AB and 73D. (a) The hypothetical *trans*-activator encoded at the 73D transband locus (rectangle) is required for basal transcription of the 33 genes co-regulated by both 73D and 30AB in the absence of lead. In the presence of lead (Pb), the *trans*-activator encoded at the 73D transband locus further increases steady-state mRNA levels when it has either the *ORE* (green) or *2B* (red) genotype. However, the trans-regulator encoded at the 30AB transband locus (oval) further increases steady-state mRNA levels only when it has the *ORE* genotype. A second possibility is that a "translational stabilizer" encoded at the 73D transband locus is required for basal stabilization of the mRNA for the 33 genes co-regulated by both 73D and 30AB in the absence of lead. In the presence of lead, the translational stabilizer encoded at the 73D transband locus further stabilizes target mRNA levels (*i.e.*, protects from degradation) when it has either the *ORE* or *2B* genotype. In this example, the translational stabilizer encoded at the 30AB transband locus further increases steady-state mRNA levels only when it has the *ORE* genotype. A third possibility is an "indirect model" in which signaling components upstream of the transcription or translation factors are regulated by lead. In this example, the signaling protein encoded at the 73D transband locus activates a transcription factor in either the presence or absence of lead. For example, the signaling protein encoded at the 30AB locus could inhibit a repressor of a transcription factor (R) better when it has the *ORE* genotype than when it has the *2B* genotype. In all three models, Pb induces the highest steady-state mRNA levels for the 33 targets genes when both 73D and 30AB loci have the *ORE* genotypes.

Table 6
Candidate genes from 30AB that mediate the behavioral response to lead. We took gene expression data for 122 probesets nearest our behavioral QTL and asked whether or not strain differences in gene expression levels for these, individually, were correlated with the behavioral change induced by lead treatment. There were seven "significant" (*e.g.* unadjusted for alpha = 0.05) correlations. Ggamma30A is represented twice by two different probesets, as indicated by the Affymetrix ID (Affy ID).

Gene name	Correlation coefficient	P-value	Protein function
Cks30A (CG3738)	$r = -0.266$	$P = 0.03$	Amino acid phosphorylation, cell cycle regulation, histoblast morphogenesis
Ggamma30A (CG3694) Affy ID 1637526_S_at	$r = 0.328$	$P = 0.007$	Signal transduction via G-protein, visual perception
Ggamma30A (CG3694) Affy ID 1628754_S_at	$r = 0.31$	$P = 0.011$	Signal transduction via G-protein, visual perception
CG3838	$r = 0.276$	$P = 0.024$	DNA binding
Nckx30C (CG18660)	$r = 0.309$	$P = 0.011$	Ion transport: K, Na, Ca; compound eye development
CG9525	$r = -0.288$	$P = 0.018$	No known function
CG9573	$r = -0.328$	$P = 0.007$	Protein binding

authors identified a group of genes with *trans*-eQTL that are induced by heat shock, so-called "plastic QTL" or pQTL (Li et al., 2006). Smith and Kruglyak recently did a detailed analysis of GxE-eQTL in yeast (which they call "*gxe*QTL") grown in either glucose or ethanol as a sole carbon source (Smith and Kruglyak, 2008). Other laboratories have identified genotype-by-environment interactions in which the 'environment' is a different tissue (*e.g.*, brain versus liver; Hovatta et al., 2007), but we believe that ours is the first "genetical toxicogenomic" study that combines genetical genomics with toxicogenomics. The benefit of adding environmental perturbations in genetical genomics studies is discussed in a recent paper by Li and colleagues (Li et al., 2008). Interestingly, we identified a GxE-eQTL transband with 814 genes that does not correspond to a main effect eQTL transband. Of these 814 genes, 59 of them are in the GO category "nervous system development," and 51 of these genes have significant GxE-eQTL when L (line) is a non-dependent covariate. We are undertaking a study to identify all of the genes with GxE-eQTL in *Drosophila*, where L (line) is a non-dependent covariate. The meaning of a GxE-eQTL without a main effect eQTL is controversial, and more work needs to be done to determine whether such GxE-eQTL are biologically meaningful.

We showed that there are 12 *trans*-eQTL transbands: 5 control transbands and 7 lead-treated transbands at 11 different loci (Table 1). The existence of *trans*-eQTL transbands has been validated in yeast (Brem et al., 2002; Yvert et al., 2003), but there is controversy as to whether *trans*-eQTL transbands identified in other model organisms are real or correlation artifacts (de Koning and Haley, 2005). This has been illustrated by a simulation study that showed that the five most populated bins of expression data contained 20% of the significant, but spurious, *trans*-eQTLs (Perez-Enciso, 2004). One suggested solution, which we did in this manuscript, is to perform permutation analyses of the eQTL data (*i.e.*, permuting the genotype assignments only and keeping the original expression values for all the samples) (Breitling et al., 2008). However, such permutation analyses, done by others, seemingly invalidate numerous of the previous studies if one only considers the size of the transband as the most important criterion (Breitling et al., 2008). Therefore, we have more extensively analyzed the permutations in our study and found that (1) the number of transbands, (2) the total number of eQTL, and (3) the number of identical genes in control and treated transbands might be more important criteria than the absolute size of the transband.

Another approach to validating transbands is to test gene expression for individual probesets within them for significant genetic correlation with variation among lines for functionally related traits assessed independently. The 30AB region of the *Drosophila* genome has been identified as both a behavioral QTL for changes in locomotion in response to lead treatment (Hirsch et al., 2009) and as a transband in the present study that regulates 194 probesets in response to lead. We can therefore search for individual probesets within the 30AB region showing variation in gene expression after lead exposure among roo lines that is

correlated with variation among the same lines for their behavioral response to lead to further identify potential candidate genes for both the behavioral and transband eQTL. Any probesets identified in this manner can be further screened by correlating gene expression from candidate probesets linked to 30AB with gene expression from candidate probesets regulated from the 30AB transband.

For present purposes, we focused on 122 probesets that include 30AB and the proximate half of the adjacent regions 29F and 30D (measured by base-pair distances). We identified seven probesets among the 122 spanning 30AB that showed correlations at a level of $P < 0.05$ (Table 6). We identified Ggamma30A (CG3694), involved in G-protein signal transduction and cell cycle regulation as the probeset in the 30AB region with gene expression most strongly correlated with both lead-induced behavioral changes and lead-induced changes in gene expression among those probesets regulated from 30AB. Ggamma30A has human and other mammalian orthologs with similar function, and has been associated in mice with gustation (Huang et al., 1999).

Although these cursory analyses for purposes of discussion have not been corrected for multiple comparisons, they illustrate a simple and direct method for correlating functional traits associated with lead exposure to gene expression levels in response to lead as an independent method for screening candidate genes of interest within identified QTL. The combination of gene expression data and data on functional traits in the same RI lines provides a more precise method for *a priori* screening of candidate loci representing significant QTL than either method alone would allow. This approach should be able to increase the probability of identifying valid candidate loci represented by eQTL to facilitate confirmation by direct genetic methods such as deletion mapping and gene silencing, at least in cases where functional trait QTL are derived from strain differences in traits that are a direct function of strain differences in gene expression.

In the absence of data on functional traits and functionally related gene expression in the same RI lines, *a priori* identification of candidate genes for a QTL is limited to what is already known about the functions of closely linked loci. In the present case, examining the known or predicted functions of the 122 genes closely linked to our behavioral QTL reveals one that is a strong candidate based on predicted function (CG3759: "laccase-like") associated with defense against toxins, iron and copper binding (Flybase). Although gene expression from this probeset is not correlated with the behavioral response to lead, it may represent genetic variation that acts developmentally to mediate behavioral responses to lead at a physiological or anatomical level that does not depend on adult gene expression.

We have also measured variation among these roo lines for lead burden in response to the same lead treatment used here to assay gene expression in response to lead treatments, and have identified three significant QTL for this trait (unpublished) that do not overlap with either the behavioral QTL (Hirsch et al., 2009) or the

transband eQTL identified here in response to lead treatment. This observation suggests that our transband pattern of gene expression in response to lead is not significantly dependent on strain variation in lead burden, although further analysis of individual probesets may reveal interesting associations.

Systematic analysis of genetic correlations among probesets for gene expression and functional trait expression in response to lead exposure may help identify specific candidate probesets of interest within each of the transband QTLs. Genetic validation of the putative *trans*-regulators in our transbands is required to authenticate this proposition, but, if validated, it represents a significant change in how to interpret eQTL analyses.

3.1. Limitations of genetical genomics studies

There are limitations common to all eQTL studies. For example, previous genetical genomics studies have determined that probeset expression levels are often apparently polarized for expression based on genotype—*i.e.*, probesets of one genotype in the RILs have higher relative expression levels compared with probesets of the other genotype. For example, in studies with BXD mouse RILs, which were made from the C57BL/6J sequenced mouse strain (for which the probes were designed) and the DBA/2J strain, a significantly larger number of the probesets with the C57BL/6J genotype have an apparently higher relative expression level than with the non-sequenced genotype (Chesler et al., 2005). The apparent polarization of expression based on genotype is now a well-known design flaw of microarray experiments because probes that fail to hybridize to a gene because of a sequence mismatch could be erroneously scored as having lower expression (Pienta and Esper, 1993). We controlled for this by eliminating probesets that have "outlier probe(s)" with significantly lower signal(s) than the other probes in the probeset.

Another major limitation of our genetical genomics study is that if the regulation of a lead-dependent master-modulatory gene is identical in the progenitor strains *ORE* and *2B*, then these master regulators would probably not be identified in our analysis. An approach that analyzes gene expression patterns in populations of flies, such as the quantitative trait transcript technique (Passador-Gurgel et al., 2007; Ruden, 2007), or an eQTL approach that uses more than two progenitor strains (Flint et al., 2005), will likely identify further GxE-eQTL transbands and lead-dependent master-modulatory genes.

A further limitation of our experimental design is that anatomical alterations are likely induced by either lead or genetics. For example, if some of the flies have a bigger brain (or any other tissue since we analyzed whole flies) then one would find up-regulation of brain (or some other tissue) associated genes. However this would not necessarily be true if only equivalent amounts of the tissue in question were compared. Therefore, as discussed as a general limitation of microarray design in whole organism studies (Chintapalli et al., 2007), the observed effect would not be a transcription effect but rather is a function of strain differences occurring on a larger anatomic scale. We and others have observed that the lead concentration that we used causes developmental delays, but these developmental delays are not accompanied by changes in the adult body weight (Akins et al., 1992). However, we and others have shown that certain traits such as triglyceride levels and body mass vary considerably in the RILs that we used in our studies (Clark and Keith, 1988; De Luca et al., 2005). Therefore, these caveats should be considered in the interpretation of our data and in similar studies.

Why did we use whole flies for these studies instead of heads or brains? The co-authors extensively discussed this issue before deciding on using whole flies. We thought that starting with whole flies is the best approach because lead has numerous physiological effects throughout the body, and not just in the nervous system. We also believed that lead could have global effects on gene expression in a conserved manner, but the neuronal effects would become manifest only in the nervous system. We also argued that if we decided to analyze only the head (of which over half is brain and eyes, which are also parts of the nervous system), then we might be overlooking interesting gene expression changes that occur in the PNS (*e.g.*, ganglia) and other parts of the body. Furthermore, a criticism of mouse genetical genomics experiments that use whole brains is that smaller regions of the brain would have been preferred. For *Drosophila*, we believe the next step should be to repeat these studies with fly heads or, even better, specific types of neurons or brain regions by using a purification technique, such as the bacTRAP technique that was developed for purifying mouse Purkinje cells and other neuronal types (Doyle et al., 2008; Heiman et al., 2008; Selimi et al., 2009).

Why did we only do one microarray experiment for each of the 75 RILs in the presence or absence of only one concentration of lead? The most obvious answer is that this would have been prohibitively expensive. However, the great advantage of genetical genomics studies over other types of gene expression studies with microarrays is that the most important unit of replication is the genotype of a particular genome region (*ORE* or *2B*) and not the RIL. Since we performed 150 microarrays, we analyzed gene expression of the *ORE* genotype ∼75 times and of the *2B* genotype ∼75 times for each region of the genome. Not many studies have run 75 replicate microarray experiments, but we have essentially done so in this study.

What is meant by a "significant" change in gene expression? We decided to go by *P*-value (via two tailed *t*-test) rather than by absolute change in gene expression (*e.g.*, a minimum 1.5-fold change used by many investigators). The main reason that some other studies use a threshold of 1.5 as being "significant" is that they wish to identify the genes that are likely important in a response, and it can be argued that a change in steady-state mRNA levels of only 1.2 is unlikely to cause a significant change in protein levels. However, this is not the case in our studies because we are primarily interested in pathways (*i.e.*, transbands) rather than individual genes, so we believe that "significance" by *t*-test is a more robust approach. Also, since we essentially replicated our studies ∼75 times for the genotype of each region of the genome, a 1.2-fold change, which is a typical change in gene expression in a probeset in the transbands, can be very significant.

What is the cost effectiveness of doing genetical genomics in *Drosophila* versus mice or some other model organism? In general, the smaller the genome, the more powerful the genetical genomic studies can be. For example, genetical genomic studies done with a similar number of RILs in the small-genome organism *Saccharomyces cerevisia* have mapped *cis*-eQTL and transbands to the exact polymorphism (Demogines et al., 2008). Genetical genomics studies in *C. elegans*, which has a similar genome size as *D. melanogaster*, can map eQTL to within 5 cM, which corresponds to dozens or hundreds of genes, but further fine-mapping studies can be quickly done to identify the exact polymorphisms responsible (McGrath et al., 2009; Reddy et al., 2009). Genetical genomics studies have also been done extensively in rats and mice (Bao et al., 2007; Druka et al., 2008; Li et al., 2005), which have a further 10–20-fold increase in genome size, but the precision of these studies is consequently much broader and fine-mapping studies are much more difficult. Fortunately, next-generation "cheap sequencing" technologies have begun to enter the genetical genomics field (Bloom et al., 2009), and these techniques might soon allow these studies to become cost effective in mammalian models for even small laboratories.

3.2. Evolution in a toxic environment

Responses to stressors must play a vital role in evolution, and our results provide insights into mechanisms that may be involved. We identified several transbands that might contain master-modulatory proteins that are affected by lead in one genotype but are insensitive to lead in the other genotype. Since we started with only two parental lines, there might be many more such master-modulatory proteins in the genome in a population of *Drosophila* that are similarly evolving. Another question is whether these same master-modulatory proteins uniquely acquired sensitivity to lead, which seems unlikely, or respond in similar manners to many environmental toxins. Toxin-induced changes in the regulation of these proteins might result in the development of a slightly "modified" nervous system that could enhance fitness in a contaminated environment. Consistent with this, lead-induced changes in mating and in fecundity in *Drosophila* (Hirsch et al., 2003) can be interpreted in terms of changes in resource allocation by the female that may, from a long-term perspective, be adaptive. The studies presented here show many promising new avenues for investigation.

There are several practical applications for the data generated in the study. The data will be deposited in the GeneNetwork website (http://www.genenetwork.org) so that other investigators can look for correlations between gene expression patterns and phenotypic traits. The GeneNetwork is an open resource and consists of a set of linked resources for systems genetics. It has been designed for integration of networks of genes, transcripts, and traits such as toxicity, cancer susceptibility, and behavior for several species. Phenotypic QTLs using the roo lines were identified in numerous other QTL mapping studies (Gurganus et al., 1998, 1999; Jordan et al., 2006; Leips and Mackay, 2000; Long et al., 1995; Pasyukova et al., 2000; Vieira et al., 2000; Wayne et al., 2001). For sets of phenotypes, particularly those in Gene Network's databases (*Drosophila* phenotypes are not yet in this database), a variety of correlation analyses can be performed with the gene expression data. Trait values entered by users or retrieved from the databases can be correlated with any other database of phenotypes from the same mapping genetic reference panel, in our case, the roo lines. The future addition of our data and other *Drosophila* QTL data to GeneNetwork will be a tremendous boon to the *Drosophila* community and to the genetics community in general.

4. Materials and methods

4.1. D. melanogaster stocks, growth, and RNA extraction

The 75 *Drosophila* roo lines were obtained from Trudy Mackay. To avoid batch effects (Zakharkin et al., 2005), the growth of the flies, the RNA extraction and the order of running the arrays, and the fluidics well used for each array was completely randomized for the 75 lines in two treatments. Control food consisted of standard cornmeal, agar, sugar, yeast, and 250 μM NaAc (Ashburner, 1989). Lead-contaminated food consisted of standard food plus 250 μM PbAc (lead exposure at this concentration has been shown to affect locomotion in adults; Hirsch et al., 2003). Flies from each of the 75 roo lines (20 males and 20 females) were placed in a vial with 10 ml of food (control or PbAc) for 3 days at 25 °C and allowed to lay eggs; the adults were subsequently discarded. Newly enclosed adult males were placed on the same medium (control or PbAc) as had been present during pre-adult development for 5–10 days before being used as subjects. Male progeny were pooled from each vial (65 males per vial) and frozen at −80 °C. RNA samples were extracted in groups of 24 and arrays hybridization run in groups of 4 with 3 groups run per day. Effects of RNA extraction and array hybridizations day were examined by ANOVA and Support Vector approaches and no obvious day effects were observed (data not shown).

4.2. RNA extraction and hybridization

Total RNA was isolated following Trizol Manual (Invitrogen Life Technologies) with modifications made by Kevin Bogart and Justen Andrews at Indiana University. Frozen flies (50 mg) were homogenized with Poly Tron (Kinematica AG) in a 5 ml tube with 2 ml of Trizol. The homogenate was incubated at room temperature for 5 min followed by centrifuge at 12,000 rcf for 10 min at 4 °C to pellet insoluble debris. 1 ml supernatant was transferred to a new centrifuge tube and 0.2 ml chloroform was added. After shaking the tube vigorously by hand for 15 s, sample was incubated for 3 min at room temperature and then was centrifuged at 10,000 rcf for 15 min at 4 °C. 0.54 ml aqueous phase was carefully transferred to a fresh RNAse-free tube and 0.45 ml isopropanol was added to each tube followed by vortexing. After 10 min incubation at room temperature, sample was centrifuged at 12,000 rcf for 10 min at 4 °C. Supernatant was then removed and the pellet was washed with 1 ml 75% ethanol followed by centrifuge at 7500 rcf for 5 min at 4 °C. After the supernatant was removed and the pellet was air dried for 10 min at room temperature, the pellet was resuspended in 40 μl RNAse-free water at the end. All RNA samples were lead-treated with DNase I in order to get rid of potential DNA contamination. The quality of each RNA sample was determined on 2100 Agilent Bioanalyser before proceeding RNA labeling.

RNA quality was assessed using the Agilent Bioanalyzer. Since *Drosophilia* generate atypical peaks (18S RNA is absent) the 18S to 28S ratio was not studied, however the bioanalyzer results were used to determine if the RNA extraction was clean. Samples that failed were re-extracted from other flies. The RNA that passed the quality checks were run on the Affymetrix *Drosophila* genome 2 Array®. This array has 18,952 probesets and each probeset is composed of on average 14 probe pairs with 11 μm feature size.

4.3. Microarray quality control

The images were initially processed with MAS 5.0 for quality control. Multiple quality control steps were used. The images of all arrays were manually inspected for defects. The 3′/5′ ratio of beta actin was verified as being between 1.5 and 3. The chip to chip Pearson's correlation was examined (mean correlation <0.91 to all other arrays set to being poor). The geography Index (Kim et al., 2006) and a deleted residual (Persson et al., 2005) approach were also used, and a cutoff of a KS-D of 0.15 was considered poor quality. These steps resulted in 3 arrays being excluded. These were the roo line 54 with lead, and roo lines 11 and 21 with control treatment. These quality control steps resulted in the eQTL analysis being conducted on 73 control roo lines and 74 lead lines with a total of 72 lines being measured in both treatments.

4.4. Expression quantification

The 147 microarrays that passed quality control were processed as a single group using the justRMA function of the Bioconductor software package.

4.5. Markers genotypes

Marker information for a total of the 75 roo lines which were typed for a total of 82 markers. Heterozygotes were set to missing in these lines.

4.6. Annotation

The annotation used in this experiment is that provided by Affymetrix's Netaffx (Cheng et al., 2004; Liu et al., 2003) for the *Drosophila* genome 2 array data March 27, 2006. A total of 18,147 probesets had physical genetic positions. 805 probesets without genetic position are included in the table of *P*-values (Tables S2 and S3), but are excluded from the eQTL figure (Fig. 2a). These 805 probesets without a position, had 7 significant locations in the control flies and 13 for the lead-treated flies.

4.7. Number of replicates

We chose not to use replicates in this study, and instead devoted more resources to more lines. As a result, we were able to obtain measurements on 75 RILs under control conditions and 75 RILs under lead-treated conditions with only one microarray per line per condition (*i.e.*, 150 microarray total analyses). If fewer than 75 RILs are used, or if one studies organisms with larger genomes, such as with the mouse BXD RILs which only have 30–40 commercially available lines (however, see Peirce et al., 2004), then duplicate or triplicate microarrays would probably be necessary for each experimental condition to obtain statistically significant results. Also, if "collaborative cross" RILs are used, which were made by crossing multiple mouse inbred lines (Churchill et al., 2004), then many more lines or replicates will be needed to match the power of using 75 RILs in simpler organisms such as yeast, *Drosophila* or *C. elegans*. What is important to understand in a genetical genomics experiment is that *the unit of replication in the microarray analyses is the RI line and treatments* and not the animal or the DNA aliquot. Therefore, if one has a sufficiently large number of RILs and a limited budget, then it is statistically more powerful to analyze more RILs with a single microarray per line than to conduct multiple microarrays on the same RI line.

4.8. Analysis

The expression values of the 147 arrays passing quality control were divided into their two groups, lead-treatment and control, with 73 or 74 in each. eQTL analysis was conducted on each of the two groups and then jointly with an interaction term between the treatment and locus (3 total analyses). The R/qtl (Broman et al., 2003) package was used to map QTL. Probabilities of putative QTL genotypes were computed in the 5 cM interval along the genome. Haley–Knott regression method (Haley and Knott, 1992) was used to map the QTL controlling the transcripts. Genome-wide significant level was established by permutation at 0.05 level (Churchill and Doerge, 1994) which corresponds to a *P*-value of 0.0001. Genome-wide significance was achieved at 2907 locations in the RMA processing control animals. The 2907 includes probesets that have significant linkage at more than one location in the genome. There were 3467 significant eQTLs in 3017 genes for the RMA processed and lead-treated flies (Fig. 1).

Fig. 2a and b shows the heritability estimates for the data. Heritability is calculated in the narrow sense:

$$\frac{1/2V_a}{(1/2V_a + V_e)}$$

where V_a is the variance due to strain and V_e is the variance due to environment (error) (Falconer and Mackay, 1996). Supplementary Fig. S1a provides the heritability without treatment effects. Mean heritability is 38.7% ($y = strain + error$). Fig. 2b provides the heritability with treatment effects. Mean heritability is 40.6% ($y = strain + treatment + error$).

Fig. 2c shows the *P*-values for all markers (82 markers), and every 5 cM between markers by all expression traits. As can be seen there are more 'significant' ($P < 0.05$) than expected at random. We calculated the conservative Benjamini-Hochberg FDR (1995) for the data, and determined the FDR based on all *P*-values for all marker and trait combinations. As can been seen, at an alpha level of 0.0001, which is extensively used in this paper, the FDR is a bit more than 4% (Supplementary Table S3).

4.9. Enrichment analysis

Trans QTL bands were defined as having a significant number of probesets having linkage with 5 cM of a genomic position. The transbands identified were in 3E, 30AB, 57F, 65A, and broadly 72A-87F in the lead-treated flies and at 27B, 50D/F, and broadly 70C–85F in the control flies. The probesets with at least a point wise significance of $P < 0.005$ were selected for enrichment analysis (Khatri et al., 2002). The DAVID (Dennis et al., 2003; Hosack et al., 2003) program was used to look for enrichment of classes of genes with linkage to these positions (3E (174 significant), 30AB (150 significant), 50DF (70 significant), and 65A (41 significant). A *P*-value based upon the hypergeometric distribution and at least 5 probesets with a particular GO term being linked to a region was required to define significant enrichment. Supplementary Tables S2 and S3 show the complete results.

4.10. Identification of potential false positive cis-eQTL

One concern with using Affymetrix short oligonucleotide arrays is that a probeset containing 14 perfect match probes could contain one or more probes that has a SNP in either *ORE* or *Russian 2b*. If this is not taken into account, this could result in false positive *cis*-eQTL. Therefore, we implemented mixed-model analysis of variance in SAS (Wolfinger et al., 2001) to identify potential false positive *cis*-eQTL in a method similar to that described by Gibson and colleagues (Hsieh et al., 2003). However, we did not treat the probe as a random effect since we needed to test the probe-by-fly line effect. The model we fit is: Probe intensity = treatment + fly line + probe + (probe) × (fly line) + error. We tested the interaction term, (probe) × (fly line), and probe intensity were ln(natural log) transformed and normalized as previously described (Hsieh et al., 2003). Finally, we eliminated *cis*-eQTLs that had a "snp_pvalue" of greater than 0.01 as a potential false positive *cis*-eQTL. This approach to improve the data resulted in elimination of ~25% of the *cis*-eQTL as potential false positive *cis*-eQTL (see Supplementary Table S2).

4.11. eQTL mapping: experimental design and data acquisition

Mackay's laboratory described the generation of recombinant inbred lines (RILs) of *D. melanogaster* from homozygous strains *ORE* and *Russian 2b* (Nuzhdin et al., 1997). Briefly, F_1 progeny were backcrossed to *Russian 2b* and the backcross progeny were randomly mated for four generations. At the fifth generation, 200 pairs were selected and progeny were inbred by full-sib mating for 25 generations.

The genetic map consisted of 80 transposable element insertion sites on chromosomes X, 2, and 3. The X chromosome was represented by 17 markers from cytological positions 1B to 19A. Chromosome 2 was represented by 20 markers from cytological positions 21E to 50DF and by three markers from cytological positions 57C to 60E. As no linkage disequilibrium was detected between markers at 50DF and 57C, chromosome 2 was artificially divided into two linkage groups, 2 and 2*. Chromosome 3 was represented by 40 markers from cytological positions 61A to 100A. Chromosome 4 was excluded from our analyses.

Conflict of interest

None.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuro.2009.08.011.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of Drosophila melanogaster. Science 2000;287:2185–95.

Akins JM, Schroeder JA, Brower DL, Aposhian HV. Evaluation of Drosophila melanogaster as an alternative animal for studying the neurotoxicity of heavy metals. Biometals 1992;5:111–20.

Alber SA, Strupp BJ. An in-depth analysis of lead effects in a delayed spatial alternation task: assessment of mnemonic effects, side bias, and proactive interference. Neurotoxicol Teratol 1996;18:3–15.

Ashburner M. Drosophila: a laboratory handbook. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 1989 Edition.

Balamurugan K, Egli D, Hua H, Rajaram R, Seisenbacher G, Georgiev O, et al. Copper homeostasis in Drosophila by complex interplay of import, storage and behavioral avoidance. EMBO J 2007;26:1035–44.

Balamurugan K, Schaffner W. Copper homeostasis in eukaryotes: teetering on a tightrope. Biochim Biophys Acta 2006;1763:737–46.

Bao L, Peirce JL, Zhou M, Li H, Goldowitz D, Williams RW, et al. An integrative genomics strategy for systematic characterization of genetic loci modulating phenotypes. Hum Mol Genet 2007;16:1381–90.

Barth M, Hirsch HV, Meinertzhagen IA, Heisenberg M. Experience-dependent developmental plasticity in the optic lobe of Drosophila melanogaster. J Neurosci 1997a;17:1493–504.

Barth M, Hirsch HVB, Heisenberg M. Rearing in different light regimes affects courtship behaviour in Drosophila melanogaster. Anim Behav 1997b;53:25–38.

Bate M, Martinez Arias A. The development of Drosophila melanogaster. Plainview, N.Y.: Cold Spring Harbor Laboratory Press; 1993 Edition.

Bloom JS, Khan Z, Kruglyak L, Singh M, Caudy AA. Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. BMC Genomics 2009;10:221.

Breitling R, Li Y, Tesson BM, Fu J, Wu C, Wiltshire T, et al. Genetical genomics: spotlight on QTL hotspots. PLoS Genet 2008;4:e1000232.

Brem RB, Yvert G, Clinton R, Kruglyak L. Genetic dissection of transcriptional regulation in budding yeast. Science 2002;296:752–5.

Broman KW. Mapping expression in randomized rodent genomes. Nature Genetics 2005;37:209–10.

Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. Bioinformatics 2003;19:889–90.

Chang YF, Teng HC, Cheng SY, Wang CT, Chiou SH, Kao LS, et al. Orai1-STIM1 formed store-operated Ca(2+) channels (SOCs) as the molecular components needed for Pb(2+) entry in living cells. Toxicol Appl Pharmacol 2008;227:430–9.

Chen WH, Wang M, Yu SS, Su L, Zhu DM, She JQ, et al. Clioquinol and vitamin B12 (cobalamin) synergistically rescue the lead-induced impairments of synaptic plasticity in hippocampal dentate gyrus area of the anesthetized rats in vivo. Neuroscience 2007;147:853–64.

Cheng J, Sun S, Tracy A, Hubbell E, Morris J, Valmeekam V, et al. NetAffx Gene Ontology Mining Tool: a visual approach for microarray data analysis. Bioinformatics 2004;20:1462–3.

Chesler EJ, Lu L, Shou S, Qu Y, Gu J, Wang J, et al. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nature Genetics 2005;37:233–42.

Chintapalli VR, Wang J, Dow JA. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. Nat Genet 2007;39:715–20.

Chisholm AD, Hardin J. Epidermal morphogenesis. WormBook 2005;1–22.

Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J, et al. The Collaborative Cross, a community resource for the genetic analysis of complex traits. Nat Genet 2004;36:1133–7.

Churchill GA, Doerge RW. Empirical threshold values for quantitative trait mapping. Genetics 1994;138:963–71.

Clark AG, Keith LE. Variation among extracted lines of Drosophila melanogaster in triacylglycerol and carbohydrate storage. Genetics 1988;119:595–607.

Coe KJ, Jia Y, Ho HK, Rademacher P, Bammler TK, Beyer RP, et al. Comparison of the cytotoxicity of the nitroaromatic drug flutamide to its cyano analogue in the hepatocyte cell line TAMH: evidence for complex I inhibition and mitochondrial dysfunction using toxicogenomic screening. Chem Res Toxicol 2007;20:1277–90.

Cohn J, Widzowski DV, Cory-Slechta DA. Lead retards development of Drosophila melanogaster. Comp Biochem Physiol C 1992;102:45–9.

Cooper GP, Manalis RS. Influence of heavy metals on synaptic transmission: a review. Neurotoxicology 1983;4:69–83.

Cooper GP, Suszkiw JB, Manalis RS. Heavy metals: effects on synaptic transmission. Neurotoxicology 1984;5:247–66.

Counter SA, Buchanan LH, Ortega F, Rifai N. Blood lead and hemoglobin levels in Andean children with chronic lead intoxication. Neurotoxicology 2000;21:301–8.

de Koning DJ, Haley CS. Genetical genomics in humans and model organisms. Trends Genet 2005;21:377–81.

De Luca M, Yi N, Allison DB, Leips J, Ruden DM. Mapping quantitative trait loci affecting variation in Drosophila triacylglycerol storage. Obes Res 2005;13:1–10.

Demogines A, Smith E, Kruglyak L, Alani E. Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast. PLoS Genet 2008;4:e1000123.

Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. Genome Biol 2003;4:P3.

Dietrich KN, Ris MD, Succop PA, Berger OG, Bornschein RL. Early exposure to lead and juvenile delinquency. Neurotoxicol Teratol 2001;23:511–8.

Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, et al. Application of a translational profiling approach for the comparative analysis of CNS cell types. Cell 2008;135:749–62.

Druka A, Druka I, Centeno AG, Li H, Sun Z, Thomas WT, et al. Towards systems genetic analyses in barley: Integration of phenotypic, expression and genotype data into GeneNetwork. BMC Genet 2008;9:73.

Egli D, Domenech J, Selvaraj A, Balamurugan K, Hua H, Capdevila M, et al. The four members of the Drosophila metallothionein family exhibit distinct yet overlapping roles in heavy metal homeostasis and detoxification. Genes Cells 2006a;11:647–58.

Egli D, Yepiskoposyan H, Selvaraj A, Balamurugan K, Rajaram R, Simons A, et al. A family knockout of all four Drosophila metallothioneins reveals a central role in copper homeostasis and detoxification. Mol Cell Biol 2006b;26:2286–96.

Falconer DS, Mackay TF. Introduction to quantitative genetics. 4th ed. 1996 Editoin Edition.

Flint J, Valdar W, Shifman S, Mott R. Strategies for mapping and cloning quantitative trait genes in rodents. Nat Rev Genet 2005;6:271–86.

Foster WR, Chen SJ, He A, Truong A, Bhaskaran V, Nelson DM, et al. A retrospective analysis of toxicogenomics in the safety assessment of drug candidates. Toxicol Pathol 2007;35:621–35.

Fradkin LG, Kamphorst JT, DiAntonio A, Goodman CS, Noordermeer JN. Genomewide analysis of the Drosophila tetraspanins reveals a subset with similar function in the formation of the embryonic synapse. Proc Natl Acad Sci USA 2002;99:13663–8.

Garavan H, Morgan RE, Levitsky DA, Hermer-Vazquez L, Strupp BJ. Enduring effects of early lead exposure: evidence for a specific deficit in associative ability. Neurotoxicol Teratol 2000;22:151–64.

Gibson G. The environmental contribution to gene expression profiles. Nat Rev Genet 2008;9:575–81.

Gilbert ME, Kelly ME, Samsam TE, Goodman JH. Chronic developmental lead exposure reduces neurogenesis in adult rat hippocampus but does not impair spatial learning. Toxicol Sci 2005;86:365–74.

Gurganus MC, Fry JD, Nuzhdin SV, Pasyukova EG, Lyman RF, Mackay TF. Genotype-environment interaction at quantitative trait loci affecting sensory bristle number in Drosophila melanogaster. Genetics 1998;149:1883–98.

Gurganus MC, Nuzhdin SV, Leips JW, Mackay TF. High-resolution mapping of quantitative trait loci for sternopleural bristle number in Drosophila melanogaster. Genetics 1999;152:1585–604.

Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 1992;69:315–24.

He T, Hirsch HV, Ruden DM, Lnenicka GA. Chronic lead exposure alters presynaptic calcium regulation and synaptic facilitation in Drosophila larvae. Neurotoxicology 2009, in press, doi:10.1016/j.neuro.2009.08.007.

Heiman M, Schaefer A, Gong S, Peterson JD, Day M, Ramsey KE, et al. A translational profiling approach for the molecular characterization of CNS cell types. Cell 2008;135:738–48.

Hirsch HV, Mercer J, Sambaziotis H, Huber M, Stark DT, Torno-Morley T, et al. Behavioral effects of chronic exposure to low levels of lead in Drosophila melanogaster. Neurotoxicology 2003;24:435–42.

Hirsch HVB, Barth M, Sun L, Sambaziotis H, Huber M, Possidente D, et al. Early visual experience affects mate choice of Drosophila melanogaster. Anim Behav 1995;50:1211–7.

Hirsch HVB, Ghiradella H. Environmental influences on behavioral development in insects. In: Capinera J, editor. Encyclopedia of entomology. Dordrecht, Netherlands: Kluwer Press; 2004.

Hirsch HVB, Possidente D, Averill S, Palmetto Despain T, Buytkins J, Thomas V, et al. Variations at a quantitative trait locus (QTL) affect development of behavior in lead-exposed Drosophila melanogaster. Neurotoxicology 2009;30:305–11.

Hosack DA, Dennis G Jr, Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome Biol 2003;4:R70.

Hovatta I, Zapala MA, Broide RS, Schadt EE, Libiger O, Schork NJ, et al. DNA variation and brain region-specific expression profiles exhibit different relationships between inbred mouse strains: implications for eQTL mapping studies. Genome Biol 2007;8:R25.

Hsieh WP, Chu TM, Wolfinger RD, Gibson G. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. Genetics 2003;165:747–57.

Huang L, Shanker YG, Dubauskaite J, Zheng JZ, Yan W, Rosenzweig S, et al. Ggamma13 colocalizes with gustducin in taste receptor cells and mediates IP3 responses to bitter denatonium. Nat Neurosci 1999;2:1055–62.

Jansen RC, Nap JP. Genetical genomics: the added value from segregation. Trends Genet 2001;17:388–91.

Jett DA, Kuhlmann AC, Guilarte TR. Intrahippocampal administration of lead (Pb) impairs performance of rats in the Morris water maze. Pharmacol Biochem Behav 1997;57:263–9.

Jordan KW, Morgan TJ, Mackay TF. Quantitative trait loci for locomotor behavior in *Drosophila melanogaster*. Genetics 2006;174:271–84.

Jusko TA, Henderson CR, Lanphear BP, Cory-Slechta DA, Parsons PJ, Canfield RL. Blood lead concentrations <10 μg/dL and child intelligence at 6 years of age. Environ Health Perspect 2008;116:243–8.

Khatri P, Draghici S, Ostermeier GC, Krawetz SA. Profiling gene expression using onto-express. Genomics 2002;79:266–70.

Kim K, Page GP, Beasley TM, Barnes S, Scheirer KE, Allison DB. A proposed metric for assessing the measurement quality of individual microarrays. BMC Bioinformatics 2006;7:35–45.

Kiraly E, Jones DG. Dendritic spine changes in rat hippocampal pyramidal cells after postnatal lead treatment: a Golgi study. Exp Neurol 1982;77:236–9.

Koh YH, Gramates LS, Budnik V. Drosophila larval neuromuscular junction: molecular components and mechanisms underlying synaptic plasticity. Microsc Res Tech 2000;49:14–25.

Kopczynski CC, Davis GW, Goodman CS. A neural tetraspanin, encoded by late bloomer, that facilitates synapse formation. Science 1996;271:1867–70.

Kordas K, Canfield RL, Lopez P, Rosado JL, Vargas GG, Cebrian ME, et al. Deficits in cognitive function and achievement in Mexican first-graders with low blood lead concentrations. Environ Res 2006;100:371–86.

Lasky RE, Laughlin NK. Exploring a partially enclosed space by lead-exposed female rhesus monkeys. Neurotoxicol Teratol 2001;23:177–83.

Leips J, Mackay TF. Quantitative trait loci for life span in *Drosophila melanogaster*: interactions with genetic background and larval density. Genetics 2000;155:1773–88.

Li H, Lu L, Manly KF, Chesler EJ, Bao L, Wang J, et al. Inferring gene transcriptional modulatory relations: a genetical genomics approach. Hum Mol Genet 2005;14:1119–25.

Li J, Burmeister M. Genetical genomics: combining genetics with gene expression analysis. Hum Mol Genet 2005;14(Spec. 2):R163–9.

Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JA, et al. Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. PLoS Genet 2006;2:e222.

Li Y, Breitling R, Jansen RC. Generalizing genetical genomics: getting added value from environmental perturbation. Trends Genet 2008;24:518–24.

Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, et al. NetAffx: Affymetrix probesets and annotations. Nucleic Acids Res 2003;31:82–6.

Long AD, Mullaney SL, Reid LA, Fry JD, Langley CH, Mackay TF. High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. Genetics 1995;139:1273–91.

Marchetti C. Molecular targets of lead in brain neurotoxicity. Neurotox Res 2003;5:221–36.

McGrath PT, Rockman MV, Zimmer M, Jang H, Macosko EZ, Kruglyak L, et al. Quantitative mapping of a digenic behavioral trait implicates globin variation in *C. elegans* sensory behaviors. Neuron 2009;61:692–9.

Mehta SQ, Hiesinger PR, Beronja S, Zhai RG, Schulze KL, Verstreken P, et al. Mutations in Drosophila sec15 reveal a function in neuronal targeting for a subset of exocyst components. Neuron 2005;46:219–32.

Min JY, Min KB, Cho SI, Kim R, Sakong J, Paek D. Neurobehavioral function in children with low blood lead concentrations. Neurotoxicology 2007;28:421–5.

Ming GL, Song H. Adult neurogenesis in the mammalian central nervous system. Annu Rev Neurosci 2005;28:223–50.

Moreira EG, Vassilieff I, Vassilieff VS. Developmental lead exposure: behavioral alterations in the short and long term. Neurotoxicol Teratol 2001;23:489–95.

Morgan RE, Levitsky DA, Strupp BJ. Effects of chronic lead exposure on learning and reaction time in a visual discrimination task. Neurotoxicol Teratol 2000;22:337–45.

Morley EJ, Hirsch HV, Hollocher K, Lnenicka GA. Effects of chronic lead exposure on the neuromuscular junction in Drosophila larvae. Neurotoxicology 2003;24:35–41.

Needham LL, Barr DB, Caudill SP, Pirkle JL, Turner WE, Osterloh J, et al. Concentrations of environmental chemicals associated with neurodevelopmental effects in U.S. population. Neurotoxicology 2005;26:531–45.

Nuzhdin SV, Keightley PD, Pasyukova EG. The use of retrotransposons as markers for mapping genes responsible for fitness differences between related *Drosophila melanogaster* strains. Genet Res 1993;62:125–31.

Nuzhdin SV, Pasyukova EG, Dilda CL, Zeng ZB, Mackay TF. Sex-specific quantitative trait loci affecting longevity in *Drosophila melanogaster*. Proc Natl Acad Sci USA 1997;94:9734–9.

Page GP, Ruden DM. Combining high dimensional biological data to study complex diseases and quantitative traits—design, analysis, and interpretation of results. Boca Raton, FL: Chapman & Hall/CRC; 2005 Editoin Edition.

Passador-Gurgel G, Hsieh WP, Hunt P, Deighton N, Gibson G. Quantitative trait transcripts for nicotine resistance in *Drosophila melanogaster*. Nat Genet 2007;39:264–8.

Pasyukova EG, Vieira C, Mackay TF. Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. Genetics 2000;156:1129–46.

Peirce JL, Lu L, Gu J, Silver LM, Williams RW. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. BMC Genet 2004;5:7.

Perez-Enciso M. In silico study of transcriptome genetic variation in outbred populations. Genetics 2004;166:547–54.

Perlstein EO, Ruderfer DM, Roberts DC, Schreiber SL, Kruglyak L. Genetic basis of individual differences in the response to small-molecule drugs in yeast. Nat Genet 2007;39:496–502.

Persson S, Wei H, Milne J, Page GP, Somerville CR. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc Natl Acad Sci USA 2005;102:8633–8.

Petit TL, LeBoutillier JC. Effects of lead exposure during development on neocortical dendritic and synaptic structure. Exp Neurol 1979;64:482–92.

Pienta K, Esper P. Risk factors for prostate cancer. Ann Int Med 1993;118:793–803.

Podbilewicz B. Cell fusion. WormBook 2006;1–32.

Reddy KC, Andersen EC, Kruglyak L, Kim DH. A polymorphism in npr-1 is a behavioral determinant of pathogen susceptibility in *C. elegans*. Science 2009;323: 382–4.

Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet 2006;7:862–72.

Ruden DM. Personalized medicine and quantitative trait transcripts. Nat Genet 2007;39:144–5.

Schneider JS, Decamp E. Postnatal lead poisoning impairs behavioral recovery following brain damage. Neurotoxicology 2007;28:1153–7.

Schubert ML. Gastric secretion. Curr Opin Gastroenterol 2008;24:659–64.

Selimi F, Cristea IM, Heller E, Chait BT, Heintz N. Proteomic studies of a single CNS synapse type: the parallel fiber/purkinje cell synapse. PLoS Biol 2009;7:e83.

Sinenko SA, Mathey-Prevot B. Increased expression of Drosophila tetraspanin, Tsp68C, suppresses the abnormal proliferation of ytr-deficient and Ras/Raf-activated hemocytes. Oncogene 2004;23:9120–8.

Smith EN, Kruglyak L. Gene-environment interaction in yeast gene expression. PLoS Biol 2008;6:e83.

Stewart PW, Sargent DM, Reihman J, Gump BB, Lonky E, Darvill T, et al. Response inhibition during differential reinforcement of low rates (DRL) schedules may be sensitive to low-level polychlorinated biphenyl, methylmercury, and lead exposure in children. Environ Health Perspect 2006;114:1923–9.

Strickler-Shaw S, Taylor DH. Lead inhibits acquisition and retention learning in bullfrog tadpoles. Neurotoxicol Teratol 1991;13:167–73.

Sun L, Zhao ZY, Hu J, Zhou XL. Potential association of lead exposure during early development of mice with alteration of hippocampus nitric oxide levels and learning memory. Biomed Environ Sci 2005;18:375–8.

Surkan PJ, Zhang A, Trachtenberg F, Daniel DB, McKinlay S, Bellinger DC. Neuropsychological function in children with blood lead levels <10 μg/dL. Neurotoxicology 2007;28:1170–7.

Tang HW, Liang YX, Hu XH. Effects of low level lead exposure on behavior of young rats. Zhongguo Yao Li Xue Bao 1994;15:316–9.

Todres E, Nardi JB, Robertson HM. The tetraspanin superfamily in insects. Insect Mol Biol 2000;9:581–90.

Toscano CD, Guilarte TR. Lead neurotoxicity: from exposure to molecular effects. Brain Res Brain Res Rev 2005;49:529–54.

Urbach R, Technau GM. Early steps in building the insect brain: neuroblast formation and segmental patterning in the developing brain of different insect species. Arthropod Struct Dev 2003;32:103–23.

Verina T, Rohde CA, Guilarte TR. Environmental lead exposure during early life alters granule cell neurogenesis and morphology in the hippocampus of young adult rats. Neuroscience 2007;145:1037–47.

Vieira C, Pasyukova EG, Zeng ZB, Hackett JB, Lyman RF, Mackay TF. Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. Genetics 2000;154:213–27.

Wayne ML, Hackett JB, Dilda CL, Nuzhdin SV, Pasyukova EG, Mackay TF. Quantitative trait locus mapping of fitness-related traits in *Drosophila melanogaster*. Genet Res 2001;77:107–16.

White LD, Cory-Slechta DA, Gilbert ME, Tiffany-Castiglioni E, Zawia NH, Virgolini M, et al. New and evolving concepts in the neurotoxicology of lead. Toxicol Appl Pharmacol 2007;225:1–27.

Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh P, Bushel P, Afshari C, Paules RS. Assessing gene significance from cDNA microarray expression data via mixed models. J Comp Biol 2001;8:625–37.

Ye HY, Ye BP, Wang DY. Molecular control of memory in nematode *Caenorhabditis elegans*. Neurosci Bull 2008;24:49–55.

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, et al. Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. Nat Genet 2003;35:57–64.

Zakharkin SO, Kim K, Mehta T, Chen L, Barnes S, Scheirer KE, et al. Sources of variation in Affymetrix microarray experiments. BMC Bioinformatics 2005;6:214.