



Identification of Splicing Quantitative Trait Loci (sQTL) in *Drosophila melanogaster* with Developmental Lead (Pb²⁺) Exposure

Wen Qu¹, Katherine Gurdziel², Roger Pique-Regi³ and Douglas M. Ruden^{1,3,4*}

¹ Laboratory of Epigenomics, Department of Pharmacology, C.S. Mott Center for Human Growth and Development, Wayne State University, Detroit, MI, United States, ² Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI, United States, ³ Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, United States, ⁴ Institute of Environmental Health Sciences, Wayne State University, Detroit, MI, United States

OPEN ACCESS

Edited by:

Naoyuki Kataoka,
The University of Tokyo, Japan

Reviewed by:

Daisuke Kaida,
University of Toyama, Japan
Kunio Inoue,
Kobe University, Japan

*Correspondence:

Douglas M. Ruden
douglasr@wayne.edu

Specialty section:

This article was submitted to
RNA,
a section of the journal
Frontiers in Genetics

Received: 04 August 2017

Accepted: 22 September 2017

Published: 24 October 2017

Citation:

Qu W, Gurdziel K, Pique-Regi R and Ruden DM (2017) Identification of Splicing Quantitative Trait Loci (sQTL) in *Drosophila melanogaster* with Developmental Lead (Pb²⁺) Exposure. *Front. Genet.* 8:145. doi: 10.3389/fgene.2017.00145

Lead (Pb) poisoning has been a major public health issue globally and the recent Flint water crisis has drawn nation-wide attention to its effects. To better understand how lead plays a role as a neurotoxin, we utilized the *Drosophila melanogaster* model to study the genetic effects of lead exposure during development and identified lead-responsive genes. In our previous studies, we have successfully identified hundreds of lead-responsive expression QTLs (eQTLs) by using RNA-seq analysis on heads collected from the *Drosophila* Synthetic Population Resource. *Cis*-eQTLs, also known as allele-specific expression (ASE) polymorphisms, are generally single-nucleotide polymorphisms in the promoter regions of genes that affect expression of the gene, such as by inhibiting the binding of transcription factors. *Trans*-eQTLs are genes that regulate mRNA levels for many genes, and are generally thought to be SNPs in *trans*-acting transcription or translation factors. In this study, we focused our attention on alternative splicing events that are affected by lead exposure. Splicing QTLs (sQTLs), which can be caused by SNPs that alter splicing or alternative splicing (AS), such as by changing the sequence-specific binding affinity of splicing factors to the pre-mRNA. We applied two methods in search for sQTLs by using RNA-seq data from control and lead-exposed w¹¹¹⁸ *Drosophila* heads. First, we used the fraction of reads in a gene that falls in each exon as the phenotype. Second, we directly compared the transcript counts among the various splicing isoforms as the phenotype. Among the 1,236 potential Pb-responsive sQTLs ($p < 0.0001$, FDR < 0.39), mostly *cis*-sQTLs, one of the most distinct genes is *Dscam1* (Down Syndrome Cell Adhesion Molecule), which has over 30,000 potential alternative splicing isoforms. We have also identified a candidate Pb-responsive *trans*-sQTL hotspot that appears to regulate 129 genes that are enriched in the “cation channel” gene ontology category, suggesting a model in which alternative splicing of these channels might lead to an increase in the elimination of Pb²⁺ from the neurons encoding these channels. To our knowledge, this is the first paper that uses sQTL analyses to understand the neurotoxicology of an environmental toxin in any organism, and the first reported discovery of a candidate *trans*-sQTL hotspot.

Keywords: developmental lead exposure, *Drosophila melanogaster*, splicing quantitative trait loci (sQTL or splice QTL), *Drosophila* synthetic population resource (DSPR), RNA-seq, toxicogenomics

INTRODUCTION

Lead Burdens

Although, the phase-out of leaded paint and gasoline has substantially reduced mean blood lead levels in the United States (White et al., 2007), lead contamination in the city of Detroit and the neighboring city of Flint have been of extreme concern in the past two years due to the Flint water crisis (Hanna-Attisha et al., 2015). In a recent study, we found multigenerational epigenetic effects on DNA methylation in Detroit children (Senut et al., 2012; Sen et al., 2015a,b). Additionally, lead exposure from environmental contamination remains a major global public health issue (Tong et al., 2000; Rauh and Margolis, 2016).

Our lab has been studying the genetics of lead neurotoxicology using the *Drosophila melanogaster* model (Hirsch et al., 2012). In our genetic and physiology studies, 250 μ M lead acetate in the standard fly food causes adult soft-tissue lead levels to be 50–100 μ g/dL (Hirsch et al., 2009). This is in the high range of human lead exposure, and the currently CDC level of concern for lead exposure is 5 μ g/dL (Bellinger, 2013). We found that lower levels of lead exposure, as low as 50 μ M, can significantly alter the *Drosophila* larval neuromuscular junction (Morley et al., 2003; He et al., 2009a,b). In the current studies, we use 250 μ M developmental exposure to lead in the food because it consistently affects synaptic (He et al., 2009a), behavioral (Hirsch et al., 2009), and gene expression changes (Ruden et al., 2009). In our previous *Drosophila* analyses, we found that lead exposure modified the uniformity of the synaptic matching between axons and muscle fibers during larval development (Morley et al., 2003) and it also changes a variety of behavioral phenotypes such as courtship (Hirsch et al., 1995) and circadian locomotor activity (Hirsch et al., 2003). Additionally, our lab has used microarrays and RNA-seq technology along with eQTL analysis to map lead-sensitive genes in *Drosophila* (Ruden et al., 2009). Therefore, we are confident that *Drosophila* is a useful model to understand some of the mechanisms for how developmental lead exposure to lead affects gene expression and splicing in neurons, some of which are likely to be conserved in humans.

Alternative Splicing

After the discovery of splicing in the Adenovirus *hexon* gene in 1977 (Sambrook, 1977), Walter Gilbert proposed in 1978 that different combinations of exons and introns, namely “alternative splicing” (AS), could produce different mRNA isoforms of a gene (Gilbert, 1978; Modrek and Lee, 2002). The disparity between the expected 150,000 or more human genes and the surprising actual report of under 32,000 later suggested an underestimated role for alternative splicing in the production of an increased variety of mRNAs and proteins (Pennisi, 2000; Venter et al., 2001). It has been estimated that AS is a crucial form of gene regulation affecting about 60–90% of human genes (Modrek and Lee, 2002) and over 40% of *Drosophila* genes (Stolc et al., 2004). Mutations that affect mRNA splicing and AS were also considered to be highly linked with disease occurrences (Singh and Cooper, 2012). In addition, it has been estimated that 15%

of human disease mutations lie within splicing sites and 22% of disease-related SNPs may affect splicing (Krawczak et al., 2007; Lim et al., 2011).

The *Drosophila Dscam1* gene exemplifies one of the most extreme examples of alternative splicing. *Dscam1* (Down Syndrome Cell Adhesion Molecule 1) is a cell surface protein which gives rise to over 30,000 potential alternatively spliced isoforms in the *Drosophila* nervous system (Schmucker and Flanagan, 2004). The human homolog *DSCAM* was first discovered as a candidate disease gene for the central and peripheral nervous system defects associated with Down syndrome (Yamakawa et al., 1998). The *Drosophila Dscam1* was later found to have extreme structural diversity and is essential for neural circuit assembly (Schmucker et al., 2000; Hattori et al., 2007). Its diversity allowed each neuron to have a unique pattern on its cell membrane, which made self-recognition possible (Zipursky and Grueber, 2013; Armitage et al., 2015). It has also been shown that *Dscam1* regulated interactions between neurons through isoform-specific homophilic binding or repulsion (Wojtowicz et al., 2004; Tadros et al., 2016). Additionally, its role in the insect cellular immune system has been suggested since 2005 (Watson et al., 2005). Even after years of study, many questions remain unanswered, such as how *Dscam1* mRNA isoforms are selectively expressed and how homophilic interactions are translated into binding or repulsing responses during neurogenesis (Schmucker and Flanagan, 2004).

Quantitative Trait Locus

A quantitative trait locus (QTL) is a sequence of DNA (the locus) that is associated with variation in a phenotype (the quantitative trait) (Miles and Wayne, 2008). In the case of expression QTLs (eQTLs) and splicing QTLs (sQTLs) (Yoo et al., 2016; Mei et al., 2017; Takata et al., 2017; Yang et al., 2017), the quantitative traits are relative gene expression levels and relative splicing isoform abundance. All QTLs, no matter the type, are typically distributed in a normal distribution in the population. Significant eQTLs and sQTLs could be categorized into two groups: *cis*- and *trans*-eQTLs and sQTLs. Here, in this paper, we focus on *cis*- and *trans*-sQTLs. By definition, *cis*-sQTLs are referred as genetic variants that affect the splicing event of a locus only on the same haplotype, while *trans*-sQTLs affect multiple haplotypes (Benzer, 1955; Hasin-Brumshtein et al., 2014). Therefore, *cis*-sQTLs tend to be “local,” adjacent to the transcript location, while *trans*-sQTLs tend to be “distal,” away from the regulator (Benzer, 1955; Hasin-Brumshtein et al., 2014). For the purposes of this paper, an sQTL is defined as genetic variants that are associated with changes in the splicing ratios of transcripts (Monlong et al., 2014).

In our previous studies, we described the identification of *cis*- and *trans*-eQTLs in *Drosophila*. We identified lead-responsive *cis*-eQTLs and *trans*-eQTLs by using Affymetrix *Drosophila* gene expression microarrays in 2009 (Ruden et al., 2009). At that time, we also found eQTLs linked with developmental behavioral effects of lead exposure (Hirsch et al., 2009). In a more recent study, we used RNA-seq technology to quantify expression profiles and ran a similar analysis to map lead-sensitive eQTLs

in *Drosophila*. We used *Drosophila* samples that may include more genetic variations to further explore the mechanisms of the Pb-responsive *cis*- and *trans*-eQTLs (manuscript submitted to *Frontiers in Genetics*). Here, in this paper, we use the same set of RNA-seq data to search for Pb-responsive sQTLs. We found hundreds of candidate *cis*-sQTLs, and one major *trans*-sQTL hotspot, among which *Dscam1* was one of the most significant *cis*-sQTLs both at the exon level and at the transcript level.

MATERIALS AND METHODS

Genotyping

The 79 *D. melanogaster* sample lines were from one set of *Drosophila* Synthetic Population resource (DSPR)—Subpopulation A2 (<http://FlyRILs.org>) and the genomic data were downloaded from <http://wfitch.bio.uci.edu/R/> (King et al., 2012). This eight-way synthetic population was first initiated with eight inbred founder lines (A1–A8) with a wide genetic variation. The first generation was by intercrossing A1 with A2 and A2 with A3 until A7 with A8. 10 offspring per genotype per sex were mixed together to establish the next generation. After 50 generations of intercrossing and 25 generations of inbreeding, ~1,600 recombinant inbred lines (RILs) were generated. The Restriction-Associated DNA (RAD) and hidden Markov model (HMM) were used to estimate the genetic contribution of parental lines in each RIL and the genetic data contained the complete set of underlying founder haplotype structure for all RILs (King et al., 2012). All the genotype information was provided by the DSPR group (King et al., 2012).

Sample Collection

Flies were cultured at 25°C in 35 ml vial containing standard *Drosophila* 10 ml medium. Medium was blended with 250 μM PbAc or 250 μM NaAc in order to mimic lead toxicity (Hirsch et al., 2009). Total RNA was extracted from 50 adult male heads

(5–10 days old) of DSPR homozygous recombinant inbred lines by using TruSeq Cluster RNA sample prep kit provided by Illumina and 1 μg of RNA was used after RNA isolation. The D1K ScreenTape on the Agilent TapeStation instrument along with the quantitative PCR on the QuantStudio 12K Flex was used to guarantee the quality of library. Libraries were standardly prepared and sequenced on the HiSeq2000™ instrument from Illumina (50-bp paired end reads). General read quality was examined by FastQC (Andrews, 2010). The average coverage is 23 million read pairs and the RNA-seq data are available on the NCBI GEO accession: GSE83141.

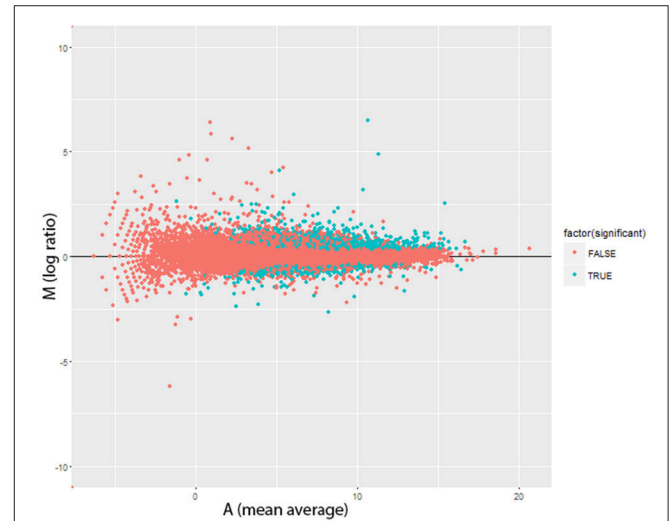


FIGURE 2 | Altered Gene Expression Levels after Lead Treatment in *Drosophila* male head. MA plots for change in transcript expression ($n = 1,682$) comparing Pb-treated ($n = 79$) with control-treated samples ($n = 79$). Red dots represent transcript expression profiles were not significantly changed and cyan dots were significantly regulated transcripts (FDR < 0.0001, 0.450 ± 0.272 mean log₂ fold changes \pm s.d). $M = \log_2(P/C)$, $A = (\log_2(C) + \log_2(P))/2$ (P, Pb-treated FPKM values; C, control FPKM values).

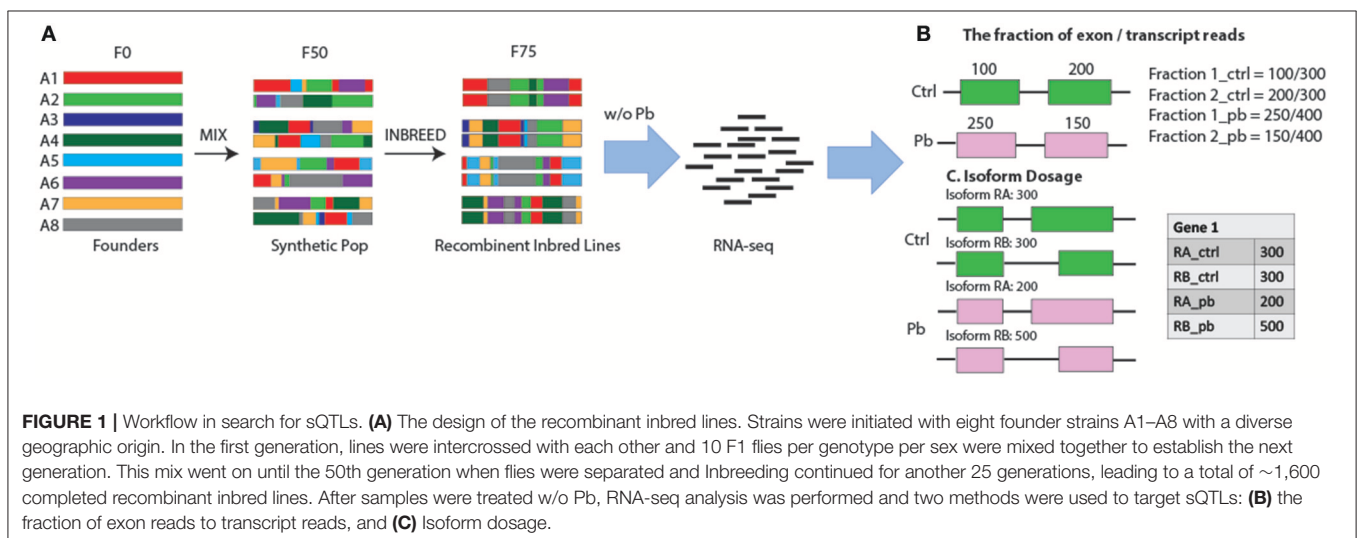


FIGURE 1 | Workflow in search for sQTLs. **(A)** The design of the recombinant inbred lines. Strains were initiated with eight founder strains A1–A8 with a diverse geographic origin. In the first generation, lines were intercrossed with each other and 10 F1 flies per genotype per sex were mixed together to establish the next generation. This mix went on until the 50th generation when flies were separated and Inbreeding continued for another 25 generations, leading to a total of ~1,600 completed recombinant inbred lines. After samples were treated w/o Pb, RNA-seq analysis was performed and two methods were used to target sQTLs: **(B)** the fraction of exon reads to transcript reads, and **(C)** Isoform dosage.

Expression Quantification

RNA-seq reads were mapped to the UCSC/dm3 *D. melanogaster* references genome (track: Flybase Genes) using TopHat (Karolchik et al., 2004; Kim et al., 2013). Upload/Convert ID tool from Flybase.org was used to convert the annotation symbols into official symbols (Tweedie et al., 2009) and Htseq was used to quantify exon and transcript expression (Anders et al., 2014). The RIL information was used as a covariate ($Y \sim \text{treatment} + \text{RIL}$), when performing the differential expression analysis. For the GO enrichment analysis, Goseq was used to search among the differentially expressed genes after lead exposure and genetic programming (GP) categories of “Molecular Function” and “Biological Process” were selected (Kent et al., 2002; Young et al., 2010).

ANOVA Test

All analyses were performed in R-3.2.3. After calculating division of exon reads to its corresponding transcript reads, quantile normalization, and confounding factors were removed by PCA (n.pc = 3), we used the following test to identify sQTLs for each transcript:

$$H_0: Y_n = \mu + \beta^E E_n + \epsilon_n$$

$$H_A: Y_n = \mu + \sum_i (\beta_i^G G_{ij} + \beta_i^{G \times E} G_{ij} E_n) + \beta^E E_n + \epsilon_n$$

where Y_n is the vector of splicing measures, G_{ij} is the i -th parental genotype probability at locus j , E_n is a vector representing two environmental conditions (control or lead-treated). The parameters μ , β^E , β^G , $\beta^{G \times E}$ represent respectively the grand mean, the genotype, the environment and the interaction effects.

For transcript expression levels, transcript reads were quantile normalized, and confounding factors were removed by PCA (n.pc = 4). Expression data for genes that have more than one isoform were subgrouped (max. 3,975). Then for each gene we test for interaction between $G \times E$ interaction in splicing isoform usage using the following model:

$$H_0: Y_{nk} = \mu_k + \sum_{ik} (\beta_{ik}^G G_{ij}) + \sum_k (\beta_k^E T_k E_n) + \epsilon_{nk}$$

$$H_A: Y_{nk} = \mu_k + \sum_{ik} (\beta_{ik}^G G_{ij} T_k + \beta_{ik}^{G \times E} T_k G_{ij} E_n) + \sum_k (\beta_k^E T_k E_n) + \epsilon_{nk}$$

where Y_{nk} is a matrix representing all normalized read counts across individuals n and isoforms k for each gene, G_{ij} is the i -th parental genotype probability at locus j , E_n is a vector representing two environmental conditions (control or lead-treated), T_k is an indicator variable for each isoform k . The parameters μ_k , β_k^E , β_{ik}^G , $\beta_{ik}^{G \times E}$, represent respectively the grand mean, the genotype, the environment and the interaction effects across different isoforms.

After obtaining all the p -values indicating the likelihood of association between each genomic location and each transcript, q -value function in R (library: q -value) was used to transform the p -values into FDRs and we defined $p \leq 0.0001$, corresponding FDRs ≤ 0.39 , as significant signals.

RESULTS

Differential Expression Modification after Lead Exposure

In order to search for sQTLs, we collected RNA-seq data from 79 fly lines from The *Drosophila* Synthetic Population Resource (DSPR) (King et al., 2012). The lines were started with eight founder strains (A1–A8) that come from diverse demographic origins and include several million genetic variations in the *D. melanogaster* species. The eight founder lines were intercrossed for 50 generations and were separated to inbred for additional 25 generations (Figure 1A) (King et al., 2012). We randomly selected 79 lines out of the ~1,600 recombinant inbred lines and treated them with standard fly food plus 250 μ M NaAc or 250 μ M PbAc. Fifty heads of male flies in each sample were collected manually by forceps and dropped into RNALater™, and RNA-seq was performed per the standard protocol (see section Materials and Methods). Thus, we had 158 RNA-seq samples (79 * 2 \pm Pb).

There were significant changes in the gene expression profiles after lead exposure: 1,682 changes among the 20,507 transcripts (14,058 genes), including 187 exhibiting over 50%

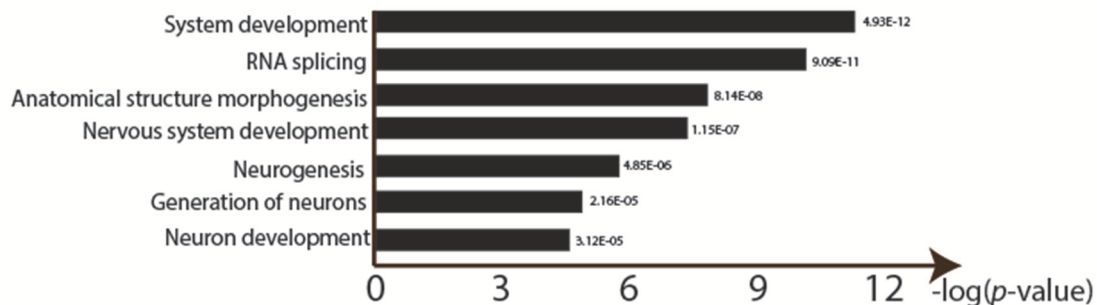


FIGURE 3 | Gene ontology enrichment analysis of lead treatment in *Drosophila* male heads. Goseq was used to detect over represented GO categories in RNA-seq data (FDR < 0.0001). Y-axis shows the logarithm of each significant GO ID’s p -value (after Bonferroni correction for multiple testing).

change in expression levels ($FDR < 0.0001$, 0.450 ± 0.272 mean absolute \log_2 change \pm s.d.; **Figure 2**; Supplemental Table 1). Among the responders, 1,338 transcripts were upregulated after lead treatment and 344 transcripts were downregulated. Among all the significantly regulated transcripts, nervous system development and neurogenesis were among the topmost enriched gene ontology (GO) categories (**Figure 3**). These results correspond with our expectations, since during sample preparation, we only collected *Drosophila* heads and the nervous system development has been regarded as the main target for Pb toxicity (Baranowska-Bosiacka et al., 2012).

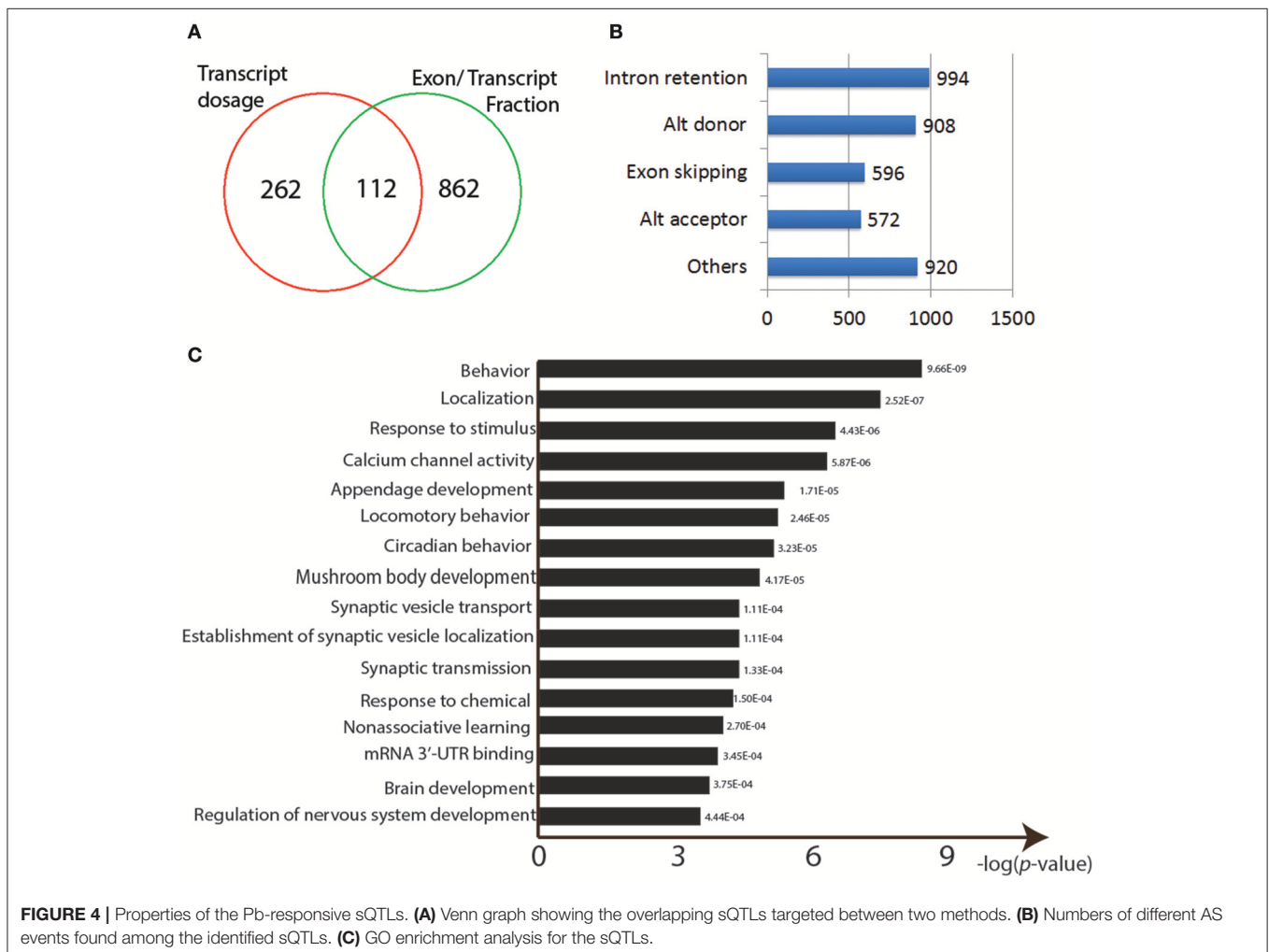
Identification of *cis*-Splicing QTLs

After detecting differentially expressed transcripts affected by Pb treatment, we continued to search for splicing QTLs—the genomic regions where genetic variants affect splicing events. In most sQTL studies (Lappalainen et al., 2013; Battle et al., 2014; Kurmangaliyev et al., 2015; Ongen and Dermitzakis, 2015), SNPs were used to reflect the genomic variations. However, in our study, each fly line was a mosaic of the eight parental lines (A1–A8; **Figure 1A**) and the genetic contribution by the parental genotypes represent the genomic variations. With this type of genotype information, the *cis*-sQTL is defined as a genomic locus

where alternative splicing is associated with a differential parental contribution.

Reads were mapped by Tophat2 (Trapnell et al., 2012; Kim et al., 2013) and quantified to exons and transcripts by Htseq (Anders et al., 2014). Our major goal in this paper is to identify splicing QTLs. However, our RNA-seq data is 50 bp paired-end, which is not optimal for studying RNA splicing. Others have suggested that the ideal length for splicing junction detection is 100 bp, so that sequences can be uniquely identified on both sides of the junction (Chhangawala et al., 2015). Additionally, King et al. mentioned that the power to map a 10% QTL with 100 DSPR lines is potentially about 10% (King et al., 2012). Therefore, it is challenging to map sQTLs, especially when their effects are relatively small, with only 79 RILs.

Keeping these issues in mind, we used two ways to detect sQTLs: (1) we used the fraction of reads in a transcript that falls in each exon as the quantitative trait (**Figure 1B**); and (2) we used the differential transcript isoform dosage in the same gene as the quantitative trait (**Figure 1C**, see section Materials and Methods). Exon/transcript fraction captures changes in reads within each exon, while the second method captures events where both exon reads and transcript reads change, might be missed by the former strategy.

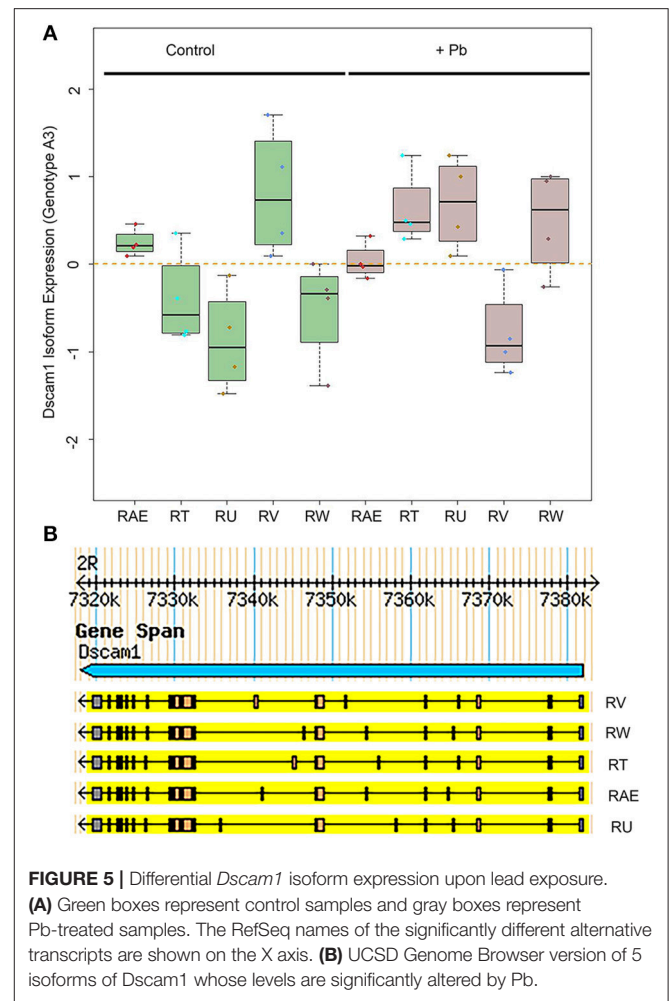


Here, we used ANOVA analysis to detect Pb-responsive sQTLs (see section Materials and Methods; Hoaglin and Welsch, 1978). In total, we obtained 974 potential Pb-responsive sQTLs by calculating exon/transcript fraction and 374 by isoform dosage, with 112 shared ones ($p < 0.0001$, FDR < 0.39; **Figure 4A**; Supplementary Table S1).

We then used the alternative splicing transcriptional landscape visualization tool (ASTALAVISTA; Foissac and Sammeth, 2007) to determine the types of events represented by the entire set of sQTLs (**Figure 4B**). The four main AS types were intron retention ($n = 994$), Alternative donor site usage ($n = 908$), exon skipping ($n = 596$), and alternative acceptor site usage ($n = 572$; **Figure 4B**).

The identified sQTLs were also run through a GO enrichment analysis. The top enriched categories were “behavior” ($p = 9.66E-09$) and “response to stimulus” ($p = 4.43E-06$) and “calcium channel activity” ($p = 5.87E-06$; **Figure 4C**). Neural developmental related GO categories were also among the most significant: “mushroom body development” ($p = 4.17E-05$), “synaptic vesicle transport” ($p = 4.17E-05$), “non-associative learning” ($p = 2.70E-04$), “brain development” ($p = 3.75E-04$), and “regulation of nervous system development” ($p = 4.44E-04$) (**Figure 4C**). Other over-represented GO categories include “locomotory behavior” ($p = 2.46E-05$), “response to chemical” ($p = 1.50E-04$), and “mRNA 3'-UTR binding” ($p = 3.45E-04$) (**Figure 4C**).

One of the most significant sQTLs is *Dscam1* linked with Chr 3L: 2,790,000 (FDR < 0.1%). Transcript alternative splicing patterns from the A3 parent were altered significantly, while others were not (**Figure 5A**). In samples that were inherited from A3 parent at Chr 3L: 2,790,000 locus, RT, RU, and RW isoforms were upregulated after lead exposure, RV was downregulated, while RAE was remained steadily. This suggested that A3 strain responded to lead poisoning differently from the rest of the parents by altering usage of the various isoforms. The structures of the alternative spliced isoforms of RT, RU, RW, RV, and RAE are shown (**Figure 5B**). In order to explore further into the exon usage in *Dscam1*, we used the Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al., 2013), which is a popular visualization tool for integrated genomic data. We noticed that reads for exon 7, 8, 10, and 11 were increased after lead treatment in A3 samples, while in other samples the expression change was in the reversed direction (**Figure 6A**). However, not all exons were affected in the same way. For example, read counts for exon 18 and 19 in all samples were upregulated after lead exposure (**Figure 6A**). We note that it is not the variable exons that are alternatively spliced by the sQTLs we identified in *Dscam1*, but rather the invariant exons that are downstream of the variant exons (**Figure 6B**). It is not known whether changing the levels of the invariant exons in *Dscam1* would alter the splicing of the variant exons, which encode a possible 38,021 different protein products (**Figure 6C**). Nevertheless, it's possible that differential exon usage of *Dscam1*, and the other genes with lead-regulated sQTL, are a part of the compensatory pathway after lead poisoning, but future research is needed to tackle this problem.



Identification of *Trans*-Splicing QTLs

We visualized the distribution of all the significant associations with a comprehensive sQTL map (**Figure 7**). Each dot represents one significant association between the genetic locus shown on the x-axis and the transcript on the y-axis (FDR < 0.39). Among the scattered dots, there was one prominent vertical band, which we call a *trans*-sQTL hotspot, located at Chr3L:18,810,000, consisting of a high density of dots representing 129 different genes located throughout the genome. The *trans*-sQTL hotspot located at Chr3L:18,810,000 is a genomic locus that regulates a cluster of transcripts regardless of their loci. This *trans*-sQTL hotspot is similar to what has been observed in many eQTL studies, where genomic loci that are linked with abnormally high numbers of eQTLs are called *trans*-eQTL hotspots (Joo et al., 2014; King et al., 2014) or *trans*-eQTL bands (Rockman and Kruglyak, 2006). Intriguingly, the *trans*-sQTL hotspot at Chr3L:18,810,000 locus is Pb-responsive ($p < 1E-10$), because it is only present in the presence of Pb and not in the control data. We also found that “ion transport domain” was the topmost GO category (FDR = 0.03) for the list of 129 genes. The six genes in this category and their ion transport functions are listed in **Table 1**. Interestingly, 58 of the 129 genes (45%) undergo

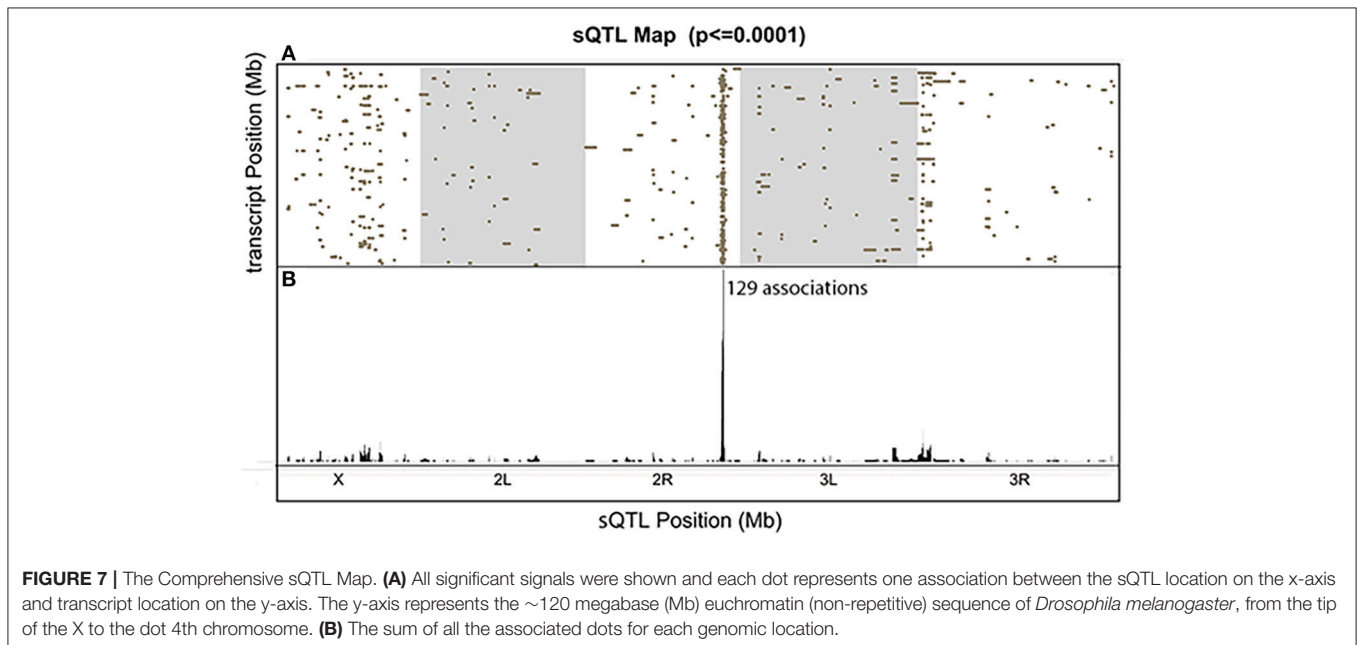
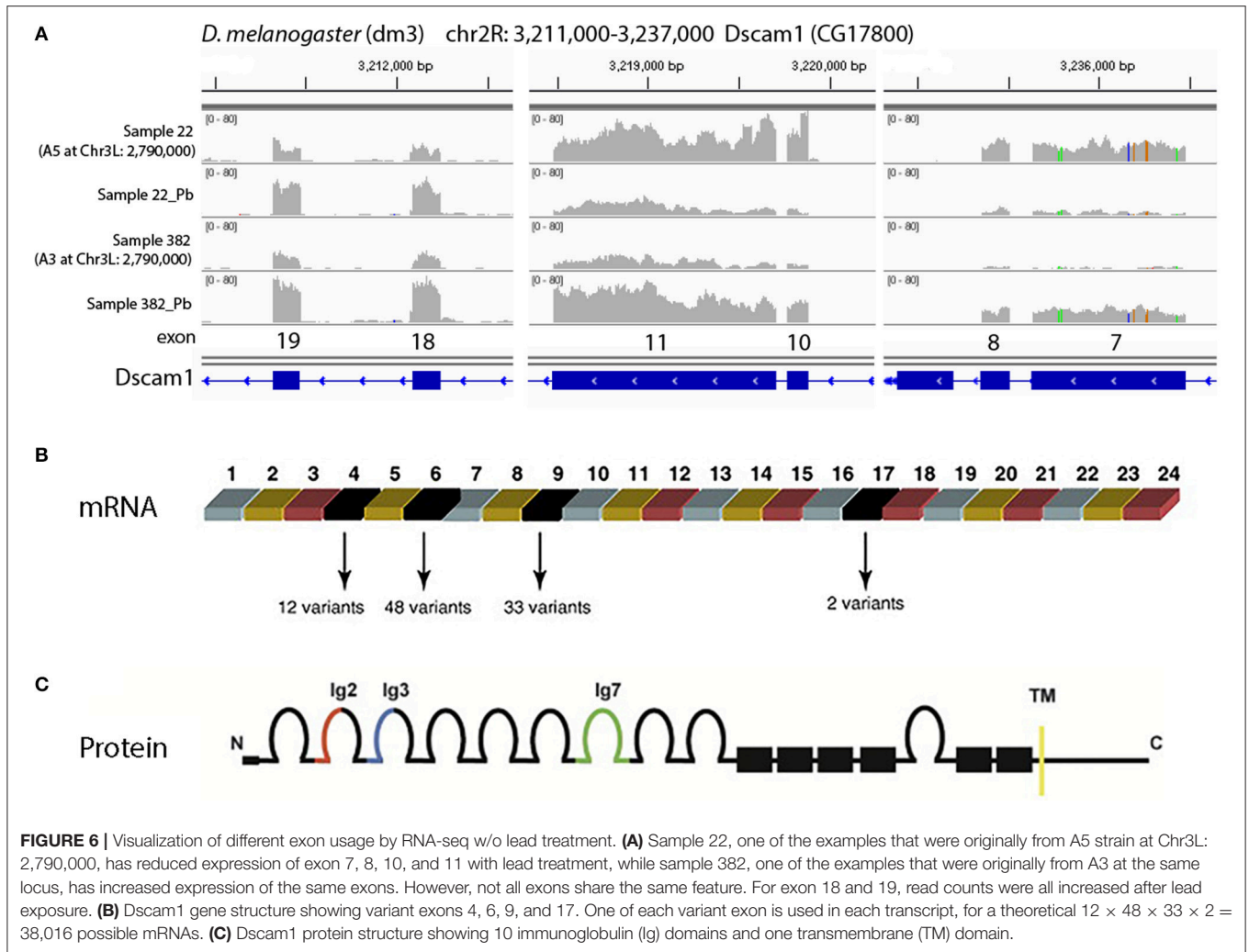


TABLE 1 | Ion transport genes in trans-splicing QTL.

Genes	Description	References
Shaw	Shaker cognate w (Shaw) is a voltage-gated potassium channel (Kv3.1) that mediates a non-inactivating potassium current open at resting membrane potential. Shaw regulates circadian rhythms and is in a pathway with a Na ⁺ K ⁺ Ca ²⁺ Co-transporter.	Hodge et al., 2005; Hodge and Stanewsky, 2008
Slo	Slowpoke (Slo) the structural alpha subunit of a BK ("maxi K") calcium-activated potassium channel. Slo channels function to regulate neurotransmitter release at the synapse and maintain electrical excitability in neurons and muscle cells.	Pallanck and Ganetzky, 1994; Lee et al., 2014
Itp-R83A	Inositol 1,4,5,-tris-phosphate receptor (Itp-r83A) is an intracellular ligand gated calcium channel. It functions downstream of G-protein coupled receptors.	Bollepalli et al., 2017
Ca-Alpha1D	Ca ²⁺ -channel protein α 1 subunit D (Ca- α 1D) is the pore-forming α subunit of an L-type voltage-gated Ca ²⁺ channel expressed in neurons.	Hara et al., 2015
CG42260	CG42260 is predicted to form the beta subunit of a cyclic nucleotide-gated cation channel.	Walkinshaw et al., 2015
Wtrw	Water witch (wtrw) is a calcium ion transporter that is involved in sensory perception of sound; and response to humidity.	Liu et al., 2007

RNA editing, which is almost 10-fold higher than the average (Supplementary Table S2).

DISCUSSIONS AND CONCLUSIONS

In this paper, we used two methods to detect Pb-responsive sQTLs. The first one, which is the fraction of exon reads to the transcript reads, searches splicing events on exon levels and all types of genetic-related AS events that cause change either in exon or in transcript after lead treatment will be selected. The other method compares the transcript counts among isoforms. It is on the transcript levels and will select those have differential isoform dosage after lead exposure. The combination of two methods resulted in 1,236 potential Pb-responsive sQTLs.

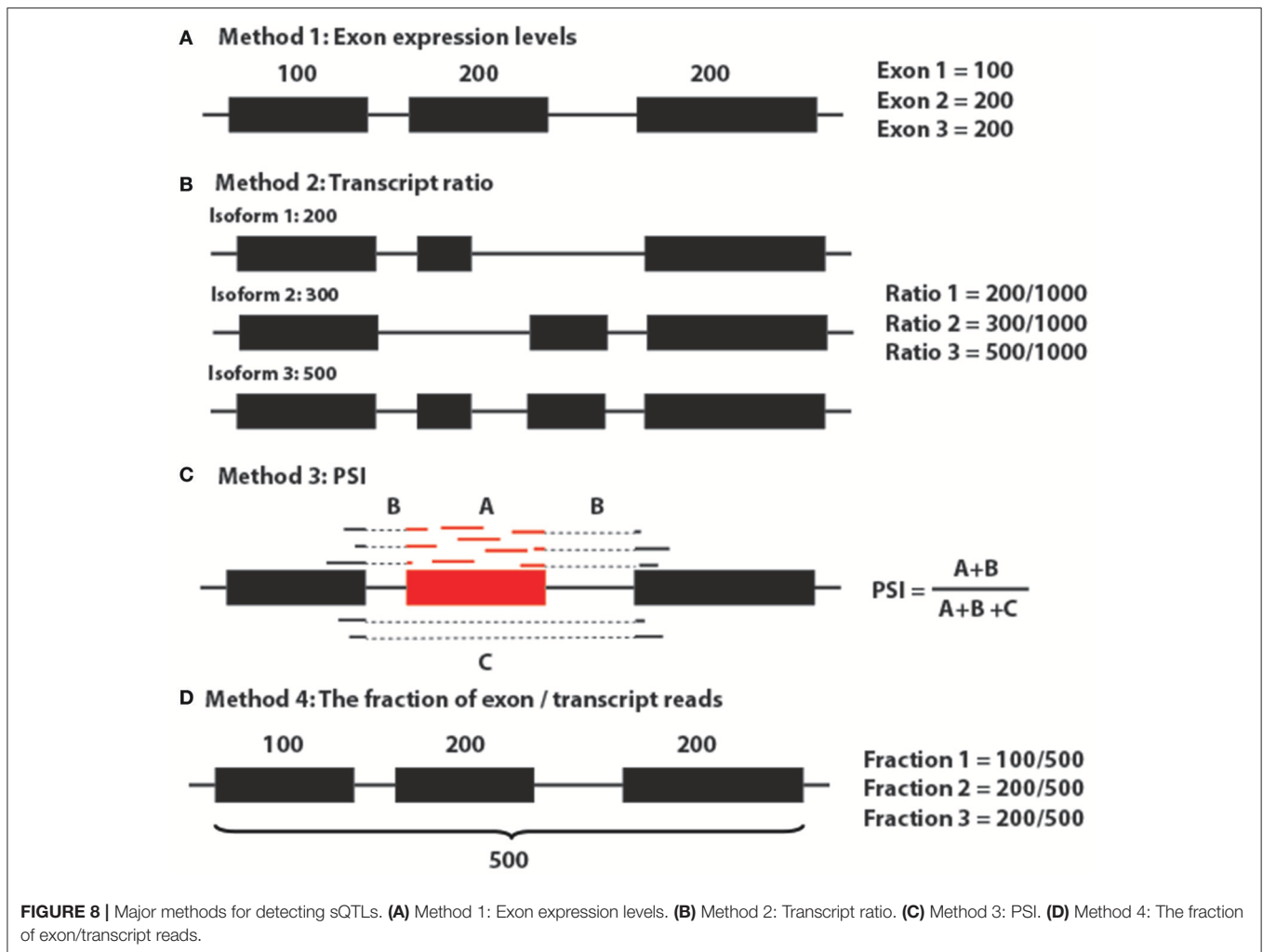
Generally, to target sQTLs, there are five major approaches (Figure 8; Gymrek, 2014): (1) use exon expression profiles as the quantitative trait and this could also be referred to as exon QTLs (Montgomery et al., 2010; Lappalainen et al., 2013; Gymrek, 2014); (2) use the proportion of each transcript quantification of the sum of all transcripts per gene (Figure 5A; Lappalainen et al., 2013; Battle et al., 2014; Gymrek, 2014); (3) use percent spliced in (PSI) (Lappalainen et al., 2013; Zhao et al., 2013; Gymrek, 2014; Kurmangaliyev et al., 2015); (4) use the fraction of reads in a gene that falls in a given exon as the phenotype (Figure 6; Pickrell et al., 2010; Gymrek, 2014); and (5) multivariate approaches, such as sQTLseeker (Gymrek, 2014; Monlong et al., 2014). The sQTLseeker is a multivariate model called for each gene consisting of the relative abundance of each transcript (Monlong et al., 2014). It calculated the variability of splicing ratios of a gene across samples by using a MANOVA-like distance-based approach and then compared the variability of the splicing ratios within genotypes with the variability among genotypes. We have run our data through the sQTLseeker pipeline. However, no significant association was returned. One of the potential explanation for this result is that the sQTLseeker was originally designed to incorporate the genotypes as SNP information (0 for ref/ref, 1 for ref/mutated, 2 for mutated/mutated), but our genotype data, which represent the original parents of evenly

distributed genomic locations (A1–A8, representing 8 parents), pose potential challenges to process the data (Monlong et al., 2014).

Most of the sQTL studies were performed in human cell lines. One study by Kurmangaliyev et al. which has claimed to be the first sQTL study in *Drosophila*, searched for genotype-specific alternative splicing donor/acceptor sites by using 81 *Drosophila* hybrid strains generated by crossing natural populations to a single inbred reference line (Kurmangaliyev et al., 2015). They found 59 AS donor/acceptor events by performing 120,240 association tests (Kurmangaliyev et al., 2015). In our study, we ran over 1,255,422,008 association tests (106,681 exon/transcript reads and 11,768 genomic loci) and our detection should not only include alternative donor/acceptor splicing but also other types of AS events.

The identification of *Dscam1* as one of the most significant sQTLs helps to further understand the isoform usage and changes after Pb exposure. We observed the expression alterations in *Dscam1* both at the exon level and at the transcript level after lead poisoning. However, we have few ideas on how to interpret these phenomena: why changes occurred after lead exposure and whether they could contribute to the neural developmental damage. Among the previous studies, Schmucker et al. have shown that the overexpression of one *Dscam1* isoform resulted in strong dominant phenotypes in mushroom body neurons (Schmucker et al., 2000). Some others have found that the diversity of *Dscam1* isoforms allows neurons to have differential patterns on their cell membrane and interaction was performed through isoform-specific hemophilic binding (Wojtowicz et al., 2004; Zipursky and Grueber, 2013; Armitage et al., 2015; Tadros et al., 2016). Schmucker and Flanagan also suggested either that different neurons express different *Dscam1* isoforms or that isoforms need to be present at a precise concentration or a certain development time period (Schmucker and Flanagan, 2004). More analysis is needed and future studies might consider combining sQTL analysis with other molecular and cellular experiments to further explore lead neurotoxicology.

One interesting finding is the preliminary identification of a *trans*-sQTL hotspot that is apparently regulated by Pb. This



trans-sQTL hotspot, located at Chr3L:18,810,000, includes 129 genes that are significantly enriched in the gene ontology category “ion channel domain” (Table 1). One possible explanation for the *trans*-sQTL hotspot is that this locus encodes a splicing factor whose activity is altered by Pb exposure. Since Pb²⁺ is a cation, the induction of alternative splicing of a cation channels could lead an increase in the elimination of lead from neurons, for instance. While this hypothesis is provocative, it would need to be confirmed experimentally by determining whether the Pb-induced isoforms of these alternatively spliced calcium channels have an increased ability to transport Pb²⁺ out of neurons. Further validation is also needed to confirm that the *Dscam1* alternative splicing isoforms that we identified are splicing QTL (Figure 5). However, since used the total reads in each isoform in our quantitative trait mapping analyses, and the alternative exons in each isoform are spread throughout the ~7 kb processed mRNA, the only way to confirm the *Dscam1* isoforms is by long RNA-seq analyses. Recently, this was done by the Graveley lab for *Dscam1* using the Minion™ nanopore sequencer (Oxford Nanopore, Inc.; Bolisetty et al., 2015), which

can sequence tens or even hundreds of kb long cDNA molecules. Attempts by our laboratory to generate long RNA-seq runs of *Dscam1* so far have been unsuccessful, but we continue to pursue this approach.

To our knowledge, this study is the first to link splicing QTL analysis with environmental toxin in *Drosophila*. However, there are multiple limitations in our study, which prevent us from reaching more meaningful results: (1) the RNA-seq data were prepared as 50 bp paired-end. However, 100 bp paired-end reads were considered to enhance splicing junction detection significantly (Chhangawala et al., 2015). (2) Both methods in this paper rely on known transcript annotation and transcript level quantifications and these are not widely trusted yet (Gymrek, 2014).

In conclusion, we have shown that sQTL analysis is a useful way in understanding alternative splicing mechanisms and the neuro-toxicology of environmental toxin. We have discovered widespread genetic variation affecting the splicing events. Our characterization of causal regulatory variation sheds light on the mechanisms of neurotoxicity of lead, and allows us to infer

putative causal variants for hundreds of potential environmental toxic-associated loci.

AUTHOR CONTRIBUTIONS

DR is the PI and supervised all aspects of the paper. WQ conducted the research and wrote the paper. KG helped with the bioinformatics and writing. RP helped with the bioinformatics and writing.

FUNDING

This work was supported by National Institutes of Health (NIH) [grant number 5R01ES012933 (DR), 5P30ES020957 (DR), 1UG3OD023285 (DR), 5R01GM109215 (RP)] and pilot funds

REFERENCES

- Anders, S., Pyl, P. T., and Huber, W. (2014). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Armitage, S. A., Peuß, R., and Kurtz, J. (2015). Dscam and pancrustacean immune memory—a review of the evidence. *Dev. Comp. Immunol.* 48 315–323. doi: 10.1016/j.dci.2014.03.004
- Baranowska-Bosiacka, I., Gutowska, I., Rybicka, M., Nowacki, P., and Chlubek, D. (2012). Neurotoxicity of lead. Hypothetical molecular mechanisms of synaptic function disorders. *Neurol. Neurochir. Pol.* 46, 569–578. doi: 10.5114/ninp.2012.31607
- Battle, A., Mostafavi, S., Zhu, X., Potash, J. B., Weissman, M. M., McCormick, C., et al. (2014). Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res.* 24, 14–24. doi: 10.1101/gr.155192.113
- Bellinger, D. C. (2013). Prenatal exposures to environmental chemicals and children's neurodevelopment: an update. *Saf. Health Wrk.* 4, 1–11. doi: 10.5491/SHAW.2013.4.1.1
- Benzer, S. (1955). Fine structure of a genetic region in bacteriophage. *Proc. Natl. Acad. Sci. U.S.A.* 41:344. doi: 10.1073/pnas.41.6.344
- Bolisetty, M. T., Rajadinakaran, G., and Graveley, B. R. (2015). Determining exon connectivity in complex mRNAs by nanopore sequencing. *Genome Biol.* 16, 204. doi: 10.1186/s13059-015-0777-z
- Bollepalli, M. K., Kuipers, M. E., Liu, C.-H., Asteriti, S., and Hardie, R. C. (2017). Phototransduction in *Drosophila* is compromised by Gal4 expression but not by *InsP₃* receptor knockdown or mutation. *eNeuro* 4:ENEURO.0143-17.2017. doi: 10.1523/ENEURO.0143-17.2017
- Chhangawala, S., Rudy, G., Mason, C. E., and Rosenfeld, J. A. (2015). The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome Biol.* 16, 1–10. doi: 10.1186/s13059-015-0697-y
- Foissac, S., and Sammeth, M. (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.* 35, W297–W299. doi: 10.1093/nar/gkm311
- Gilbert, W. (1978). Why genes in pieces? *Nature* 271, 501.
- Gymrek, M. (2014). *The Complicated World of Splice QTLs*. Available online at: Blog.melissagymrek.com
- Hanna-Attisha, M., LaChance, J., Sadler, R. C., and Champney Schnepf, A. (2015). Elevated blood lead levels in children associated with the Flint drinking water crisis: a spatial analysis of risk and public health response. *Am. J. Public Health* 106, 283–290. doi: 10.2105/AJPH.2015.303003
- Hara, Y., Koganezawa, M., and Yamamoto, D. (2015). The Dmca1D channel mediates Ca(2+) inward currents in *Drosophila* embryonic muscles. *J. Neurogenet.* 29, 117–123. doi: 10.3109/01677063.2015.1054991

from the Wayne State University Office of the Vice President for Research (DR).

ACKNOWLEDGMENTS

The *Drosophila* stocks are from the Bloomington, IN, stock center. We thank the High-Performance Computing Services offered by Wayne State University for making the entire statistical analysis fast and efficient.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2017.00145/full#supplementary-material>

- Hasin-Brumshtein, Y., Hormozdiari, F., Martin, L., Van Nas, A., Eskin, E., Lusis, A. J., et al. (2014). Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics* 15:471. doi: 10.1186/1471-2164-15-471
- Hattori, D., Demir, E., Kim, H. W., Viragh, E., Zipursky, S. L., and Dickson, B. J. (2007). Dscam diversity is essential for neuronal wiring and self-recognition. *Nature* 449, 223–227. doi: 10.1038/nature06099
- He, T., Hirsch, H. V., Ruden, D. M., and Lnenicka, G. A. (2009a). Chronic lead exposure alters presynaptic calcium regulation and synaptic facilitation in *Drosophila* larvae. *Neurotoxicology* 30, 777–784. doi: 10.1016/j.neuro.2009.08.007
- He, T., Singh, V., Rumpal, N., and Lnenicka, G. (2009b). Differences in Ca 2+ regulation for high-output Is and low-output Ib motor terminals in *Drosophila* larvae. *Neuroscience* 159, 1283–1291. doi: 10.1016/j.neuroscience.2009.01.074
- Hirsch, H., Barth, M., Luo, S., Sambaziotis, H., Huber, M., Possidente, D., et al. (1995). Early visual experience affects mate choice of *Drosophila melanogaster*. *Anim. Behav.* 50, 1211–1217. doi: 10.1016/0003-3472(95)80038-7
- Hirsch, H. V., Lnenicka, G., Possidente, D., Possidente, B., Garfinkel, M. D., Wang, L., et al. (2012). *Drosophila melanogaster* as a model for lead neurotoxicology and toxicogenomics research. *Front. Genet.* 3:68. doi: 10.3389/fgene.2012.00068
- Hirsch, H. V., Mercer, J., Sambaziotis, H., Huber, M., Stark, D. T., Torno-Morley, T., et al. (2003). Behavioral effects of chronic exposure to low levels of lead in *Drosophila melanogaster*. *Neurotoxicology* 24, 435–442. doi: 10.1016/S0161-813X(03)00021-4
- Hirsch, H. V., Possidente, D., Averill, S., Despain, T. P., Buytkins, J., Thomas, V., et al. (2009). Variations at a Quantitative Trait Locus (QTL) affect development of behavior in lead-exposed *Drosophila melanogaster*. *Neurotoxicology* 30, 305–311. doi: 10.1016/j.neuro.2009.01.004
- Hoaglin, D. C., and Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *Am. Stat.* 32, 17–22.
- Hodge, J. J., Choi, J. C., O'Kane, C. J., and Griffith, L. C. (2005). Shaw potassium channel genes in *Drosophila*. *J. Neurobiol.* 63, 235–254. doi: 10.1002/neu.20126
- Hodge, J. J., and Stanewsky, R. (2008). Function of the Shaw potassium channel within the *Drosophila* circadian clock. *PLoS ONE* 3:e2274. doi: 10.1371/journal.pone.0002274
- Joo, W. J. J., Sul, J. H., Han, B., Ye, C., and Eskin, E. (2014). Effectively identifying regulatory hotspots while capturing expression heterogeneity in gene expression studies. *Genome Biol.* 15:r61. doi: 10.1186/gb-2014-15-4-r61
- Karolchik, D., Hinrichs, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., et al. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32, D493–D496. doi: 10.1093/nar/gkh103
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006. doi: 10.1101/gr.229102
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36

- King, E. G., Merkes, C. M., McNeil, C. L., Hooper, S. R., Sen, S., Broman, K. W., et al. (2012). Genetic dissection of a model complex trait using the *Drosophila* Synthetic Population Resource. *Genome Res.* 22, 1558–1566. doi: 10.1101/gr.134031.111
- King, E. G., Sanderson, B. J., McNeil, C. L., Long, A. D., and MacDonald, S. J. (2014). Genetic dissection of the *Drosophila melanogaster* female head transcriptome reveals widespread allelic heterogeneity. *PLoS Genet.* 10:e1004322. doi: 10.1371/journal.pgen.1004322
- Krawczak, M., Thomas, N. S., Hundrieser, B., Mort, M., Wittig, M., Hampe, J., et al. (2007). Single base-pair substitutions in exon–intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. *Hum. Mutat.* 28, 150–158. doi: 10.1002/humu.20400
- Kurmangaliyev, Y. Z., Favorov, A. V., Osman, N. M., Lehmann, K.-V., Campo, D., Salomon, M. P., et al. (2015). Natural variation of gene models in *Drosophila melanogaster*. *BMC Genomics* 16:198. doi: 10.1186/s12864-015-1415-6
- Lappalainen, T., Sammeth, M., Friedländer, M. R., 't Hoen, P. A., Monlong, J., Rivas, M. A., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501, 506–511. doi: 10.1038/nature12531
- Lee, J., Ueda, A., and Wu, C. F. (2014). Distinct roles of *Drosophila* cacophony and Dmca1D Ca(2+) channels in synaptic homeostasis: genetic interactions with slowpoke Ca(2+) -activated BK channels in presynaptic excitability and postsynaptic response. *Dev. Neurobiol.* 74, 1–15. doi: 10.1002/dneu.22120
- Lim, K. H., Ferraris, L., Filloux, M. E., Raphael, B. J., and Fairbrother, W. G. (2011). Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 11093–11098. doi: 10.1073/pnas.1101135108
- Liu, L., Li, Y., Wang, R., Yin, C., Dong, Q., Hing, H., et al. (2007). *Drosophila* hygrosensation requires the TRP channels water witch and nanchung. *Nature* 450, 294–298. doi: 10.1038/nature06223
- Mei, W., Liu, S., Schnable, J. C., Yeh, C. T., Springer, N. M., Schnable, P. S., et al. (2017). A comprehensive analysis of alternative splicing in paleopolyploid maize. *Front. Plant Sci.* 8:694. doi: 10.3389/fpls.2017.00694
- Miles, C. M., and Wayne, M. (2008). Quantitative trait locus (QTL) analysis. *Nat. Educ.* 1, 208. Available online at: <https://www.nature.com/scitable/topicpage/quantitative-trait-locus-qtl-analysis-53904>
- Modrek, B., and Lee, C. (2002). A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19. doi: 10.1038/ng0102-13
- Monlong, J., Calvo, M., Ferreira, P. G., and Guigó, R. (2014). Identification of genetic variants associated with alternative splicing using sQTLseeker. *Nat. Commun.* 5:4698. doi: 10.1038/ncomms5698
- Montgomery, S. B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R. P., Ingle, C., Nisbett, J., et al. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* 464, 773–777. doi: 10.1038/nature08903
- Morley, E. J., Hirsch, H. V., Hollocher, K., and Lnenicka, G. A. (2003). Effects of chronic lead exposure on the neuromuscular junction in *Drosophila* larvae. *Neurotoxicology* 24, 35–41. doi: 10.1016/S0161-813X(02)00095-5
- Ongen, H., and Dermitzakis, E. T. (2015). Alternative splicing QTLs in European and African populations. *Am. J. Hum. Genet.* 97, 567–575. doi: 10.1016/j.ajhg.2015.09.004
- Pallanck, L., and Ganetzky, B. (1994). Cloning and characterization of human and mouse homologs of the *Drosophila* calcium-activated potassium channel gene, slowpoke. *Hum. Mol. Genet.* 3, 1239–1243. doi: 10.1093/hmg/3.8.1239
- Pennisi, E. (2000). Human genome project. And the gene number is? *Science* 288, 1146–1147. doi: 10.1126/science.288.5469.1146
- Pickrell, J. K., Marioni, J. C., Pai, A. A., Degner, J. F., Engelhardt, B. E., Nkadori, E., et al. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 464, 768–772. doi: 10.1038/nature08872
- Rauh, V. A., and Margolis, A. E. (2016). Research review: environmental exposures, neurodevelopment, and child mental health—new paradigms for the study of brain and behavioral effects. *J. Child Psychol. Psychiatry* 57, 775–793. doi: 10.1111/jcpp.12537
- Rockman, M. V., and Kruglyak, L. (2006). Genetics of global gene expression. *Nat. Genet.* 7, 862–872. doi: 10.1038/nrg1964
- Ruden, D. M., Chen, L., Possidente, D., Possidente, B., Rasouli, P., Wang, L., et al. (2009). Genetical toxicogenomics in *Drosophila* identifies master-modulatory loci that are regulated by developmental exposure to lead. *Neurotoxicology* 30, 898–914. doi: 10.1016/j.neuro.2009.08.011
- Sambrook, J. (1977). Adenovirus amazes at cold spring harbor. *Nature* 268:101. doi: 10.1038/268101a0
- Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., et al. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101, 671–684. doi: 10.1016/S0092-8674(00)80878-8
- Schmucker, D., and Flanagan, J. G. (2004). Generation of recognition diversity in the nervous system. *Neuron* 44, 219–222. doi: 10.1016/j.neuron.2004.10.004
- Sen, A., Cingolani, P., Senut, M.-C., Land, S., Mercado-Garcia, A., Tellez-Rojo, M. M., et al. (2015a). Lead exposure induces changes in 5-hydroxymethylcytosine clusters in CpG islands in human embryonic stem cells and umbilical cord blood. *Epigenetics* 10, 607–621. doi: 10.1080/15592294.2015.1050172
- Sen, A., Heredia, N., Senut, M. C., Land, S., Hollocher, K., Lu, X., et al. (2015b). Multigenerational epigenetic inheritance in humans: DNA methylation changes associated with maternal exposure to lead can be transmitted to the grandchildren. *Sci. Rep.* 5:14466. doi: 10.1038/srep14466
- Senut, M.-C., Cingolani, P., Sen, A., Shaik, A., Hirsch, H., et al. (2012). Epigenetics of early-life lead exposure and effects on brain development. *Epigenomics* 4, 665–674. doi: 10.2217/epi.12.58
- Singh, R. K., and Cooper, T. A. (2012). Pre-mRNA splicing in disease and therapeutics. *Trends Mol. Med.* 18, 472–482. doi: 10.1016/j.molmed.2012.06.006
- Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M. F., Rifkin, S. A., et al. (2004). A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* 306, 655–660. doi: 10.1126/science.1101312
- Tadros, W., Xu, S., Akin, O., Yi, C. H., Shin, G. J., Millard, S. S., et al. (2016). Dscam proteins direct dendritic targeting through adhesion. *Neuron* 89, 480–493. doi: 10.1016/j.neuron.2015.12.026
- Takata, A., Matsumoto, N., and Kato, T. (2017). Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat. Commun.* 8:14519. doi: 10.1038/ncomms14519
- Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinformatics* 14, 178–192. doi: 10.1093/bib/bbs017
- Tong, S., von Schirnding, Y., and Prapamontol, T. (2000). Environmental lead exposure: a public health problem of global dimensions. *Bull. World Health Organ.* 78, 1068–1077. Available online at: <http://apps.who.int/iris/handle/10665/57599>
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., and Kelley, D. R. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with Tophat and Cufflinks. *Nat. Protocols* 7, 562–578. doi: 10.1038/nprot.2012.016
- Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., et al. (2009). FlyBase: enhancing *Drosophila* gene ontology annotations. *Nucleic Acids Res.* 37, D555–D559. doi: 10.1093/nar/gkn788
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351. doi: 10.1126/science.1058040
- Walkinshaw, E., Gai, Y., Farkas, C., Richter, D., Nicholas, E., Keleman, K., et al. (2015). Identification of genes that promote or inhibit olfactory memory formation in *Drosophila*. *Genetics* 199, 1173–1182. doi: 10.1534/genetics.114.173575
- Watson, F. L., Pittmann-Holgado, R., Thomas, F., Lamar, D. L., Hughes, M., Kondo, M., et al. (2005). Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* 309, 1874–1878. doi: 10.1126/science.1116887
- White, L., Cory-Slechta, D., Gilbert, M., Tiffany-Castiglioni, E., Zawia, N., and Virgolini, M. (2007). New and evolving concepts in the neurotoxicology of lead. *Toxicol. Appl. Pharmacol.* 225, 1–27. doi: 10.1016/j.taap.2007.08.001
- Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L., and Clemens, J. C. (2004). Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* 118, 619–633. doi: 10.1016/j.cell.2004.08.021
- Yamakawa, K., Huo, Y.-K., Haendel, M. A., Hubert, R., Chen, X.-N., Korenberg, J. R., et al. (1998). DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development

- of the nervous system. *Hum. Mol. Genet.* 7, 227–237. doi: 10.1093/hmg/7.2.227
- Yang, Q., Hu, Y., Li, J., and Zhang, X. (2017). ulfasQTL: an ultra-fast method of composite splicing QTL analysis. *BMC Genomics* 18:963. doi: 10.1186/s12864-016-3258-1
- Yoo, W., Kyung, S., Han, S., and Kim, S. (2016). Investigation of splicing quantitative trait loci in *Arabidopsis thaliana*. *Genomics Inform.* 14, 211–215. doi: 10.5808/GI.2016.14.4.211
- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Method Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14
- Zhao, K., Lu, Z.-X., Park, J. W., Zhou, Q., and Xing, Y. (2013). GLiMMPS: robust statistical model for regulatory variation of alternative splicing using RNA-seq data. *Genome Biol.* 11:R74. doi: 10.1186/gb-2013-14-7-r74
- Zipursky, S. L., and Grueber, W. B. (2013). The molecular basis of self-avoidance. *Annu. Rev. Neurosci.* 36, 547–568. doi: 10.1146/annurev-neuro-062111-150414

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Qu, Gurdziel, Pique-Regi and Ruden. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.