Global Genetic Robustness of the Alternative Splicing Machinery in *Caenorhabditis elegans*

Yang Li,* Rainer Breitling,* L. Basten Snoek,† K. Joeri van der Velde,* Morris A. Swertz,*‡ Joost Riksen,† Ritsert C. Jansen*‡,1 and Jan E. Kammenga†

*Groningen Bioinformatics Centre, University of Groningen, 9751 NN, Haren, The Netherlands, †Laboratory of Nematology, Wageningen University, 6708 PB, Wageningen, The Netherlands and ‡Department of Genetics, University Medical Center Groningen and University of Groningen, 9700 RB, Groningen, The Netherlands

Manuscript received June 7, 2010
Accepted for publication June 21, 2010

ABSTRACT

Alternative splicing is considered a major mechanism for creating multicellular diversity from a limited repertoire of genes. Here, we performed the first study of genetic variation controlling alternative splicing patterns by comprehensively identifying quantitative trait loci affecting the differential expression of transcript isoforms in a large recombinant inbred population of *Caenorhabditis elegans*, using a new generation of whole-genome very-high-density oligonucleotide microarrays. Using 60 experimental lines, we were able to detect 435 genes with substantial heritable variation, of which 36% were regulated at a distance (in trans). Nonetheless, we find only a very small number of examples of heritable variation in alternative splicing (22 transcripts), and most of these genes colocalize with the associated genomic loci. Our findings suggest that the regulatory mechanism of alternative splicing in *C. elegans* is robust toward genetic variation at the genome-wide scale, which is in striking contrast to earlier observations in humans.

**MATERIALS AND METHODS**

**Worm samples, genotyping, and Affymetrix GeneChips:** We used *C. elegans* recombinant inbred lines that were generated from a cross of N2 and CB4856 and were genotyped by Li *et al.* (2006). Age-synchronized *C. elegans* was cultured at 24°C and the total RNA was isolated from the late L3 stage using the Trizol method. The RNA was cleaned using the QIAGEN (Valencia, CA) RNEasy Micro RNA cleanup kit. Double-stranded cDNA synthesis was done with the Affymetrix GeneChip WT double-stranded cDNA synthesis kit. We...
cleaned the cDNA using the GeneChip Sample Cleanup Module also from Affymetrix. For fragmentation and labeling, the GeneChip WT double-stranded DNA terminal labeling kit was used. The concentrations of RNA and cDNA were measured with a Nanodrop. After the fragmentation we determined the fragment size on a Nusieve 3:1 agarose gel. mRNA was hybridized to Affymetrix 1.0 C. elegans tiling arrays (2.9 million probes on each array) and the hybridization was done by ServiceXS (Leiden, The Netherlands). Since polymorphisms in the probe region can lead to spurious local eQTL (Alberts et al. 2007), 80,903 probes with known SNP (including predicted SNP; WS195 release) were removed for subsequent analysis. Each probe is annotated as exonic, intronic, or intergenic, when the entire probe of 25 bp falls in one of the three regions, respectively. Probes spanning exon–intron boundaries are labeled as boundary probes.

Data analysis: Preprocessing of raw data: The raw gene expression data from 60 microarrays (one RIL per array) were taken base-2 log transformed and then quantile normalized. expression data from 60 microarrays (one RIL per array) were variable, and hybridization and measurement and treated as a categorical variable, and e is the residual error.

Differential expression between genotypes (eQTL): We used a robust and powerful statistical approach to associate microarray probe intensity and genotype data in the face of widely different hybridization properties of individual probes. Instead of computing significance of a statistical test, we evaluated a nonparametric effect size [Cliff's $D$ (Cliff 1996)] for all 3 million probes at each genomic marker. For each probe on the array we compute the eQTL effect size using Cliff's nonparametric $\Delta$-statistic

$$\Delta = \frac{\#(X_1 > X_2) - \#(X_1 < X_2)}{m \times n},$$

where $n_1$ and $n_2$ are the numbers of carriers of the N2 and the CB4856 allele, and $\#(X_1 > X_2)$ is the number of possible pairwise comparisons where the expression level of gene $i$ in an N2 carrier is larger than in a CB4856 carrier. The genotype information of the 60 RILs was previously described (Li et al. 2006). For an individual probe, a value of $\Delta = 0.45$ corresponds to a $P$-value = 0.001 in a Wilcoxon rank sum test (Del Rosal et al. 2003).

As several positions in the genome show a strongly imbalanced genotype ratio (i.e., the number of RILs carrying the N2 allele is far larger than the number of RILs carrying the CB4856 allele at a particular locus), the corresponding threshold (Wilcoxon's $U$-value) for each marker at significance level $P = 0.001$ was obtained first, taking the locus-specific imbalance into account. Then, these values were converted into the corresponding threshold for the effect size (Cliff's $D$) on the basis of $D = 2U/(n_1n_2) - 1$ (Del Rosal et al. 2003). The threshold of distorted genome regions is expected to be larger than that of balanced marker positions. These marker-dependent thresholds were applied in further analysis.

Summarizing the eQTL effect for exons: To increase the robustness of the procedure, the median effect size of probes within each exon was taken as representing the expression QTL effect size of this exon for each genomic marker. Subsequently, the eQTL profile at the marker with maximal summarized eQTL effect was obtained. To achieve a reliable estimate of eQTL effect size, only exons covered by more than three probes were considered here. Transcripts with a summarized eQTL effect larger than the threshold for at least one exon were declared as having a significant eQTL and were used for further analysis.

Classification of eQTL: There are 435 transcripts with a significant eQTL in total. They were examined in greater detail and manually classified as shown in Figure 1. By visualizing the intensity level and eQTL size of the entire transcript, we first classified transcripts as having a consistent eQTL if all annotated exons show the same eQTL pattern at a threshold of $\Delta = 0.45$ and there is no additional eQTL signal in the presumed intron regions. In addition, there are eQTL patterns that indicate the need for revised gene definitions (but no evidence for difference in splicing), which can be subdivided into five subcategories: (1) new exons (at least two consecutive intron probes showing a similar expression level and eQTL size as the exon probes of the gene), (2) new introns (at least two consecutive exon probes showing a clear decrease of expression level and eQTL size compared to the other exon probes of the gene), (3) intron inclusions (all probes corresponding to an intron showing the same expression and eQTL size as the exon probes), (4) exon extensions (at least two intron probes next to an exon showing similar expression levels and eQTL size as the adjacent exon), and (5) intron extensions (at least the first or the last two exon probes showing a decrease of expression level and eQTL
results and discussion

Here, we performed the first genome-wide analysis of genetic variation of alternative splicing in C. elegans using a comprehensive tiling microarray. We used 60 recombinant inbred lines of a cross between two very diverse strains, Bristol (N2) and Hawaii (CB4856), which have been genotyped using 121 markers (Li et al. 2006). By using tiling array data, with multiple probes targeting every exon of each gene, we obtained a more comprehensive and sensitive picture of heritable variation of gene expression than possible with previous technologies. It also allows us to dissect the genetic component for differences in isoform-specific gene expression. Thus we can detect asQTL, the genome regions controlling variation in isoform-specific expression. Two categories of asQTL can be distinguished, i.e., those that map in close vicinity to the gene itself (local) and those that map elsewhere in the genome (distant).
Local activity can be explained, for example, by altered functional motifs in exonic splicing enhancers that will affect the splicing activity. The mechanism of distant regulation is often more complicated and can possibly be explained by a polymorphism in an auxiliary splicing factor (e.g., SR protein) that modulates the activity of the spliceosome. In this case we would expect to see a genetic master regulator at the locus of the splicing factor controlling isoform ratios for large groups of transcripts.

Using nonparametric effect size estimates, corrected for genotype imbalance (materials and methods) and corresponding to a P-value of 0.001 (Wilcoxon’s test), we detected 435 genes with substantial heritable variation for at least one exon. The comparison of gene position and associated polymorphisms shows that most eQTL map in close proximity to the affected gene (local eQTL: 277 genes or 64%; Figure 3). There are 158 eQTL mapping to another chromosome (distant eQTL). Two hundred sixty-seven genes show higher expression in carriers of the N2 allele than in CB4856 carriers, including 53 cases of known gene deletions in the CB4856 strain (Maydan et al. 2007).

A large majority of eQTL (319 or 70.4%) lead to a consistent differential expression across all exons of the affected gene. Interestingly, the genetic effects (eQTL size) of these consistent eQTL show a strong correlation (Spearman’s $r = 0.78$) with a previous experiment using cDNA micorarrays (Li et al. 2006). As shown in Figure 1, 8.7% of cases show evidence for a necessary refinement of existing gene definitions, predominantly by expanding known exons (plotted results for all genes are available at www.wormplot.org for a detailed examination). In contrast to the large number of consistent eQTL, we find only 22 genes that show evidence for genetic variation of alternative splicing, i.e., an exon-specific asQTL (Figure 4). This genome-wide evidence for the genetic robustness of the alternative splicing machinery is consistent with the earlier indication that individual alternative splicing events in C. elegans are highly conserved and hardly tolerate genetic variation (Barberan-Soler and Zahler 2008a). Note, however, that variation in alternative splicing events restricted to a specific cell or tissue type can be diluted in measurements on whole-worm mRNA. In addition, 77% of asQTL were found to be locally regulated. This agrees with recent findings that alternative splicing can be regulated without involvement of an auxiliary splicing factor, by cis-acting RNA sequences that can function as a splicing silencer (Yu et al. 2008).

Most of the reported asQTL detected in our study have strong genetic effects (qualitative on–off patterns) and we found only a few cases of subtle quantitative effects on alternative splicing. However, this does not mean that this on–off behavior is a general property of alternative splicing patterns, but rather that despite the large population used in this study, technical noise and biological variation might limit our ability to detect subtle shifts in isoform proportions. To detect more quantitative effects (Figure 2), more precise technology such as deep-sequencing would be required. Even then, reliable detection of changes in isoform proportions will depend on extremely large read numbers.

Our genome-wide study provides the first genome-wide evidence supporting earlier hypotheses that in C. elegans the alternative splicing machinery exhibits a general genetic robustness, and only a minor fraction of genes show heritable variation in splicing forms and
relative abundance. This observation points to a profound difference in the regulation of the alternative splicing machinery compared to that in humans (Kwan et al. 2008), which parallels the differences in cellular diversity and developmental flexibility in the two species and has important consequences for interpreting future studies using C. elegans as a model organism for metazoan splicing.

Figure 4.—Expression intensity and eQTL effect per probe along the genome for selected genes. (A) Detecting consistent heritable differences in gene expression with high resolution. Nearly 300 probes cover the area of this gene, Y87G2A.5. Exon probes show consistently high expression (median intensity = 9.64), compared to intron probes. However, there is huge variation between probes, which makes the clear delimitation of exon boundaries challenging. In contrast, there is a clear and highly consistent differential expression between carriers of different alleles (N2 and CB4856). This so-called eQTL effect, indicated by red bars, is highly reproducible across all exon probes within the gene. In this example, the average expression difference between the two alleles is ~2.4-fold. In total, there are 306 genes with similar consistent expression differences. It should be noted that a majority of genes show consistently lower intensity (and thus lower eQTL effect) in the 3′-untranslated region (UTR) indicating the decaying end of transcript (Kolmogorov–Smirnov test, $P$-value, $2.2 \times 10^{-10}$). (B) Refining existing gene definition. The exon probes within gene T21E8.1 show consistently higher expression for individuals carrying the N2 allele than for those carrying the CB4856 allele. Additionally, several adjacent probes within the sixth intron show the same differential expression pattern, suggesting that this intron contains a previously unannotated additional exon. This would not have been detectable based on the absolute expression levels, due to the high interprobe variability. We find a total of 41 genes that require refined definition according to the eQTL pattern, mostly extensions of known exons and redefinitions of the transcript start and end sites. (C and D) Detecting heritable variation in alternative splicing. These genes do not show heritable expression differences in general, but individual exons show consistently lower signal for carriers of the CB4856 allele. This suggests that these exons are specifically removed by alternative splicing in one of the two alleles. In both cases, this alternative splicing variation is determined by a local sequence variation (QTL mapping in cis). The first example (Y69H2.3) was confirmed experimentally by Barberan-Soler and Zahler (2008a). We find 22 comparable instances of heritable differences in splicing patterns.
This work was supported by European Union grant FP7 PANACEA 222936 and The Netherlands Organization for Scientific Research, grant NWO-86504001.

LITERATURE CITED


Communicating editor: J. Borevitz
Global Genetic Robustness of the Alternative Splicing Machinery
in \textit{Caenorhabditis elegans}

Yang Li, Rainer Breitling, L. Basten Snoek, K. Joeri van der Velde, Morris A. Swertz, Joost Riksen, Ritsert C. Jansen and Jan E. Kammenga

Copyright © 2010 by the Genetics Society of America
DOI: 10.1534/genetics.110.119677
TABLE S1

Classification membership for genes with significant eQTL

Table S1 is available for download as a .csv file at http://www.genetics.org/cgi/content/full/genetics.110.119677/DC1.