4308–4314 *Nucleic Acids Research, 2009, Vol. 37, No. 13* *Published online 18 May 2009*
doi:10.1093/nar/gkp334

Detection of intergenic non-coding RNAs expressed in the main developmental stages in *Drosophila melanogaster*

Zheng Li[1], Min Liu[1], Li Zhang[2], Wenxia Zhang[1], Ge Gao[2], Zuoyan Zhu[1], Liping Wei[2], Qichang Fan[1,*] and Manyuan Long[2,3,*]

[1]Center of Developmental Biology and Genetics, School of Life Sciences, Peking University, Beijing, China, [2]Center for Bioinformatics, School of Life Sciences, Peking University, Beijing, China and [3]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

Received January 14, 2009; Revised April 20, 2009; Accepted April 21, 2009

ABSTRACT

How many intergenically encoded non-coding RNAs (ncRNAs) are expressed during various developmental stages in *Drosophila*? Previous analyses in one or a few developmental stages indicated abundant expression of intergenic ncRNAs. However, some reported that ncRNAs have been recently falsified, and, in general, the false positive rate for ncRNA detection is unknown. In this report, we used reverse transcription-PCR (RT-PCR), a more robust method, to detect ncRNAs from the intergenic regions that are expressed in four major developmental stages (6–8 h embryo, 20–22 h embryo, larvae and adult). We tested 1027 regions, ~10% of all intergenic regions, and detected transcription by RT–PCR. The results from 18 342 RT–PCR experiments revealed evidence for transcription in 72.7% of intergenic regions in the developmental process. The early developmental stage appears to be associated with more abundant ncRNAs than later developmental stages. In the early stage, we detected 43.6% of intergenic regions that encode transcripts in the triplicate RT–PCR experiments, yielding an estimate of 5006 intergenic regions in the entire genome likely encoding ncRNAs. We compared the RT–PCR-related approach with previous tiling array-based approach and observed that the latter method is insensitive to short ncRNAs, especially the molecules less than 120 bp. We measured false positive rates for the analyzed genomic approaches including the RT–PCR and tiling array method.

INTRODUCTION

Increasing numbers of non-coding RNA (ncRNA) genes with important functions have been reported in various organisms (1–9). Determining the abundance and distribution of ncRNA genes is critical for further understanding of the functions and properties of genes and the composition of genomes. Taking advantage of the high-quality annotation of the *Drosophila melanogaster* genome (10–13), numerous ncRNA genes were detected from this species (3,8). The microarray methods were also used to detect ncRNA candidates only expressed in one or few developmental stages (14).

These previous efforts indicated a potentially large number of ncRNAs encoded in the *D. melanogaster* genome expressed in various developmental stages. These analyses raised interesting problems in both methodology and biology. For example, technically, what are false positive rates for various genomic measurements? Biologically, what is the distribution and variation of ncRNA encoding genes across developmental stages? Indeed, both recently reported falsification and confirmation of previously predicted ncRNA candidates identified by the tiling array method (14–16) that indicate the importance of assessing the false positive rates for these methods.

To address these issues, we developed a sensitive approach to detect ncRNA genes. For the sensitivity and simplicity desirable for genomic search, we chose widely used RT–PCR (17) as the technique to detect ncRNAs (see 'Materials and Methods' section) for the first large validation of global ncRNA gene detection in all major developmental stages.

Our analyses reveal that a high proportion of intergenic regions in *D. melanogaster* encode ncRNAs. It is also

*To whom correspondence should be addressed. Tel/Fax: 773 702 0577; Email: mlong@midway.uchicago.edu
Correspondence may also be addressed to Qichang Fan. Tel/Fax: +86 10 6275 2438; E-mail: qfan@pku.edu.cn

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

shown that the early stage of development is associated with the most abundant species of ncRNAs. These results cast new insights into the role of non-coding genomic regions such as intergenic regions in organism function and evolution, with ncRNAs participating with protein-coding genes and non-coding functional genes during development.

MATERIALS AND METHODS

Data sources

We downloaded tables of 'Genes on sequence' for fly chromosomes X, 2, 3 and 4 from NCBI database (http://www.ncbi.nlm.nih.gov), determined the intergenic regions according to the gene lists and obtained the intergenic region coordinates from Flybase (http://www.flybase.org). We chose one in every ten intergenic regions; so about 10% of intergenic regions were collected. We fetched sequences of the chosen regions from Ensemble (http://www.ensembl.org). ESTs were obtained from NCBI (http://www.ncbi.nlm.nih.gov/). Tiling array data (14) were obtained online (http://www.transcriptome.affymetrix.com/publication/drosophila_development/).

RT–PCR primer design

Using the sequence comparison function of VISTA (http://www.pipeline.lbl.gov/cgi-bin/GenomeVista) (18), we generated sequence alignments between *D. melanogaster* and five other closely related species (*D. simulans, D. yakuba, D. erecta, D. pseudoobscura and D. virillis*) (19,20). We set parameters as follows: the size of the sliding window used to calculate conservation scores of each base pair was 100 bp; minimum width of a conserved region was 100 bp and minimum percent conservation identity was 70%. (We compared the subject sequences of *D. melanogaster* with those of *D. simulans, D. yakuba, D. erecta, D. pseudoobscura* and *D. virillis* and defined the overlap region with the identity above 70% in each comparison). The conservation scores were generated by the phastCons15way (21). With the results of alignments, we first collected intergenic fragments that were conserved in all six species, then the fragments that were less conserved (conserved in five, four, three or two species). In some especially short regions (<500 bp), there were no significant alignments; the whole region was taken as a single intergenic fragment. We designed primers manually in the conserved intergenic fragments, with the help of the software program 'Primer Premier 5'. We optimized our experimental conditions as follows: PCR product size was 95–200 bp and the primer length was 20 ± 3 bp. Generally, we collected 1–3 conserved intergenic fragments for regions 100 bp to 2 kbp long, 3–6 for regions 2–5 kbp long and 6–10 for regions larger than 5 kbp. We designed one pair of primers in each conserved intergenic fragment. We tested primers on *D. melanogaster* genomic DNA. We only accepted primers with proper PCR bands in agarose gel. If not, we redesigned the primers. The tiling array from the previous study had a 25 bp probe every 35 bp along the genome with variation in different locations. Most intergenic fragments overlapped with three probes (34%), four probes (57%) or five probes (6%) (Figure S1). We chose the housekeeping gene *Gapdh2* as a positive control and II171a, a chromosome II-linked intergenic region that has been determined to be silent for transcription, as a negative control. The primer sequences for positive and negative controls were designed as

Gapdh2 forward primer: 5′ AAGAACACTACCCACCC ACA 3′;
Gapdh2 reverse primer: 5′ ATTTCAGAGCATAGCGCA TA 3′;
II171a forward primer: 5′ GTTGGAGAAACTAAGAGC C 3′ and
II171a reverse primer: 5′ CCATACTCAATCGGACC 3′.

Designing the RT–PCR primers of pre-miRNA

There are 152 known microRNAs (miRNAs) in *D. melanogaster* (from miRBase, http://www.microrna.sanger.ac.uk/) (22). Ten were chosen for RT–PCR analysis, including mir-184, mir-277, mir-279, mir-100, mir-286, mir-31b, mir-125, mir-2c, mir-219 and mir-1017 (Table S1). Stem-loop sequences were obtained from miRBase (http://www.microrna.sanger.ac.uk/). Ten pairs of primers for the miRNAs were designed manually with the help of software 'Primer Premier 5'.

Material preparation

Drosophila tissues. About 400 adults were kept in a bottle and allowed to lay eggs at 25°C on grape juice-agar plates. The flies hatched during the first hour were discarded. The eggs were collected for 2 h on new plates. Flies were removed and the new plates were incubated at the temperature of 25°C for 6 h and 20 h, respectively, yielding embryos of 6–8 h and 20–22 h stages. After incubation, embryos were washed with diethylpyrocarbonate (DEPC) water (DEPC-treated and RNase-free), and collected for RNA extraction (23).

Larvae of second to third instar stage were collected in a tube and rinsed three times with DEPC water. Adult flies 7–30 days old were collected in tubes and killed by keeping the tubes at −20°C for 5–10 min.

RNA extraction. Specimens were transferred into a mortar with liquid nitrogen and ground into powder with a pestle. RNAs were extracted from powder with TRIzol (Invitrogen Company) following TRIzol reagent instructions. About 10 μg of sample RNA was mixed with 20 μl RQ1 RNase-free DNase (1 unit/μl; Promega Company), 10 μl 10× DNase buffer, 2 μl RNase inhibitor HPR1 (Takara Company) and DEPC water (up to 100 μl), and incubated for 3 h at 37°C. Contamination of RNAs with DNA was ruled out by PCR amplification of two pairs of primers for *Gapdh2* and II171a on the extracted RNAs.

cDNA preparation. Total 0.5 μg 6-mer random primer (Takara Company) and 5 μl dNTPs (2.5 mM, Takara Company) per microgram of RNA sample were mixed in a total volume of ≤15 μl in a tube; the tube was heated to 70°C for 5 min to melt secondary structure;

the tube was cooled immediately on ice for at least 3 min to prevent secondary structure from reforming; 5 µl M-MLV 5× Reaction Buffer (Promega Company), 5 µl dNTPs, 0.5 µl RNase Inhibitor (HPR I, Takara Company) and 1 µl M-MLV RTase (Promega Company) were added to the annealed primer/template. Then the DEPC water was added to the volume of 25 µl. PCR of primer II171a was conducted to guarantee no genomic DNA contamination; and PCR of *Gapdh2* was conducted to guarantee the quality of cDNA.

PCR and RT–PCR amplification

DNAs from the extractions of adult flies were used for PCR, which was performed as described in the protocol provided with Tiangen DNA Extraction Kit. PCR and RT–PCR amplifications were performed with Transgen MasterMix (including Taq polymerase, buffer and dNTPs) according to the manufacturer's instructions. The total number of PCR cycles was 40. Cycle conditions included a 95°C denaturation (30 s, first cycle of 5 min), a 54°C annealing (30 s) and a 72°C extension (40 s). After the final cycle, tubes were placed at the extension temperature for 10 min. Machine model used was DNA Engine Dyad and Tetrad 2. PCR and RT–PCR amplification products were loaded on 3% (wt/vol) agarose gels, and were stained with ethidium bromide.

To confirm the consistency of RT–PCR products with the expected intergenic sequences, 33 RT–PCR products amplified from the expected fragments from four chromosomes (X, 2, 3 and 4) were chosen and sequenced using the cloning-sequencing approach with the pMD18-T cloning vector (23).

To control for the reliability, we chose 453 fragments on chromosome 2L to repeat the RT–PCR three times independently. Each time, we extracted the total RNA from different flies, and prepared the cDNA templates. We considered the fragments that showed positive signals in all three experiments as true positive.

Comparison of miRNAs. On stem-loop sequences of the 10 pre-miRNAs, primers were designed for RT–PCR analysis that amplified fragments longer than 100 bp. Nine out of the ten were detected in at least one of the four stages (6–8 h embryo, 20–22 h embryo, larva and adult stage). Probe signals in the same areas were also analyzed. Then we compared the RT–PCR results with the previous tiling array transfrags (Table S3; Figure S2).

Elongation of RT–PCR-detected fragments. Embryo ESTs were extracted from original *D. melanogaster* ESTs, and then BLATed (24) to map to whole genome (FlyBase 4.2). ESTs with sequence identity higher than 0.98 and coverage higher than 0.9 were kept. Those mapped to more than one genomic location were ignored. We, thereby, identified 241 975 and 198 125 ESTs that matched the previously detected tiling-array transfrag in the 6–8 h and 20–22 h embryos (14), respectively. Eighty-seven to ninety-one per cent of these matches in the two embryo stages were longer than 100 bp. Then overlapping ESTs of sense and anti-sense strand were merged together. After that, RT–PCR-detected fragments from 6–8 h and 20–22 h were combined and then elongated by merging overlapping ESTs. In this step, we altogether identified 404 ESTs that matched the RT–PCR-detected fragments in the two embryo stages, with 72.5% overlap length longer than 30 bp. Finally, RT–PCR-detected fragments that overlapped or did not overlap with transfrags were analyzed separately.

RESULTS

On the basis of the annotated genomic sequences of *D. melanogaster*, we chose ~10% (1027) of all intergenic regions, evenly distributed on chromosomes X, 2, 3 and 4, and tested these regions for the presence of conserved ncRNA genes. Because a conserved region might more likely be under selective constraint for a functional genetic element despite the exceptions (25), we focused on conserved intergenic regions. By comparing genomic sequences of six closely related *Drosophila* species with the VISTA tool (26) (see 'Materials and Methods' section), we chose 2755 conserved intergenic fragments for primer design and RT–PCR detection (Table S2).

Despite repeated and much improved annotations of the *D. melanogaster* genomic sequences (20,27), annotated intergenic regions in the genome today are still not completely free from undetected transcribed regions from the annotated genes. In fact, a previous tiling array analysis (14) found that many annotated genes have undetected distal 5′ exons in intergenic regions. To avoid these potential exon regions from the protein-coding genes, we designed primers in the middle of intergenic regions, with 65.2% intergenic fragments falling in the middle 80% of intergenic regions. We prepared cDNAs by PCR, reverse-transcribing total RNAs extracted from four developmental stages (6–8 h embryo, 20–22 h embryo, larvae and adult).

To determine the reliability of our RT–PCR protocol, we first chose 453 intergenic fragments on chromosome 2L to repeat the RT–PCR three times independently (Table S3). We extracted the total RNAs from three groups of flies from the same strain (Canton S) in *D. melanogaster* in the 6–8 h embryo stage. Then we prepared the cDNA templates from the extracted RNAs and conducted the RT–PCR experiments with the three batches of cDNAs. We detected the expression signals in 207, 177, 208 of 453 intergenic fragments in the 6–8 h embryo stage. The coefficient of variation, defined as standard variation over mean, is estimated to be low (CV = 8.9%), suggesting a high reliability of our protocol.

Furthermore, 98 out of 453 intergenic fragments were detected in all three batches. This provides a very conservative estimate for the ncRNA genes that are expressed in this stage: at least 22% of the intergenic fragments encode ncRNAs in the early developmental stage. On average of the three repeats, 197 intergenic fragments were detected to have expression signals. This average percentage, 43.6%, is close to the expression percentage (44%) calculated from the expressed fragment numbers that were detected in at least two experiments (98 detected in all

three batched with another 98 detected in two batches). 102 were detected in only one batch. Using this approach, we investigated the ncRNAs encoded in intergenic regions and made a comparison with the published results from the analysis of tiling arrays (14) (see 'Discussion' section).

We then extended this approach to all the sampled intergenic regions in the whole genome in all four developmental stages. On total, we conducted 18 342 successful RT–PCR experiments with both negative and positive control experiments. To confirm that the RT–PCR products were amplified from the expected regions, we sequenced RT–PCR products of 33 intergenic fragments from the adult flies (Table S4 and Figure S3); the results indicate that 21 of 33 RT–PCR products are 100% identical to the expected genomic sequences, 5 are 99% identical, 4 out of the remaining 7 are 93–98%. One has a lower but significant similarity (72%) whereas only two do not match expected sequences. The changes in the 30 fragments reflect nucleotide polymorphisms between the strain used (Canton S) and the strain used for the previous genome sequencing (y; cn bw sp) and one fragment is likely from a replicate copy of intergenic region in the particular strain we used. The nucleotide polymorphism between different lines of *D. melanogaster* was commonly observed previously (e.g. 28). The result indicated that the vast majority (94%) of the RT–PCR products are consistent with the expectation, providing additional strong evidence in support of our approach.

Six examples of detectable intergenic fragments with detectable expression were shown in Figure 1. Table 1, A and B showed the proportions of detectable expression in the intergenic regions and fragments, respectively. In total, we detected expression of 72.7% (747) of the intergenic regions, and 53.3% (1469) of the intergenic fragments at some point during the development processes (Table S2). Therefore, in the entire genome, we estimated that 8350 intergenic regions would show positive amplification signals, and likely encode ncRNAs.

Furthermore, different chromosomes and developmental stages showed significantly different levels of ncRNA expression. Intergenic transcription was more common on chromosomes 2 and 4 than on chromosomes X and 3. The early stages of development, the embryo stages had the highest numbers of detectable intergenic fragments, whereas the later stages had lower numbers of detectable intergenic fragments. These data revealed high numbers of expressed ncRNA genes in the early developmental stage and drastically reduced numbers of the expressed ncRNA genes in the late developmental processes (Table 1, panel B; Figure 2).

Using these large datasets, we conducted a systematic comparison with a large dataset from one developmental stage with less laborious tiling array approaches (14). Using the publicly available EST databases, we independently checked the correspondence of both our RT–PCR-generated candidates and the tiling array-generated candidates with EST sequences from the same intergenic regions. We discuss below the comparative analyses and reported estimates for the false positive rates of both the current RT–PCR and previous tiling array approaches.

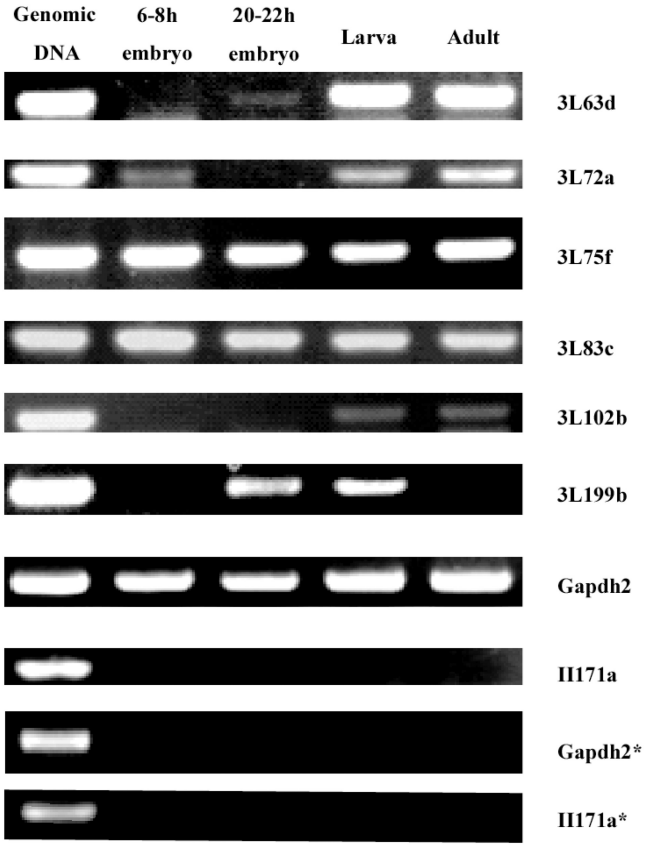

Figure 1. Examples of the ncRNA detected in the intergenic regions. 3L63d, 3L72a, 3L75f, 3L83c, 3L102b and 3L199b were six detected intergenic fragments in either of four developmental stages (6–8 h stage, 20–22 h stage, larva stage and adult stage). Gapdh2 was positive control and II171a was negative control. Genomic DNA and cDNAs of four stages were used as templates for PCR amplification. For rows of Gapdh2* and II171a*, genomic DNA and DNase I-digested RNAs of four stages were used as templates for PCR amplification, and the results ruled out contamination of RNAs with genomic DNA.

DISCUSSION

Our analyses detected transcription of putative ncRNAs from a high proportion of intergenic regions in *D. melanogaster*, and revealed that ncRNA transcription was more common in early developmental stages. The ncRNA genes we identified provided new rich databases for various development stages in fruit flies for further analyses of functions, structures and evolution of ncRNAs. One interesting issue is the detection power for previously published methods, e.g. the tiling array method (14), and the false positive rates for these methods. Especially, previous estimates by tiling array hybridization in one development stage (14) did not consider false positive rate and some of the detected ncRNA candidates were falsified recently (15,16). Thus, it is necessary to compare our method with the early methods and data and measure false positive rates for the ncRNA detection using genomic techniques.

We compared the RT–PCR results in our investigated intergenic fragments with the previous published tiling array data (1) in the same embryo stages, 6–8 h and

Table 1. Detection of intergenic fragments and intergenic regions in various developmental stages

Stage[a]/Chromosome[b]								
A: Proportions of detected intergenic regions								
	X(165) (in %)	2L(172) (in %)	2R(191) (in %)	2(363) (in %)	3L(180) (in %)	3R(243) (in %)	3(423) (in %)	4(76) (in %)
6–8 h	52.1	64.0	63.4	63.6	32.2	53.9	44.6	94.7
20–22 h	46.1	57.6	58.6	58.1	33.9	47.3	41.5	60.5
Embryo[c]	62.4	74.4	71.2	72.7	44.4	61.7	54.2	96.1
Larvae	37.6	29.7	29.3	29.5	32.2	33.3	32.8	42.1
Adult	35.2	35.5	25.7	30.3	34.4	40.3	37.7	60.5
Sum	70.3	79.1	75.4	77.1	56.7	72.0	65.3	97.4
B: Proportions of detected intergenic fragments								
	X(492) (in %)	2L(478) (in %)	2R(504) (in %)	2(982) (in %)	3L(447) (in %)	3R(646) (in %)	3(1093) (in %)	4(188) (in %)
6–8 h	29.1	45.2	39.3	42.2	19.7	31.4	26.6	81.9
20–22 h	23.8	36.4	34.7	35.5	19.5	28.3	24.7	41.0
Embryo[c]	36.8	56.5	47.2	51.7	28.4	41.2	36.0	86.7
Larvae	17.9	13.6	15.5	14.6	17.0	15.6	16.2	22.3
Adult	17.7	20.3	13.1	16.6	18.6	19.0	18.8	38.8
Sum	45.7	63.2	52.0	57.4	39.6	51.9	46.8	89.4

[a]The developmental stages we chose to present RT–PCR experiments.
[b]Different chromosomes are listed separately.
[c]Results of 6–8 h embryo stage and 20–22 h embryo stage were combined and shown as Embryo stage.

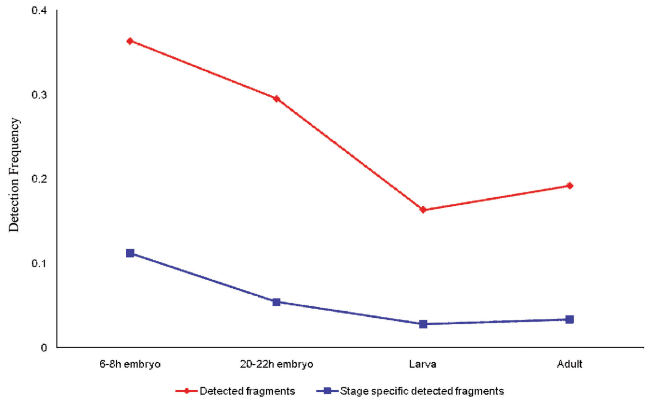

Figure 2. ncRNA amount in different developmental stages. The distribution of detected ncRNAs that were detected from the 2755 regains in the four developmental stages.

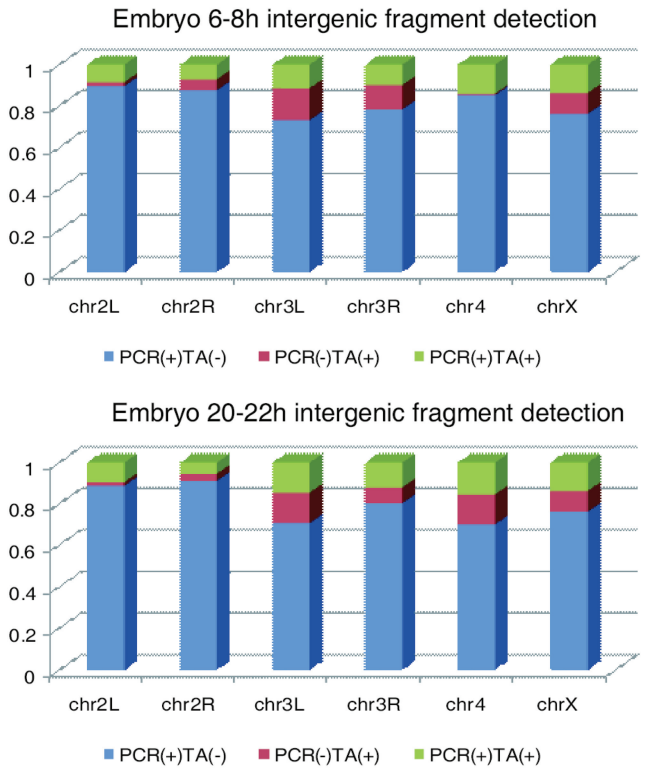

Figure 3. Comparison of results (the relative proportions of the three categories above) from RT–PCR and tiling array. RT–PCR detected intergenic fragments with tiling array-detected transfrags in the 6–8 h embryo stage and the 20–22 h embryo stage.

20–22 h embryo stages. Because transcripts might or might not be present in an intergenic fragment in both the tiling array and the RT–PCR methods, we categorized three possible cases, excluding the case that both methods detected no transcripts: (i) the RT–PCR method detected transcripts when the tiling array method detected no transcripts, (ii) the tiling array detected transcripts when the RT–PCR method detected no transcripts and (iii) both methods detected the same transcript. Figure 3 lists the distribution for these comparisons.

Figure 3 reveals that the two methods showed significant differences. Figure 3 indicates that in case (i), only RT–PCR method that detected transcripts (PCR(+)TA(−)) accounted for 72–89% on different chromosomes, whereas there are only 1–15% of tiling array-detected transcripts that cannot be verified by RT–PCR (PCR(−)TA(+)). In only 7–14% of the cases, the two methods can detect the same transcripts (PCR(+)TA(+)). In general, these comparisons showed a surprisingly low degree of overlapping. Thus we further asked what factors

Table 2. RT–PCR-detected intergenic fragments matched with transfrags before and after elongation of intergenic fragments across the assembled EST sequences

	Before	After
PCR(+)TA(−)	70	10
PCR(+)TA(+)	43	103

contribute to such a difference and which method was more reliable.

We infer that because of the different methodology involved, the RT–PCR and the tiling arrays would have different levels of sensitivity. Because the transfrags identified by the tiling array required a minimal run of 90 bp with no more than 40 bp (1) gaps in it thus the ncRNAs, the probes with less than four consecutive oligos would have a significant chance to be missed. The application of a sliding window, as installed in the tiling array analysis, increased the risk when the neighbouring signal is low. Take miRNAs for an example. miRNAs range from 21 to 23 nt, and are excised from pre-miRNAs (stem-loop RNA) with the length of around 100 nt. We tested 10 miRNAs that were previously reported (19), using the RT–PCR method (Table S1). Nine of them were detected (e.g. see Figure S2). However, the tiling array technique (14) detected no transfrags that were overlapping with these nine pre-miRNAs, although the array probes covered these genes. This analysis suggests that the tiling array is less sensitive for transcripts shorter than 120 nt.

We also considered the second possibility that the RT–PCR only detected a small part (~100 bp) of the ncRNA genes whose other regions were detected by the tiling array method. *Drosophila* EST data (see 'Materials and Methods' section) were used to test this possibility. We searched the EST database (Web address) (ftp://ftp.ncbi.nih.gov/blast/db/) and identified EST sequences that matched 113 RT–PCR-detected fragments. We observed that 61.9% of these RT–PCR-detected fragments were not detected by the tiling arrays (Table 2). Then we elongated these RT–PCR-detected fragments by assembled EST sequences. However, when we used the elongated fragments to search the predicted transfrags, we observed that only 8.8% (10) of the RT–PCR-detected fragments were not detectable by the tiling array method, i.e. 91.2% of the RT–PCR-detected fragments were detected by the tiling array method. These support that, despite likely missing short ncRNAs, tiling array method may detect a majority of ncRNA genes that have reasonably large lengths. Moreover, we identified 268 tiling array-detected intergenic fragments that had also been tested by our RT–PCR experiments. We found that 205 (76.5%) of these fragments that had either been detected by RT–PCR (188) or EST evidence (60), with 43 detected by both. This once again revealed the vast majority of the tiling array-identified fragments may be authentic.

With these data of abundant transcripts from intergenic regions, some of which were confirmed by both PCR-based and tiling array-based methods (14), we should ask whether they represent alternatively spliced exons encoding 5′ UTR of protein-coding genes or independently transcripted ncRNAs or the both cases. First, the tiling array analyses showed that only a small proportion (15.6%) of the detected intergenic transcripts might be 5′ exons of protein-coding genes and the vast majority had no evidence to be associated with the protein-coding genes (14). Second, our approach tried to avoid the regions close to flanking genes so were less likely to detect the 5′ exons than the tiling array method. Third, our approach indeed detected previously reported ncRNAs (miRNAs) with a very high detection rate. These observations, with the fine protein-coding gene annotations in this genome, favour a conclusion that the most detected intergenic transcripts more likely represent ncRNAs, although we should not reject possibility that a low percentage of the detected transcripts are likely unknown 5′ UTR exons of certain protein-coding regions.

The RT–PCR method added the first detection of intergenic ncRNAs in *D. melanogaster* that was independent of the tiling array method. We observed that this approach appeared to be more sensitive than previous tiling array method, especially, for detecting short ncRNAs, which were almost undetectable using tiling array method. However, the high proportion of matches (91.2%) of RT–PCR-detected fragments with the tiling array candidates was in consistence with the credible rate of false positives for duplicate polymorphisms (14%) in the independent PCR tests for genomic DNA hybridization with the tiling arrays recently detected in the populations of *D. melanogaster* (20).

Our results revealed that there was widespread expression of intergenic regions. This striking observation suggests the functional importance of non-coding regions in genomes. It has been known that many ncRNAs were associated with regulatory functions of protein-coding genes and non-protein coding genes (e.g. 2,4). Thus, in the light of our finding and these previous observations, the intergenic regions were no longer merely linked regions of adjacent genes. Instead, the intergenic regions may provide regulatory functions or various other molecular functions through RNA–protein complexes. Furthermore, the intergenic regions were conventionally viewed in evolution as neutral regions because of the belief of the lack of functional elements in the regions, except for a recent report that detected evolutionary signal from the possible coding potential in non-coding regions of the *Drosophila* genomes (29). Our findings further discounted the conventional notion and detected abundant ncRNAs encoded in the intergenic regions, which likely contributed to the functional evolution of non-coding regions. Our data also uncovered a large variation of ncRNAs in different developmental stages and provided credible estimates of false positive rates associated with the genomic measurement of ncRNAs. The rich ncRNAs at the early stage manifested the more frequent turning-on events of the genes important for the start of many developmentally important processes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Yajuan Lin, Long Qian, Rong Wang, Mengling Yang, Xianwei Zhang and Shuang Li participated in part of experimental works. We are indebted to the support from National Key Basic Research Program of China (973 Program) founded by Ministry of Science and Technology (MOST) (2007CB946900, 2007CB946904).

FUNDING

Director-funds from National Natural Science Foundation of China (NSFC) (#30550004 & #30750001), and the Changjiang Adjunct Professorship Funds of Peking University and a National Institutes of Health, USA, Grant No. (R01 GM078070-01A1) to M.L. Funding for open access charge: Ministry of Science and Technology of China 2007CB946900 & 2007CB946904.

Conflict of interest statement. None declared.

REFERENCES

1. Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
2. Szymanski,M., Erdmann,V.A. and Barciszewski,J. (2003) Noncoding regulatory RNAs database. *Nucleic Acids Res.*, **31**, 429–431.
3. Inagaki,S., Numata,K., Kondo,T., Tomita,M., Yasuda,K., Kanai,A. and Kageyama,Y. (2005) Identification and expression analysis of putative mRNA-like non-coding RNA in Drosophila. *Genes Cells*, **10**, 1163–1173.
4. Mattick,J.S. and Makunin,I.V. (2006) Non-coding RNA. *Hum. Mol. Genet.*, **1**, R17–R29.
5. Deng,W., Zhu,X., Skogerbo,G., Zhao,Y., Fu,Z., Wang,Y., He,H., Cai,L., Sun,H., Liu,C. *et al.* (2006) Organization of the Caenorhabditis elegans small non-coding transcriptome: genomic features, biogenesis, and expression. *Genome Res.*, **16**, 20–29.
6. Ravasi,T., Suzuki,H., Pang,K.C., Katayama,S., Furuno,M., Okunishi,R., Fukuda,S., Ru,K., Frith,M.C., Gongora,M.M. *et al.* (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.*, **16**, 11–19.
7. Pollard,K.S., Salama,S.R., Lambert,N., Lambot,M.A., Coppens,S., Pedersen,J.S., Katzman,S., King,B., Onodera,C., Siepel,A. *et al.* (2006) An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, **443**, 167–172.
8. Deng,X. and Meller,V.H. (2006) Non-coding RNA in fly dosage compensation. *Trends. Biochem. Sci.*, **31**, 526–532.
9. Martinho,R.G., Kunwar,P.S., Casanova,J. and Lehmann,R. (2004) A noncoding RNA is required for the repression of RNApolII-dependent transcription in primordial germ cells. *Curr. Biol.*, **14**, 159–165.
10. Misra,S., Crosby,M.A., Mungall,C.J., Matthews,B.B., Campbell,K.S., Hradecky,P., Huang,Y., Kaminker,J.S., Millburn,G.H., Prochnik,S.E. *et al.* (2002) Annotation of the Drosophila melanogaster euchromatic genome: a systematic review. *Genome Biol.*, **3**, RESEARCH0083.
11. Brunner,E., Ahrens,C.H., Mohanty,S., Baetschmann,H., Loevenich,S., Potthast,F., Deutsch,E.W., Panse,C., de Lichtenberg,U., Rinner,O. *et al.* (2007) A high-quality catalog of the Drosophila melanogaster proteome. *Nat. Biotechnol.*, **25**, 576–583.
12. Stolc,V., Gauhar,Z., Mason,C., Halasz,G., van Batenburg,M.F., Rifkin,S.A., Hua,S., Herreman,T., Tongprasit,W., Barbano,P.E. *et al.* (2004) A gene expression map for the euchromatic genome of Drosophila melanogaster. *Science*, **306**, 655–660.
13. Celniker,S.E. and Rubin,G.M. (2003) The Drosophila melanogaster genome. *Annu. Rev. Genomics Hum. Genet.*, **4**, 89–117.
14. Manak,J.R., Dike,S., Sementchenko,V., Kapranov,P., Biemar,F., Long,J., Cheng,J., Bell,I., Ghosh,S., Piccolboni,A. *et al.* (2006) Biological function of unannotated transcription during the early development of Drosophila melanogaster. *Nat. Genet.*, **38**, 1151–1158.
15. Nakamura,T., Naito,K., Yokota,N., Sugita,C. and Sugita,M. (2007) A cyanobacterial non-coding RNA, Yfr1, is required for growth under multiple stress conditions. *Plant Cell Physiol.*, **48**, 1309–1318.
16. Halasz,G., van Batenburg,M.F., Perusse,J., Hua,S., Lu,X.J., White,K.P. and Bussemaker,H.J. (2006) Detecting transcriptionally active regions using genomic tiling arrays. *Genome Biol.*, **7**, R59.
17. Oliver,B. (2006) Tiling DNA microarrays for fly genome cartography. *Nat. Genet.*, **38**, 1101–1102.
18. Frazer,K.A., Pachter,L., Poliakov,A., Rubin,E.M. and Dubchak,I. (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res.*, **32(Web Server issue)**, W273–W279.
19. Ashburner,M., Golic,K.G. and Hawley,R.S. (2005) *Drosophila: A Laboratory Handbook*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
20. Clark,A.G., Eisen,M.B., Smith,D.R., Bergman,C.M., Oliver,B., Markow,T.A., Kaufman,T.C., Kellis,M., Gelbart,W., Iyer,V.N. *et al.* (2007) Evolution of genes and genomes on the Drosophila phylogeny. *Nature*, **450**, 203–218.
21. Siepel,A., Bejerano,G., Pedersen,J.S., Hinrichs,A.S., Hou,M., Rosenbloom,K., Clawson,H., Spieth,J., Hillier,L.W., Richards,S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
22. Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34(Database issue)**, D140–D144.
23. Sambrook,J. and Russell,D.W. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
24. Kent,W.J. (2002) BLAT – the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
25. Pang,K.C., Frith,M.C. and Mattick,J.S. (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.*, **22**, 1–5.
26. Frazer,K.A., Pachter,L., Poliakov,A., Rubin,E.M. and Dubchak,I. (2004) VISTA: computational tools for comparative genomics. *Nucleic Acids Res.*, **32(Web Server issue)**, W273–W279.
27. Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F. *et al.* (2000) The genome sequence of Drosophila melanogaster. *Science*, **287**, 2185–2195.
28. Anderson,J.A., Gilliland,W.D. and Langley,C.H. (2009) Molecular population genetics and evolution of Drosophila meiosis genes. *Genetics*, **181**, 177–185.
29. Andolfatto,P. (2005) Adaptive evolution of non-coding DNA in Drosophila. *Nature*, **437**, 1149–1152.