Profiling Caenorhabditis elegans non-coding RNA expression with a combined microarray

Housheng He1,5, Lun Cai2,5, Geir Skogerbo1, Wei Deng1, Tao Liu1,5, Xiaopeng Zhu1,5, Yudong Wang1, Dong Jia1, Zhihua Zhang1,5, Yong Tao5,6, Haipan Zeng7, Muhammad Nauman Aftab1,5, Yan Cui4, Guozhen Liu7 and Runsheng Chen1,2,3,*

1Bioinformatics Laboratory and National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China, 2Computational Biology Research Group, Division of Intelligent Software Systems, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China, 3Chinese National Human Genome Center, Beijing 100176, China, 4Department of Molecular Sciences/Center of Genomics and Bioinformatics, University of Tennessee Health Science Center, Memphis, TN 38163, USA, 5Graduate School of the Chinese Academy of Sciences, Beijing 100080, China, 6Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China and 7Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 101300, China

Received January 17, 2006; Revised March 8, 2006; Accepted April 27, 2006

ABSTRACT

Small non-coding RNAs (ncRNAs) are encoded by genes that function at the RNA level, and several hundred ncRNAs have been identified in various organisms. Here we describe an analysis of the small non-coding transcriptome of Caenorhabditis elegans, microRNAs excepted. As a substantial fraction of the ncRNAs is located in introns of protein-coding genes in C.elegans, we also analysed the relationship between ncRNA and host gene expression. To this end, we designed a combined microarray, which included probes against ncRNA as well as host gene mRNA transcripts. The microarray revealed pronounced differences in expression profiles, even among ncRNAs with housekeeping functions (e.g. snRNAs and snoRNAs), indicating distinct developmental regulation and stage-specific functions of a number of novel transcripts. Analysis of ncRNA–host mRNA relations showed that the expression of intronic ncRNA loci with conserved upstream motifs was not correlated to (and much higher than) expression levels of their host genes. Even promoter-less intronic ncRNA loci, though showing a clear correlation to host gene expression, appeared to have a surprising amount of ‘expressional freedom’, depending on host gene function. Taken together, our microarray analysis presents a more complete and detailed picture of a non-coding transcriptome than hitherto has been presented for any other multicellular organism.

INTRODUCTION

Non-protein-coding RNAs (ncRNAs) are encoded by genes that function at the RNA level. Over the years it was believed that there were few ncRNAs and that they were mainly accessory components to aid protein functioning. However, over the last decade it has become apparent that there are numerous ncRNAs and that their cellular functions are varied and important (1–3). Several strategies have been employed to detect and discover novel ncRNAs, including both experimental and computational screening (3,4). Results from the recently developed tilling arrays also indicated that much larger portions of eukaryote transcriptomes represent non-coding transcripts than believed previously (5).

Our group has recently cloned and verified a set of 161 ncRNAs (corresponding to 198 genetic loci) in Caenorhabditis elegans, including nearly all known and predicted ncRNAs as well as 100 novel transcripts (6). Analysis of this material revealed several novel aspects of the C.elegans small non-coding transcriptome. The genomic organization of small ncRNAs in C.elegans is peculiar, in that nearly half of the loci, corresponding to a variety of different ncRNA types, are intronic. Three putative ncRNA-specific core promoters were identified, located at intronic as well as intergenic loci. Among the novel transcripts that could not be assigned to any known class of ncRNAs, we identified two putatively novel functional classes of ncRNAs.

*To whom correspondence should be addressed at Bioinformatics Laboratory, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. Tel: +86 10 64888543; Fax: +86 10 64889892; Email: crs@sun5.ibp.ac.cn

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

© 2006 The Author(s).
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
The functions of most ncRNAs remain unknown, and their numbers are steadily increasing. Though most commonly used for measuring mRNA expression, microarrays have been applied recently to analyse genome wide expression of ncRNAs; most of the recent expression profile analysis of ncRNAs, however, have focused on microRNAs (7–10), which regulate the expression of their mRNA targets. Only very few microarray analyses have been carried out on other ncRNAs. As ncRNA and mRNA are often functionally related, a combined analysis of mRNA and ncRNA expression would potentially provide much more information concerning their function than separate analyses (10,11).

We developed a combined microarray for unbiased simultaneous analysis of ncRNA and mRNA levels. The combined microarray was modified from a commonly used cDNA microarray for mRNA expression analysis designed with 48mer oligonucleotides. Using this combined microarray strategy, we carried out a developmental expression analysis of nearly the entire ncRNA complement in *C. elegans*, excluding microRNAs. The expressional relations between intronic ncRNAs and their corresponding host mRNAs were also analysed.

MATERIALS AND METHODS

Microarray construction

All the 161 cloned ncRNAs (6) (Supplementary Data) were selected for probe design using Oligoarray 2.1 (12). A set of stringent criteria was used to design the probes. Oligo length was set to 48 nt, and the GC content confined to a range from 35 to 55% to narrow the distribution of predicted melting temperatures. The thresholds to reject secondary structures and cross-hybridization were set to 65, and sequences containing tracts of more than six identical nucleotides were excluded. Using these criteria, 127 probes were designed. In a few cases one probe was complementary to two or more very similar transcripts (Supplementary Table 1), such that the 127 probes represented 134 ncRNAs.

Probes against 205 mRNAs, including 74 host genes, 3 housekeeping genes and 128 pathway-related genes were also designed using the same criteria; the pathway-related mRNAs were later used for normalization of raw data. SSC buffer and an oligo with no homology (<9 bp identity) to the *C. elegans* genome were used as negative hybridization controls, and two rice mRNA oligos with 14 bp identity to two *C. elegans* mRNAs were used as non-specific hybridization controls. The three housekeeping genes were used as positive internal controls. Oligos were printed in triplicate on the microarray (some controls were printed in sextuplicate) by PE SpotArray72, UV crosslinked at 3000 mJ, and stored at 4°C.

Sample preparation and microarray hybridization

RNA was extracted from heat shock-treated worms (mixed stage wild-type, N2 strain worms treated at 30°C for 3 h), starved L1 larvae (L1s) and from worms at seven different developmental stages [Egg, L1, L2, L3, L4 and mature adult (MA) and dauer]. All developmental stages were determined by time of culture since feeding of L1s worms at 20°C. In addition, a batch of reference RNAs used for all experiments was extracted from mixed staged N2 worms. mRNA was first isolated from total RNA using the MICROBExpress™ Bacterial mRNA Purification Kit, and non-poly(A) RNA was isolated from the remaining fraction as described previously (6) with minor modifications. Briefly, the Ambion MicroBExpress kit was adapted to remove remaining rRNAs by hybridizing the non-poly(A) RNA fraction to a mixture of specifically designed oligos, which targeted and removed unwanted RNA molecules with a magnetic bead based protocol (Ambion). Thereafter, the purified non-poly(A) RNA fraction was dephosphorylated with calf intestine alkaline phosphatase (Fermentas), and ligated to a 3’ adapter oligonucleotide with T4 RNA ligase (Fermentas). The ligated ncRNAs were reverse transcribed using an oligonucleotide complementary to the 3’ adapter, while the mRNA was reverse transcribed using an oligodT12–18 primer. The cDNAs from the ncRNA and mRNA fractions were combined and labelled with Cy5 (or Cy3) using the Ambion Amino Alky1 cDNA labeling kit (Supplementary Figure 1).

The microarrays were prehybridized at 50°C for 2 h and hybridized at 42°C for 14–16 h. Microarrays were scanned using Genepix 4000B scanner, and raw data were acquired and quantified with GenePix software.

Computational methods

The raw data were processed using the MIDAS (TIGR TM4) software. Background was subtracted from the median pixel intensity values for Cy3 and Cy5, and data points were removed if intensities did not exceed 2-fold of background levels for both Cy3 and Cy5. Total intensity normalization and LocFit normalization were applied with housekeeping genes as controls. The MIDAS in-slide replicate analysis was applied to merge replicates of each gene. The Cy5/Cy3 ratios were log-transformed (base 2), and TMEV (TIGR TM4 software) was used for hierarchical clustering (ncRNA and host gene mRNA expression data are available in Supplementary Data). A Z-score was calculated for each gene under each condition (Supplementary Data). Genes with both a Z-score and a sample/reference ratio exceeding (or equal to) ±2 were identified as differentially expressed. Permutation P-values of the Pearson correlation coefficient were computed by fixing one profile and randomly permuting the entries of the other profile. To search for conserved motif in host gene upstream sequences, the MEME motif discovery tool (version 3.0.13) (13) was used.

RESULTS

Microarray specificity

To examine the expression of ncRNA and mRNA simultaneously, we adapted methodology commonly used for quantifying mRNA expression by designing a combined microarray. Probes of 48mer oligos were designed for both ncRNAs and mRNAs with the same criteria, and various specificity controls were applied to validate the microarray data. Negative and non-specific hybridization controls introduced to examine the stringency of the combined microarray system all showed very low signal/noise ratios (<1 on average). To assay slide reproducibility, duplicate hybridizations were carried out (Supplementary Data), giving an average
Pearson correlation coefficient (henceforth ‘r’) between slide replicates of 0.93.

To further validate the data, normalized signal intensities of ncRNAs extracted from mixed stage worms were compared to the number of clones identified in an ncRNA library (6), resulting in an r-value of 0.83 ($P < 10^{-6}$) between chip signal intensities and library clone number. The mRNA expression data were compared with published data for mRNA expression at the dauer stage (14). For the 137 mRNAs assayed in both experiments, the r-value was 0.67, despite the differences in probe design and experimental conditions [the r-values between the four replicates of the published data (14) vary from 0.7 to 0.96].

In general, the hybridization stringency was sufficient to distinguish members of closely related ncRNAs families, and the differences in their signal intensities corresponded well to their differences in library clone numbers, i.e. the frequency of each ncRNA among the ~2000 library clones from which they were identified (6). For example, CeN25-1 and CeN25-4 are 66% similar over a span of 77 nt; however, the difference in their chip signal intensities closely matched their difference in clone number (Supplementary Figure 2), demonstrating that the microarray was able to distinguish between closely related ncRNA species. However, for probes representing more than one member of a family (e.g. probe ‘CeN36-1’; Supplementary Table 1) or ncRNAs with multiple genetic loci (e.g. snRNA U6), the observed intensity was likely to represent the combined expression of more than one locus.

To validate the normalization method we used for the combined microarray data, we carried out a pilot experiment of self versus self hybridization in which an RNA sample was divided into two equal parts and labelled with Cy3 and Cy5, respectively, before hybridization. The raw data show slight deflection upwards of Cy5 samples at high intensities and a somewhat stronger deflection downwards of the Cy3 sample at low intensities (Supplementary Figure 3a). After normalization (Supplementary Figure 3b and c), the deflections were greatly reduced, yielding an r-value between the two samples of 0.98.

**ncRNA expression profiles**

Using the combined microarray, we examined the expression of 127 ncRNAs across seven developmental stages and two stimulated conditions. All the ncRNAs were robustly expressed with signal intensities >10-fold over negative controls. Thirty-six transcripts showed >2-fold variation in at least one development stage or stimulated condition, and of these, 25 were identified as significantly over- or under-expressed (Supplementary Data). These 25 ncRNAs broadly fell into two groups (Table 1), one comprising ncRNAs (mostly snoRNAs) whose expression reached high levels under stress (heat shock and starvation), while the other included ncRNAs that were expressed at high levels mainly at the egg-embryo or mature adult stages. The latter is a mixed group composed of spliceosomal snRNAs, spliced leader RNAs and snRNA-like RNAs (snlRNAs, see below). Most of the snoRNAs showed low expression levels at the egg-embryo stage, and one (CeN79) was significantly under-expressed.

<table>
<thead>
<tr>
<th>ncRNA-A</th>
<th>Functional class</th>
<th>Stage</th>
<th>Z-score</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>CeN19</td>
<td>snRNA SL2</td>
<td>HS</td>
<td>2.34</td>
<td>Under stress</td>
</tr>
<tr>
<td>CeN90</td>
<td>snoRNA H/ACA</td>
<td>HS</td>
<td>2.56</td>
<td></td>
</tr>
<tr>
<td>CeN5</td>
<td>snoRNA C/D</td>
<td>L1s</td>
<td>2.46</td>
<td></td>
</tr>
<tr>
<td>CeN113</td>
<td>snoRNA C/D</td>
<td>L1s</td>
<td>2.21</td>
<td></td>
</tr>
<tr>
<td>CeN40</td>
<td>snoRNA C/D</td>
<td>L1s</td>
<td>2.50</td>
<td></td>
</tr>
<tr>
<td>CeN103</td>
<td>snoRNA C/D</td>
<td>L1s</td>
<td>2.43</td>
<td></td>
</tr>
<tr>
<td>CeN104</td>
<td>snoRNA H/ACA</td>
<td>L1s</td>
<td>2.07</td>
<td></td>
</tr>
<tr>
<td>CeN94</td>
<td>snoRNA H/ACA</td>
<td>L1s</td>
<td>2.03</td>
<td></td>
</tr>
<tr>
<td>CeN105</td>
<td>snoRNA H/ACA</td>
<td>L1s</td>
<td>2.38</td>
<td></td>
</tr>
<tr>
<td>CeN48</td>
<td>snoRNA H/ACA</td>
<td>Dauer</td>
<td>2.08</td>
<td></td>
</tr>
<tr>
<td>CeN68</td>
<td>snoRNA H/ACA</td>
<td>L4</td>
<td>2.13</td>
<td>Developmental stage</td>
</tr>
<tr>
<td>CeN79</td>
<td>snoRNA H/ACA</td>
<td>Egg</td>
<td>−2.08</td>
<td></td>
</tr>
<tr>
<td>CeN75</td>
<td>sbRNA</td>
<td>Egg</td>
<td>−2.01</td>
<td></td>
</tr>
<tr>
<td>CeN12</td>
<td>snRNA SL2</td>
<td>Egg</td>
<td>2.48</td>
<td></td>
</tr>
<tr>
<td>CeN16-4</td>
<td>snRNA SL2</td>
<td>Egg</td>
<td>2.10</td>
<td></td>
</tr>
<tr>
<td>CeN16-1</td>
<td>snRNA SL2</td>
<td>Egg</td>
<td>2.22</td>
<td></td>
</tr>
<tr>
<td>CeN7</td>
<td>snRNA SL2</td>
<td>Egg</td>
<td>2.10</td>
<td></td>
</tr>
<tr>
<td>CeN3-6</td>
<td>snoRNA U5</td>
<td>Egg</td>
<td>2.35</td>
<td></td>
</tr>
<tr>
<td>CeN3-5</td>
<td>snoRNA U5</td>
<td>Egg</td>
<td>2.55</td>
<td></td>
</tr>
<tr>
<td>CeN9</td>
<td>snoRNA YRNA</td>
<td>Egg</td>
<td>2.26</td>
<td></td>
</tr>
<tr>
<td>CeN31</td>
<td>Sm Y/snlRNA</td>
<td>Egg</td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td>CeN25-4</td>
<td>Sm Y/snlRNA</td>
<td>Egg</td>
<td>2.58</td>
<td></td>
</tr>
<tr>
<td>CeN115</td>
<td>Sm Y/snlRNA</td>
<td>MA</td>
<td>2.12</td>
<td></td>
</tr>
<tr>
<td>CeN112</td>
<td>Sm Y/snlRNA</td>
<td>MA</td>
<td>2.08</td>
<td></td>
</tr>
<tr>
<td>CeN20</td>
<td>snRNA SL2</td>
<td>MA</td>
<td>2.22</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Developmental stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>CeN19 (2.34)</td>
</tr>
<tr>
<td>CeN90 (2.56)</td>
</tr>
<tr>
<td>CeN5 (2.46)</td>
</tr>
<tr>
<td>CeN113 (2.21)</td>
</tr>
<tr>
<td>CeN40 (2.50)</td>
</tr>
<tr>
<td>CeN103 (2.43)</td>
</tr>
<tr>
<td>CeN104 (2.07)</td>
</tr>
<tr>
<td>CeN94 (2.03)</td>
</tr>
<tr>
<td>CeN105 (2.38)</td>
</tr>
<tr>
<td>CeN48 (2.08)</td>
</tr>
<tr>
<td>CeN68 (2.13)</td>
</tr>
<tr>
<td>CeN79 (−2.08)</td>
</tr>
<tr>
<td>CeN75 (−2.01)</td>
</tr>
<tr>
<td>CeN12 (2.48)</td>
</tr>
<tr>
<td>CeN16-4 (2.10)</td>
</tr>
<tr>
<td>CeN16-1 (2.22)</td>
</tr>
<tr>
<td>CeN7 (2.10)</td>
</tr>
<tr>
<td>CeN3-6 (2.35)</td>
</tr>
<tr>
<td>CeN3-5 (2.55)</td>
</tr>
<tr>
<td>CeN9 (2.26)</td>
</tr>
<tr>
<td>CeN31 (2.40)</td>
</tr>
<tr>
<td>CeN25-4 (2.58)</td>
</tr>
<tr>
<td>CeN115 (2.12)</td>
</tr>
<tr>
<td>CeN112 (2.08)</td>
</tr>
<tr>
<td>CeN20 (2.22)</td>
</tr>
</tbody>
</table>

They broadly fall into two groups, one comprising ncRNAs (mostly snoRNAs) that reached high levels under stress, while the other includes ncRNAs that were highly expressed at the egg-embryo (Egg) or mature adult (MA) stage. The latter is a mixed group composed mainly of spliceosomal snRNAs, spliced leader RNAs and Sm Y/snlRNA.

To further analyze the ncRNA expression patterns, the expression datasets of all ncRNAs were hierarchically clustered using TMEV3.0 (TM4 software; Supplementary Figure 4), resulting in 12 clusters with distinct expression patterns, generally overlapping with ncRNA functional groups. Clusters 3 and 4 (Figure 1a and b) were dominated by snRNAs, a recently detected group of ncRNAs which may correspond to the Sm Y RNA of *Ascaris lumbricoides* (Supplementary Table 3) (15). The Sm Y/snlRNAs are found at eight loci in *C. elegans* (of which six are represented in the microarray), and probably at six additional loci (Supplementary Table 2). The Sm Y/snlRNAs resemble snRNAs both in that they contain the Sm protein-binding site and in that their loci, with a few exceptions, share the same upstream motifs (6). The expression profiles of both clusters peaked at the MA stage, with cluster 3 also showing high expression at the egg-embryo stage. The fact that snoRNA U5 is included in cluster 3 reinforces the impression that the Sm Y/snlRNAs may be functionally related to snRNAs.

Cluster 6 and 7 (Figure 1c and d) were dominated by SL2 RNAs, which are involved in splicing of oporonic genes. There are 12 variants of the SL2 RNAs in *C. elegans*, corresponding to ~20 loci (16) (http://www.wormbook.org), of which 11 (9 probes) were included in the microarray. The SL2 RNAs were highly expressed at early developmental stages, and showed variable expression patterns across other developmental stages or stimulated conditions. The SL1 RNA, which only occurs in one transcribed variant despite the existence of ~110 loci (16) (http://www.wormbook.org), showed a quite different, unvaried expression pattern.
through the entire *C. elegans* life cycle and under stimulated conditions (Supplementary Figure 5a). Most trans-splicing in the worm occurs with SL1 RNA (17,18) (http://www.wormbook.org), and the expression level of SL1, as measured by signal intensities or northern blots, was higher than those of any individual SL2 RNA, in accordance with earlier findings of 7–10 times higher levels of SL1 compared with SL2 RNA (18) (http://www.wormbook.org). However, the total levels of SL2 RNAs may be higher than that of SL1 (Supplementary Figure 5b and c), which has not been reported previously.

Seven clusters were dominated by C/D box and H/ACA box snoRNAs, which usually function as 2'-O-ribose methylation and pseudouridylation guides. Although the expression patterns of the seven clusters are different, they share similarity in showing low expression at the egg-embryo stage, which is different from most of the RNAs involved in splicing processes (snRNA, SL RNA). In addition to their high expression levels at the egg-embryo and L1s stages, a small group of four snoRNAs (part of cluster 8; Figure 1e) is interesting since three of them have possible modification sites on snRNAs U5 and U6 (Figure 2; we did not find modification sites for the others), while most of the other snoRNAs have modification sites on RNAs (data not shown). Taking into account that snRNA U5 and U6 both showed high expression levels at the earlier developmental stages, this out-group of RNAs might function as small Cajal body RNAs (scaRNAs) guiding modification of bases on snRNAs U5 and U6.

**ncRNA versus host gene expression**

Of the 127 ncRNAs analysed, 77 have intronic loci, and probes against their corresponding host gene mRNAs were included in our microarray. One might expect to find some degree of positive correlation between the expression level of an ncRNA embedded in an intron, and the expression level of its host gene mRNA; however, for the total dataset no such trend was discernible ($r = 0.08$). ncRNA expression levels (as indicated by normalized signal intensities) were also on average more than six times higher than those for the host gene mRNAs. When, on the other hand, the data were broken down according to whether the respective (intronic) ncRNA loci contained any of the two major conserved upstream motifs (UM1 or UM2) (6) clear differences emerged (Figure 3a). The signal intensities of ncRNAs from UM1 and UM2 loci (henceforth ‘motif-loci’) were on average of the order 55-fold higher than those of host gene mRNAs, and no obvious correlation between ncRNA and mRNA expression was visible (UM1 loci, $r = 0.17$; UM2 loci, $r = 0.12$).

For intronic ncRNAs with no upstream motif (henceforth, ‘non-motif loci’) the picture was quite different. There was a clear positive correlation between ncRNA and mRNA signal...
We describe here a microarray analysis of the expression of the known inventory of short ncRNAs (microRNA excepted) in *C. elegans*, including an investigation of the expressional relationships between intronic ncRNA loci and their host genes. About 20% of the ncRNAs were differentially expressed across developmental stages and stimulated conditions, and ncRNAs within the same functional groups tended to have similar expression patterns. The analysis of ncRNA and host mRNA expression revealed that their relationship is clearly related to the absence or presence of conserved upstream elements at the ncRNA loci, and to some extent also to host mRNA function.

**Table 2.** Average signal intensities of intronic ncRNA and host gene mRNA pairs

<table>
<thead>
<tr>
<th>ncRNA locus</th>
<th>No. of pairs</th>
<th>ncRNAs</th>
<th>mRNAs</th>
<th>Corr(r)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upstream motif 1</td>
<td>23</td>
<td>35 308</td>
<td>480</td>
<td>−0.17</td>
<td>0.01</td>
</tr>
<tr>
<td>Upstream motif 2</td>
<td>20</td>
<td>25 287</td>
<td>650</td>
<td>−0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>No upstream motif</td>
<td>33</td>
<td>14 367</td>
<td>7574</td>
<td>0.47</td>
<td>&lt;10^{-8}</td>
</tr>
<tr>
<td>All RNAs</td>
<td>76</td>
<td>23 591</td>
<td>3561</td>
<td>−0.07</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Values are normalized signal intensities for Cy-labelled samples from mixed *C. elegans* populations. The expression levels of ncRNAs with upstream motifs were on average 55-fold of their host gene mRNAs, while that for ncRNAs without upstream motif were only 2-fold. Corr(r) represented the correlation expression level between ncRNAs and their host gene mRNAs. *P*-value is a significant indicator for corr(r).

The advantage of the combined chip

We have demonstrated that for certain types of expression studies, similar to the intronic ncRNA–host gene mRNA

**Figure 3.** Relationship between intronic ncRNA and host gene expression levels. The values were normalized signal intensities for Cy-labelled samples from mixed *C. elegans* populations. (a) Expression levels of intronic ncRNA genes with either no upstream motif (Non-UM) or with upstream motif 1 (UM1) or 2 (UM2). ncRNAs with UM1 and UM2 showed no obvious correlation with their host gene mRNAs, while the non-motif ncRNAs showed a positive correlation with their host gene mRNAs. (b) Host gene function influenced the expressional relationship between ncRNA and host mRNA. The correlation between the expression levels of ribosomal host genes (Ribo) and their intronic ncRNAs was higher (*r* = 0.83) than that of non-ribosomal hosts (Non-ribo) and their intronic ncRNAs (*r* = 0.39).
relationship studied here, a combined microarray that can
detect ncRNA and mRGN levels simultaneously can be a
useful tool. Identical probe design strategies for ncRNA and
mRNA ensure similar experimental conditions, and although
ncRNA and mRNA can be analysed using separate micro-
arrays under the same experimental conditions, a combined
microarray is superior to this strategy in two ways. Even
under identical experimental conditions there will be
variation caused by array operation, so that between-slide
variation is generally greater than within-slide variation.
Another problem is that commonly used normalization
methods, which were developed for analysis of a large
number of genes (22) are not suitable for a chip with a smal-
er number of ncRNAs probes; however, in the combined
microarray, the ncRNAs can be normalized against a larger
number of mRNAs. Several systems have been described
recently for expression profiling of microRNAs (7–10).
Baskerville and Bartel (10) and Thompson et al. (8) used a
synthetic reference set, which provides a uniform positive
control for hybridization and a valuable internal standard
for normalization, but might not be appropriate for longer
ncRNAs and mRNAs.

**Upstream motifs signify independent ncRNA expression**

Apart from early work on tRNAs (23), little direct
experimental validation of ncRNA promoters has been
carried out in *C. elegans*, and assumptions regarding the
nature of ncRNA transcription therefore either rest on infer-
ences from data on other organisms (24), or on sequence
homology analyses of upstream and internal elements (6).
A putative ‘proximal sequence element (PSE)’ was reported
for several spliceosomal snRNA genes based on sequence
homology between upstream flanks of different loci (25);
however, little sequence homology was found between the
*C. elegans* PSE variant and those of snRNAs in other organ-
isms. Though the *C. elegans* PSE does not appear to have
been systematically investigated, there is a report on expres-
sion of an SL2 RNA driven by its own promoter, and directed
mutagenesis of some of the most conserved bases of the PSE
strongly reduced expression of the downstream gene (26).
The PSE is also embedded in a longer upstream motif (UM1) found recently at 82 intergenic and intronic loci of
verified ncRNAs, including snRNAs and a number of other
ncRNAs, and a second upstream motif (UM2) shows strong
similarity to the (internal) tRNA promoter (6). On the other
hand, differences in steady state expression levels observed
from a microarray do not constitute direct evidence for
differences in transcription, and may in principle be derived
from differences in synthesis rates, differences degradation
rates or both. However, given the larger number of transcripts
studied simultaneously in a microarray assay, it is still
possible to make a few deductions regarding which of these
factors are most plausible. A simple explanation for the
elevated expression levels of intrinsic motif-loci ncRNAs
compared with their corresponding host mRNAs might be
that the motifs act as a promoter elements, driving inde-
pendent transcription of the intronic ncRNA transcript to
much higher levels. This explanation also fits with the fact
that for intronic non-motif ncRNAs, differences in expression
levels between ncRNAs and host mRNAs were far more
moderate, and the expression levels of ncRNAs and host
mRNAs correlated well, as one would expect if the ncRNAs
were derived from splicing, and subsequent intron processing,
of a common pre-mRNA transcript.

Alternatively, the differences in expression levels between
ncRNAs and host mRNAs might be due a higher stability of the
ncRNAs. This is quite plausible, as most of these ncRNAs
(e.g. snRNAs, snoRNAs) form larger ribonucleoprotein
complexes, which are likely to shield them from exonuclease
degradation, and thus may confer lower turnover rates
compared with those for host mRNAs. It may also explain
why expression levels of both motif and non-motif ncRNAs
are somewhat higher that those of their corresponding host
mRNAs. However, differences in RNA stability do not easily
explain why this difference is much higher for motif ncRNAs
than for non-motif ncRNAs. This difficulty is particularly
apparent when intrinsic UM2 loci are compared to their
non-motif counterparts, since both these groups consist
mainly of snoRNAs, which form the same type of RNP
complexes, and should, on average, be equally protected
from degradation. Therefore, unless one is to assume that
the motifs are somehow able to confer long-term stability
on the ncRNAs transcripts after splicing and processing, it
is difficult to account for the observed differences in expres-
sion levels by differential transcript stability alone. A third
possibility is that the motifs might influence pre-mRNA
splicing and/or subsequent processing of excised intron lari-
tots to mature ncRNAs. This could clearly produce differences in
ncRNA and host mRNA expression levels, either by increas-
ing the efficiency of the intron lariat processing, or alterna-
tively by reducing the efficiency of mRNA formation.

Combined with the higher stability of ncRNAs in general,
increased lariat processing might explain the generally higher
expression levels of non-motif ncRNAs. However, the major-
ity of the intrinsic UM1 ncRNAs are snRNAs or other types
of transcripts for which no mechanism for snRNA release
from intron lariats are known, and in the single case where
an snRNA is known to be encoded within an intron (27),
its locus was shown to be fully equipped with an active pro-
moter driving independent transcription of this RNA.

Reduced mRNA formation might explain why host mRNAs
of motif ncRNAs generally have rather low expression levels,
but would not explain the higher expression levels of motif
versus non-motif ncRNAs, unless we also assume that
genes hosting motif ncRNAs have very high transcription
rates.

Therefore, though differences in stability may contribute
to the observed differences in expression levels between
ncRNAs, and influence from the upstream motifs on splicing
or ncRNA processing cannot be excluded as a hypothetical
possibility, we believe that the most parsimonious
explanation for the differences in expression among motif
and non-motif ncRNAs and their corresponding host
mRNAs is that the motifs have a role in transcriptional
activation of their respective loci. This is more valid when
we consider that these motifs occupy similar positions as
do verified promoter elements of corresponding ncRNAs in
other organisms (24), and that these motifs are also found
at a number of intergenic loci (6), which are bound to be
independently transcribed.
Transcriptionally regulated transcription

The combination of a high number of potentially independently transcribed intronic loci (6) and the tendency of the expression of intronic motif-loci to be negatively correlated to host gene expression opens the possibility for an additional level of transcriptional control in *C. elegans*. An inverse variation in transcriptional activity between a host gene and an intronic ncRNA makes biological sense, since a transcriptional peak in one would be of less risk to interfere with a transcriptional peak in the other. Alternatively, highly expressed ncRNAs might prefer to locate to host genes with low expression. Recent analysis of the ‘transcriptional landscape’ of the mouse genome indicates that a huge number of loci may actually be ‘forests’ of transcripts overlapping in both sense and antisense direction, this providing an additional level of transcriptional control (28). Transfer RNA genes cluster in the nucleolus during transcription by RNA polymerase III (29), and if a similar localization takes place when an intronic UM2-locus is transcribed, possibilities for transcription of the host gene could be substantially impaired. However, although a compartmentalization of an intronically located ncRNA probably would constitute a particularly strong form of inhibition of host gene transcription, constant engagement of an intronic promoter element with a transcription factor complex could potentially also reduce the transcriptional activity of the host gene. Similarly, high transcriptional activity of the host gene could prevent engagement of an intronic promoter as well as localization of the intronic locus to the nucleolus, thus exerting control over expression of the ncRNA.

snoRNAs released by cleavage or splicing—the putative TOP promoter

In vertebrates, snoRNA loci tend to concentrate in ribosomal and other genes featuring a 5′ end oligopyrimidine tract (also called a TOP promoter) which controls both transcription and translation of these genes (30,31). It has been shown recently that the TOP promoter also direct the ratio between splicing and cleavage of the pre-mRNA (21). In non-vertebrates, no analysis of ribosomal gene promoter elements appears to have been published; however, a search for 5′ end features of such genes in *C. elegans* identified pyrimidine-rich tracts around the (putative) transcription start site of several ribosomal host genes. For the ribosomal host genes, the correlation between ncRNA and host mRNA expression was far weaker, in good agreement with data indicating that 5′ end elements of such host genes may allow for ncRNA generation by both splicing and cleavage of the pre-mRNA (21), thus yielding far more variable ncRNA-host mRNA ratios.

Although the complement of small ncRNAs analysed here has been studied intensively for the better part of the last two decades or more, a full-scale expression analysis across the entire developmental course of a multicellular organism has nevertheless yielded interesting insights into how these molecules interact with the greater molecular apparatus of the cell. The genomic organization of the *Caenorhabditis* small non-coding transcriptome is quite peculiar, and the simultaneous analysis of both ncRNA and host gene expression produced strong evidence for independent transcription of a large fraction of intronic ncRNA loci from their respective ncRNA-specific promoter elements. This in turns leaves open the possibility for an additional level of transcriptional control in the nematodes, in which both ncRNA and host gene transcription may be reciprocally regulated through mutual inhibition.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Rob Wilson for careful reading and correction of the manuscript. This work was supported by the National High Technology Development Program of China under Grant No. 2002AA231031, National Sciences Foundation of China under Grant No. 60496324, National Key Basic Research and Development Program 973 under Grant Nos. 2002CB713805 and 2003CB715907. Funding to pay the Open Access publication charges for this article was provided by National Key Basic Research and Development Program 973 under Grant No. 2003CB715907.

Conflict of interest statement. None declared.

REFERENCES


