2976–2983 *Nucleic Acids Research, 2006, Vol. 34, No. 10*
doi:10.1093/nar/gkl371

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

Housheng He[1,5], Lun Cai[2,5], Geir Skogerbø[1], Wei Deng[1], Tao Liu[1,5], Xiaopeng Zhu[1,5], Yudong Wang[1], Dong Jia[1], Zhihua Zhang[1,5], Yong Tao[5,6], Haipan Zeng[7], Muhammad Nauman Aftab[1,5], Yan Cui[4], Guozhen Liu[7] and Runsheng Chen[1,2,3,*]

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

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 oligod$T_{12–18}$ primer. The cDNAs from the ncRNA and mRNA fractions were combined and labelled with Cy5 (or Cy3) using the Ambion Amino Allyl 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 was 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 1. Twenty-five ncRNAs identified as significantly over- or under-expressed

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

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 analyse 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 snlRNAs, 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 snRNA 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 operonic 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

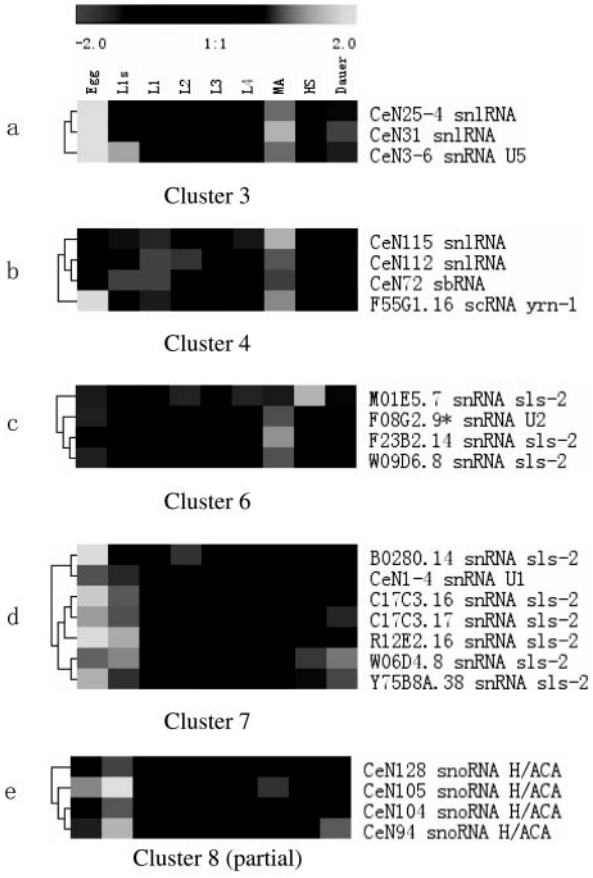

Figure 1. Selected clusters of ncRNA expression profiles. ncRNAs within functional groups usually showed similar expression patterns. The cluster numbers are identical to Supplementary Figure 4. An asterix indicates that the ncRNA has more than one locus in the genome. (**a** and **b**) Two clusters dominated by Sm Y/snlRNAs. (**c** and **d**) Clusters dominated by SL2 RNAs. (**e**) A group of snoRNAs that showed high expression levels at early developmental stages.

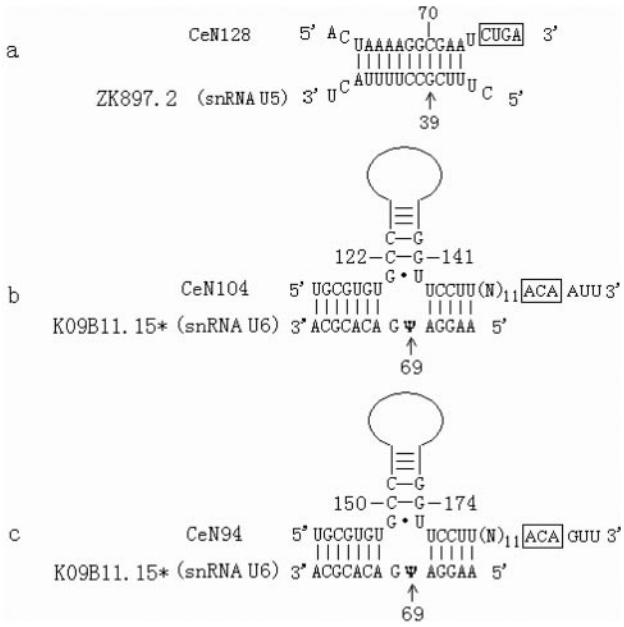

Figure 2. Predicted methylation and pseudouridylation guide duplex between snoRNAs and snRNAs. The snoRNA sequences in a 5′–3′ orientation are shown in the upper strands, while snRNA sequences in a 3′–5′ orientation are shown in the lower strands. Sequence motifs are boxed and the positions of modification sites are indicated by arrows and numbers. The upper parts of the hairpins of the H/ACA snoRNAs are represented by continuous lines. (**a**) Predicted methylation guide duplex between C/D box snoRNA CeN128 and U5 snRNA. (**b** and **c**) Predicted pseudouridylation guide duplex between H/ACA box snoRNAs CeN104 and CeN94 and U6 snRNA.

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 rRNAs (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

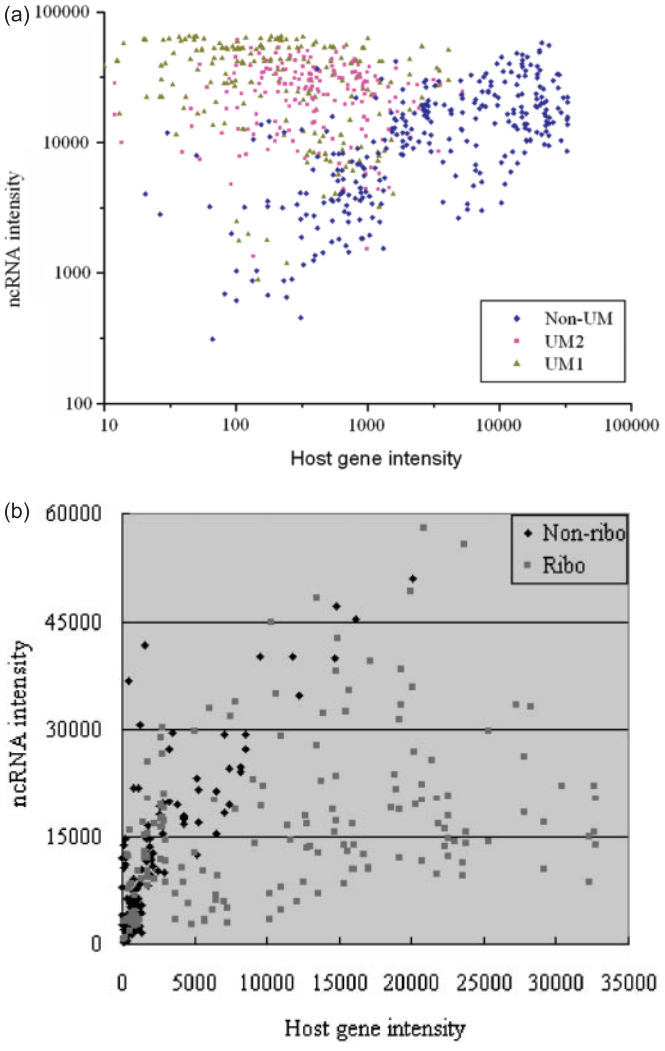

Figure 3. Relationship between intronic ncRNA and host gene expression levels. The values were normalized signal intensities for Cy-labelled samples from mixed *C.elegans* populations. (**a**) Expression levels of intronic ncRNA genes with either no upstream motif (Non-UM) or with upstream motif 1 (UM1) 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 ($r = 0.39$).

intensities ($r = 0.47$, $P < 0.01$). The ratio of ncRNA to mRNA average signal intensities was no more than 2, one to two orders of magnitude lower than that for motif loci ncRNAs (Table 2), and strongly indicative of tight relationships between host gene transcription and ncRNA biogenesis. Examining the expression profiles of individual ncRNA–host mRNA pairs over developmental stages and stimulated conditions revealed similar differences. Among the non-motif loci, about half of ncRNA–host mRNA pairs showed similar expression profiles (i.e. $r > 0.4$), whereas such pairs made up only 17 and 15% of UM1-loci and UM2-loci pairs, respectively (Supplementary Table 4). These results indicate that the expression of the motif-loci RNA is independent of that of the host gene. Adding the fact that both motifs are found

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

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

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 moitf 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).

at both intronic and intergenic loci (the latter necessarily being independently transcribed), these data strongly suggest that independent expression is associated with the UM1 and UM2 sequences.

In vertebrates, co-transcribed intronic ncRNAs are released from host pre-mRNAs by splicing or cleavage, or by a combination of both. In vertebrates, all ribosomal and some other snoRNA host genes commonly feature a 5′ terminal oligopyrimidine tract, also called a TOP promoter (19,20), shown to influence the choice of pre-mRNA processing pathway (21). When the non-motif group was divided according to whether the host gene protein is associated with the ribosome or not, we found that the correlation between ncRNA and host gene mRNA expression levels (signal intensities) increased to 0.83 for the non-ribosomal host-genes, whereas as it fell to only 0.39 for the ribosomal group (Figure 3b). Examination of expression profiles for individual ncRNA–host mRNA pairs indicated a similar tendency. Among such pairs with a ribosomal host gene, 32% (6/19) had a correlation coefficient >0.4, whereas the corresponding figure for pairs with a non-ribosomal host gene was 71% (10/14). A search for a putative TOP promoter among ribosomal proteins showed that 9 of the 14 ribosomal host genes share similar sequence elements around their transcription initiation sites (Supplementary Figure 6), suggesting that TOP promoters may also regulate transcription and processing of snoRNA host genes in *C.elegans*.

DISCUSSION

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.

The advantage of the combined chip

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

relationship studied here, a combined microarray that can detect ncRNA and mRNA 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 microarrays 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 smaller 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 inferences 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 organisms. Though the *C.elegans* PSE does not appear to have been systematically investigated, there is a report on expression 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 intronic motif-loci ncRNAs compared with their corresponding host mRNAs might be that the motifs act as a promoter elements, driving independent 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 intronic 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 expression 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 lariats to mature ncRNAs. This could clearly produce differences in ncRNA and host mRNA expression levels, either by increasing the efficiency of the intron lariat processing, or alternatively 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 majority of the intronic 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 promoter 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 directs 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 expressional 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

1. Eddy,S.R. (2002) Computational genomics of noncoding RNA genes. *Cell*, **109**, 137–140.
2. Gottesman,S. (2002) Stealth regulation: biological circuits with small RNA switches. *Genes*, **16**, 2829–2842.
3. Huttenhofer,A., Brosius,J. and Bachellerie,J.P. (2002) RNomics: identification and function of small, non-messenger RNAs. *Curr. Opin. Chem. Biol.*, **6**, 835–843.
4. Nam,J.W., Shin,K.R., Han,J., Lee,Y., Kim,V.N. and Zhang,B.T. (2005) Human microRNA prediction through a probabilistic co-learning model of sequence and structure. *Nucleic Acids Res.*, **33**, 3570–3581.
5. Rinn,J.L., Euskirchen,G., Bertone,P., Martone,R., Luscombe,N.M., Hartman,S., Harrison,P.M., Nelson,F.K., Miller,P., Gerstein,M. *et al.* (2003) The transcriptional activity of human Chromosome 22. *Genes Dev.*, **17**, 529–540.
6. Deng,W., Zhu,X., Skogerbo,G., Zhao,Y., Fu,Z., Wang,Y., He,H., Cai,L., Sun,H., Liu,C. *et al.* (2006) Organisation of the *Caenorhabditis elegans* small non-coding transcriptome: genomic features, biogenesis and expression. *Genome Res.*, **16**, 20–29.
7. Miska,E.A., Alvarez-Saavedra,E., Townsend,M., Yoshii,A., Sestan,N., Rakic,P., Constantine-Paton,M. and Horvitz,H.R. (2004) Microarray analysis of microRNA expression in the developing mammalian brain. *Genome Biol.*, **5**, R68.
8. Thomson,J.M., Parker,J., Perou,C.M. and Hammond,S.M. (2004) A custom microarray platform for analysis of microRNA gene expression. *Nature Methods*, **1**, 47–53.
9. Barad,O., Meiri,E., Avniel,A., Aharonov,R., Barzilai,A., Bentwich,I., Einav,U., Gilad,S., Hurban,P., Karov,Y. *et al.* (2004) MicroRNA expression detected by oligonucleotide microarrays: system establishment and expression profiling in human tissues. *Genome Res.*, **14**, 2486–2494.
10. Baskerville,S. and Bartel,D.P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, **11**, 241–247.
11. 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.

12. Rouillard,J.M., Zuker,M. and Gulari,E. (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. *Nucleic Acids Res*., **31**, 3057–3062.
13. Bailey,T.L. and Elkan,C. (1995) The value of prior knowledge in discovering motifs with MEME. *Proc. Int. Conf. Intell. Syst. Mol. Biol*., **3**, 21–29.
14. Wang,J. and Kim,S.K. (2003) Global analysis of dauer gene expression in *Caenorhabditis elegans*. *Development*, **130**, 1621–1634.
15. Maroney,P.A., Yu,Y.T., Jankowska,M. and Nilsen,T.W. (1996) Direct analysis of nematode *cis*- and *trans*-spliceosomes: a functional role for U5 snRNA in spliced leader addition *trans*-splicing and the identification of novel Sm snRNPs. *RNA*, **2**, 735–745.
16. Stricklin,S.L., Griffiths-Jones,S. and Eddy,S.R. (2005) *C.elegans* noncoding RNA genes. In WormBook (ed.), *The C.elegans Research Community*. WormBook, doi/10.1895/wormbook.1.1.1.
17. Zorio,D.A., Cheng,N.N., Blumenthal,T. and Spieth,J. (1994) Operons as a common form of chromosomal organization in *C.elegans*. *Nature*, **372**, 270–272.
18. Blumenthal,T. (2005) trans-splicing and operons. In WormBook (ed.), *The C.elegans Research Community*. WormBook, doi/10.1895/wormbook.1.5.1.
19. Smith,C.M. and Steitz,J.A. (1998) Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5′-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol. Cell. Biol*., **18**, 6897–6909.
20. Kruszka,K., Barneche,F., Guyot,R., Ailhas,J., Meneau,I., Schiffer,S., Marchfelder,A. and Echeverria,M. (2003) Plant dicistronic tRNA–snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z. *EMBO J*., **22**, 621–632.
21. De Turris,V., Di Leva,G., Caldarola,S., Loreni,F., Amaldi,F. and Bozzoni,I. (2004) TOP promoter elements control the relative ratio of intron-encoded snoRNA versus spliced mRNA biosynthesis. *J. Mol. Biol*., **344**, 383–394.
22. Quackenbush,J. (2002) Microarray data normalization and transformation. *Nature Genet*., **32**, 496–501.
23. Ciliberto,G., Castagnoli,L., Melton,D.A. and Cortese,R. (1982) Promoter of a eukaryotic tRNAPro gene is composed of three noncontiguous regions. *Proc. Natl Acad. Sci. USA*, **79**, 1195–1199.
24. Hernandez,N. (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J. Biol. Chem*., **276**, 26733–26736.
25. Thomas,J., Lea,K., Zucker-Aprison,E. and Blumenthal,T. (1990) The spliceosomal snRNAs of *Caenorhabditis elegans*. *Nucleic Acids Res*., **18**, 2633–2642.
26. Evans,D. and Blumenthal,T. (2000) *trans* splicing of polycistronic *Caenorhabditis elegans* pre-mRNAs: analysis of the SL2 RNA. *Mol. Cell. Biol*., **20**, 6659–6667.
27. Dominski,Z., Yang,X.-C., Purdy,M. and Marzluff,W.F. (2003) Cloning and characterization of the *Drosophila* U7 small nuclear RNA. *Proc. Natl Acad. Sci. USA*, **100**, 9422–9427.
28. The FANTOM Consortium (2005) The transcriptional landscape of the Mammalian Genome. *Science*, 1559–1563.
29. Thompson,M., Haeusler,R.A., Good,P.D. and Engelke,D.R. (2003) Nucleolar clustering of dispersed tRNA genes. *Science*, **302**, 1399–1401.
30. Shibui-Nihei,A., Ohmori,Y., Yoshida,K., Imai,J., Oosuga,I., Iidaka,M., Suzuki,Y., Mizushima-Sugano,J., Yoshitomo-Nakagawa,K. and Sugano,S. (2003) The 5′ terminal oligopyrimidine tract of human elongation factor 1A-1 gene functions as a transcriptional initiator and produces a variable number of Us at the transcriptional level. *Gene*, **311**, 137–145.
31. Zhu,J., Hayakawa,A., Kakegawa,T. and Kaspar,R.L. (2001) Binding of the La autoantigen to the 5′ untranslated region of a chimeric human translation elongation factor 1A reporter mRNA inhibits translation *in vitro*. *Biochim. Biophys. Acta*, **1521**, 19–29.