

Research article

Open Access

## ***In vivo* analysis of *Caenorhabditis elegans* noncoding RNA promoter motifs**

Tiantian Li<sup>†1,4</sup>, Housheng He<sup>†1,4</sup>, Yunfei Wang<sup>1,4</sup>, Haixia Zheng<sup>1,5</sup>, Geir Skogerbø\*<sup>1</sup> and Runsheng Chen\*<sup>1,2,3</sup>

Address: <sup>1</sup>Bioinformatics Laboratory and National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences Beijing 100101, PR China, <sup>2</sup>Bioinformatics Research Group, Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Science, Beijing 100080, PR China, <sup>3</sup>Chinese National Human Genome Center, Beijing 100176, PR China, <sup>4</sup>Graduate School of the Chinese Academy of Science, Beijing 100080, PR China and <sup>5</sup>Shanxi Agricultural University, Taigu, Shanxi, 030801, PR China

Email: Tiantian Li - litiantian\_222@hotmail.com; Housheng He - johnhsh1@hotmail.com; Yunfei Wang - tszn1984@163.com; Haixia Zheng - zhenghaixia722@163.com; Geir Skogerbø\* - zgb@moon.ibp.ac.cn; Runsheng Chen\* - crs@sun5.ibp.ac.cn

\* Corresponding authors †Equal contributors

Published: 5 August 2008

Received: 22 April 2008

BMC Molecular Biology 2008, 9:71 doi:10.1186/1471-2199-9-71

Accepted: 5 August 2008

This article is available from: <http://www.biomedcentral.com/1471-2199/9/71>

© 2008 Li et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### **Abstract**

**Background:** Noncoding RNAs (ncRNAs) play important roles in a variety of cellular processes. Characterizing the transcriptional activity of ncRNA promoters is therefore a critical step toward understanding the complex cellular roles of ncRNAs.

**Results:** Here we present an *in vivo* transcriptional analysis of three *C. elegans* ncRNA upstream motifs (UM1-3). Transcriptional activity of all three motifs has been demonstrated, and mutational analysis revealed differential contributions of different parts of each motif. We showed that upstream motif 1 (UM1) can drive the expression of green fluorescent protein (GFP), and utilized this for detailed analysis of temporal and spatial expression patterns of 5 SL2 RNAs. Upstream motifs 2 and 3 do not drive GFP expression, and termination at consecutive T runs suggests transcription by RNA polymerase III. The UM2 sequence resembles the tRNA promoter, and is actually embedded within its own short-lived, primary transcript. This is a structure which is also found at a few plant and yeast loci, and may indicate an evolutionarily very old dicistronic transcription pattern in which a tRNA serves as a promoter for an adjacent snoRNA.

**Conclusion:** The study has demonstrated that the three upstream motifs UM1-3 have promoter activity. The UM1 sequence can drive expression of GFP, which allows for the use of UM1::GFP fusion constructs to study temporal-spatial expression patterns of UM1 ncRNA loci. The UM1 loci appear to act in concert with other upstream sequences, whereas the transcriptional activities of the UM2 and UM3 are confined to the motifs themselves.

### **Background**

Genome wide analyses have in recent years revealed an increasing number of noncoding RNAs (ncRNAs) [1-12], however, the functional roles of these ncRNAs are mostly still unknown. Characterizing the transcriptional activity

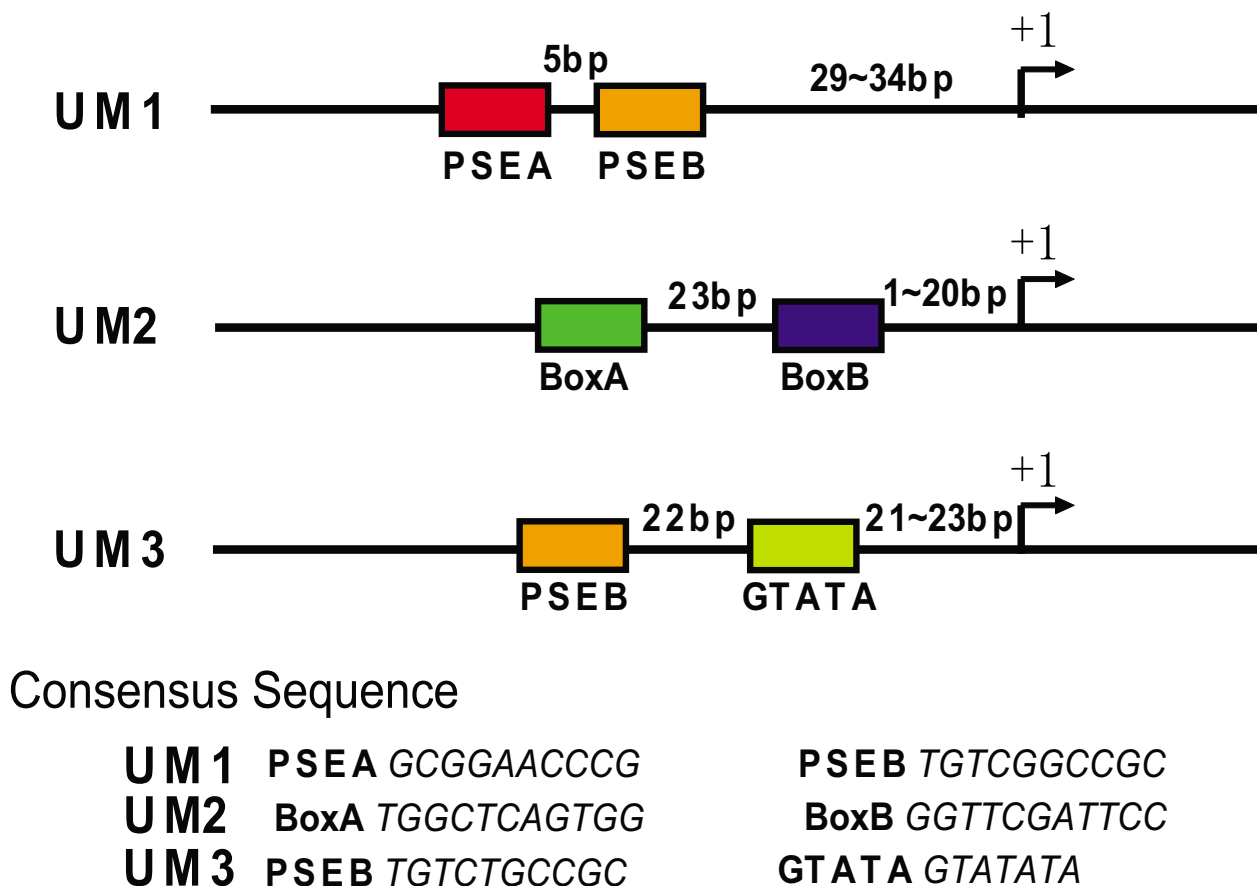
of the promoters of these loci could be a useful step towards revealing their functional roles. Extensive analysis of ncRNA promoters have been carried out in human, *Drosophila* and yeast [13-18]. There are more than 200 known short ncRNAs loci reported in *C. elegans* (microR-

NAs and tRNAs not included) and a recent tiling microarray study suggests the existence of an additional 1200 short transcripts with unknown function (TUFs) [1]. However, compared to the extensive work on promoter of protein coding genes, few analyses of ncRNA promoter activity have been performed in this species [19-24].

Analysis of the 100 bp upstream sequences of 161 *C. elegans* ncRNAs using the MEME software [25] detected three distinct 50 bp upstream motifs (upstream motifs 1~3, henceforth UM1-3) [6]. Among the 161 ncRNAs, UM1 is found at the loci of 54 ncRNAs, including 23 snRNAs, 11 snoRNAs and 11 snlRNAs. UM2 is found at the loci of 47 ncRNAs, of which 40 are snoRNAs. UM3 is found at the loci of 9 ncRNAs, all are stem-bulge RNAs (sbRNAs). Of the 1222 transcripts of unknown function (TUF)[1], UM1-3 are found at 76, 44 and 4 loci, respectively.

The core sequence of the 50 bp long UM1 is the 21 bp long snRNA proximal sequence element (PSE) [26]. This core sequence is composed of two sub-motifs spaced by 5 bp (Figure 1). In the *Drosophila* PSE the corresponding sub-motifs have been denoted as PSEA and PSEB [27] and we have used the same denotation here. Most of the *C. elegans* PSE/UM1 loci are TATA-less, and transcripts generated from such loci generally carry a 5'-end cap, suggesting transcription by polymerase II [6].

In the second motif (UM2) the most invariant sub-motifs are spaced by 33 bp and strongly resemble the Box A and B motifs of the tRNA internal promoter (Figure 1), which is known to bind transcription factor IIIC (TFIIIC) [28]. It is possible that UM2 is derived from tRNA genes that have served as promoters for downstream ncRNA genes, as similar tRNA-snoRNA dicistronic transcriptional structures have been described in plants and yeast [29-32]. Most of



**Figure 1**  
**Structures of the three putative promoters of *C. elegans* noncoding RNAs.** The most invariant submotifs in each promoter have been indicated by boxes. "+1" corresponds to the first base of the mature endogenous RNA sequence.

the UM2 loci encode snoRNAs which generally produce uncapped transcripts terminated at an oligo-T tract, and are thus likely to be transcribed by polymerase III [6], though the lack of a cap could also be due to processing of the primary snoRNA transcript [32,33].

The third motif (UM3) resembles the PSE/UM1 in that they both contain the PSEB sub-motif, but UM3 lacks PSEA, and has in addition a structure with the consensus motif GTATA located closer to the ncRNA transcription start site (TSS; Figure 1). UM3 is exclusively found at stem-bulge RNA (sBRNA) loci. The sBRNAs are terminated at an oligo-T tract, and most appear to be uncapped, indicating transcription by RNA polymerase III [6].

Short ncRNA loci in *C. elegans* are frequently found in introns of protein coding genes [2,6]. Such loci may or may not have an upstream motif. Previous analyses have found that for ncRNAs located in introns, the expression levels of intronic ncRNA loci not containing any obvious upstream motifs are closely correlated to the expression levels of the host genes [34]. On the other hand, for intronic ncRNA locus containing an upstream motif, the expression of the ncRNA locus is uncorrelated to the host gene expression, indicating that ncRNA loci with upstream motifs are independently transcribed [34].

No specific analysis of the ncRNA promoters appears to have been carried out in *C. elegans*, however, previous analysis has shown that a fragment including the first 162 bp upstream of the SL4 RNA (a variant of SL2 RNA, CeN7, which locus contains PSE/UM1) is sufficient to drive *LacZ* expression [35], and a mutation of four bases in the PSEB submotif resulted in a 10-fold reduction in transcription of an SL2 RNA [36]. We here demonstrate the transcriptional activity of the three *C. elegans* ncRNA promoters. The roles of the most invariant sub-motifs were investigated by mutation analysis, and the extent of upstream sequence with influence on the ncRNA transcriptional activity was analysed.

## Results

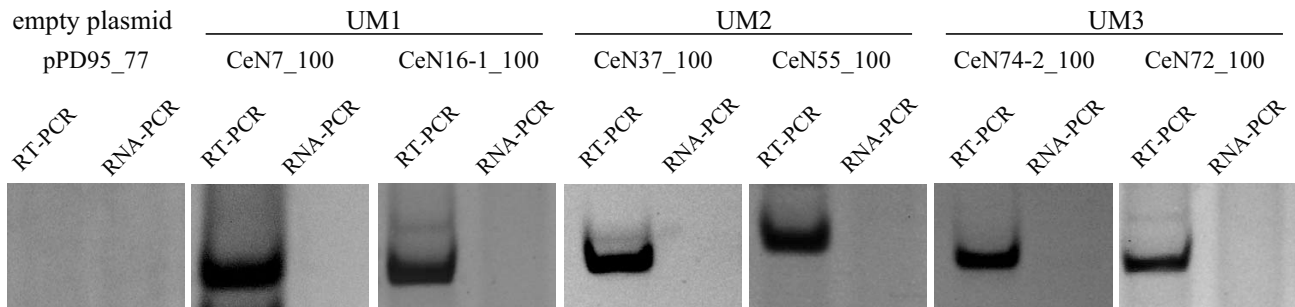
In order to analyse the transcriptional activities of the three common ncRNA motifs, we made constructs containing varying amounts (~100 bp, ~300 bp and ~1 kb) of upstream sequences (including 30–70 bp of transcribed sequence from each selected ncRNA locus) fused to the green fluorescent protein (GFP) open reading frame (ORF). Constructs containing approximately 100 bp upstream sequences (denoted by LOCUS NAME\_100) were used to test the inherent transcriptional activity of each upstream motif, whereas constructs with longer upstream sequences (LOCUS NAME\_300, LOCUS NAME\_1kb) were employed to investigate the possibility of additional regulatory elements.

### All the three upstream motifs are transcriptionally active

Cloning of approximately 100 bp upstream sequence encompassing each of the three upstream motifs 1–3 in front of chimeric ncRNA::GFP reporter genes suggested that all three motifs have independent transcriptional activity. To test whether the PSE/UM1 is sufficient for ncRNA expression, we made a reporter construct consisting of a fragment of the CeN7 locus including 90 bp upstream and 67 bp transcribed (i.e. -90 to +67 bp; Additional file 1) sequence. The CeN7 locus encodes an SL2 RNA with a TATA-less PSE/UM1 upstream sequence. The fragment was inserted into plasmid pPD95\_77, which contains a GFP ORF, thereby creating a CeN7::GFP chimeric reporter gene (henceforth CeN7\_100). The recombinant gene was co-injected with the rol-6 marker gene into young adult hermaphrodite gonads, and transgenic strains were selected based on the presence of the roller phenotype. Reporter gene expression was examined by reverse transcription polymerase chain reaction (RT-PCR) with one primer located in the CeN7 fragment and the other primer in the plasmid sequence, using RNA extracted from transgenic *C. elegans* strains (Figure 2). To confirm that the observed transcriptional activation was not a spurious result, we further selected an additional PSE/UM1 SL2 RNA locus (CeN16-1). The construct CeN16-1\_100 was made as above from a fragment containing 119 bp upstream sequence and 60 bp transcribed sequence, and expression of the fusion reporter (CeN16-1::GFP) was confirmed with RT-PCR (Figure 2). Transgenic strains containing the empty plasmid pPD95\_77 were also obtained the same way as described above, but no reporter gene expression was observed by RT-PCR.

The upstream motif 2 is composed of two sub-motifs with sequence and spacing similar to that of the A and B boxes of the tRNA promoter. To address the transcriptional activity of UM2, two constructs were made. One (CeN37\_100) was made from a fragment containing 130 bp upstream sequence and 56 bp transcribed sequence from the snoRNA locus CeN37, the other (CeN55\_100) from a fragment containing 134 bp upstream sequence and 89 bp transcribed sequence from the snoRNA locus CeN55. RT-PCR with one primer located in the CeN37 and CeN55 fragment and the other primer in the plasmid sequence demonstrated the expression of transgenes CeN37::GFP and CeN55::GFP, respectively (Figure 2).

Upstream motif 3 is exclusively found at stem-bulge RNA (sBRNA) loci. To determine the transcriptional activity of UM3, 140 bp and 120 bp of the upstream sequence of the sBRNA loci CeN74-2 and CeN72, respectively, was cloned in front of the GFP ORF in pPD95\_77 to yield constructs CeN74-2\_100 and CeN72\_100 (the constructs also included 32 and 36 bp of the respective transcribed sBRNA sequences, fused to the GFP reporter gene). RT-

**Figure 2**

**Upstream motifs drive reporter gene expression.** Expression of the reporter gene was detected by RT-PCR using DNase I digested RNA extracted from transgenic worms as template. For the negative controls (RNA-PCR) RT-PCR was performed using the same template without adding reverse transcriptase.

PCR verified the expression of both reporter fusions (Figure 2).

#### **Additional regulatory elements**

To assay whether the sequence upstream of each motif might contain additional regulatory elements, we also constructed plasmids containing approximately 300 to 1000 bp fragments of the 5'-flanking sequence from the CeN7, CeN37 and CeN74-2 loci, respectively. Transgene expression levels were assayed by quantitative RT-PCR (qRT-PCR), and for each construct, 2 to 5 transgenic lines were tested.

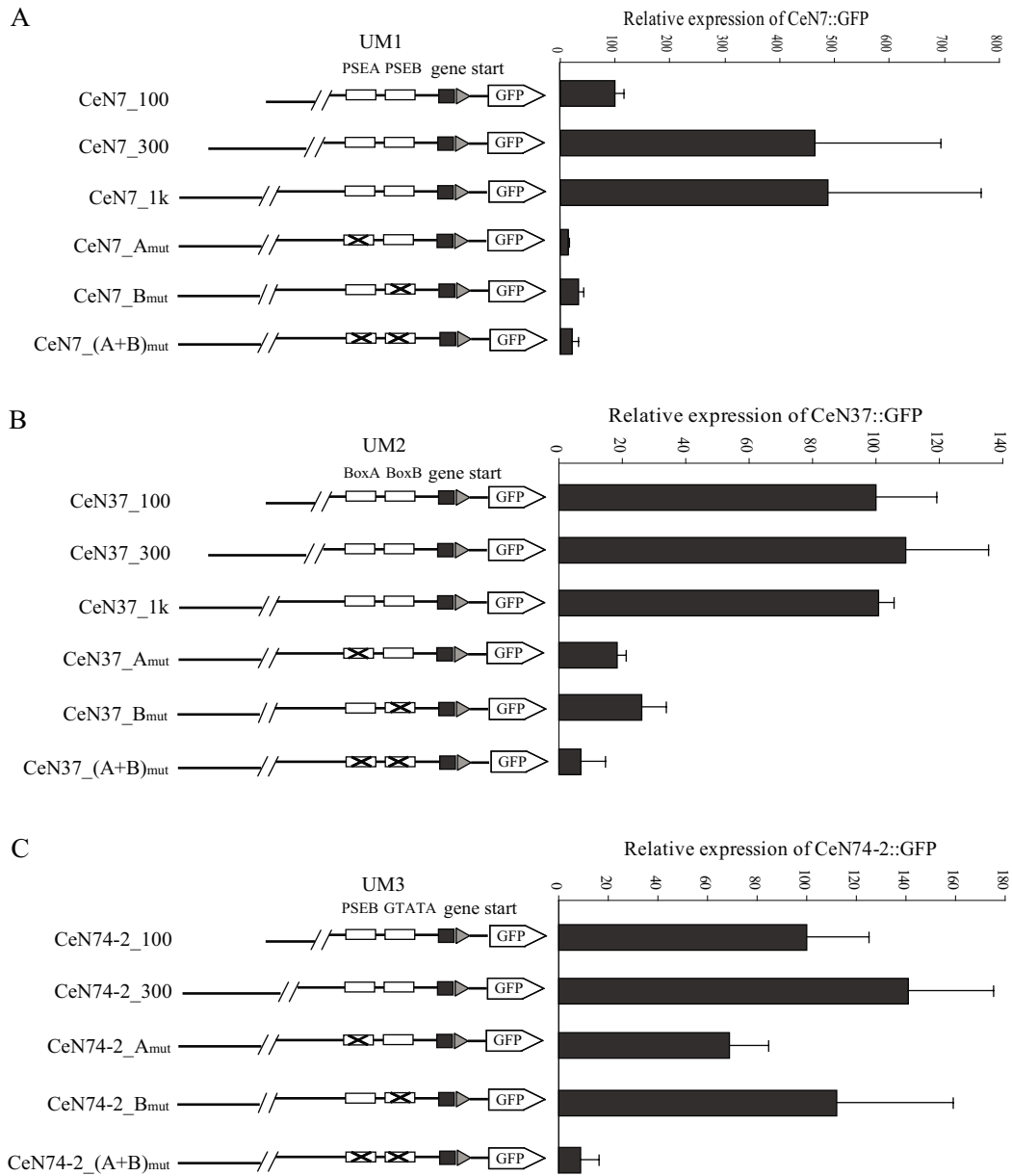
In the case of UM1, qRT-PCR of RNA from transgenic strains showed that the expression driven by CeN7\_300 (containing a 264 bp 5' flanking fragment) was 4–5 fold higher than expression driven by CeN7\_100 (Figure 3A), strongly suggesting the existence of an enhancer located within the 90~264 bp upstream region. However, a search for 5' end features in this region of PSE/UM1 ncRNAs in *C. elegans* failed to yield any common motifs (see Methods for details). The expression driven by CeN7\_1k showed an almost identical level to that of CeN7\_300, indicating that no additional regulatory elements exist within the 264~1000 bp upstream of the ncRNA locus. In the case of UM3, no significant increase in expression was found by increasing the length of the upstream fragment to 245 bp (CeN74-2\_300; Figure 3B). Similarly, qRT-PCR indicated no increase in expression of the UM2 locus when the upstream region was extended to 294 bp (CeN37\_300) and 989 bp (CeN37\_1k; Figure 3C). These observations suggest that for the two assumed RNA polymerase III-driving promoters (i.e. UM2 and UM3) most or all the promoter activity resides within 100 bp of the 5' end of the respective annotated loci.

#### **Mutational analysis of promoter sub-motifs**

Each of the three upstream motifs contains two sub-motifs whose base pairs are particularly invariant among loci. To investigate the contribution of the PSEA and PSEB sub-motifs to the overall PSE/UM1 transcriptional activity, we mutated each of the two sub-motifs mainly by converting each purine and pyrimidine residue to its opposite purine and pyrimidine, respectively (i.e. A to G, and *vice versa*), in the context of CeN7\_1k. The corresponding constructs were labeled as CeN7\_A<sub>mut</sub>, CeN7\_B<sub>mut</sub> and CeN7\_(A+B)<sub>mut</sub>. Mutation of either of the two sub-motifs (PSEA and PSEB) reduced expression to 3~7 % of that of CeN7\_1k, suggesting that both PSEA and PSEB are required for the transcription of PSE/UM1 loci (Figure 3A).

The sub-motifs in UM2 promoter strongly resemble the A and B box motifs of the tRNA internal promoter. Mutation of either of these two motifs in the context of CeN37\_1k caused a strong reduction in the expression levels of the mutant constructs (18 % for CeN37\_A<sub>mut</sub> and 26 % for CeN37\_B<sub>mut</sub> compared with CeN37\_1k; Figure 3B). Concomitant mutation of both sub-motifs (CeN37\_(A+B)<sub>mut</sub>) reduced the expression level to 7 % compared with CeN37\_1k.

Of the two sub-motifs in UM3, one strongly resembles the PSEB of the PSE/UM1, whereas the second sub-motif has the consensus sequence GTATA. We mutated both motifs in the same fashion as described above, and observed the effect on expression in the context of CeN74-2\_300. Abolishing the PSEB-like sub-motif (CeN74-2\_A<sub>mut</sub>) reduced the expression level to 49 % compared with that of non-mutated CeN74-2\_300. Mutation of the GTATA sub-motif produced an even more modest reduction (79 % compared to that of non-mutated CeN74-2\_300) than PSEB. Mutating both sub-motifs simultaneously, however,



**Figure 3**

**Analysis of upstream motif promoter activity.** The panels show the effects of additional upstream sequence length, and of mutation of sub-motifs, on the transcriptional activity of each of the three promoters PSE/UM1, UM2 and UM3, respectively. (A) UM1 from the SL2 RNA locus CeN7. "PSEA" and "PSEB" are the most invariant sequence elements in the PSE/UM1. (B) UM2 from the snoRNA locus CeN37. "BoxA" and "BoxB" show the two sub-motifs in UM2. (C) UM3 from the sbRNA locus CeN74-2. "PSEB" and "GTATA" are the most invariant sequence elements of the UM3. "Gene start" corresponds to the first 69, 32 and 56 bp of the CeN7, CeN74-2 and CeN37 mature transcripts, respectively. Relative expression levels are calculated by normalizing the qRT-PCR intensities of each CeNX\_100 construct to 100. Sub-motif mutations were performed on the constructs containing the longest upstream sequences (e.g. CeN7\_1k) mainly by converting each purine and pyrimidine residue to the opposite purine and pyrimidine, respectively (i.e. A to G, and vice versa).

resulted in near abolishment of expression of the reporter (CeN74-2\_(A+B)<sub>mut</sub>; Figure 3C). The results suggest a certain redundancy in regulatory activity between the two sub-motifs of UM3, but also that at least one of the sub-motifs must be present for expression of the downstream locus to occur at an appreciable rate.

#### **Transcription start sites of the upstream motif loci**

5' RACE performed on non-mutated constructs found that the TSS of CeN7::GFP and CeN74-2::GFP were identical to 5' ends of the endogenous RNAs. For CeN37::GFP, however, the TSS was apparently located 70 bp upstream of the 5' end of mature endogenous RNA and 9 bp upstream of UM2 sequence itself (Figure 4A).

#### **UM2 – a possible remnant of tRNA-snoRNA dicistronic loci**

As shown above, 5' RACE of the reporter transcript of a UM2 construct (CeN37\_1k) found that transcription was initiated 70 bp upstream of the 5' end of the mature endogenous RNA, and even 9 bp upstream of the UM2 sequence. In plants, dicistronic tRNA-snoRNA transcripts, which are subsequently cleaved to yield mature tRNA and snoRNA have been described [32]. To investigate whether the UM2 is also internally located with regards to the primary transcript of the endogenous RNAs, we visually inspected the tiling microarray data [1] for possible evidence of transcription of the UM2 site itself. Although we found no indication of expression of the UM2 sequence at the CeN37 locus, we found other 21 instances with indications of some level of expression, 16 at known snoRNA loci and five preceding unannotated TUFs. In addition to CeN37, we therefore selected five candidate loci (CeN50-2, CeN39, CeN53, CeN55 and CeN119) for analysis by 3' RACE as indicated in Figure 5. If the dicistronic tRNA-snoRNA model also applies to the UM2-snoRNA transcripts, two RACE bands should be expected, one corresponding to the UM2-snoRNA primary transcript, the other to the UM2 fragment remaining after cleavage. Four (CeN39, CeN53, CeN55 and CeN119) of the 6 loci yielded RACE fragments of length expected if transcription initiation occurred at the start of, or upstream of the UM2 sequence, and sequencing confirmed a joint UM2-snoRNA transcript in all four cases. A band corresponding to a smaller fragment was observed in three cases, but none of the sequences included the UM2 fragment. Thus, the part of the primary transcript that contains the UM2 sequence is either rapidly degraded after cleavage, or the UM2-containing fraction of the primary transcript is removed by 5' exonuclease digestion during maturation of the snoRNA. 5' RACE analysis showed that the transcription of the UM2 loci starts 9~13 bp upstream of the first base in the box A-like sub-motif.

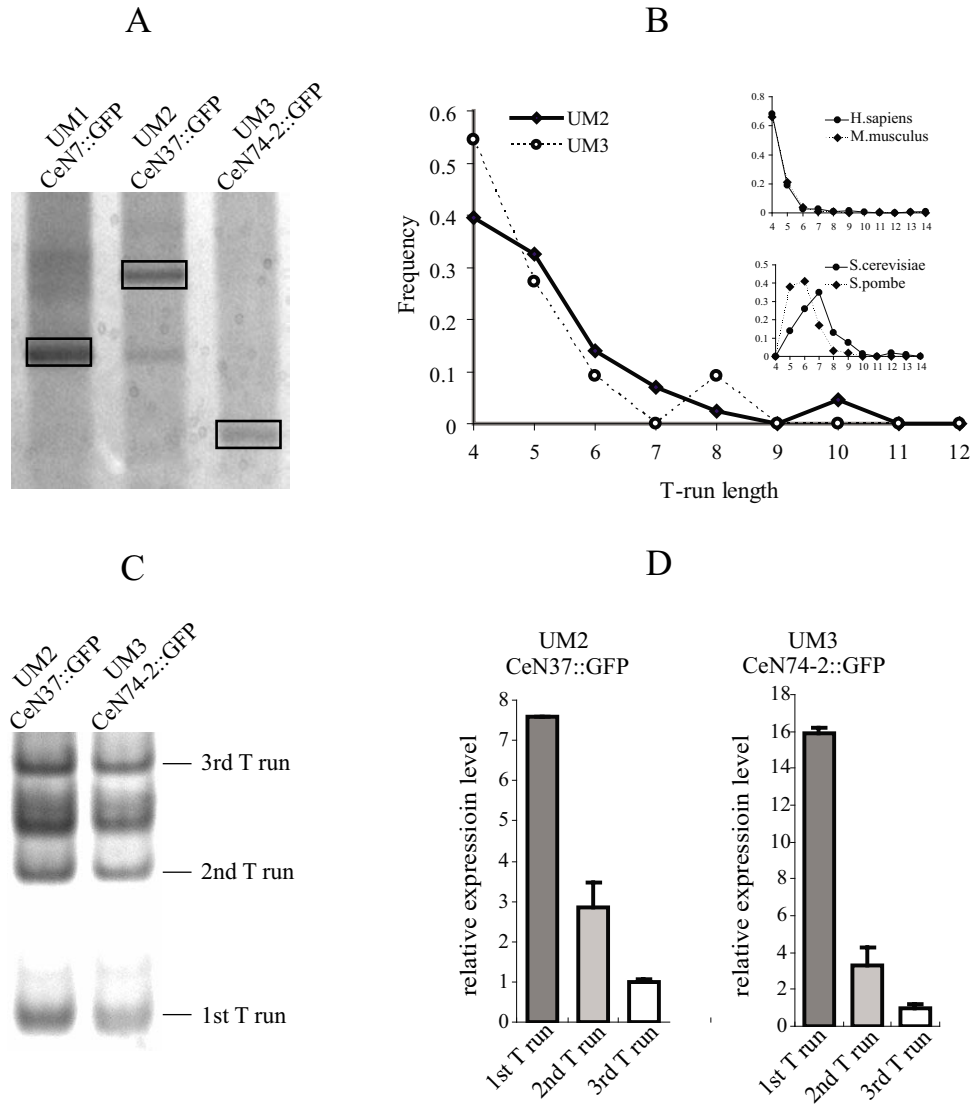
#### **Transcriptional termination**

Transcripts derived from UM2 and UM3 loci are usually terminated at a run of several consecutive thymidine (T) residues, which is a property of loci transcribed by RNA polymerase III. The minimal number of consecutive T residues sufficient for termination of RNA polymerase III transcription varies among organisms, but analysis of known UM2 and UM3 loci suggests 4 consecutive Ts are sufficient for termination in *C. elegans* (Figure 4B). This is similar to what has been found in human and mouse, but is less than the 5–7 normally needed for termination in the yeast species (Figure 4B) [14]. In the plasmid pPD95\_77 there are runs of four (or more) Ts located at variable distances from the plasmid multicloning site (MCS). To determine the actual termination sites of the UM2 and UM3 constructs, we performed 3'RACE on the resulting reporter transcripts. For CeN37\_1k (UM2) the reporter transcript terminated at the first, second and third "TTTT" tracts (located 122, 177 and 236 nt downstream of TSS; Figure 4C). For the CeN74-2\_300 (UM3) construct, reporter transcript termination was identical to what has been observed in CeN37\_1k (Figure 4C). qRT-PCR of fragments corresponding to the different termination sites suggested that most of the UM2 and UM3 reporter transcripts terminated at the first "TTTT" tract (Figure 4D).

#### **PSE/UM1 drive expression of GFP**

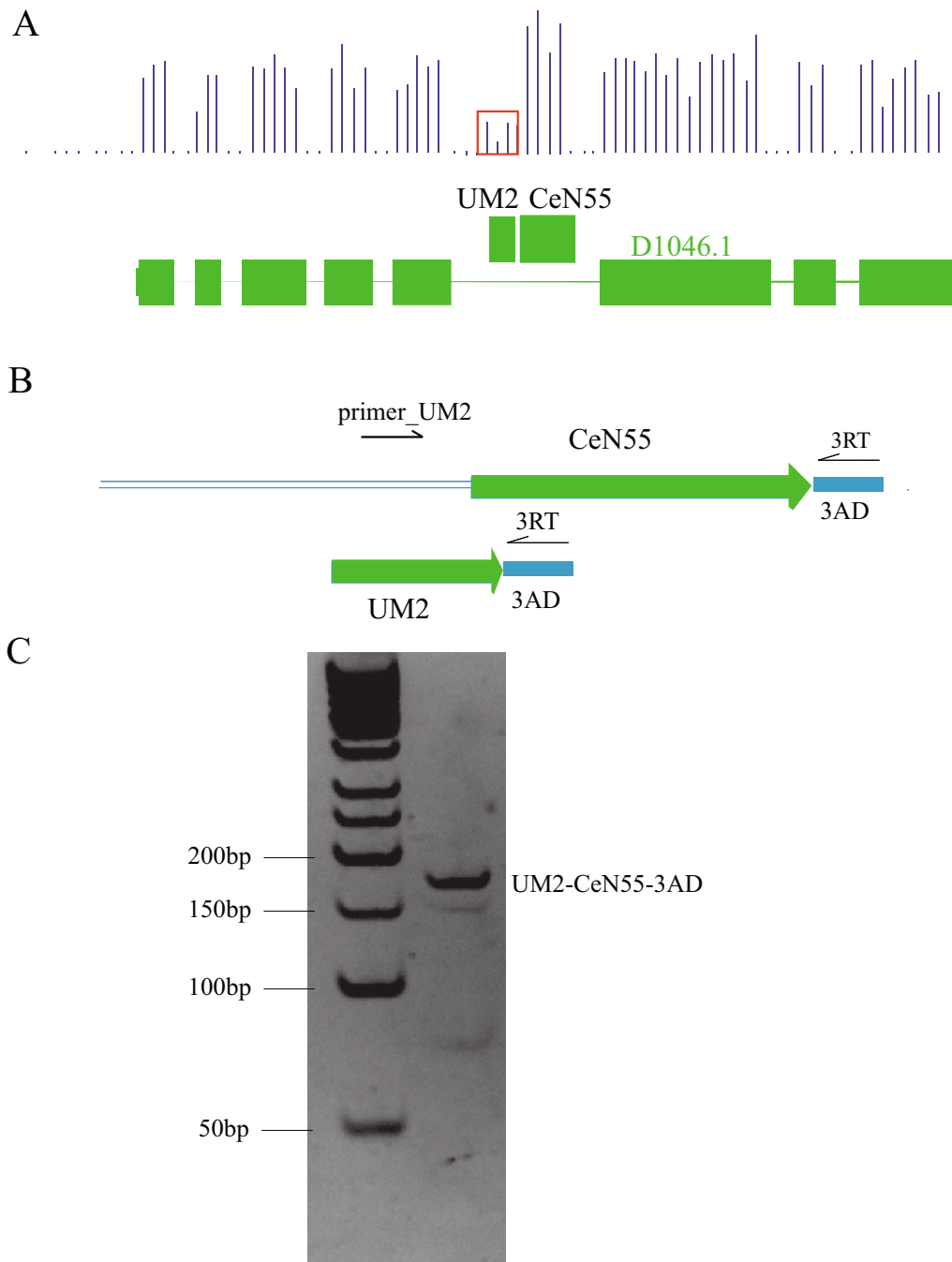
3'RACE of the UM1 reporter transcript gave no specific result, but RT-PCR indicated that at least a fraction of the GFP ORF was included in the transcript. We further investigated the expression of GFP in worms from 2–3 independent transgenic strains from each of the different constructs. No GFP expression was observed after genetic transformation with any of the UM2 (CeN37) and UM3 (CeN74-2) constructs. However, marked GFP expression was observed under fluorescent microscope for the CeN7\_1k and CeN7\_300 transgenic worms, and even the CeN7\_100 transgenic worms showed weak GFP expression. To eliminate the possibility that the observed GFP expression was an effect of this particular PSE/UM1 locus, we tested another PSE/UM1 ncRNA locus, CeN16-1, whose upstream sequence was also able to drive GFP expression. 5'-end RACE performed on RNA extracted from CeN7\_1k, CeN7\_300 and CeN16-1\_1k transgene worms showed that the transcription start site of the reporters were identical to that of the wild type ncRNA loci, suggesting that the transcriptional activity resulting in the observed GFP expression was the same as that driving the transcription of the endogenous ncRNAs.

Since PSE/UM1 could drive expression of GFP, we used UM1::GFP fusions to analyse the temporal-spatial expression pattern of related ncRNA loci having this upstream motif. One group of such loci are the *C. elegans* SL2 RNAs, which are a nematode-specific group of ncRNAs that func-



**Figure 4**

**UMI-3 transcript start and terminator analysis.** (A) 5'RACE analysis of the UMI-3 reporter gene transcripts (CeN7::GFP, CeN37::GFP and CeN74-2::GFP). Boxed bands correspond to the 5'RACE products of each reporter. The transcription start sites indicated by the 5'RACE bands of reporter constructs CeN7::GFP (UM1) and CeN74-2::GFP (UM3) are identical to the 5'ends of the mature endogenous CeN7 and CeN74-2 transcripts, respectively. The 5'RACE band from the CeN37::GFP (UM2) reporter constructed corresponds to a transcript starting 70 bp upstream of the 5'end of the mature endogenous CeN37 transcript (The lower band in the same lane is an unspecific band). (B) Distribution of *C. elegans* UM2 and UM3 ncRNAs terminator lengths. Frequency distribution analysis of poly(dT) terminator length was conducted on UM2 and UM3 ncRNA loci possessing a single RNA polymerase III termination signal starting within the first 40 bp downstream of the genomic sequence corresponding to the mature ncRNA. Inserted figures show data based on the analysis in ref [14]. (C) 3' RACE of CeN74-2\_300 (UM3) and CeN37\_1k (UM2). Bands corresponding to transcripts terminated at the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> "TTTT" run are indicated. (Sequence analysis suggests that the band between the second and third T run corresponds to transcript terminated at a T rich tract "TTCTTGTT"). (D) Relative expression level of transcripts terminated at the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> T run. qRT-PCR was performed by reverse transcription using the primer located in front of the 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> T run, respectively, followed by a nested PCR.(see Methods for details).



**Figure 5**

**The UM2-snoRNA transcript structure.** (A) The H/ACA snoRNA locus CeN55 is located in an intron of the protein coding gene D1046.1. Tiling array probe signal intensities indicate that the UM2 sequence upstream of snoRNA CeN55 might also be expressed (red box). (B) Model of RACE amplification of the putative UM2-snoRNA dicistronic fragments. 3AD is a 3' end adaptor ligated to the ncRNAs; 3RT is the primer complementary to 3AD, primer\_UM2 located in the region of UM2. The RACE analysis was performed by reverse transcription using the 3RT as primer followed by PCR with primers primer\_UM2 and 3RT. (C) Gel analysis of the RACE fragments.



tion in trans-splicing of operonic mRNAs. There are around 20 SL2 RNA loci in *C. elegans* with slightly variable sequence characteristics. As far as is known, all participate in the same function, i.e. joining of an additional "exon" to the 5' end of internal (or non-first) mRNAs in operonic loci, but nothing appears to be known about the background for the numerous and variable SL2 RNA loci in the *C. elegans* genome. Previous experiments have demonstrated that different SL2 RNA genes show different temporal expression patterns [6,34,35,37], but little is known about their spatial expression patterns. We therefore examined the temporal-spatial expression of five SL2 RNA genes using promoter::GFP fusion constructs. The tested ncRNAs showed considerable variation in expression both in time and space. CeN7 is principally expressed in hypodermal cells, which is in agreement with previous *in situ* hybridization result [35], but also showed expression in skin muscles, excretory cells, head and ventral neuron cells (Figure 6A & 6B; Additional file 2). The CeN16-1 locus was also active in excretory cells, and showed strong expression in the pharynx (Figure 6C; Additional file 2). Two other loci, CeN6 and CeN11, showed marked expression in intestinal muscles near the tail (Figure 6D & 6F; Additional file 2), however, their temporal expression pattern differed, CeN11 showing visible expression from larval stage 3, while CeN6 showed no expression until the mature adult stage. The fifth SL2 RNA locus, CeN19, is expressed in intestinal muscles near the vulva (Figure 6E; Additional file 2). The temporal expression pattern is generally identical with previously northern and microarray analysis [6,34,35,37]. CeN7, CeN16-1 and CeN19 showed strong expression from an early stage of embryo development (Additional file 2) through the mature adult stage, whereas CeN6 and CeN11 both showed stage specific expression.

An interesting question is whether the transcriptional activity of the various SL2 RNA loci correlates with their target loci (i.e., the mRNAs to which the SL2 exons are spliced). To this end we downloaded spliced leader [38] and expression data [19,39]. However, mRNA spliced to a specific SL2 RNA do not show any well correlated spatial expression patterns (Additional file 3), and even though it is possible to find SL2 RNA loci whose expression resembles one or a few of SL2 RNA [34] and mRNA temporal expression data [20,40,41] did not show any significant correlation between specific SL2 RNAs and their target mRNAs (Additional file 3).

We also observed that mutations in the PSEA and PSEB sub-motifs greatly changed the expression patterns of the CeN7 locus constructs (i.e. CeN7<sub>A<sub>mut</sub></sub>, CeN7<sub>B<sub>mut</sub></sub> and CeN7<sub>(A+B)<sub>mut</sub></sub>) compared with that of non-mutated constructs (i.e. CeN7<sub>1k</sub> and CeN7<sub>300</sub>). The CeN7<sub>1k</sub> and CeN7<sub>300</sub> constructs were clearly expressed in hypo-

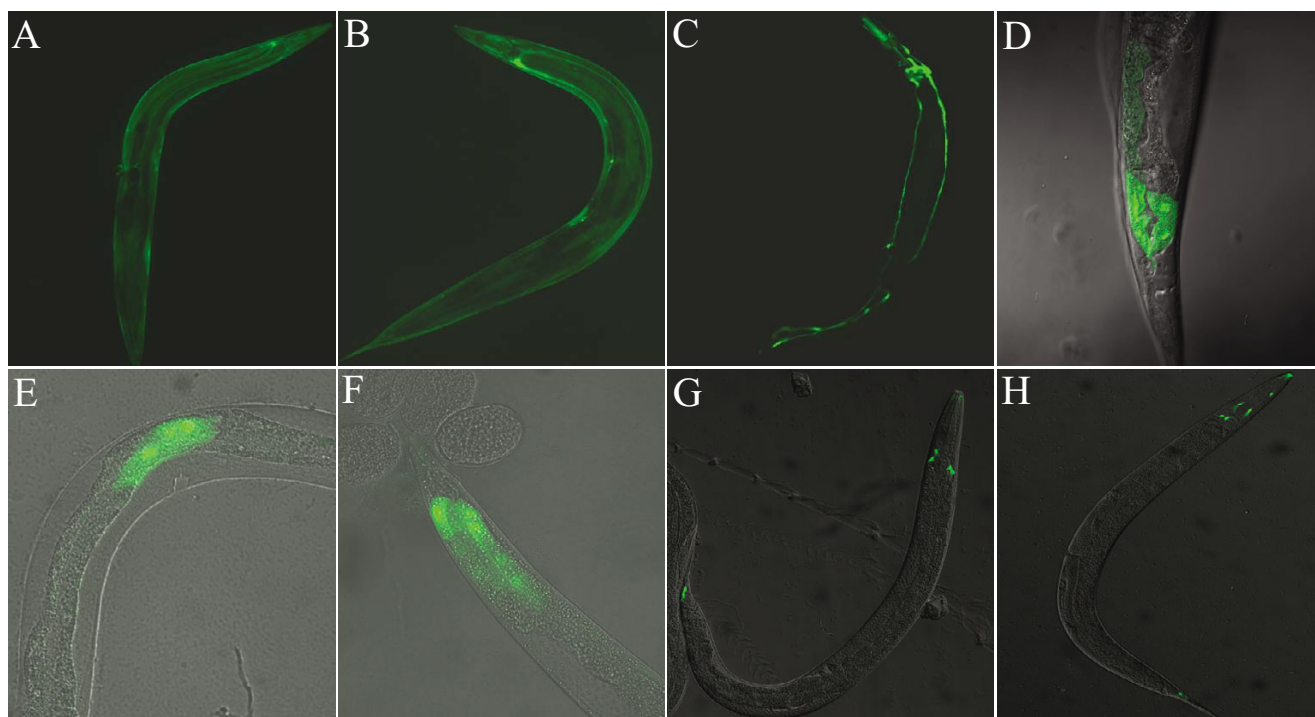
dermal cells, skin muscles, excretory cells, head and ventral neurons (Figure 6A & 6B; Additional file 2), however, the expression driven by CeN7 mutations was restricted to amphid and tail neuron cells (Figure 6G & 6H). This suggests that the sequence characteristics of PSEA and PSEB are not only important for the general expression level of a locus, but may also influence where and when a locus is expressed.

## Discussion

Non-protein-coding RNAs are gaining in importance as functional elements in eukaryote cellular and organismal development [42-46]. *C. elegans* is one of the most important biological model systems for genetic and developmental studies, and it has recently been demonstrated that around 50 % of the transcriptional output in this organism cannot be identified as arising from protein-coding genes [34]. Among this vast amount of transcripts there are approximately 1400 relatively well-defined short non-coding transcripts. A notable fraction of these have strongly invariant upstream sequence motifs and this study has demonstrated that three of these motifs are able to activate *in vivo* transcription of otherwise inactive reporter genes.

The PSE/UM1 sequence is found at about 10 % of the 1400 known or putative noncoding loci, and is the most common promoter structure of *C. elegans* noncoding RNAs. Most of the PSE/UM1 ncRNAs are highly expressed, indicating this motif has relatively strong promoter activity. However, the expression levels of the PSE/UM1 loci vary greatly, even within the same functional class. For example, within the SL2 RNAs the expression levels of different loci can differ more than 10 fold [34]. Our analysis indicated that sequence elements within -90 to -264 bp relative to the transcription start site are also important for expression from the UM1-type promoter. At human snRNA loci a distal sequence element (DSE) is commonly found around 200 bp upstream of transcription start site [47,48], but a search in the region upstream of the PSE/UM1 in *C. elegans* failed to identify any common sequence motif (see Methods for details).

The PSE/UM1 promoter was also shown to drive expression of the protein (GFP) encoded in the reporter construct, suggesting that this motif activates RNA polymerase II expression. This agrees with previous evidence that the *C. elegans* PSE can drive expression of a *lacZ* mRNA [35], but is the first demonstration in *C. elegans* of an active protein expressed under an ncRNA promoter. The transcriptional activity of two other upstream motifs (UM2 and UM3) was clearly demonstrated by RT-PCR, but no GFP expression was observed from these two promoters. The UM2 motif clearly resembles the tRNA promoter which is known to activate RNA polymerase III

**Figure 6**

**In vivo expression patterns of different SL2 loci.** Expression of GFP under the UM1 promoters from the loci of (A) CeN7::GFP (CeN7\_300), (B) CeN7::GFP (CeN7\_1k), (C) CeN16-I::GFP (CeN16-I\_1k), (D) CeN6::GFP (CeN6\_1k), (E) CeN19::GFP (CeN19\_1k), (F) CeN11::GFP (CeN11\_1k) are shown. (G) GFP expression of CeN7\_A<sub>mut</sub>, (H) GFP expression of CeN7\_B<sub>mut</sub>.

transcription through binding to the transcription factor IIC [14]. Although polymerase specificity was not interrogated in this study, accumulated evidence [6] suggests that both UM2 and UM3 activate RNA polymerase III transcription. This is further corroborated by the finding that both endogenous loci and reporter constructs activated by these two promoters terminate transcription at runs of four (or more) consecutive T residues.

The observation that the PSE/UM1 promoter is able to drive GFP expression could allow for detailed analysis of the spatial-temporal expression of PSE/UM1 loci. Much genome wide data on the temporal expression of coding and noncoding genes in *C. elegans* have been obtained through Serial Analysis of Gene Expression (SAGE) and microarrays [3,6,20,34,40,41,49-51], and large scale expression profiling aimed at the spatial expression pattern of protein coding genes have been performed by several groups [19,22,23,52-54]. However, there are almost no data available on the spatial expression patterns of ncRNAs. As demonstrated in this study, a GFP expression under PSE/UM1 promoters from SL2 RNA agreed well with reported *in situ* hybridization data [35], and was able to specify the detailed expression characteristics of several

SL2 RNA loci, in some cases down to the cellular level. Determination of the spatial and temporal expression patterns of ncRNAs can be the key to their function, and this assay could be a very convenient tool for *in vivo* analysis of the expression pattern of the TATA-less UM1 loci. An additional aspect is that there appear to be few reports on embryo-specific promoters in *C. elegans*, and the finding that some of the PSE/UM1 promoters are active at this stage may be of practical use to other research within the field.

As a promoter the UM2 sequence represents a particularly interesting case. The sub-motifs of the UM2 resemble the tRNA internal promoter elements box A and B, and 5'RACE analysis showed that the UM2 sequence was embedded in the primary reporter transcript. Detailed re-analysis of the *C. elegans* tiling microarray data [34] indicated the existence of similar primary transcripts arising from several genomic UM2 loci, and subsequent 3'RACE verified this. The fact that a UM2-containing primary transcript could not be identified at all inspected loci (including the endogenous CeN37 locus used for the reporter construct) could owe to the primary transcripts being inherently unstable and rapidly degraded during snoRNA

maturation [33]. The *C. elegans* UM2 primary transcripts resemble the dicistronic tRNA-snoRNA transcripts found at a few plant and yeast loci [29,32,55,56]. Recently, several *Drosophila* snoRNAs were found to derive from longer RNA polymerase III transcripts, some of which were shown to contain an element similar to the B box of the tRNA promoter [57], and closer inspection of the same sequences in *Drosophila* also suggested the presence of an A box-like element upstream of the B box-like sequence at several loci. In the yeast genome there has also been reported one snoRNA whose RNA polymerase III transcription is driven by an A+B box configuration [13,55]. *In vitro* experiments in yeast demonstrated that box A alone can direct efficient TFIIC-dependent transcription, while box B is dispensable [13], however, *in vivo* experiments found that both box A and B are required for the downstream transcripts accumulation [13]. The *in vivo* mutational analysis reported here also suggests a requirement for both box A and B for efficient transcription of UM2 loci in *C. elegans*. Given that similar snoRNA promoter characteristics are found in animals, plants and fungi could point to a very old promoter strategy that utilised a tRNA-like promoter to drive expression of snoRNA genes.

Although the promoter activities of ncRNAs, in particular those of snRNAs, have been analysed in great detail in a variety of organisms such as human, *Drosophila* and yeast for a couple of decades, little work has been done in *C. elegans*. Our work demonstrated that all the three investigated upstream motifs in *C. elegans* are transcriptionally active. However, this work has only concentrated on promoters with distinctive sequence characteristics, and the great majority of intergenic ncRNA loci show no obvious upstream motifs. What sequence elements are important and which protein complexes are recruited to initiate the transcription of such loci is still not known, and further efforts are needed for better understanding of the *C. elegans* ncRNA transcriptional mechanism.

## Conclusion

We demonstrate here the transcriptional activities of three putative ncRNA promoters. Mutational analysis found that the most invariant sub-motifs of the UM1 and UM2 sequences are required for the downstream genes transcription, while the two sub-motifs of UM3 show redundancy with respect to transcriptional activity. We also show that UM1 can drive expression of GFP, suggesting that this promoter drive RNA polymerase II transcription, and the UM1::GFP fusions have shown to be useful in determine the temporal-spatial expression patterns of UM1 ncRNAs. UM2 and UM3 can not drive the expression of GFP, and termination at "TTTT" tract strongly suggests RNA polymerase III transcription. Several cases of tRNA-snoRNA dicistronic transcription pattern have been

found, and it is likely that most of the UM2 snoRNAs apply this model of transcription.

## Methods

### Worm strains

Standard nematode cultivating conditions were used as described in ref [58]. The strains used were N2 and UNC-76.

### Construction of plasmids

Individual PCR reactions were performed in 50 or 100 ul reaction volume using wild type *C. elegans* (N2 strain) genomic DNA as template. PCR products corresponding to the desired fragments were digested with enzyme HindIII and BamHI, and then cloned into the promoter-less vector pPD95\_77 (kindly provided by Andrew Fire) upstream of the GFP reporter gene. The constructs were verified by sequencing. Mutations were performed using standard PCR procedures. The primers used in this work are listed in Additional file 4.

### Transgenic *C. elegans* lines

Reporter constructs were injected at 50 ng/ul together with 50 ng/ul transformation marker pRF4 [rol-6(su1006)] or unc-76. Stable lines of transgenic worms were established as described previously [59]. For each construct, 2~5 transgenic lines were analysed.

### RT-PCR/qRT-PCR

RNA was extracted from transgenic worms according to Trizol (Invitrogen) protocol. RNA digested with DNase I (Fermentas) was used as template for RT-PCR (Qiagen one step RT-PCR kit). For qRT-PCR, reverse transcription was performed using primers 95\_77\_2\_R and U2\_R, and the cDNA was used as template for the qPCR according to qPCR mix protocol (Qiagen QuantiTect SYBR Green PCR Kit). When the expression levels of transcripts of different length were to be compared, a nested approach using internal primer yielding identical length PCR products were used for the qPCR step (following reverse transcription). The reactions were carried out on an MJ Research Opticon TM 2.

### 3'-RACE

3'-RACE was performed as described [1] with minor modifications. Briefly, DNase I digested total RNA was ligated to the 3' end adaptor 3AD, and the ligated RNA was reverse transcribed into complementary DNA (cDNA) with a primer (3RT) complementary to the 3' adaptor 3AD. First round PCRs were performed by using primers 3RT and pPD95\_77\_1\_F. Nested PCR was then performed, using 3RT and pPD95\_77\_2\_F as primers, and the first round PCR products as template. The PCR products were analysed on a PAGE gel, and candidate bands were recovered and sequenced.

## 5'-RACE

DNase I digested total RNA was reverse transcribed into cDNA with primers pPD95\_77\_2\_R, pPD95\_77\_3\_R and pPD95\_77\_4\_R, respectively. The cDNA was treated with terminal DNA transferase (Fermentas) to add a 3'end poly(A) tail. First round PCR was performed with 3'CDS primer and the corresponding primer used for reverse transcription. Second round PCR was then carried out using the diluted first round PCR products as template, and 3'CDS and pPD95\_77\_1\_R as primers. The second round PCR products were analysed, cloned and sequenced as above.

## Computational analysis

*C. elegans* genome annotation and sequence data were downloaded from Wormbase (version WS140) [39]. The MEME motif discovery tool (version 3.0.13) [25] was used to search for conserved motifs within and upstream of the ncRNA loci. The search for common motifs acting as enhancer on 54 UM1/PSE ncRNAs was carried out on 100~300 bp upstream sequences of these loci.

## Authors' contributions

TL, HH, YW and HZ performed the experiments. TL and HH interpreted the results. TL, HH and GS drafted the manuscript. RC and GS directed the design of the study. All authors have read and approved the final manuscript.

## Additional material

### Additional file 1

*Genomic location. The data provided describes the genomic locations of all ncRNA loci used in this work.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-71-S1.pdf>]

### Additional file 2

*In vivo expression patterns. The data provided shows the expression patterns of multiple lines of each SL2 RNAs.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-71-S2.pdf>]

### Additional file 3

*trans-spliced genes and related SL2 RNAs. The data provided shows the expression patterns of trans-spliced genes and related SL2 RNAs.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-71-S3.xls>]

### Additional file 4

*Primers. The data provided shows the primers used in this work.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-71-S4.pdf>]

## Acknowledgements

The *C. elegans* strains used in this work were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources. This work was supported by the National Sciences Foundation of China Grant No. 30630040; National Key Basic Research & Development Program 973 under Grant Nos. 2002CB713805 and 2003CB715907.

## References

1. He H, Wang J, Liu T, Liu XS, Li T, Wang Y, Qian Z, Zheng H, Zhu X, Wu T, Shi B, Deng W, Zhou W, Skogerbo G, Chen R: **Mapping the *C. elegans* noncoding transcriptome with a whole-genome tiling microarray.** *Genome Res* 2007, **17(10)**:1471-1477.
2. Zemann A, op de Bekke A, Kiefmann M, Brosius J, Schmitz J: **Evolution of small nucleolar RNAs in nematodes.** *Nucl Acids Res* 2006, **34(9)**:2676-2685.
3. Lall S, Grun D, Krek A, Chen K, Wang YL, Dewey CN, Sood P, Colombo T, Bray N, Macmenamin P, Kao HL, Gonsalus KC, Pachter L, Piano F, Rajewsky N: **A genome-wide map of conserved microRNA targets in *C. elegans*.** *Curr Bio* 2006, **16**:460-471.
4. Li L, Wang X, Stolc V, Li X, Zhang D, Su N, Tongprasit W, Li S, Cheng Z, Wang J, Deng XW: **Genome-wide transcription analyses in rice using tiling microarrays.** *Nature genetics* 2006, **38(1)**:124-129.
5. David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM: **A high-resolution map of transcription in the yeast genome.** *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103(14)**:5320-5325.
6. Deng W, Zhu X, Skogerbo G, Zhao Y, Fu Z, Wang Y, He H, Cai L, Sun H, Liu C, Li B, Bai B, Wang J, Jia D, Sun S, He H, Cui Y, Wang Y, Bu D, Chen R: **Organization of the *Caenorhabditis elegans* small non-coding transcriptome: Genomic features, biogenesis, and expression.** *Genome Res* 2006, **16(1)**:20-29.
7. Stolc V, Samanta MP, Tongprasit W, Sethi H, Liang S, Nelson DC, Hegeman A, Nelson C, Rancour D, Bednarek S, Ulrich EL, Zhao Q, Wrobel RL, Newman CS, Fox BG, Phillips GN Jr, Markley JL, Sussman MR: **Identification of transcribed sequences in *Arabidopsis thaliana* by using high-resolution genome tiling arrays.** *PNAS* 2005, **102(12)**:4453-4458.
8. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, Sementchenko V, Piccolboni A, Bekiranov S, Bailey DK, Ganesh M, Ghosh S, Bell I, Gerhard DS, Gingeras TR: **Transcriptional Maps of 10 Human Chromosomes at 5-Nucleotide Resolution.** *Science* 2005, **308(5725)**:1149-1154.
9. Stolc V, Gauthar Z, Mason C, Halasz G, van Batenburg MF, Rifkin SA, Hua S, Herremans T, Tongprasit W, Barbano PE, Bussemaker HJ, White KP: **A Gene Expression Map for the Euchromatic Genome of *Drosophila melanogaster*.** *Science* 2004, **306(5696)**:655-660.
10. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M: **Global Identification of Human Transcribed Sequences with Genome Tiling Arrays.** *Science* 2004, **306**:2242-6.
11. Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, Hartman S, Harrison PM, Nelson FK, Miller P, Gerstein M, Weissman S, Snyder M: **The transcriptional activity of human Chromosome 22.** *Genes & Development* 2003, **17(4)**:529-540.
12. Huttenhofer A, Brosius J, Bachelier JP: **RNomics: identification and function of small, non-messenger RNAs.** *Current Opinion in Chemical Biology* 2002, **6(6)**:835-843.
13. Guffanti E, Ferrari R, Preti M, Forloni M, Harismendy O, Lefebvre O, Dieci G: **A minimal promoter for TFIIC-dependent in vitro transcription of snoRNA and tRNA genes by RNA polymerase III.** *J Biol Chem* 2006, **281(33)**:23945-23957.
14. Braglia P, Percudani R, Dieci G: **Sequence context effects on oligo(dT) termination signal recognition by *Saccharomyces cerevisiae* RNA polymerase III.** *J Biol Chem* 2005, **280(20)**:19551-19562.
15. Li C, Harding GA, Parise J, McNamara-Schroeder KJ, Stumph WE: **Architectural arrangement of cloned proximal sequence element-binding protein subunits on *Drosophila* U1 and U6 snRNA gene promoters.** *Mol Cell Biol* 2004, **24(5)**:1897-1906.

16. Wang Y, Stumph WE: **Identification and topological arrangement of Drosophila proximal sequence element (PSE)-binding protein subunits that contact the PSEs of U1 and U6 small nuclear RNA genes.** *Mol Cell Biol* 1998, **18(3)**:1570-1579.
17. Martin MP, Gerlach VL, Brow DA: **A novel upstream RNA polymerase III promoter element becomes essential when the chromatin structure of the yeast U6 RNA gene is altered.** *Mol Cell Biol* 2001, **21(19)**:6429-6439.
18. Saxena A, Ma B, Schramm L, Hernandez N: **Structure-function analysis of the human TFIIIB-related factor II protein reveals an essential role for the C-terminal domain in RNA polymerase III transcription.** *Mol Cell Biol* 2005, **25(21)**:9406-9418.
19. Hunt-Newbury R, Viveiros R, Johnsen R, Mah A, Anastas D, Fang L, Halfnight E, Lee D, Lin J, Lorch A, McKay S, Okada HM, Pan J, Schulz AK, Tu D, Wong K, Zhao Z, Alexeyenko A, Burglin T, Sonnhammer E, Schnabel R, Jones SJ, Marra MA, Baillie DL, Moerman DG: **High-throughput in vivo analysis of gene expression in Caenorhabditis elegans.** *PLoS Biol* 2007, **5(9)**:e237.
20. GuhaThakurta D, Palomar L, Stormo GD, Tedesco P, Johnson TE, Walker DW, Lithgow G, Kim S, Link CD: **Identification of a novel cis-regulatory element involved in the heat shock response in Caenorhabditis elegans using microarray gene expression and computational methods.** *Genome Res* 2002, **12(5)**:701-712.
21. Huang P, Pleasance ED, Maydan JS, Hunt-Newbury R, O'Neil NJ, Mah A, Baillie DL, Marra MA, Moerman DG, Jones SJ: **Identification and analysis of internal promoters in Caenorhabditis elegans operons.** *Genome Res* 2007, **17(10)**:1478-1485.
22. Dupuy D, Li QR, Deplanche B, Boxem M, Hao T, Lamesch P, Sequerra R, Bosak S, Doucette-Stamm L, Hope IA, Hill DE, Walhout AJ, Vidal M: **A first version of the Caenorhabditis elegans Promoterome.** *Genome Res* 2004, **14(10B)**:2169-2175.
23. Dupuy D, Bertin N, Hidalgo CA, Venkatesan K, Tu D, Lee D, Rosenberg J, Svrzikapa N, Blanc A, Carnec A, Carvunis AR, Pulak R, Shingles J, Reece-Hoyes J, Hunt-Newbury R, Viveiros R, Mohler WA, Tasan M, Roth FP, Le Peuch C, Hope IA, Johnsen R, Moerman DG, Barabasi AL, Baillie D, Vidal M: **Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans.** *Nat Biotechnol* 2007, **25(6)**:663-668.
24. Raharjo I, Gaudet J: **Gland-specific expression of C. elegans hhl-6 requires the combinatorial action of three distinct promoter elements.** *Biol Cell* 2007, **302(1)**:295-308.
25. Bailey TL, Elkan C: **The value of prior knowledge in discovering motifs with MEME.** *Proc Int Conf Intell Syst Mol Biol* 1995, **3**:21-29.
26. Thomas J, Lea K, Zucker-Aprison E, Blumenthal T: **The spliceosomal snRNAs of Caenorhabditis elegans.** *Nucl Acids Res* 1990, **18(9)**:2633-2642.
27. Hernandez N: **Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription.** *J Biol Chem* 2001, **276(29)**:26733-26736.
28. Li L, Linning RM, Kondo K, Honda BM: **Differential expression of individual suppressor tRNA(Trp) gene family members in vitro and in vivo in the nematode Caenorhabditis elegans.** *Mol Cell Biol* 1998, **18(2)**:703-709.
29. Moqtaderi Z, Struhl K: **Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes.** *Mol Cell Biol* 2004, **24(10)**:4118-4127.
30. Schattner P, Decatur WA, Davis CA, Ares M Jr, Fournier MJ, Lowe TM: **Genome-wide searching for pseudouridylation guide snoRNAs: analysis of the Saccharomyces cerevisiae genome.** *Nucleic Acids Res* 2004, **32(14)**:4281-4296.
31. Martens JA, Laprade L, Winston F: **Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene.** *Nature* 2004, **429(6991)**:571-574.
32. Kruszka K, Barneche F, Guyot R, Ailhas J, Meneau I, Schiffer S, Marchfelder A, Echeverria M: **Plant dicistronic tRNA-snoRNA genes: a new mode of expression of the small nucleolar RNAs processed by RNase Z.** *EMBO J* 2003, **22(3)**:621-632.
33. Preti M, Guffanti E, Valitutto E, Dieci G: **Assembly into snoRNP controls 5'-end maturation of a box C/D snoRNA in Saccharomyces cerevisiae.** *Biochem Biophys Res Commun* 2006, **351(2)**:468-473.
34. He H, Cai L, Skogerboe G, Deng W, Liu T, Zhu X, Wang Y, Jia D, Zhang Z, Tao Y, Zeng H, Aftab MN, Cui Y, Liu G, Chen R: **Profiling Caenorhabditis elegans non-coding RNA expression with a combined microarray.** *Nucl Acids Res* 2006, **34(10)**:2976-2983.
35. Ross LH, Freedman JH, Rubin CS: **Structure and expression of novel spliced leader RNA genes in Caenorhabditis elegans.** *J Biol Chem* 1995, **270(37)**:22066-22075.
36. Evans D, Blumenthal T: **trans Splicing of Polycistronic Caenorhabditis elegans Pre-mRNAs: Analysis of the SL2 RNA.** *Mol Cell Biol* 2000, **20(18)**:6659-6667.
37. MacMorris M, Kumar M, Lasda E, Larsen A, Kraemer B, Blumenthal T: **A novel family of C. elegans snRNPs contains proteins associated with trans-splicing.** *Rna* 2007, **13(4)**:511-520.
38. Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, Kim SK: **A global analysis of Caenorhabditis elegans operons.** *Nature* 2002, **417(6891)**:851-854.
39. Harris TW, Chen N, Cunningham F, Tello-Ruiz M, Antoshechkin I, Bastiani C, Bieri T, Blasiar D, Bradnam K, Chan J, Chen CK, Chen WJ, Davis P, Kenny E, Kishore R, Lawson D, Lee R, Muller HM, Nakamura C, Ozersky P, Petcherski A, Rogers A, Sabo A, Schwarz EM, Van Auken K, Wang Q, Durbin R, Spieth J, Sternberg PW, Stein LD: **WormBase: a multi-species resource for nematode biology and genomics.** *Nucleic Acids Res* 2004:D411-417.
40. Wang J, Kim SK: **Global analysis of dauer gene expression in Caenorhabditis elegans.** *Development* 2003, **130(8)**:1621-1634.
41. Jiang M, Ryu J, Kiraly M, Duke K, Reinke V, Kim SK: **Genome-wide analysis of developmental and sex-regulated gene expression profiles in Caenorhabditis elegans.** *PNAS* 2001, **98(1)**:218-223.
42. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Bruggmann SA, Good-nough LH, Helms JA, Farnham PJ, Segal E, Chang HY: **Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs.** *Cell* 2007, **129(7)**:1311-1323.
43. Petruk S, Sedkov Y, Riley KM, Hodgson J, Schweisguth F, Hirose S, Jaynes JB, Brock HW, Mazo A: **Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference.** *Cell* 2006, **127(6)**:1209-1221.
44. Umlauf D, Fraser P: **The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme.** *Biol Chem* 2008.
45. Mercer TR, Dinger ME, Sunken SM, Mehler MF, Mattick JS: **Specific expression of long noncoding RNAs in the mouse brain.** *Proc Natl Acad Sci USA* 2008, **105(2)**:716-721.
46. Hongay CF, Grisafi PL, Galitski T, Fink GR: **Antisense transcription controls cell fate in Saccharomyces cerevisiae.** *Cell* 2006, **127(4)**:735-745.
47. Kunkel GR, Pederson T: **Upstream elements required for efficient transcription of a human U6 RNA gene resemble those of U1 and U2 genes even though a different polymerase is used.** *Genes Dev* 1988, **2(2)**:196-204.
48. Skuzeski JM, Lund E, Murphy JT, Steinberg TH, Burgess RR, Dahlberg JE: **Synthesis of human U1 RNA. II. Identification of two regions of the promoter essential for transcription initiation at position +1.** *J Biol Chem* 1984, **259(13)**:8345-8352.
49. Ruzanov P, Jones SJ, Riddle DL: **Discovery of novel alternatively spliced C. elegans transcripts by computational analysis of SAGE data.** *BMC Genomics* 2007, **8**:447.
50. Reinke V, Gil IS, Ward S, Kazmer K: **Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans.** *Development* 2004, **131(2)**:311-323.
51. Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS: **A gene expression map for Caenorhabditis elegans.** *Science* 2001, **293(5537)**:2087-2092.
52. Hope IA, Arnold JM, McCarroll D, Jun G, Krupa AP, Herbert R: **Promoter trapping identifies real genes in C. elegans.** *Mol Gen Genet* 1998, **260(2-3)**:300-308.
53. Lynch AS, Briggs D, Hope IA: **Developmental expression pattern screen for genes predicted in the C. elegans genome sequencing project.** *Nat Genet* 1995, **11(3)**:309-313.
54. Hope IA: **'Promoter trapping' in Caenorhabditis elegans.** *Development* 1991, **113(2)**:399-408.
55. Harismendy O, Gendrel CG, Soularue P, Gidrol X, Sentenac A, Werner M, Lefebvre O: **Genome-wide location of yeast RNA polymerase III transcription machinery.** *Embo J* 2003, **22(18)**:4738-4747.
56. Roberts DN, Stewart AJ, Huff JT, Cairns BR: **The RNA polymerase III transcriptome revealed by genome-wide localization and**

- activity-occupancy relationships.** *Proc Natl Acad Sci USA* 2003, **100(25)**:14695-14700.
57. Isogai Y, Takada S, Tjian R, Keles S: **Novel TRFI/BRF target genes revealed by genome-wide analysis of Drosophila Pol III transcription.** *Embo J* 2007, **26(1)**:79-89.
58. Brenner S: **The genetics of Caenorhabditis elegans.** *Genetics* 1974, **77(1)**:71-94.
59. Freedman JH, Slice LW, Dixon D, Fire A, Rubin CS: **The novel metallothionein genes of Caenorhabditis elegans. Structural organization and inducible, cell-specific expression.** *J Biol Chem* 1993, **268(4)**:2554-2564.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

