*6304–6318* *Nucleic Acids Research, 2012, Vol. 40, No. 13* *Published online 29 March 2012*
*doi:10.1093/nar/gks282*

# Analysis of *C. elegans* intestinal gene expression and polyadenylation by fluorescence-activated nuclei sorting and 3′-end-seq

**Simon Haenni[1], Zhe Ji[2], Mainul Hoque[2], Nigel Rust[3], Helen Sharpe[1], Ralf Eberhard[4,5], Cathy Browne[1], Michael O. Hengartner[4], Jane Mellor[1], Bin Tian[2,*] and André Furger[1,*]**

[1]Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK, [2]Department of Biochemistry and Molecular Biology, UMDNJ-New Jersey Medical School, 185 South Orange Avenue, Newark, NJ 07101-1709, USA, [3]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK, [4]Institute of Molecular Life Sciences, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich and [5]Institute of Neuropathology, Schmelzbergstrasse 12, CH 8091 Zürich, Switzerland

Received February 20, 2012; Revised March 13, 2012; Accepted March 14, 2012

## ABSTRACT

**Despite the many advantages of *Caenorhabditis elegans*, biochemical approaches to study tissue-specific gene expression in post-embryonic stages are challenging. Here, we report a novel experimental approach for efficient determination of tissue-specific transcriptomes involving the rapid release and purification of nuclei from major tissues of post-embryonic animals by fluorescence-activated nuclei sorting (FANS), followed by deep sequencing of linearly amplified 3′-end regions of transcripts (3′-end-seq). We employed these approaches to compile the transcriptome of the developed *C. elegans* intestine and used this to analyse tissue-specific cleavage and polyadenylation. In agreement with intestinal-specific gene expression, highly expressed genes have enriched GATA-elements in their promoter regions and their functional properties are associated with processes that are characteristic for the intestine. We systematically mapped pre-mRNA cleavage and polyadenylation sites, or polyA sites, including more than 3000 sites that have previously not been identified. The detailed analysis of the 3′-ends of the nuclear mRNA revealed widespread alternative polyA site use (APA) in intestinally expressed genes. Importantly, we found that intestinal polyA sites that undergo APA tend to have U-rich and/or A-rich upstream auxiliary elements that may contribute to the regulation of 3′-end formation in the intestine.**

## INTRODUCTION

The transcribed genome equips cells and tissues with the necessary tools to complete the required biological processes and execute its function. Analysis of tissue-specific gene expression is particularly informative if tissues are isolated from a whole living organism rather than performed with *in vitro* cultured somatic cells that must have undergone some reprogramming to maintain proliferation.

Although tissue-specific profiling is straightforward for most large multicellular animals, this can be challenging for smaller species such as *Caenorhabditis elegans* which, despite its obvious advantages for genetic approaches (1), has limited amenability for biochemical and tissue-specific approaches. The tough cuticle of the nematodes makes it very difficult to isolate intact organs and if possible, requires laborious hand dissection (2). Significant advancements have recently been made with the cultivation and subsequent fluorescence activated cell sorting (FACS)-based purification of tissue-specific embryonic cells (3–6). In contrast, the analysis of mature tissues is more complex as post-embryonic cells can't be cultivated. The only currently available approach for high-throughput transcriptome analysis is based on the tissue-specific expression of a tagged poly(A) binding protein, followed by formaldehyde cross-linking and immunoprecipitation of polyadenylated RNAs (7–10). Although this represents an elegant and successful approach, quicker and simpler methods that are not limited to the analysis of polyadenylated mRNAs, are highly desirable.

The roles of chromatin remodelling, transcriptional regulation and alternative splicing in establishing tissue-specific

*To whom correspondence should be addressed. Tel: +44 1865 613261; Fax: +44 1865 613276; Email: andre.furger@bioch.ox.ac.uk
Correspondence may also be addressed to Bin Tian. Tel: +1 973 972 3615; Fax: +1 973 972 5594; Email: btian@umdnj.edu

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

mRNA expression have been well recognized (11–13). In contrast, the contribution of alternative cleavage and polyadenylation (APA) to tissue-specific gene expression is only beginning to be fully appreciated (14–21). It is currently believed that more than half of the mammalian genes have multiple polyadenylation sites, or polyA sites (22), resulting in mRNA isoforms with different 3′-untranslated regions (3′-UTR) and/or protein-coding sequences. Specific sequences located in the 3′-UTRs, via association with RNA binding proteins and miRNAs, affect multiple steps of gene expression (23) including nuclear cytoplasmic mRNA export, localization, translation and stability (24). Thus, modulating 3′-UTR length regulates the scope of potential interaction partners with the mRNA and so APA is a major contributor to tuning gene expression for tissue-specific needs.

Pre-mRNA cleavage and polyadenylation is carried out by the multi-subunit 3′-end processing complex (25,26), which executes the two coupled steps, cleavage and polyadenylation (27). *Cis*-elements located near the cleavage site govern the reaction, and vary in sequence and placement across species (28).

Here we present a novel experimental approach to study tissue-specific gene expression and polyadenylation in post-embryonic nematode tissues. We isolated and purified green fluorescent protein (GFP)-tagged intestinal nuclei by fluorescence-activated nuclei sorting (FANS) and determined the intestinal transcriptome by a novel deep sequencing method named 3′-end-seq. This approach allowed us to identify and determine expression levels for about 10 000 nematode genes and compile the most comprehensive analysis of the transcriptome in the *C. elegans* intestine to date. Consistently, genes highly expressed in the intestine are significantly associated with the GATA element in the promoter region and are annotated with gene ontology (GO) functions characteristic of the intestine (2). In addition, we provide the first global and quantitative analysis of tissue-specific polyA site usage in *C. elegans*. We identified 3282 polyA sites which have not been reported before. We show widespread APA in the intestine that correlates with the presence of specific auxiliary *cis*-elements positioned upstream of the alternatively used cleavage sites.

## MATERIALS AND METHODS

### Nematode strains

Strain JM149 expresses a nuclear GFP-H2B fusion protein from the *elt-2* promoter (without coding sequence of the *elt-2* gene) in intestinal cells {Is[pRF4 (*rol-6*); pJM459 (*elt-2*p*::nls::gfp-h2b*)]} and was kindly provided by Jim McGhee. Strain AW60 [*him-5(e1490); wIs51*] expresses a nuclear GFP marker in seam cells (*scm::nls::gfp*). Strain PK2011 {*unc-119*(*ed3*)*; crEx37* [*sur-5p::GFP, unc-119*(+)]} expressing a nuclear GFP marker in most somatic cells (short name sur5-GFP) was kindly provided by Jonathan Hodgkin, as well as the strain GFPF with the *myo-3*-p*::nls::gfp* reporter that expresses GFP in the nuclei of body wall muscle cells. Strain MS604 [*unc-119*(*ed4*)*III; him-8*(*e1489*)*IV; irIs37*] expressing a nuclear YFP (yellow fluorescent protein) marker in neuronal cells from the *unc-119* promoter (*unc-119*p*::nls::yfp::lacZ*) was obtained from Morris Maduro. Strain CB4974 [*Is*(*su1006 Ce MyoD::β-Gal*)] was used for the purity assessment of FANS (note: *MyoD* is annotated as *hlh-1*).

### Nuclei isolation for microscopy and flow cytometry analysis

Nuclei preparations for microscopy analysis were prepared at small scale [two to four 9-cm NGM (nematode growth medium) plates] and those for flow cytometry experiments at large scale. For large scale assays, mixed stage 150 ml liquid cultures, grown at 25°C, were harvested by three washes in M9 buffer (22 mM $KH_2PO_4$, 33.71 mM $Na_2HPO_4$, 85.56 mM NaCl, 1 mM $MgSO_4$) and incubated for 30–45 min at 25°C in M9 buffer for elimination of intestinal bacteria. The worm solution was subsequently passed through a 70 µm cell strainer (BD Biosciences) and the filtrate collected in 10 ml NBP (10 mM HEPES pH 7.6, 10 mM KCl, 1.5 mM $MgCl_2$, 1 mM EGTA, 0.25 mM sucrose) per 250 µl worm slurry (estimated before filtration) for quick buffer exchange and enrichment of adults. All subsequent steps were performed at 4°C/on ice. Nuclei were released in 10 ml batches by 5–10 strokes in a pre-chilled Wheat on stainless-steel tissue grinder (clearance 0.0005 inches = 12.5 µm) (29). After sequential filtration through 100 µm and 40 µm nylon cell strainers (BD Biosciences), the solution was centrifuged at 2500 *g* for 5–10 min at 4°C. For run-on experiments, see below. For flow cytometry analyses, the pellet was resuspended in 3 ml cold NPB and split into two Eppendorf tubes. Large worm fragments were pelleted for 1 min at 4°C at 300 g and the supernatants were transferred to fresh tubes. A quantity of 100 µl of this unsorted sample was transferred into 1 ml Trizol® Reagent, mixed well and left on ice for later RNA isolation (see below). A quantity of 50–100 µl were stained with 1 µg/ml propidium iodide, mounted on microscope slides and analysed using an Axioplan 2 microscope, the AxioCam digital camera and the AxioVision software (Zeiss). Half of the remaining nuclei preparations (of both wild-type control and a reporter strain) was stained with 1 µg/ml propidium iodide and all samples were immediately analysed by flow cytometry.

### Nuclear run-on analysis

The nuclear run-on (NRO) protocol is based on similar experimental approaches used in mammalian systems (30). The nuclei pellet from the NPB precipitation step were resuspended in 4 ml cold hypotonic lysis buffer (HLB: 10 mM Tris–HCl pH 7.5, 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% igepal), put on ice for 5 min and underlaid with 1 ml HLB/10% sucrose. The two-layered mix was spun at 2500 *g* at 4°C for 10 min and the nuclei pellet resuspended in an equivalent volume of 2x transcription mix (40 mM Tris–HCl pH 7.9, 300 mM KCl, 10 mM $MgCl_2$, 40% glycerol, 2 mM DTT). For pol II specific

gene detection, nuclei isolated from an entire 150 ml liquid worm culture were used. For pol I and pol III transcription assays, a third of the nuclei preparation was used. Nuclei were pre-incubated, in the absence of nucleotides, on ice for 20 min with or without α-amanitin. α-Amanitin was added to a final concentration of 100 µg/ml. Unlabelled ATP, CTP and GTP (0.57 mM final concentration each) plus 50 µCi [α-$^{32}$P]-UTP (400 Ci/mmol) were added (15% of the total volume) and transcription reactions were incubated at 20°C for 15 min. Nuclei were directly subjected to hot phenol treatment, then RNA was extracted with phenol–chloroform and subsequently ethanol precipitated. The RNA pellet was resuspended in 60 µl water. Transcripts were partially hydrolysed by addition of 15 µl of 1 M NaOH and incubation on ice for 5 min. The reaction was stopped by adding 30 µl of 0.5 M Tris/0.5 M HCl. Radio-labelled transcripts were hybridized to antisense riboprobes (bound to a nylon filter, see below) overnight at 42°C in hybridization solution (6x SSPE, 50% formamide, 5 x Denhardt's solution, 0.1% SDS, 50 µg/ml herring sperm DNA and 0.1 mg/ml tRNA). Hybond N+ nylon membranes were prepared and using a slot-blotter: 1 µg of RNA per slot was applied for each probe and the filters were subsequently pre-hybridized at 42°C for 90–120 min in hybridization solution. After hybridization, filters were washed in 5x SSPE at room temperature for 5 min and in 1x SSPE/0.1% SDS. Quantitation was carried out using a Fuji phosphorimager FLA-3000 (software provided): signal refers to photo stimulated luminescence (PSL) units.

### Riboprobes

The riboprobes for NRO assays were made by insertion of PCR fragments into linearized pGEM-T plasmids (see riboprobe table in Supplementary Data). Antisense riboprobes were synthesized by *in vitro* transcription from linearized plasmid DNA as a template in 80 mM HEPES pH 7.6, 2 mM spermidine, 40 mM DTT, 3 mM each of rATP, rUTP, rCTP and rGTP, 5 U/ml pyrophosphatase (Sigma), 1000 U/ml RNase Out, and either 12 mM $MgCl_2$ plus 1,800 U/ml T7 polymerase, or 16 mM $MgCl_2$ plus 1800 U/ml SP6 polymerase. Transcription reactions were carried out at 37°C (T7) or 40°C (SP6) for 4 h. DNA was removed by the addition of DNase I buffer and 40 U/µl DNase I and incubation at 37°C for 1 h. The RNA was phenol–chloroform extracted, ethanol precipitated and resuspended in 80% formamide, 10 mM EDTA pH 8.0. Samples were stored at −70°C.

### FANS

Flow cytometry and sorting experiments were performed according to standard procedures using Dako Cytomation MoFlo Legacy sorters equipped with either a 488 nm argon ion laser (200 mW) or a 488 nm solid-state laser (100 mW). Filters: 530/40 bandpass filter (FL-1, green fluorescence), 670/30 bandpass filter (FL-3, red channel for PI fluorescence). Sort rates were ∼20 000 events per second and the sort mode was set to purity 1 [highest purity (∼99%) with best recovery]. We used a log/log scale for FSC/SSC to detect small particles. The threshold was set to only remove electronic noise signals. About 10 000 or 50 000 events were sorted onto a microscope slide and the rest of double-positive events (100 000—1 Mio events, depending on the batch) was gated for RNA isolation using Trizol® Reagent. The software to monitor and analyse the flow cytometry experiments was Summit v4.3 (Dako). The whole data sets with all the controls to set up the gates and instrument parameters are available on request.

### Purity assessment of FANS

Liquid worm cultures of strains JM149 and CB4974 were mixed 1:1 and used for a nuclei preparation followed by FANS for the JM149-specific GFP marker. RNA isolated from one-twentieth of the pre-sorted material (unsorted) and of ∼120 000 sorted events (sorted) was linearly amplified using the MessageAmp ™ II aRNA amplification kit (Applied Biosystems, see next section) and analysed by real-time RT–PCR using gene-specific RT primers [the forward primers had to be used because of the antisense nature of the amplified antisense RNA (aRNA)]. *Rpl-43* was used as a normalization control (primers 5′-GAAGGTCGGAATCGTCGGAA-3′, 5′-GGTGACGGTTCCGTAGACGTA-3′). Primers for the *LacZ* (*β-Gal*) marker of CB4974: 5′-CGCCGGTCGCTACCATTACC-3′, 5′-GAGCACAGGGAGAAAGAGCATG-3′. Primers for the JM149 GFP::H2B marker: 5′-CAGGAGAACTTGCCAAGCACG-3′, 5′-TCATTCACAGGACAAAGAGAGG-3′.

### 3′-end-seq

RNA isolated from unsorted or sorted JM149 nuclei (∼65 ng) was amplified using the MessageAmp™ II aRNA amplification kit (Applied Biosystems) to create aRNA. Briefly, the RNA was reverse transcribed using a primer (R1, Figure 3A) containing the T7 promoter sequence and 24 Ts. After the second strand synthesis, double-stranded cDNAs were *in vitro* transcribed by T7 RNA polymerase to give rise to aRNAs. About 30 ng of aRNA per sample was used for ligation with a 3′ adapter (Bioo Scientific), which is 5′ adenylated and 3′ blocked. Ligation was carried out at 22°C for 1 h using truncated RNA ligase2 (Bioo Scientific). Ligated RNA was then reverse transcribed using Superscript II reverse transcriptase (Invitrogen), followed by 12 cycles of PCR amplification using a three primer mix (P1, P2 and P3). Primers P1 and P2 were used at the ratio of 1:10. P2 and P3 contained sequences for cluster generation on the Illumina flow cell. P3 also contained an index sequence for multiplex sequencing. The amplified PCR product was run in an 8% acrylamide gel, and the product corresponding to insert size of ∼50–60 nt were excised from the gel and eluted overnight. Eluted DNA was purified by ethanol precipitation and was checked in an Agilent Bioanalyzer using the high sensitivity DNA kit (Agilent technologies). The cDNA were

then sequenced on an Illumina GA IIx. Oligo sequences are as follows:

R1: 5′-TAATACGACTCACTATAGGGAGA$(T)_{24}$
3′ Adapter: 5′-rAppTGGAATTCTCGGGTGCCAAGGddC
R2: 5′-GCCTTGGCACCCGAGAATTCCA
P1: 5′-GTTCAGAGTTCTACAGTCCGA$(T)_{12}$VN
P2: 5′-AATGATACGGCGACCACCGAGATCTACAC GTTCAGAGTTCTACAGTCCGA
P3: 5′-CAAGCAGAAGACGGCATACGAGATNNNN NNGTGACTGGAGTTCCTTGGCACCCGAGAAT TCCA, in which 'NNNNNN' is an index sequence.
S1: 5′-CGACAGGTTCAGAGTTCTACAGTCCGA$(T)_{11}$
S2: 5′-GGAATTCTCGGGTGCCAAGGAACTCCAGT CAC

### Align reads to the *C. elegans* genome

Sequencing reads (72 nt) were first trimmed to 50 nt and then aligned to the WS190 genome using TopHat (31) allowing two mismatches. Only reads uniquely aligned to the genome were used for subsequent analyses. Overall, we obtained 9.24 million uniquely mapped reads for the unsorted sample and 4.71 million uniquely mapped reads for the sorted sample.

### Gene expression analysis

Gene expression levels were calculated as the number of reads within the gene boundary, with the 5′-end defined by Wormbase and the 3′-end defined by Wormbase or the 3′-most polyA site identified in this study, whichever is further downstream. The reads per million (RPM) value, calculated as the number of reads assigned to a gene per million mapped reads in the sample, was used to indicate the gene expression level. The Fisher's exact test was used to examine whether a gene had a significant difference in gene expression between the unsorted and sorted samples.

### Promoter and GO analyses

The promoter sequences, −500 to +100 nt around mapped 5′-ends of genes, were used in this analysis. The Fisher's exact test was used to calculate enrichment of hexamers for highly expressed genes in the intestine, as compared with lowly expressed genes. In addition, we divided the promoter sequence into 10 regions and calculated the fraction of genes containing GATA elements in each region, including TGATAA and its antisense TTATCA. For GO analysis, we used NCBI GO annotation of genes. The GO Parser program of BioPerl was used to get all genes associated with a GO term. We used the Fisher's exact test to assess whether a GO term is significantly associated with highly or lowly expressed genes.

### PolyA site identification

The genomic position corresponding to the 5′-end of aligned 3′-end-seq reads was considered as the −1 position relative to the cleavage site. Since the cleavage reaction often is not precise, leading to multiple cleavage sites located adjacent to each other (22), we progressively clustered cleavage sites located within 20 nt from one another. When a cluster size was ≤20 nt, the position with the highest number of supporting reads was used as the representative cleavage site for the polyA site, and all other cleavage sites were assigned to the same polyA site. When a cluster was >20 nt, we split the cluster into multiple polyA site clusters. This was carried out by: (i) identifying the cleavage site with the greatest number of supporting reads in the cluster and assigning other cleavage sites within 20 nt to the cluster and (ii) repeating (i) until no cleavage site was unassigned. We further required that a representative cleavage site had at least three supporting reads. PolyA sites with the usage level <5% after clustering in both unsorted and sorted samples were not used in this study. In addition, we examined the −10 to +10 nt region surrounding each polyA site for indication of false polyA site identification due to internal priming of A-rich sequence (32). If there were ≥6 consecutive As or ≥7 As in a 10 nt window, the polyA site was considered as an internal priming candidate and was not used for further analysis.

### Analysis of polyA sites

For genes with more than one polyA site, we calculated the usage level for each polyA site as the number of supporting reads for the polyA site divided by the total number of reads assigned to the gene. The Fisher's exact test and absolute usage level difference between samples were used to examine whether a polyA site had significant difference in usage between unsorted and sorted samples. The Fisher's exact test was used to assess whether a *cis*-element is significantly associated with polyA sites more or less used in the intestine. To identify PAS in a set of polyA sites, we selected the hexamer with the highest frequency in the −40 to −1 nt region of the polyA site, and removed polyA sites associated with the hexamer and repeated the process for the remaining polyA sites, until not a single hexamer occurred in more than 5% of the remaining polyA sites.

### *Trans*-splicing gene annotation

The operon information was obtained from Ref. (33). Operon genes were further divided into first, middle and last genes, based on the location in operon. The SL1 gene annotation was obtained from the modENCODE Spliced Leaders track (7). Genes without spliced leader annotation were annotated as No SL.

## RESULTS

### Purification of *C. elegans* nuclei by FANS

In order to analyse tissue-specific gene expression in postembryonic stages of *C. elegans*, we explored the possibility of purifying nuclei from transgenic strains expressing nuclear GFP markers from tissue-specific promoters, rather than attempting to isolate whole intact cells. As the isolation of nuclei from postembryonic stages was not established, we first developed an experimental approach that enabled us to mechanically release

nuclei from the nematodes. As shown in Figure 1A and B, the consequent protocol results in the release of free nuclei of varied diameters ranging from 3 to 10 μm. To confirm that the isolated nuclei remain structurally intact, we subjected them to nuclear run-on analysis (Figure 1D–F). To that end, we prepared filters containing antisense riboprobes complementary to regions in the polymerase I (pol I) and polymerase III (pol III) transcribed rDNA genes and the polymerase II (pol II) transcribed *rps-6* and *vit-2* genes (Figure 1C). Hybridization efficiency of all the antisense probes was first verified by exposing control filters to T7 *in vitro* transcribed radio-labelled sense transcripts (Figure 1D, T7 panel). Isolated nuclei were subsequently incubated in transcription buffer containing radio-labelled UTP either in the presence (+α) or absence (−α) of α-amanitin (Figure 1D). The pre-incubation of the nuclei with α-amanitin significantly reduced the pol II and pol III derived signals (Figure 1D–F) but largely had no effect on the α-amanitin resistant pol I transcribed rRNA genes. Thus, our results show that the isolated nuclei pools are transcriptionally active, indicating that the isolation procedures results in structurally intact nuclei.

To determine whether the nuclei release protocol is suitable to isolate tissue-specific nuclei, we focused on the worm strain JM149 which expresses a nuclear GFP-H2B fusion protein under the control of the intestinal specific *elt-2* promoter. We selected this strain because the JM149 phenotype has strongly fluorescent intestinal nuclei (Figure 2A, top panel) and, since the intestinal transcriptome has previously been analysed by other methods (2,9,10), reference data was readily available for comparison.

Nuclei pools released from JM149 cultures also included strongly green fluorescent nuclei (Figure 2A, bottom panel). In addition, to gauge the scope of the technique, we also tested the nuclei release method using worm strains expressing nuclear GFP in all somatic cells (Supplementary Figure S1A), seam cells (Supplementary Figure S1B), neuronal cells (Supplementary Figure S1C) and body wall muscle cells (Supplementary Figure S1D). In all cases, intact fluorescent nuclei were released and thus the method can be applied to nuclei from many different cell types from developed tissues (Supplementary Figure S1).

Since it has been successfully applied in other systems (34,35), we next explored the possibility of further purifying released fluorescent nuclei from the pools by subjecting them to fluorescence-activated-nuclei sorting, a process we named FANS. To that end, we scaled the procedure and isolated nuclei from a 150 ml liquid JM149 culture (Figure 2A). The released nuclei were subsequently labelled with the non-specific nucleic acid stain propidium iodide (PI) and subjected to FANS (Figure 2B). To control if this approach is feasible to purify intestinal nuclei, we first gated the highly GFP and PI positive events (Figure 2B, gate R3) onto a microscope slide. Gating was restrictive, which is evidenced by the fact that only 0.08% of all detectable double events qualified for selection (Figure 2B, bottom panel). Although only few nuclei were GFP positive before FANS (Figure 2C, top panel), the sorted material exclusively contained double (PI red and green) positive events (Figure 2C, bottom panel) which confirmed amenability of the approach. We next FANS-purified intestinal nuclei at a large scale using the same parameters, subsequently isolated total RNA from the selected material and confirmed the presence of coding and non-coding RNA by RT–PCR (data not shown). This provided us with proof of principle that our approach allows isolation of intact RNA from selected nuclei. As the ultimate goal was to determine the intestinal transcriptome and assess tissue specific polyadenylation, it was critical to establish the purity of the final sorted RNA material.

We therefore designed an approach that enabled us to assess the contamination level of the final RNA preparations by non-intestinal RNA. To that end, we mixed JM149 nematodes with an equal number of a different transgenic strain expressing β-galactosidase (*MyoD::β-GAL*). Nuclei of this hybrid culture were released and a sample was used to isolate 'unsorted' RNA. The remaining nuclei were subjected to FANS and total RNA was retrieved from the sorted material. We then compared the relative β-Gal mRNA levels present in RNA pools from nuclei before and after FANS by real time RT–PCR. This analysis revealed a contamination level of <5% in RNA isolated from FANS purified nuclei (Figure 2D), confirming that the FANS approach is a feasible method to analyse intestinal-specific nuclear gene expression of postembryonic *C. elegans* stages.

### Analysis of gene expression by 3′-end-seq

We next wanted to globally analyse intestinal gene expression by deep sequencing. However, our initial analysis was complicated by the relatively low amounts (<1 μg) of RNA extracted from sorted samples. This technical issue and the fact that many *C. elegans* genes overlap at the 3′-end (36), which makes it difficult to resolve the sequencing reads generated by regular non-strand-specific RNA-seq methods, prompted us to develop a new method suited for our samples.

We first ameliorated the issue of low quantity of unsorted RNA by including an *in vitro* transcription step by T7 polymerase to amplify the source material in a linear fashion (37). As outlined in Figure 3A, a T7-oligo-d(T) primer is used to synthesize double-stranded cDNAs, followed by *in vitro* transcription with T7 RNA polymerase generating RNA that is antisense to the original RNA (aRNA). This step also ensures that only poly(A)+ RNA is amplified. An adapter was subsequently ligated to the 3′-end of the *in vitro* transcribed RNA, which provided the 3′-end target-sequence for priming subsequent reverse transcription. After first strand synthesis, we conducted 12 cycles of PCR reaction to amplify cDNA using a mixture of 3 PCR primers named P1, P2 and P3 (Figure 3A; see Materials and Methods section for details). P1 contained 12 Ts followed by a V (non-T), and an N at the 3′-end. This primer ensures that (i) cDNAs containing ≥12 As at the 3′-end are preferentially amplified and (ii) only 12 As of the original poly(A) tail remain in the final PCR product. In essence, this primer makes the sequence of one end of the PCR product come from the region directly upstream of a polyA site. P2 and P3

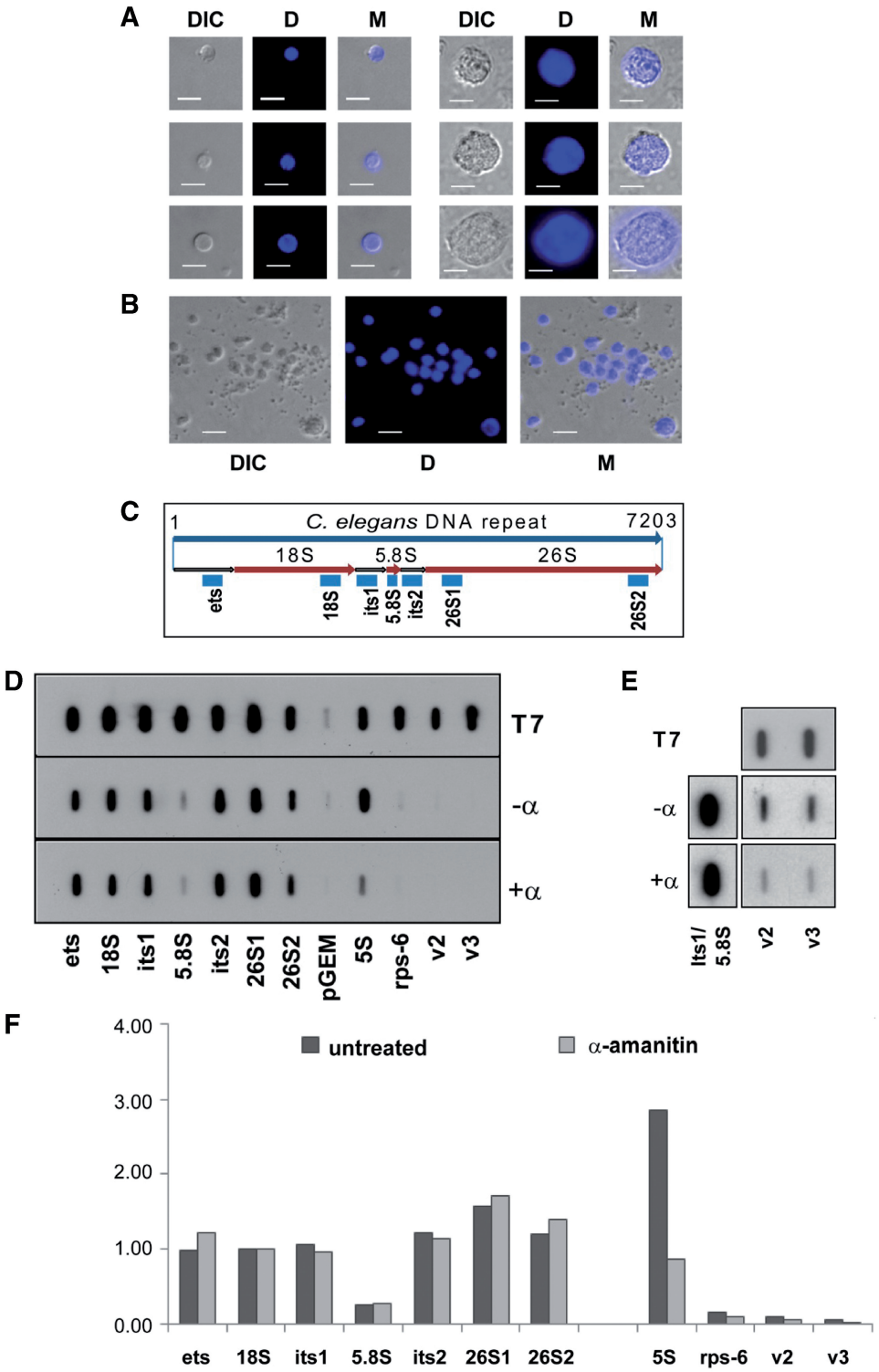


**Figure 1.** Release of transcriptionally active nuclei from postembryonic stages of *C. elegans*. (**A**) Representative pictures of isolated nuclei with different sizes and (**B**) of a wider field of nuclei to indicate the size distribution. DIC: differential interference contrast; D: DAPI staining; M: merged picture (DIC and D). Size bar: 5 µm. (**C**) Schematic of a 7-kb long ribosomal transcription unit. Antisense riboprobes are indicated by blue boxes; ets, external transcribed spacer; its1 and its2, internal transcribed spacers; 26S1 and 26S2, two different probes for the 26S rRNA. (**D**) Top panel (T7): control hybridization of filters with radiolabelled T7 sense transcripts; middle panel (−α): nuclear run-on analysis without α-amanitin; bottom panel (+α): run-on analysis with nuclei pre-incubated with α-amanitin. pGEM, control probe from the empty vector; Rps-6, v2 and v3, probes for the pol II transcribed *rps-6* (rps-6) and *vit-2* (v2 and v3) genes; 5S, probe for the pol III transcribed 5S rRNA (**E**) Nuclear run-on analysis for *vit-2* detection plus (+α) and minus (−α) α-amanitin with a larger volume of nuclei preparation. (**F**) Quantitation of signals from (D). Values from each filter were normalized to 18S signals.

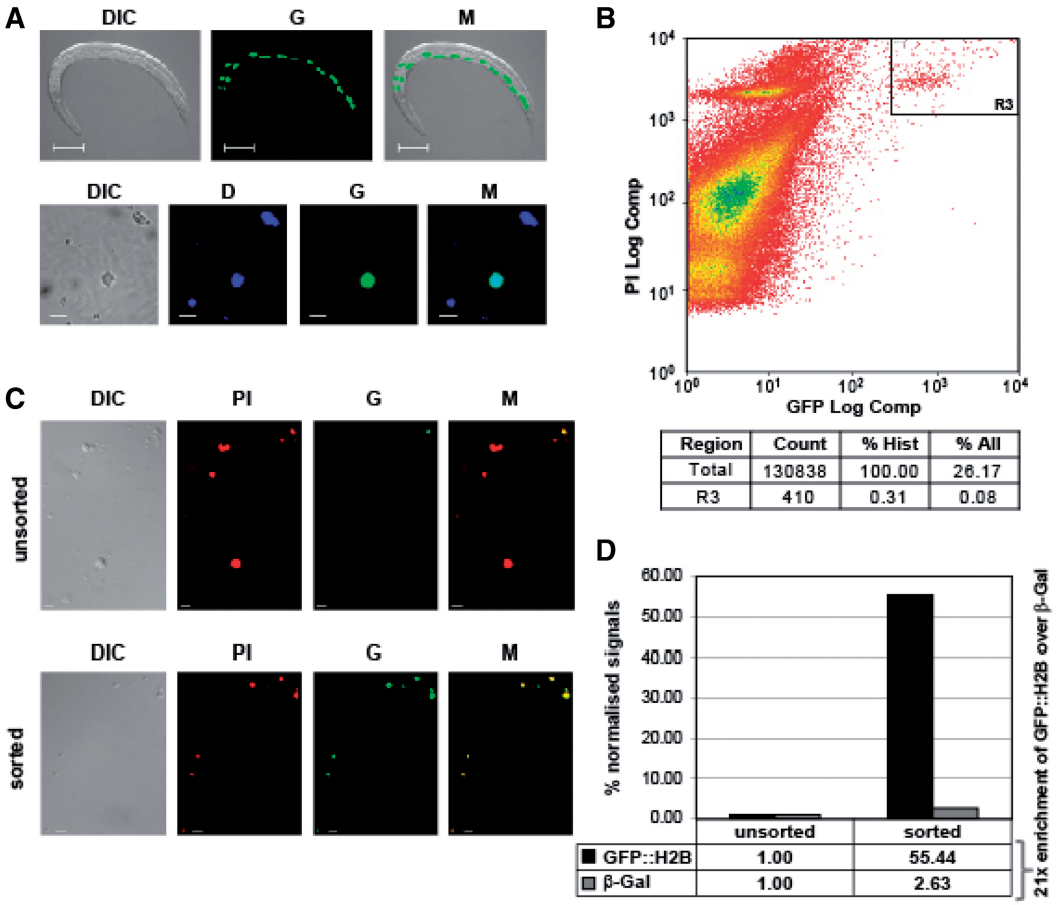


**Figure 2.** FANS purification of intestinal nuclei. (**A**) Strain JM149 with an intestinal nuclear marker (*elt-2*p*::nls::gfp-h2b*). Top panels: display of whole animal. Size bar: 100 µm. Bottom panels: representative pictures of isolated labelled nuclei. Size bar: 5 µm. DIC: differential interference contrast; G: green fluorescence; M: merged picture (DIC and G); D: DAPI staining. (**B**) Flow cytometry analysis of the isolated nuclei and the R3 gate used for sorting. Top: scatterplot of all events; bottom: summary of data. Count: number of events; % Hist: percentage of events displayed; % All: percentage of all events detected by the flow cytometer. (**C**) Representative pictures of nuclei prior to (unsorted) and after (sorted) sorting. DIC, G, and M are as in (A); PI, propidium iodide. (**D**) Purity assessment of FANS: strains JM149 (*GFP::H2B* marker) and CB4974 (*MyoD::β-Gal* marker) were equally mixed and subjected to FANS for GFP. RNA from unsorted and sorted material was isolated, amplified to aRNA and quantitatively analysed by real-time RT–PCR (see Materials and Methods section for details). The graph displays real-time PCR signals normalized against *rpl-43* with unsorted values set to 1. The 21-fold enrichment of *GFP::H2B* mRNA over *β-Gal* mRNA indicates about 4.7% contamination.

contained sequences for cluster generation on the Illumina flowcell. In addition, P3 contained an index region that allowed multiplexing in sequencing.

Finally, the so-created samples were sequenced by an Illumina GAIIx instrument. The sequencing reaction was initiated at the 3′-most A of the 12As at the end of the PCR product (Figure 3A). Thus, our read sequences, in theory, correspond to the region directly upstream of the polyA site. Indeed, we found that reads mapped to the 3′-end of transcripts annotated on Wormbase (Figure 3B and C). Notably, the reads are strand-specific, allowing unequivocal assignment of reads to genes even when they are overlapping at the 3′-end. For the subsequent gene expression analysis, we normalized read numbers mapped to each gene to the total mapped reads, and the normalized value was called RPM. We named our method 3′-end-seq.

### Gene expression analysis

Using our 3′-end-seq data, we set out to examine gene expression in the sorted intestinal nuclei. We first compared gene expression between unsorted and sorted samples using RPM. Overall, similar numbers of genes were considered as expressed in both samples based on different cut-offs (Supplementary Figure S2A and B), indicating the overall number of genes transcribed in the intestine does not differ greatly from other cell types. As expected, gene expression is not evenly distributed along the chromosomes (Supplementary Figure S2C). Using $P$ (Fisher's exact test) $< 0.01$ and fold change $> 2$, we identified 2456 genes that had higher expression in the sorted sample and 1053 genes that had lower expression in the sorted sample, as compared with the unsorted sample (Figure 4A and Supplementary Table S1 for the full list). Examples of two representative genes respectively are shown in Figure 4B.

We next carried out GO analysis to functionally characterize genes that are highly and lowly expressed in the intestine relative to the unsorted sample. Consistent with the functions of the intestine, this analysis revealed that highly expressed genes were associated with various

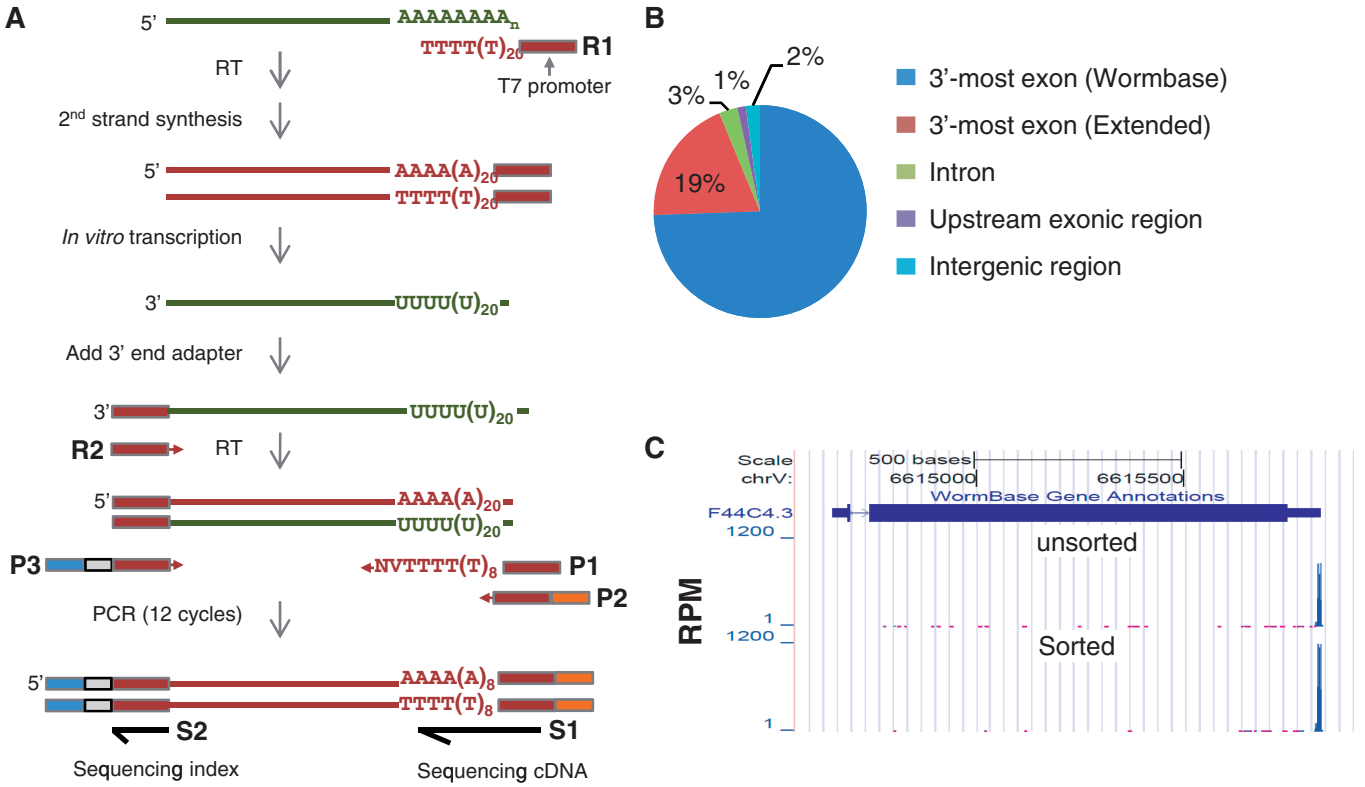


**Figure 3.** 3′-end-seq. (**A**) Schematic of 3′-end-seq. See Materials and Methods section for details. (**B**) Distribution of 3′-end-seq reads in the genome. 3′-most exons, introns and exons are defined by Wormbase (WS190); Extended refers to the region up to 3 kb downstream of the Wormbase annotated 3′-UTR. (**C**) An example gene (*F44C4.3*) showing that the reads are mapped to the 3′-end. Y-axis is the RPM value. Top, unsorted sample; bottom, sorted sample.

metabolic processes, such as 'regulation of macromolecule metabolic process', 'RNA metabolic process', 'lipid metabolic process', and 'carbohydrate metabolic process'; and several other processes related to various functions of the intestine, such as 'transmembrane transport', 'defense response', 'response to chemical stimulus' and 'oxidative reduction' (Table 1). In good agreement with the nature of a non-dividing somatic tissue, genes involved in development, differentiation, cell cycle and sexual reproduction were significantly lowly expressed in the intestine as compared with the unsorted sample (Table 1).

To further assess the tissue-specificity of the FANS-generated RNA sample, we examined the promoter regions (−500 to +100 nt around the 5′-ends of mapped genes) of genes that were highly expressed in the sorted sample as compared with the unsorted sample. As shown in Figure 4C, we found that hexamers containing the consensus GATA or its antisense TATC were significantly enriched in promoter regions of genes highly expressed in the intestine. Further analysis showed that the GATA elements are enriched by more than 2-fold in the region up to 200 nt upstream of the transcript start site (TSS) in genes that are enriched in the intestine compared with genes enriched in the unsorted sample (Figure 4D). The enrichment shows a clear trend gradually increasing towards the TSS and collapsing in the downstream region. Since the GATA element with the consensus sequence A[A/C/T]TGATAARR is considered to play a major role in activating intestinal gene expression (2,9,38), this finding provides significant support for the tissue-specificity of our samples.

Finally, we compared our data with a recent tiling array dataset for the intestinal tissue (10). For the 7781 genes considered as expressed by both our 3′-end-seq data and tiling data, a moderate but nevertheless significant correlation was discernible between up- and down-regulated genes (Figure 4D). Importantly, the genes that are highly and lowly expressed in the intestine are very significantly correlated between the two data sets shown by a Venn diagram analysis (Figure 4F) or gene density map analysis (Supplementary Figure S2D). Notably, 3′-end-seq data had a wider dynamic range for gene expression differences than the tiling array-based results (see the difference in scale between x-axis and y-axis in Figure 4D).

In addition, comparison of the genes expressed in FANS with previous studies by Pauli *et al.* (9), Spencer *et al.* (10) and McGhee *et al.* (2), showed a high degree of overlap (>72% for RPM > 0 and >62% for RPM > 1) (Supplementary Figure S3) Importantly, the genes commonly detected in our study and other ones tend to be expressed at high levels (Supplementary Figure S3B), indicating that some lowly expressed genes in the intestine were uniquely detected by different studies. Interestingly, when we examined the list of 80 genes that are considered not to be expressed in the intestine by Pauli *et al.* (9), we found that 31, 10 and 53 genes had detectable expression

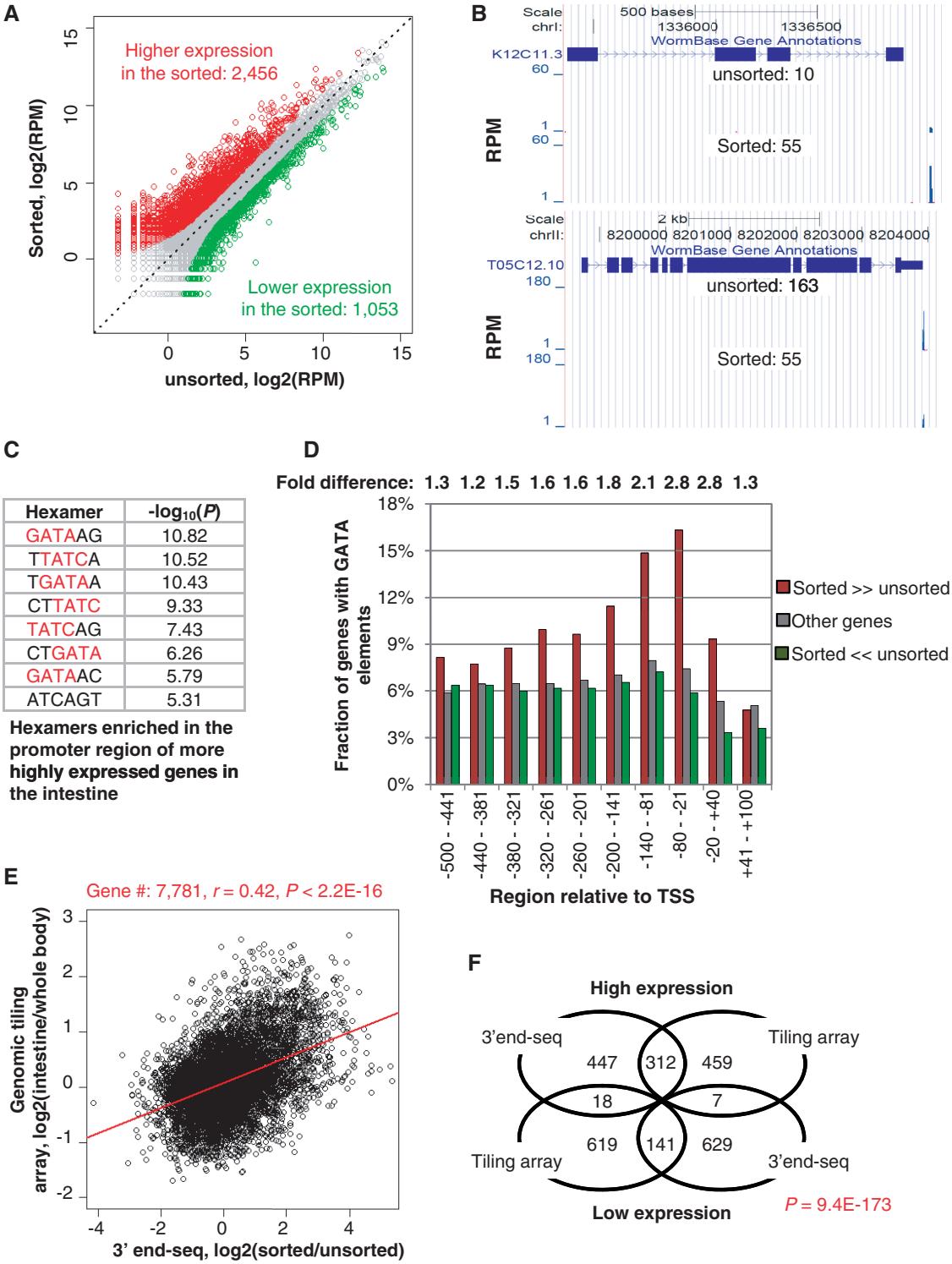


**Figure 4.** Analysis of gene expression in the intestine using 3′-end-seq reads. (**A**) Scatterplot of reads in unsorted (x-axis) and sorted (y-axis) samples. Genes higher and lower expressed ($P < 0.01$, Fisher's exact test; fold change > 2) in the sorted sample compared with the unsorted sample are shown in red and green, respectively. (**B**) Example genes having differential expression in the sorted and unsorted samples. *K12C11.3* (top) and *T05C12.10* (bottom) have higher and lower expression, respectively, in the sorted sample than the unsorted sample. Gene expression values (RPM) are indicated. (**C**)Top 8 hexamers significantly enriched for the promoter regions (−500 to +100 nt surrounding the TSS) of genes highly expressed in the intestine. *P*-values were based on the Fisher's exact test comparing genes with higher expression in the sorted sample with those with higher expression in the unsorted, as showed in (A). (**D**) The fractions of genes containing GATA elements in different regions surrounding the TSS. Genes were divided into three groups: (i) with higher expression in the sorted (sorted >>unsorted); (ii) with higher expression in the unsorted (sorted <<unsorted); (iii) other genes. Sorted/unsorted indicates the fold difference between genes in groups 1 and 2. (**E**) Comparison of gene expression using 3′-end-seq with genomic tiling array. A total of 7781 genes detected by tiling array and with expression level > 1 RPM in the unsorted and sorted samples were used. The correlation (*r*, Pearson correlation) and its *P*-value are shown on the top. (**F**) Venn diagram showing that the overlap of the most regulated genes (top 10%) identified by 3′-end-seq with those by tiling array is significant. The *P*-value was based on the Fisher's exact test.

**Table 1.** Gene Ontology (GO) analysis of differentially expressed genes

| | GO_ID, GO_term | -log(*P*) |
|---|---|---|
| GO significantly associated with genes highly expressed in the intestine | GO:0060255, regulation of macromolecule metabolic process | 8.23 |
| | GO:0016070, RNA metabolic process | 6.05 |
| | GO:0006629, lipid metabolic process | 5.73 |
| | GO:0055085, transmembrane transport | 4.40 |
| | GO:0006952, defence response | 2.63 |
| | GO:0042221, response to chemical stimulus | 2.61 |
| | GO:0005975, carbohydrate metabolic process | 2.42 |
| | GO:0055114, oxidation reduction | 2.20 |
| GO significantly associated with genes lowly expressed in the intestine | GO:0003006, reproductive developmental process | 6.28 |
| | GO:0010171, body morphogenesis | 5.63 |
| | GO:0007049, cell cycle | 4.46 |
| | GO:0000910, cytokinesis | 4.26 |
| | GO:0042692, muscle cell differentiation | 3.91 |
| | GO:0022607, cellular component assembly | 3.40 |
| | GO:0016192, vesicle-mediated transport | 3.19 |
| | GO:0006006, glucose metabolic process | 3.11 |
| | GO:0030036, actin cytoskeleton organization | 2.86 |
| | GO:0035188, hatching | 2.66 |
| | GO:0040012, regulation of locomotion | 2.54 |
| | GO:0019953, sexual reproduction | 2.21 |
| | GO:0005996, monosaccharide metabolic process | 2.13 |

GO terms were analysed using the Fisher's exact test, based on highly and lowly expressed genes in the sorted sample versus unsorted (Figure 4A). GO terms associated with more than 1000 genes or more than 1000 child terms were discarded. To eliminate redundancy, we required that each reported GO term had at least 25% of associated genes not associated with any GO term with a more significant *P*-value.

in the FANS dataset, McGhee *et al*.'s and Spencer *et al*.'s studies, respectively. Notably, of the 31 genes detected by FANS, 28 were also found to be expressed by Spencer *et al*. (Supplementary Figure S4) and the remaining 3 were not studied by Spencer *et al*. Therefore, while biological variations leading to differences in gene expression between different studies cannot be ruled out, it appears that the method used for gene expression analysis may account for some discrepancies. In addition, it may not be trivial to determine truly negative genes due to pervasive transcription and regulation at the post-transcriptional level (39).

### Analysis of polyA site usage in the intestine

Since our deep sequencing reads correspond to the 3′-end region of genes (Figure 3B) and originated from nuclear mRNA, our data created a unique opportunity to examine polyA site usage in the intestine. Using a stringent algorithm to exclude false positives due to priming at internal A-rich sequences (see Materials and Methods section for details), we mapped 19 324 polyA sites for 11 171 genes (Figure 5A). Compared with polyA sites recently reported by Mangone *et al*. (40) and those by Jan *et al*. (36) using whole worm RNA, 3282 polyA sites were unique to this study (Figure 5A and supplementary Table 2).

Using all the polyA sites identified in our data, we analysed polyadenylation signals, or PAS, in the −40 to −1 nt region relative to the polyA site. Consistent with the result reported by Mangone *et al*. (40) and Jan *et al*.(36), we found that AAUAAA, the canonical PAS in metazoans, was the most significant hexamer, associated with 43% of all intestinal polyA sites (Supplementary Figure S5A). The PAS variant AAUGAA was the second-most prominent intestinal poly(A) hexamer, associated with 10% of the sites. Another 14 hexamers, most of which are close variants of AAUAAA or AAUG AA, were associated with 34% of the sites, and about 7% of the polyA sites had no prominent hexamer. Also consistent with the findings reported by Mangone *et al*., our data showed that polyA sites with different PAS are surrounded by similar nucleotide profiles (Supplementary Figure S5B). However, sites with weaker PAS appear to have a higher U-rich content, suggesting that U-rich sequences can complement functions of PAS, as we previously observed with human polyA sites (41). Notably, polyA sites uniquely identified by either of the three studies had similar nucleotide profiles surrounding the PAS (Supplementary Figures S6 and S7) and they tend to be associated with weak PAS (Supplementary Figure S6B).

Multiple polyA sites in a gene lead to alternatively polyadenylated mRNA isoforms. Overall, we found that about 35% of the genes expressed in sorted or unsorted samples were represented by different APA isoforms with each isoform expressed ≥5% in at least one sample (Figure 5B). About 23% of the genes have APA in the 3′-most exon only, leading to 3′-UTR isoforms, and another 12% of the genes have polyA sites located upstream of the 3′-most exon, potentially leading to changes of the encoded proteins (Figure 5B). Examples for each type are shown in Supplementary Figure S8.

About 15% of *C. elegans* genes are organized in operon-like structures that are transcribed into polycistronic pre-mRNAs (42). Maturation of individual mRNAs involves cleavage and polyadenylation at the 3′-end of upstream genes and *trans*-splicing of a small nuclear spliced leader RNA (SL snRNA) to the 5′-end of downstream genes. The nature of the spliced leader used for a particular gene depends on its position in the operon.

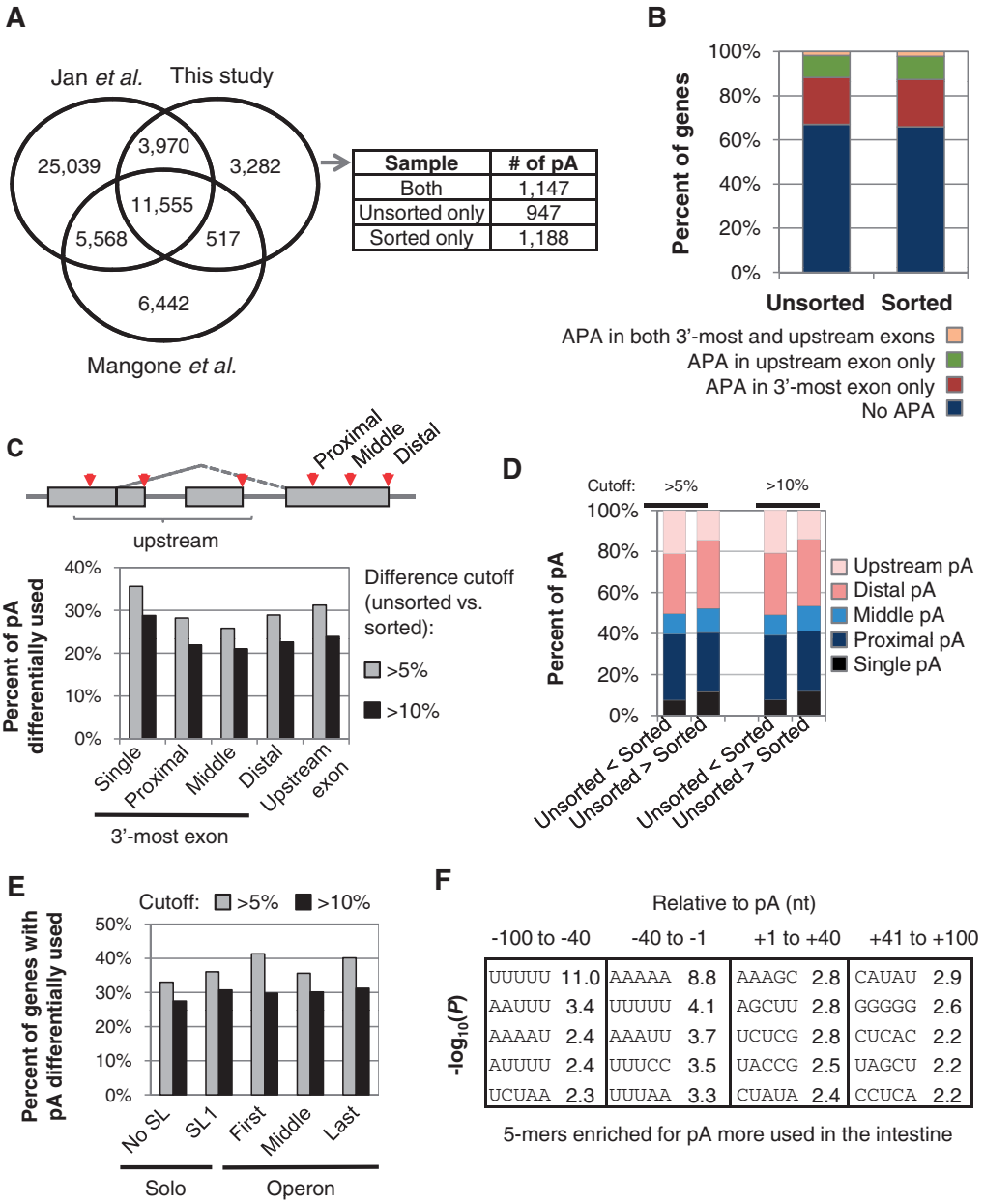


**Figure 5.** Analysis of alternative polyadenylation by 3′-end-seq. (**A**) Venn diagram comparing polyA sites identified in this study and those reported by Mangone *et al*. and Jan *et al*. PolyA sites unique to this study were further separated based on detection in unsorted and/or sorted samples, as shown in the table next to the Venn diagram. PolyA sites from different studies that are located within 20 nt from one another were considered identical sites. pA, polyA site. (**B**) Percentage of genes having alternative polyA sites in different regions of genes. (**C**) Schematic of alternative polyA sites (top) and percentage of polyA sites with differential usage (bottom). PolyA sites were grouped into different types. Two cutoffs were used: ≥5% change and ≥10% change. 'Single' refers to genes that contain a single polyA site in the 3′-most exon and further polyA sites in upstream exons. (**D**) Percentage of polyA sites of different types with more usage in the unsorted sample (unsorted > sorted) or in the sorted sample (unsorted < sorted). Two cutoffs were used, i.e. 5 and 10%, as indicated on the top. (**E**) Percentage of APA genes showing significant difference in polyA site usage in sorted versus unsorted samples. Genes were grouped based on the *trans*-splicing structure. Operon genes were divided into first, middle and last genes, based on the gene location in operons. SL1 genes have the SL1 spliced leader, and genes without spliced leader annotation are shown as No SL genes. (**F**) Significant 5-mers associated with polyA sites more used in the sorted versus unsorted samples. Differentially used polyA sites were selected using $P < 0.05$ (Fisher's exact test) and difference in usage >5%. Four regions around the polyA site were analysed. *P*-values were derived from the Fisher's exact test.

The first genes in operons receive a leader called SL1 and downstream positioned genes are *trans*-spliced to leader sequences of the SL2 type (33). SL1 is also added to most, but not all monocistronic (Solo) genes. Thus, *C. elegans* genes can be subdivided into several groups based on the pre-mRNA processing structure (Supplementary Figure S9A), including first, middle and last in an operon, and Solo genes with or without *trans*-splicing, i.e. SL1 and No SL. As reported by Mangone *et al*. (40) using whole worm RNA, we found that genes located in operons are more likely to have APA in the intestine than 'Solo' genes, and 'conventional' genes coding for transcripts that are not

*trans*-spliced (No SL) are the least likely to have APA (Supplementary Figure S9B). Also consistent with previous findings, our tissue-specific result indicates that polyA sites located in different types of genes with respect to *trans*-splicing and operon structures, and different locations within a gene differ widely in PAS usage (Supplementary Figure S9C). Some of these differences may be due to the high level of transcription generally seen in trans-spliced genes (Supplementary Figure S10).

We next focused on usage of polyA sites in genes with multiple polyA sites in the sorted vs. unsorted samples. As shown in Figure 5C, when 5% change of usage was used as the cutoff, we found 25-35% of polyA sites in different groups (Figure 5C, top panel) were differentially used comparing sorted versus unsorted samples. About 20–30% of the sites were found to be differentially used when a cutoff of 10% was applied. However, overall we did not observe global shifts towards promoter-proximal or -distal polyA sites (Figure 5D). In addition, about 25–40% of the genes showed differential expression of APA isoforms depending upon the cutoff, the *trans*-splicing structure and/or position in operons (Figure 5E). This indicates widespread regulation of alternative polyA site usage in the intestine. Importantly, we found that the distance between the two most regulated polyA sites in the 3′-most exon is >40 nt for most genes (Supplementary Figure S9D), suggesting significant potential for APA to influence gene expression by 3′-UTR-mediated regulation.

We further analysed nucleotide frequency around the polyA sites with different usage in the sorted versus unsorted samples. As shown in Figure 5F, a number of 5-mers were found to be significantly enriched for polyA sites more used in the intestine indicating *cis*-element differences around polyA sites can govern usage in the intestine. The most significant elements are UUUUU in the −100 to −41 region and AAAAA in the −40 to −1 region, suggesting specific regulation of intestinal APA through *cis*-elements. The detailed mechanism(s) and *trans*-acting factors that are associated with these sequences are to be explored in the future.

## DISCUSSION

We present a straightforward and effective approach to isolate nuclei from postembryonic nematode tissues. Combining the nuclei isolation procedure with a flow cytometry approach enabled us to purify large quantities of tissue-specific GFP-tagged nuclei. These nuclei are competent for nuclear run-on, indicating that their structure is largely intact. Since we successfully applied our method to the intestine tissue which is composed of fewer than 35 cells (nuclei) per animal, it could be applied to many other tissues in the worm. We believe this approach will be highly useful for the field because it enables tissue-specific gene expression analysis of nuclear transcriptomes using living nematodes and will complement a similar experimental approach published, whereas this manuscript was under review (43), which relies on immunopurification of doubly tagged nuclei.

The so-far most widely used method is based on the tissue-specific expression of a tagged poly(A) binding protein (FLAG-PAB-1) which allows isolation of tissue-specific mRNAs by co-immunoprecipitation (CoIP) (7,8). We believe that our FANS-based procedure not only complements this method, but also has several major advantages. Most importantly, FANS is not limited to the analysis of mRNA and by using state of the art flow cytometers, a very high degree of purity is achieved (44). By using tissue-specific fluorescently tagged histones, the desired expression of the marker can easily be monitored in a large number of animals in the growing culture at any time and during all critical stages of the isolation procedure. In addition, the use of a GFP-histone fusion marker minimizes diffusion of the tag after the nuclei are released (45). Importantly, FANS as demonstrated in Figure 2D, has the big advantage that the degree of cross-contamination of any sample can readily be assessed in each individual experiment.

Furthermore, the tagged PAB-1 competes with endogenous PAB-1, which may skew data towards highly expressed mRNAs and mRNA degradation during the CoIP mediated isolation is a constant risk. In contrast, RNA degradation is kept minimal in the FANS approach since nuclei remain intact and the whole procedure is carried out on ice.

FANS is not static but represents an important new approach with the potential for adaptation to analyse other aspects of tissue-specific gene expression including chromatin status and possibly the nuclear proteome. The method is particularly attractive as several strains are already available that express tissue-specific nuclear markers. Furthermore, as the isolated nuclei are transcriptionally active (Figure 1), FANS can be used for whole worm and tissue-specific global run on sequencing (GRO-seq) approaches to map the positions of transcriptionally active polymerases genome-wide and in a tissue-specific context. Thus, FANS is an ideal approach to study nuclear gene expression events in developed *C. elegans* tissues.

Furthermore, our deep sequencing method, 3′-end-seq, is a novel approach to study gene expression. It needs only low abundant unsorted material, is strand-specific, and provides quantitative measurement of gene expression with a good dynamic range. Since the reads correspond to the 3′-end region, data normalization is simple without the complication of gene size or splicing structure. Data are analysed in RPM instead of density values, such as RPKM (46).

### The purity issue and intestine-specific gene expression

For tissue-specific gene expression analysis, a high degree of sample purity is critical. Consequently, it was important to put several measures in place to determine and assess the purity of the FANS sample. The purification step itself is highly reliable as it is based on fluorescence-activated cell sorting, which achieves purities of ≥99%. In addition, the sorting process is fully automated and thus highly reproducible. There is, however, the possibility of cross-contamination by RNA-containing particles that

are below the detection threshold of the flow cytometer but are abundant enough to be stochastically present in our sorting droplets. For example, FANS is likely to co-purify the abundant mitochondria that are too small to be detected by the flow cytometer. Given the large number of mitochondria in the starting sample, even after several rounds of dilutions and precipitations, statistically, these organelles can be expected to be present in the sorting droplets. However, sequence reads from mitochondrial RNA in the sample can easily be filtered out and do not compromise the gene expression analysis as long as sufficient sequencing depth is reached. In addition, it is possible that the nuclei droplets can get contaminated by non-intestinal cellular RNA and/or endoplasmic reticulum (ER) associated mRNAs from non-intestinal sources. To obtain a measure for this potential source of false positives, we designed and implemented the contamination experiment presented in Figure 2D. This approach allowed us to experimentally and quantitatively establish the level of cross-contamination of non-intestinal mRNA purified by FANS which we found was <5%. Thus, cross-contamination of the nuclei droplets by non-intestinal RNA can be considered as marginal.

To globally validate the tissue specificity of the obtained data, we performed GO analysis of the highly and lowly expressed genes in our sample. This widely used analysis to verify tissue-specific gene expression revealed clear categories of the highly expressed genes that are in good agreement with the functions assigned to the intestine (Table 1). Furthermore, genes that were identified to be down-regulated compared with the whole worm unsorted nuclei, are associated with processes that are expected to be of low importance in the intestine, such as cell cycle regulation, cytokinesis, reproductive processes, body morphogenesis and muscle development (Table 1). Thus, this GO analysis provides strong support for the tissue specificity of FANS.

Moreover, since we focused on gene expression in the intestine, we were able to further globally scrutinize our data by analysing GATA promoter binding sites (Figure 4C and D) which are known to be critical for the control of intestinal gene expression (2,9,38). The analysis showed an overrepresentation of the GATA element in promoter regions of up-regulated genes in the sorted sample, in particular in the region around the TSS (Figure 4C and D). The overrepresentation of the GATA elements consequently provides strong additional evidence for the tissue specificity of our sorted mRNA.

In addition, we compared our expression data with the most recently described intestinal gene expression profile (10) and found a moderate but nevertheless significant correlation. Contributing factors that may prevent a stronger correlation could be the differences between strains (JM149 used in this study versus SD1084 used for the tiling array), comparison of embryonic and L2 larvae data (10) with mixed stage gene expression data in this study. Importantly, for highly and lowly expressed genes the two datasets are very much in agreement (Figure 4F). Finally, a cross-comparison with all available large-scale intestinal data sets (2,9,10) revealed a high degree of overlap (>70% for RPM > 0; Supplementary Figure S3).

Conclusively, based on our experimental controls and three independent global bioinformatic analyses (GO and GATA-element analyses, as well as comparison with available datasets) we conclude that FANS is a valid new experimental approach for the analysis of tissue-specific gene expression in postembryonic stages of *C. elegans*.

### The use of nuclear RNA for gene expression analysis

A potential drawback of the FANS method for gene expression analysis may be its reliance on nuclear rather than whole cell or cytoplasmic RNA. However, it appears that relative mRNA levels in nuclear and cytoplasmic compartments are highly concordant (35,47,48). Nevertheless, for some genes the focus on nuclear RNA may mask important post-transcriptional regulatory steps. In fact, regulation at the post-transcriptional level may explain the presence of potential false positive hits and could reveal important, previously unknown regulatory steps that control the expression patterns of such genes. However, as our GO and GATA analyses demonstrate, masking potential post-transcriptional regulated genes is not a major problem for the global analysis of nuclear mRNA expression profiles in nematodes, at least for the intestine. This is further supported by the fact that the analysis of nuclear RNA is increasingly used to profile tissue-specific gene expression in several organisms, including nematodes (43,49,50).

### The polyA analysis

Alternative polyadenylation has recently been recognized as an important mechanism to regulate gene expression in response to cell proliferation and tissue-specific cues (51,52). Our polyA analysis showed that up to 40% of intestinal expressed genes are subjected to alternative polyA site usage and thus suggests that APA contributes significantly to the fine-tuning of intestinal gene expression and the establishment of the intestinal transcriptome. Importantly, our intestinal APA data does not show global shifts between proximal and distal site usage that has been observed in several human tissues (13,19), a subset of *Drosophila* neuronal expressed genes (53) and during *C. elegans* development (40). This suggests that APA usage in the *C. elegans* intestine is unlikely to be controlled by modulating expression levels of key cleavage and polyadenylation factors such as CstF and CPSF as has been implicated in APA control during mammalian cell differentiation (16). It appears more likely that in *C. elegans*, the expression of specific auxiliary factors contributes to polyA selection in the intestine. Interestingly, increased usage of polyA sites in the intestine does correlate with the presence of specific *cis*-elements including 'UUUU U' and 'AAAAA' motifs in the 3′-UTR near the polyA cleavage sites. These sites could act as target sequences for intestine-specific regulators of APA.

However, it is important to highlight that the data presented in this study cannot be directly compared with previous analyses as the former is based on nuclear

mRNA versus whole cell mRNA in the latter. Whole cell mRNA represents the net output including cleavage efficiencies at different APA sites (APA usage) and differential stability of resulting mRNA isoforms. In contrast, by focusing on nuclear mRNA, our data is more likely to reflect actual polyA usage. As we were unable to detect global shifts in APA usage in our nuclear analysis, we propose that establishment of tissue-specific 3′-UTRomes, similar to those associated with different states of proliferation (54), may be more complex than initially thought. Intestine-specific APA in the nematode may largely depend on the presence of gene-specific auxiliary *cis*-elements and the expression of tissue-specific *trans*-factors.

## ACCESSION NUMBERS

Public database accession number: GSE32165 can be accessed via the following link: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32165

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1 and 2, Supplementary Figures 1–10 and Supplementary Methods.

## ACKNOWLEDGEMENTS

We thank other members of the Furger and Tian labs for helpful discussions. We thank Jonathan Hodgkin and Jonathan Ewbank for valuable suggestions and discussions. We are grateful to Jim McGhee for the JM149 strain and helpful comments.

## FUNDING

EPA Cephalosporin Fund (to A.F. and S.H.), MRC (to H.S.), Swiss National Science Foundation (SNSF, to S.H.), the Ernst Hadorn Foundation (to R.E. and M.H.), and NIH (GM084089 and HG005129, to B.T., M.H., and Z.J.). Funding for open access charge: Department of Biochemistry, University of Oxford (to A.F.).

*Conflict of interest statement*. None declared.

## REFERENCES

1. Antoshechkin,I. and Sternberg,P.W. (2007) The versatile worm: genetic and genomic resources for Caenorhabditis elegans research. *Nat. Rev. Genet.*, **8**, 518–532.
2. McGhee,J.D., Sleumer,M.C., Bilenky,M., Wong,K., McKay,S.J., Goszczynski,B., Tian,H., Krich,N.D., Khattra,J., Holt,R.A. *et al.* (2007) The ELT-2 GATA-factor and the global regulation of transcription in the *C. elegans* intestine. *Dev. Biol.*, **302**, 627–645.
3. Zhang,Y., Ma,C., Delohery,T., Nasipak,B., Foat,B.C., Bounoutas,A., Bussemaker,H.J., Kim,S.K. and Chalfie,M. (2002) Identification of genes expressed in *C. elegans* touch receptor neurons. *Nature*, **418**, 331–335.
4. Christensen,M., Estevez,A., Yin,X., Fox,R., Morrison,R., McDonnell,M., Gleason,C., Miller,D.M. 3rd and Strange,K. (2002) A primary culture system for functional analysis of *C. elegans* neurons and muscle cells. *Neuron*, **33**, 503–514.
5. Strange,K., Christensen,M. and Morrison,R. (2007) Primary culture of *Caenorhabditis elegans* developing embryo cells for electrophysiological, cell biological and molecular studies. *Nat. Protoc.*, **2**, 1003–1012.
6. Strange,K. and Morrison,R. (2006) *In vitro* culture of *C. elegans* somatic cells. *Methods Mol. Biol.*, **351**, 265–273.
7. Celniker,S.E., Dillon,L.A., Gerstein,M.B., Gunsalus,K.C., Henikoff,S., Karpen,G.H., Kellis,M., Lai,E.C., Lieb,J.D., MacAlpine,D.M. *et al.* (2009) Unlocking the secrets of the genome. *Nature*, **459**, 927–930.
8. Roy,P.J., Stuart,J.M., Lund,J. and Kim,S.K. (2002) Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. *Nature*, **418**, 975–979.
9. Pauli,F., Liu,Y., Kim,Y.A., Chen,P.J. and Kim,S.K. (2006) Chromosomal clustering and GATA transcriptional regulation of intestine-expressed genes in *C. elegans*. *Development*, **133**, 287–295.
10. Spencer,W.C., Zeller,G., Watson,J.D., Henz,S.R., Watkins,K.L., McWhirter,R.D., Petersen,S., Sreedharan,V.T., Widmer,C., Jo,J. *et al.* (2011) A spatial and temporal map of *C. elegans* gene expression. *Genome Res.*, **21**, 325–341.
11. Minard,M.E., Jain,A.K. and Barton,M.C. (2009) Analysis of epigenetic alterations to chromatin during development. *Genesis*, **47**, 559–572.
12. D'Alessio,J.A., Wright,K.J. and Tjian,R. (2009) Shifting players and paradigms in cell-specific transcription. *Mol. Cell*, **36**, 924–931.
13. Wang,E.T., Sandberg,R., Luo,S., Khrebtukova,I., Zhang,L., Mayr,C., Kingsmore,S.F., Schroth,G.P. and Burge,C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
14. Neilson,J.R. and Sandberg,R. (2010) Heterogeneity in mammalian RNA 3' end formation. *Exp. Cell Res.*, **316**, 1357–1364.
15. Lutz,C.S. and Moreira,A. (2011) Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. *Wiley Interdiscip. Rev. RNA*, **2**, 23–31.
16. Ji,Z., Lee,J.Y., Pan,Z., Jiang,B. and Tian,B. (2009) Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc. Natl Acad. Sci. USA*, **106**, 7028–7033.
17. Mayr,C. and Bartel,D.P. (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, **138**, 673–684.
18. Sandberg,R., Neilson,J.R., Sarma,A., Sharp,P.A. and Burge,C.B. (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science*, **320**, 1643–1647.
19. Zhang,H., Lee,J.Y. and Tian,B. (2005) Biased alternative polyadenylation in human tissues. *Genome Biol.*, **6**, R100.
20. Liu,D., Brockman,J.M., Dass,B., Hutchins,L.N., Singh,P., McCarrey,J.R., MacDonald,C.C. and Graber,J.H. (2007) Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis. *Nucleic Acids Res.*, **35**, 234–246.
21. Flavell,S.W., Kim,T.K., Gray,J.M., Harmin,D.A., Hemberg,M., Hong,E.J., Markenscoff-Papadimitriou,E., Bear,D.M. and Greenberg,M.E. (2008) Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron*, **60**, 1022–1038.
22. Tian,B., Hu,J., Zhang,H. and Lutz,C.S. (2005) A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res.*, **33**, 201–212.
23. Keene,J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.*, **8**, 533–543.
24. Fabian,M.R., Sonenberg,N. and Filipowicz,W. (2010) Regulation of mRNA translation and stability by microRNAs. *Ann. Rev. Biochem.*, **79**, 351–379.
25. Millevoi,S. and Vagner,S. (2010) Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res.*, **38**, 2757–2774.
26. Shi,Y., Di Giammartino,D.C., Taylor,D., Sarkeshik,A., Rice,W.J., Yates Iii,J.R., Frank,J. and Manley,J.L. (2009) Molecular

architecture of the human pre-mRNA 3' processing complex. *Mol. Cell*, **33**, 365–376.
27. Proudfoot,N.J., Furger,A. and Dye,M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501–512.
28. Tian,B. and Graber,J.H. (2012) Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdiscip. Rev. RNA.*, **3**, 385–396.
29. Lichtsteiner,S. and Tjian,R. (1995) Synergistic activation of transcription by UNC-86 and MEC-3 in *Caenorhabditis elegans* embryo extracts. *EMBO J*, **14**, 3937–3945.
30. Furger,A., O'Sullivan,J.M., Binnie,A., Lee,B.A. and Proudfoot,N.J. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev.*, **16**, 2792–2799.
31. Trapnell,C., Pachter,L. and Salzberg,S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
32. Lee,J.Y., Park,J.Y. and Tian,B. (2008) Identification of mRNA polyadenylation sites in genomes using cDNA sequences, expressed sequence tags, and Trace. *Methods Mol. Biol.*, **419**, 23–37.
33. Blumenthal,T., Evans,D., Link,C.D., Guffanti,A., Lawson,D., Thierry-Mieg,J., Thierry-Mieg,D., Chiu,W.L., Duke,K., Kiraly,M. *et al.* (2002) A global analysis of *Caenorhabditis elegans* operons. *Nature*, **417**, 851–854.
34. Okada,S., Saiwai,H., Kumamaru,H., Kubota,K., Harada,A., Yamaguchi,M., Iwamoto,Y. and Ohkawa,Y. (2011) Flow cytometric sorting of neuronal and glial nuclei from central nervous system tissue. *J. Cell. Physiol.*, **226**, 552–558.
35. Zhang,C., Barthelson,R.A., Lambert,G.M. and Galbraith,D.W. (2008) Global characterization of cell-specific gene expression through fluorescence-activated sorting of nuclei. *Plant Physiol.*, **147**, 30–40.
36. Jan,C.H., Friedman,R.C., Ruby,J.G. and Bartel,D.P. (2011) Formation, regulation and evolution of *Caenorhabditis elegans* 3[prime]UTRs. *Nature*, **469**, 97–101.
37. Hoeijmakers,W.A., Bartfai,R., Francoijs,K.J. and Stunnenberg,H.G. (2011) Linear amplification for deep sequencing. *Nat. Protoc.*, **6**, 1026–1036.
38. McGhee,J.D., Fukushige,T., Krause,M.W., Minnema,S.E., Goszczynski,B., Gaudet,J., Kohara,Y., Bossinger,O., Zhao,Y., Khattra,J. *et al.* (2009) ELT-2 is the predominant transcription factor controlling differentiation and function of the *C. elegans* intestine, from embryo to adult. *Dev. Biol.*, **327**, 551–565.
39. Merritt,C., Rasoloson,D., Ko,D. and Seydoux,G. (2008) 3' UTRs are the primary regulators of gene expression in the *C. elegans* germline. *Curr. Biol.*, **18**, 1476–1482.
40. Mangone,M., Manoharan,A.P., Thierry-Mieg,D., Thierry-Mieg,J., Han,T., Mackowiak,S.D., Mis,E., Zegar,C., Gutwein,M.R., Khivansara,V. *et al.* (2010) The landscape of *C. elegans* 3'UTRs. *Science*, **329**, 432–435.
41. Nunes,N.M., Li,W., Tian,B. and Furger,A. (2010) A functional human Poly(A) site requires only a potent DSE and an A-rich upstream sequence. *EMBO J.*, **29**, 1523–1536.
42. Blumenthal,T. (2005) Trans-splicing and operons (June 25, 2005), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.5.1, http://www.wormbook.org.
43. Steiner,F.A., Talbert,P.B., Kasinathan,S., Deal,R.B. and Henikoff,S. (2012) Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Res.*, **22**, 766–777.
44. Shapiro,H.M.P.f.c. (2003) *Practical Flow Cytometry*. Wiley & Sons, Inc., Hoboken, NJ.
45. Zhang,C., Gong,F.C., Lambert,G.M. and Galbraith,D.W. (2005) Cell type-specific characterization of nuclear DNA contents within complex tissues and organs. *Plant Methods*, **1**, 7.
46. Mortazavi,A., Williams,B.A., McCue,K., Schaeffer,L. and Wold,B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
47. Barthelson,R.A., Lambert,G.M., Vanier,C., Lynch,R.M. and Galbraith,D.W. (2007) Comparison of the contributions of the nuclear and cytoplasmic compartments to global gene expression in human cells. *BMC Genomics*, **8**, 340.
48. Jacob,Y. and Michaels,S.D. (2008) Peering through the pore: The role of AtTPR in nuclear transport and development. *Plant Signal Behav.*, **3**, 62–64.
49. Deal,R.B. and Henikoff,S. (2011) The INTACT method for cell type-specific gene expression and chromatin profiling in *Arabidopsis thaliana*. *Nat. Protoc.*, **6**, 56–68.
50. Deal,R.B. and Henikoff,S. (2010) A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Dev. Cell*, **18**, 1030–1040.
51. Di Giammartino,D.C., Nishida,K. and Manley,J.L. (2011) Mechanisms and consequences of alternative polyadenylation. *Mol. Cell*, **43**, 853–866.
52. Proudfoot,N.J. (2011) Ending the message: poly(A) signals then and now. *Genes Dev.*, **25**, 1770–1782.
53. Hilgers,V., Perry,M.W., Hendrix,D., Stark,A., Levine,M. and Haley,B. (2011) Neural-specific elongation of 3' UTRs during *Drosophila* development. *Proc. Natl Acad. Sci. USA*, **108**, 15864–15869.
54. Fu,Y., Sun,Y., Li,Y., Li,J., Rao,X., Chen,C. and Xu,A. (2011) Differential genome-wide profiling of tandem 3′ UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res.*, **21**, 741–747.