Published online 12 November 2010 *Nucleic Acids Research, 2011, Vol. 39, No. 6* ***2393–2403***
doi:10.1093/nar/gkq1158

Expression of distinct RNAs from 3′ untranslated regions

Tim R. Mercer[1], Dagmar Wilhelm[1], Marcel E. Dinger[1], Giulia Soldà[1,2], Darren J. Korbie[1], Evgeny A. Glazov[1,3], Vy Truong[1], Maren Schwenke[1], Cas Simons[1,4], Klaus I. Matthaei[5,6], Robert Saint[7,8], Peter Koopman[1] and John S. Mattick[1,*]

[1]Institute for Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia, [2]Dipartimento di Biologia e Genetica per le Scienze Mediche, Università degli Studi di Milano, Italy, [3]Diamantina Institute for Cancer, Immunology and Metabolic Medicine, The University of Queensland, Brisbane, QLD 4102, [4]Queensland Facility for Advanced Bioinformatics, The University of Queensland, Brisbane, QLD 4072, [5]John Curtin School of Medical Research, The Australian National University, Canberra, ACT 2601, Australia, [6]Stem Cell Unit, Department of Anatomy, King Saud University, Riyadh, Saudi Arabia, [7]Research School of Biological Sciences, The Australian National University, Canberra, ACT 2601 and [8]Faculty of Science, The University of Melbourne, Parkville VIC 3010, Australia

Received October 22, 2010; Revised and Accepted October 26, 2010

ABSTRACT

The 3′ untranslated regions (3′UTRs) of eukaryotic genes regulate mRNA stability, localization and translation. Here, we present evidence that large numbers of 3′UTRs in human, mouse and fly are also expressed separately from the associated protein-coding sequences to which they are normally linked, likely by post-transcriptional cleavage. Analysis of CAGE (capped analysis of gene expression), SAGE (serial analysis of gene expression) and cDNA libraries, as well as microarray expression profiles, demonstrate that the independent expression of 3′UTRs is a regulated and conserved genome-wide phenomenon. We characterize the expression of several 3′UTR-derived RNAs (uaRNAs) in detail in mouse embryos, showing by *in situ* hybridization that these transcripts are expressed in a cell- and subcellular-specific manner. Our results suggest that 3′UTR sequences can function not only in *cis* to regulate protein expression, but also intrinsically and independently in *trans*, likely as noncoding RNAs, a conclusion supported by a number of previous genetic studies. Our findings suggest novel functions for 3′UTRs, as well as caution in the use of 3′UTR sequence probes to analyze gene expression.

INTRODUCTION

The 3′ untranslated regions (3′UTRs) of messenger RNAs (mRNAs) affect the expression of eukaryotic genes by regulating mRNA translation, stability and subcellular localization (1). 3′UTRs are typically defined by cDNA cloning, which shows they are contiguous with the upstream protein-coding region in the mRNA. The length of 3′UTRs has undergone a massive expansion during metazoan evolution, with annotated 3′UTRs in human and mouse rivaling the average size of protein-coding sequences and in some cases exceeding 10 kb (2,3). Furthermore, 3′UTRs are highly conserved and contain some of the most conserved elements within the mammalian genome (4). Together, these observations suggest that 3′UTRs have assumed an increasingly important role in the evolution of the eukaryotic genome.

The control of mRNA expression by 3′UTRs is mediated by *trans*-acting factors, including RNA-binding proteins and microRNAs (miRNAs), which interact with *cis*-regulatory elements within the 3′UTR (1). The post-transcriptional regulation mediated by 3′UTRs is crucial for the correct spatial and temporal expression of the protein encoded by the mRNA. Indeed, the importance of regulation by 3′UTRs was recently highlighted by the finding that 3′UTRs are reduced in length in proliferating cells, which in some cases was shown to mediate an increased expression of the associated mRNA (5). Interestingly, the analysis of transcription

*To whom correspondence should be addressed. Tel: +61 7 3346 2079; Fax: +61 7 3346 2101; Email: j.mattick@uq.edu.au

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

start sites has suggested that transcription may also be initiated from within 3′UTR sequences, and therefore act as a source of independent transcripts (6) that may exhibit expression patterns different from their upstream protein-coding sequences.

Here, we show that the 5′ termini of many RNAs map within 3′UTRs of genes in human, mouse and fly, and verify the separate and developmentally regulated expression of 3′UTR-associated RNAs (which we have termed uaRNAs) by a range of *in silico* and molecular biology approaches, including *in situ* hybridization (ISH). Furthermore, we present evidence that a portion of these distinctively expressed 3′UTRs arises by post-transcriptional processing rather than new transcription initiation. Our results, supported by previous genetic studies of several individual genes, suggest that there is *trans*-acting embedded genetic information in 3′UTRs with potential biological function.

MATERIALS AND METHODS

Capped analysis gene expression/serial analysis of gene expression analysis

Analyses were performed using RefSeq (7) gene annotations and the hg18, mm8 and dm3 genome assemblies provided within the UCSC Genome Browser (8). Human and mouse capped analysis gene expression (CAGE) retrieved from RIKEN (http://fantom3.gsc.riken.jp/) and fruity fly Serial Analysis of Gene Expression (SAGE) tags retrieved from MachiBase (9) were mapped to the genome with ZOOM requiring exact and unique matches (10). Syntenic locations of mouse 3′UTR CAGE tags in the human genome were identified using the LiftOver utility (8). Mouse CAGE tags that mapped to the same site as human CAGE tags were defined as conserved.

Full-length cDNA analysis

Full-length human and mouse cDNA sequences were retrieved from RIKEN (http://fantom3.gsc.riken.jp/). Putative uaRNAs were identified by intersecting 5′ cDNA coordinates with RefSeq-annotated 3′UTRs. The CRITICA algorithm (11) was used to identify non-protein-coding from the RIKEN FANTOM3 full-length mouse cDNA library as described previously (12).

UaRNA transcription initiation analysis

Deep sequencing tags derived from H3K4me1, H3K4me2, H3K4me3 and H3K27ac and RNAPII immunoprecipitation for resting CD4+ cells (13) were obtained from the NCBI short read archive (accession ID SRA000234 and SRA000287) and mapped to the human genome (hg18) with ZOOM requiring exact and unique matches (10). To determine enrichment of chromatin marks with uaRNA or mRNA initiation sites, the relative mapping position of sequencing tags to the nucleotide associated with the highest CAGE tag frequency within the 3′UTR or promoter was plotted over a ±50-nt window. CAGE tags spanning exon–exon junctions (EEJs) were identified by mapping tags without a perfect match to the genome to EEJ sequences, which comprise 20 nt on either side of the splice site, located within RefSeq-annotated 3′UTRs.

CAGE expression analysis

To determine the dynamic expression of 3′UTR CAGE tags across eight mouse tissues (embryo, lung, liver, visual cortex, somatosensory cortex, cerebellum and hippocampus) (14) and six time points during the differentiation of the human THP1 myelomonocytic leukemia cell line (15), we summed the total normalized CAGE tag frequency for each 3′UTR. The 500 genes that contained the highest frequency of 3′UTR CAGE tags were clustered using the Cluster utility (16). For human genes, 3′UTR CAGE tag frequency was normalized to the median across the time series. CAGE tag frequencies were log transformed and visualized as a heat map. CAGE tag frequencies in 3′UTRs were compared to the CAGE tag frequency in the promoter for the gene subset. Promoter expression levels were defined as the sum of CAGE tags within the promoter region (±50-nt window around RefSeq-annotated transcription start site). The ratio of promoter and 3′UTR expression levels were calculated and visualized as a heat map alongside the expression clusters.

In situ hybridization

Section *in situ* hybridization (ISH) on paraffin-embedded, sectioned at 7 μm, whole-mouse embryos was performed as described previously (17). The genomic coordinates and length of the different ISH probes used are shown in the Supplementary Data.

RESULTS

Identification of 3′UTRs with independent expression

To survey 3′UTRs with evidence of independent expression, we used publicly available CAGE (capped analysis of gene expression) and cDNA libraries that were generated from a wide variety of embryonic and adult mouse tissues (18,19). CAGE uses the 5′ cap of RNA transcripts to identify the first 20–25 nt of polyadenylated RNAs. We found 175 916 (13% of the total) CAGE tags in mouse and 57 400 (5.2%) CAGE tags in human mapped to 3′UTRs (Supplementary Figure S1). The difference in the proportion of 3′UTR CAGE tags mapping to 3′UTRs in mouse and human may reflect differences in the tissue sources represented in the libraries. In total, we found 4960 mouse genes and 1518 human genes contained at least one high confidence CAGE mapping site (defined by at least three tags mapping to the same 5′ nt) within their 3′UTRs (Supplementary Table S1). These genes were not enriched for any gene ontology classes, suggesting this phenomenon is a common characteristic of mammalian genomes.

With the exception of gene promoters, the density of CAGE tags was higher in 3′UTRs than in other genomic regions, being enriched 136-fold relative to mouse

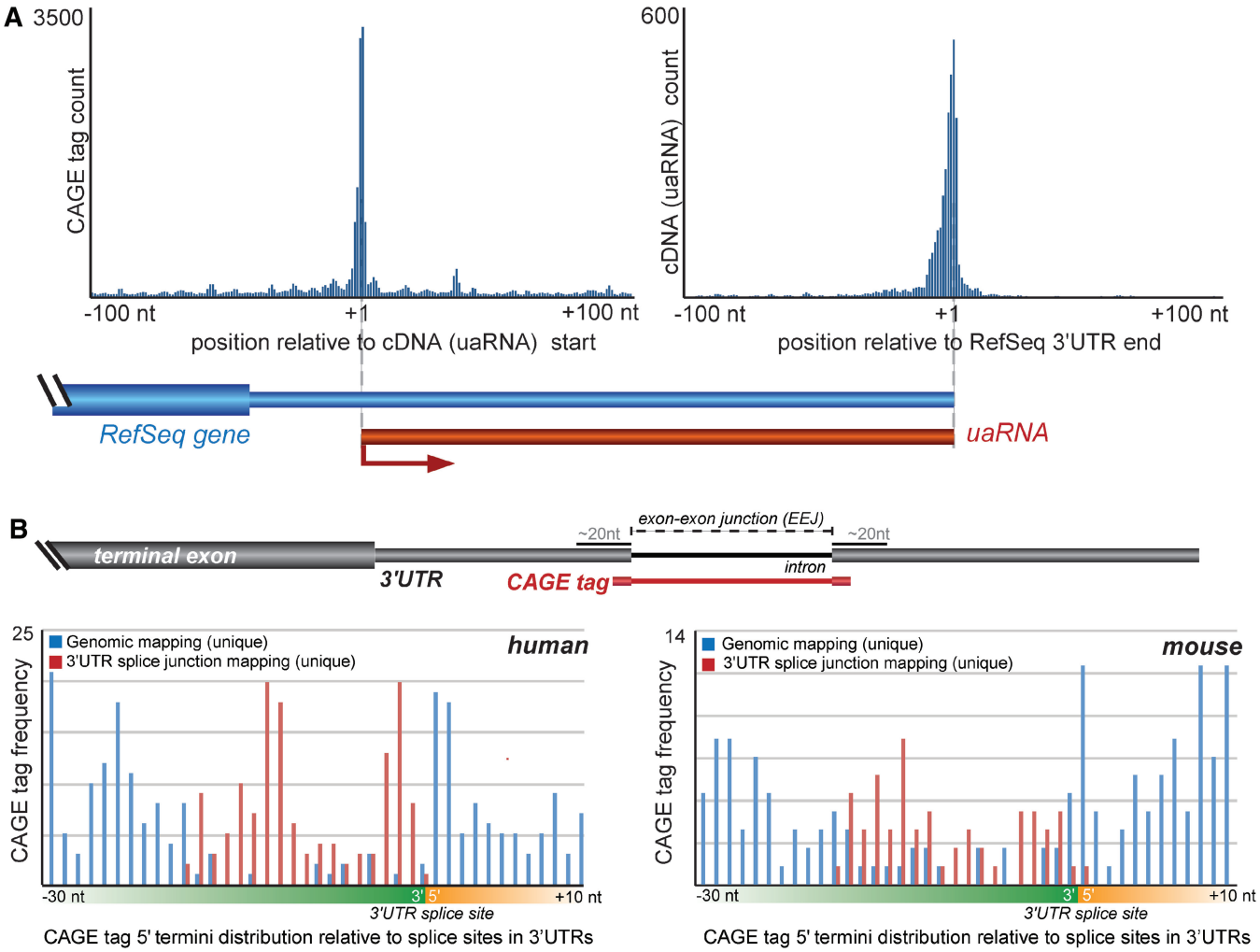

Figure 1. Post-transcriptional processing of 3′UTRs. (**A**) Full-length cDNA transcripts associated with 3′UTRs. Histogram (left) shows enrichment of CAGE tags with the 5′ termini of cDNAs mapping within 3′UTRs (i.e. uaRNAs). Histogram (right) shows correlated enrichment of the terminal regions of these cDNA transcripts with the 3′ end of the host RefSeq gene. (**B**) The top panel shows a schematic representation of the mapping strategy employed to discern CAGE tags that span exon–exon junctions (EEJs) and therefore suggest post-transcriptional processing. The lower panels show the distribution of CAGE tags mapping proximal to EEJs in human (left) and mouse (right). The frequencies of CAGE tags that map uniquely to the genome (blue) are under-represented adjacent to EEJs. This under-representation is reconciled by considering CAGE tags that map across EEJs (red).

intergenic regions and 1.6-fold relative to mouse coding regions. CAGE tags within 3′UTRs were generally organized into clusters, rather than being evenly distributed. These clusters ranged from a broad distribution over many nucleotides to a single peak, where a large number of tags mapped to a single nucleotide. Consistent with a previous report (20), we also observed enrichment for a GGG motif at the 5′ end of 3′UTR CAGE tags, indicating underlying sequence specificity (Supplementary Figure S1).

We next examined the conservation of 3′UTR CAGE sites between human and mouse. We identified 2076 homologous sites where CAGE tags occur at syntenic nucleotide positions in annotated 3′UTRs of both species, which accounted for ~20% of total 3′UTR CAGE tags (Supplementary Table S2). Compared to non-conserved sites, these conserved sites are enriched (2.7-fold, $P < 0.01$ *t*-test) for high-frequency CAGE tag mappings. Although we did not observe greater evolutionary conservation of the sequences at syntenic 3′UTR CAGE mapping sites, we did observe a distinct peak of conservation ~40 nt downstream of CAGE sites consistent with a previous report (20). We were unable to detect any enriched motifs within the conserved 3′UTR CAGE mapping sites.

To independently verify the occurrence of transcripts derived from 3′UTRs, we also examined full-length cDNA sequences in mouse. We identified 3718 full-length mouse cDNAs (~3.6% of total cDNAs) whose 5′ end maps within 2766 3′UTRs, and 1227 (33%) of these had 5′ start sites directly supported by a CAGE cluster in sense direction within ±50 bp (Figure 1A, Supplementary Table S3). Furthermore, 92% of these transcripts shared the polyadenylation site with the host RefSeq gene, suggesting that they are not 5′ ends of longer transcripts that extend past the end of the RefSeq gene (Figure 1A). To assess the potential for the 3′UTR transcripts to encode proteins or polypeptides, we analyzed their sequences

using the CRITICA algorithm (11). CRITICA reported that only 2.9% (108/3718) of 3′UTRs were likely to encode proteins (Supplementary Table S3). In addition, an examination of the PeptideAtlas (21) revealed 22 of the transcripts intersected with peptide-mapping regions (16 of which substantiated CRITICA's protein-coding predictions; Supplementary Table S3). Together, these data suggest that these transcripts are, as anticipated for UTRs, predominantly noncoding.

Putative mechanisms for uaRNA biogenesis

Next, we considered possible molecular mechanisms underlying the biogenesis of uaRNAs. Given that CAGE tags indicate 5′-capped ends, it has been previously assumed that these tags correspond to RNA polymerase II (RNAPII)-dependent transcription start sites. This conclusion was supported by evidence that the sequences upstream of a small sample of 3′UTR CAGE tags could drive expression of a reporter gene and a recent annotation of human promoters found that a considerable portion occur within 3′UTRs (22). Dynamic chromatin domains have been shown to be reliable indicators of transcription start sites (23). We identified several 3′UTRs that contained signatures of dynamic chromatin domains, such as in the 3′UTRs of the *Klhl31* and *Notch1* (Supplementary Figure S2). Within mouse embryonic stem cells, we identified 19 3′UTRs that contain multiple transcription factor binding sites and 17 3′UTRs that fully encompass chromatin domains indicative of transcription initiation (i.e. H3K4me3, H3K27me3) (Supplementary Table S4).

To determine whether dynamic chromatin domains were a general feature associated with uaRNAs, we examined deep sequencing tags from chromatin immunoprecipitation using antibodies against various histone modifications that mark active promoters, including H3K4me1, H3K4me2 and H3K4me3 (13,24). Despite the instances described above, we did not observe any enrichment of modified histones with 3′UTR CAGE sites in stark contrast to strong enrichment of modified histones at conventional gene promoters (Supplementary Figure S3). Furthermore, there was no enrichment for RNAPII occupancy at 3′UTR CAGE sites. Together, these results suggest that uaRNAs generally did not originate from active transcription from RNAPII promoters located within 3′UTRs, and that they are not derived via the same mechanisms as mRNAs.

In agreement with this hypothesis, it was recently found that many CAGE tags mapping across exon–exon junctions occurred in such close proximity to the splice junction that they were unlikely to be efficiently spliced, suggesting that 5′ caps on transcripts could also result from end-modification associated with post-transcriptional and post-splicing cleavage of a longer precursor (25–28). Although introns are rare in 3′UTRs, because they are thought to trigger nonsense-mediated decay (29), we were able to identify 59 mouse and 141 human CAGE tags spanning exon–exon junctions (Figure 1B). Together with the lack of signatures of active promoters, this suggests that some uaRNAs arise as a consequence of post-transcriptional cleavage rather than conventional transcription initiation.

Comparison of the expression of 3′UTRs and coding exons

We employed custom microarrays to compare directly the expression of 138 uaRNAs and their associated protein-coding sequences in two developmental systems in mouse: embryonic stem (ES) cell differentiation, and ovary and testis formation (see 'Materials and Methods' section). We found that for 54% (74 of 138; $R^2 < 0$; Supplementary Table S5) of genes, the expression levels between 3′UTRs and the associated coding exons were discordant in one (67%) or more (33%) developmental systems.

The complex relationships between the expression of coding exons and the associated 3′UTRs are illustrated by transcriptional profiling during mouse ES cell differentiation (Supplementary Figure S4). Consistent with the conventional case where the 3′UTR is part of the same mRNA, 50% of the 93 genes expressed above background showed concordant expression ($r \geq 0.5$) of the 3′UTR and upstream protein-coding sequence (Supplementary Table S5). However, 34% of genes showed no correlation ($-0.5 < r < 0.5$) and 16% showed a negative correlation ($r \leq -0.5$). For example, the 3′UTR of *Tmpo*, a gene with roles in cell differentiation and proliferation (30), exhibits an inverse expression profile relative to the upstream coding exons (Supplementary Figure S4). Furthermore, 12 genes show increased expression of the 3′UTR compared to the CDS at specific stages during differentiation (Supplementary Table S5). There are three interpretations for these results, which are not mutually exclusive: (i) differential expression of alternatively spliced mRNAs that contain different combinations of coding exons and 3′UTRs; (ii) 3′UTRs can be transcribed independently of the coding region; or (iii) 3′UTRs are post-transcriptionally processed from the mRNA and the resulting RNAs are differentially regulated.

Developmental stage- and tissue-specific expression of uaRNAs

To determine whether the tissue-specific expression observed in the microarrays is the result of alternative splicing or distinct expression, we analyzed differential CAGE tag frequency in 3′UTRs. Because CAGE tags correspond to the 5′ RNA termini, their differential frequencies are unlikely to arise from alternative splicing. To compare the frequency of CAGE tags across samples, we examined CAGE libraries from eight different mouse tissues (14,20). We found 4491 genes containing CAGE tags within their 3′UTR. Clustering of the top 500 genes containing the highest frequency of 3′UTR CAGE tags clearly showed that there was a wide dynamic range of tissue-specific expression of uaRNAs (Figure 2A). We did not observe a correlation between the number of CAGE tags mapping to the 3′UTR (of determination; $r^2 = 0.0064$) and the number of CAGE tags mapping to the promoter of the same gene, but rather a range of ratios that varied across tissue types

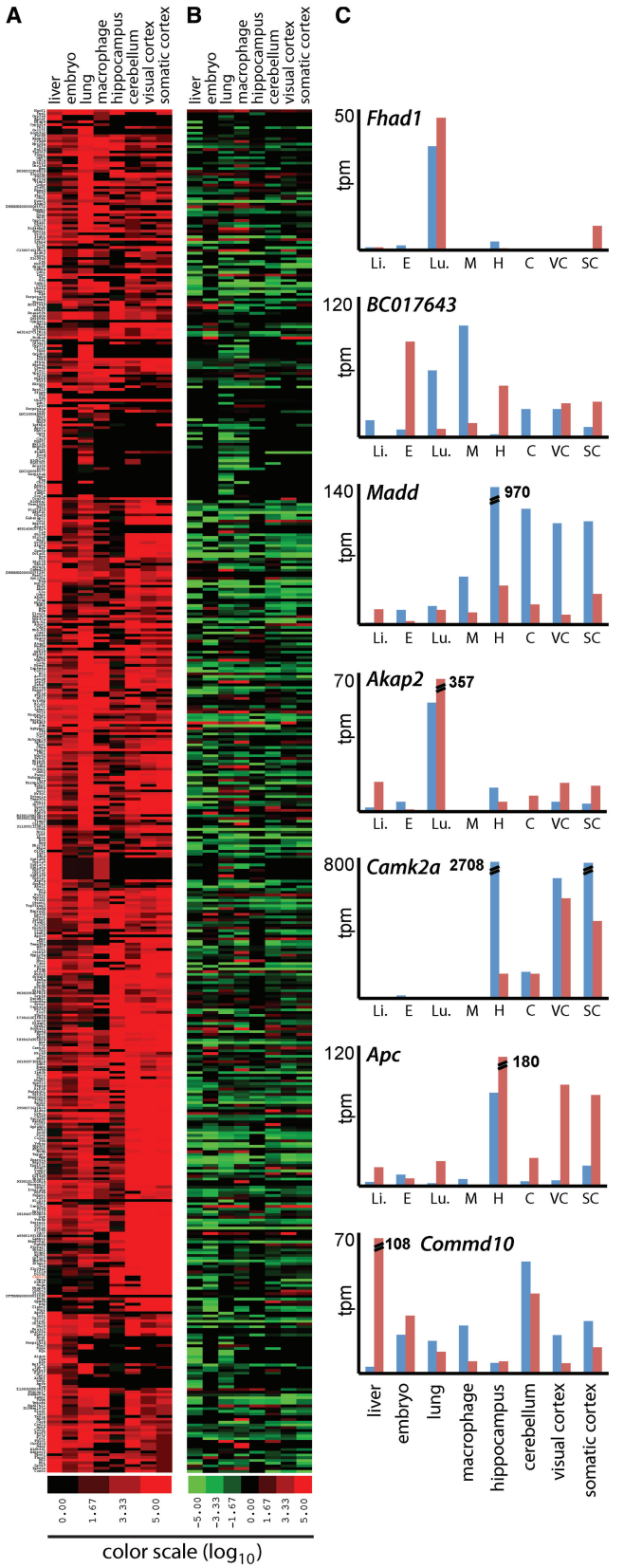

Figure 2. Specific and distinct expression profiles for 3′UTRs in mouse tissues. (**A**) Cluster analysis of the expression of the 500 genes containing the highest 3′UTR CAGE tag frequency shows the differential expression levels (no expression, black; high expression, red) of 3′UTRs in various mouse tissues. Expression level was determined as the normalized CAGE tag frequency mapping to 3′UTRs. (**B**) The diverse ratios of promoter to 3′UTR CAGE frequency (high, green; low, red) for each tissue indicates independent expression of mRNAs and 3′UTRs. (**C**) Illustrative examples indicating promoter (blue) and 3′UTR (red) CAGE frequency. tpm, tags per million; Li, liver; E, embryo; Lu, lung; M, macrophage; C, cerebellum; H, hippocampus; VC, visual cortex; SC, somatic cortex.

(Figure 2, Supplementary Table S6). This suggests uaRNA expression is a regulated process rather than a byproduct of host gene expression. In this analysis, we also observed instances where uaRNA initiation occurred preferentially at different sites in different tissues. For example, in the *Camk2a* gene (31), uaRNAs are preferentially initiated at a different site in the hippocampus than in the somato-sensory and visual cortex (Figure 3A).

We also analyzed the temporal expression pattern of uaRNAs during the differentiation of the human THP-1 myelomonocytic leukemia cell line upon stimulation with phorbol myristate acetate (PMA) at six time points between 0 and 96 h (15). We identified 2726 genes that contained CAGE tags within their 3′UTRs. Considering the expression of the top 500 genes, we again observed dynamic expression profiles that could not be accounted for by the host gene expression (Supplementary Figure S5, Supplementary Table S7). Together with the observation that several of the identified genes have important roles in macrophage activation, such as *ITGB2* (15) and *SRGN* (32), these results raise the possibility that the regulated expression of uaRNAs has biological consequences.

Independent regulated 3′UTR expression in *Drosophila melanogaster*

To examine whether uaRNAs occur outside of vertebrates, we examined *Drosophila melanogaster* SAGE libraries, which, like CAGE, define the 5′ termini of RNA transcripts. We found 27 832 SAGE tags (1.2% of the total) in four libraries (embryo, S2 cells, young and old females) that mapped within RefSeq annotated 3′UTRs of 275 genes (Supplementary Table S1). Unlike the human and mouse 3′UTR CAGE tags, we did not identify the GGG motif or the 40-nt downstream conserved region associated in the *D. melanogaster* 3′UTR SAGE tags. However, we did observe elevated evolutionary conservation downstream of 3′UTR SAGE tags (Supplementary Figure S6). During this analysis, we also noted that the 3′UTR of *oskar* (*osk*), a gene whose 3′UTR had previously been indicated to function independently to the host gene (33), showed large numbers of SAGE tags specific to adult stages (Figures 3B and S6; see 'Discussion' section). Together, these results provide independent validation for the existence of uaRNAs in an evolutionarily distant lineage, indicating their possible widespread biological role(s).

Candidate uaRNAs are specifically expressed during mouse embryogenesis

To further characterize the discordant expression observed between the coding sequence (CDS) and 3′UTR in differentiating mouse ES cells (see above), we

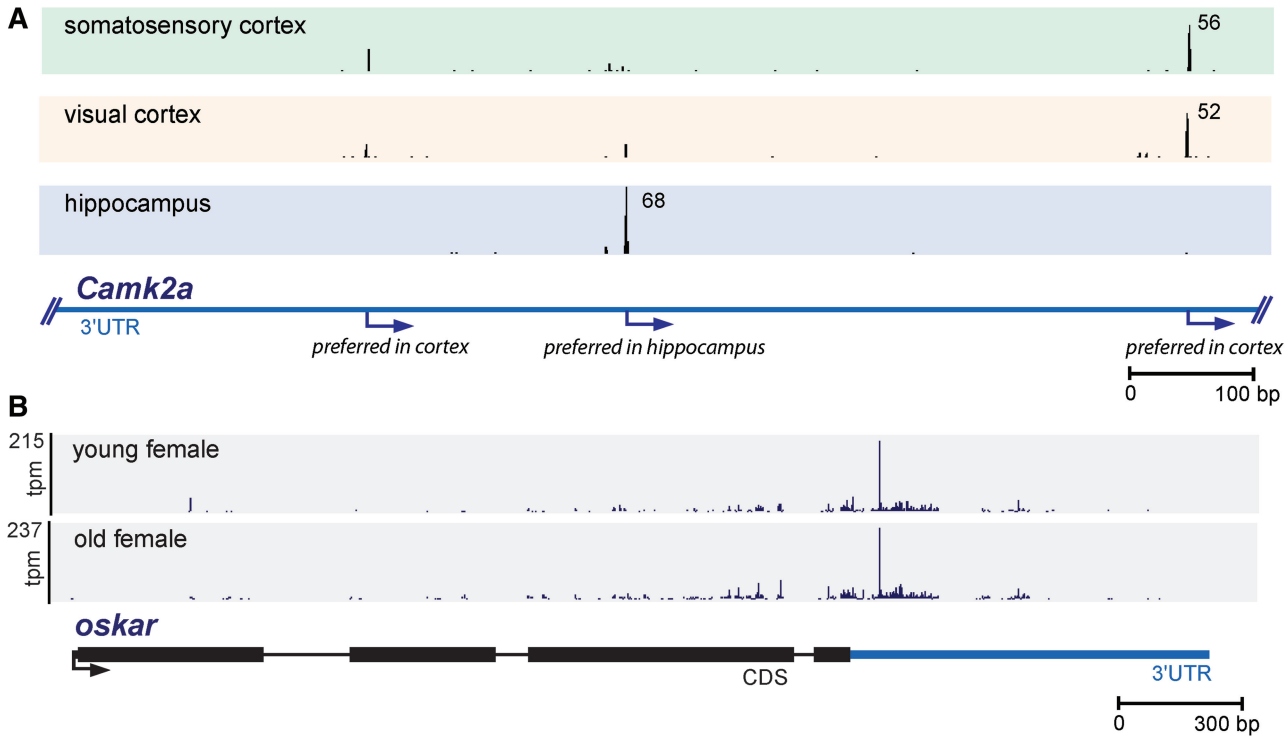

Figure 3. Illustrative examples of 3′UTR transcripts in mouse and fruit fly. (**A**) Schematic representation of CAGE tag clusters in the 3′UTR of the mouse *Camk2a* gene from somatosensory cortex (green), visual cortex (orange) and hippocampus (blue). Different sites with high frequency of CAGE tags map preferentially in hippocampus (68 tags per million; tpm), the visual cortex (52 tpm) and somato-sensory cortex (56 tpm). (**B**) Schematic representation of the fruit fly gene *oskar* showing specific SAGE tags in the 3′UTR in young and old females.

performed 5′ rapid amplification of complementary DNA ends (RACE) on six uaRNA candidates (*Col1a1*, *Mef2c*, *Mical2*, *Nfia*, *Mylk* and *Myadm*) and compared the expression of their coding regions and corresponding 3′UTRs using ISH. These candidates were selected on the basis of their discordant expression by microarray profiling, 3′UTR conservation and presence of high-confidence 3′UTR CAGE tags. The RACE analysis identified 5′ termini within the 3′UTRs of *Col1a1*, *Mef2c*, *Mical2*, *Nfia* and *Mylk* (Figure 4A; Supplementary Table S8), confirming the presence of independent transcripts for these loci in whole mouse embryos. Interestingly, in most cases, several distinct 5′ RACE products were cloned at different frequencies, suggesting that multiple uaRNAs can arise from the same 3′UTR and that these can be differentially expressed (Supplementary Table S8). To confirm that these uaRNAs were indeed encompassed within the 3′UTRs of the respective full-length mRNAs, we performed gene and strand-specific reverse transcription (RT) followed by polymerase chain reaction (PCR) of the terminal exon (see Supplementary Data). Consistent with expectations based on the RefSeq annotations for these genes, in all cases examined, we confirmed that the 3′UTRs were connected to the terminal coding exon (Figures 4B and S7). ISH in whole-mouse embryos revealed discordant expression between the CDS and 3′UTR in three of the candidates, *Col1a1*, *Nfia* and *Myadm*, which are discussed in further detail below.

The tissue and cellular expression of *Col1a1*, which encodes the procollagen type I alpha 1 chain, was detected using an RNA probe targeting the constitutive terminal coding exon and two probes targeting the 3′UTR sequences, one downstream of each of the two major 5′ RACE clusters (Figure 4A). The specificity of all three probes to *Col1a1* was confirmed by northern blot (Supplementary Figure S7). Interestingly, the most distal 3′UTR probe also detected a ∼120-nt transcript, which may represent a processed uaRNA (Supplementary Figure S7). Section ISH on whole-mouse embryos at 13.5 days *post coitum* (dpc; Figure 4C) demonstrated that all three regions are expressed at sites of chondrogenesis, such as the otic vesicle and the developing ribs. The expression of the coding region persisted during ossification, whereas no expression of the uaRNAs was detected. Moreover, the most distal uaRNA probe revealed distinct expression of the targeted sequence in the dorsal root ganglia, i.e. expression was not detected in these areas by the coding probe or the other uaRNA probe. Higher magnification of the dorsal root ganglia indicated that the uaRNA is localized to the nucleus (Figure 4C). Together, the different expression profiles of the *Col1a1* exons and the uaRNAs verified that uaRNAs are distinctly expressed in a tissue-specific manner. To confirm the discordant expression profiles observed between the 3′UTR and coding regions were not due to alternative splicing, we performed ISH with two additional independent probes targeting the coding

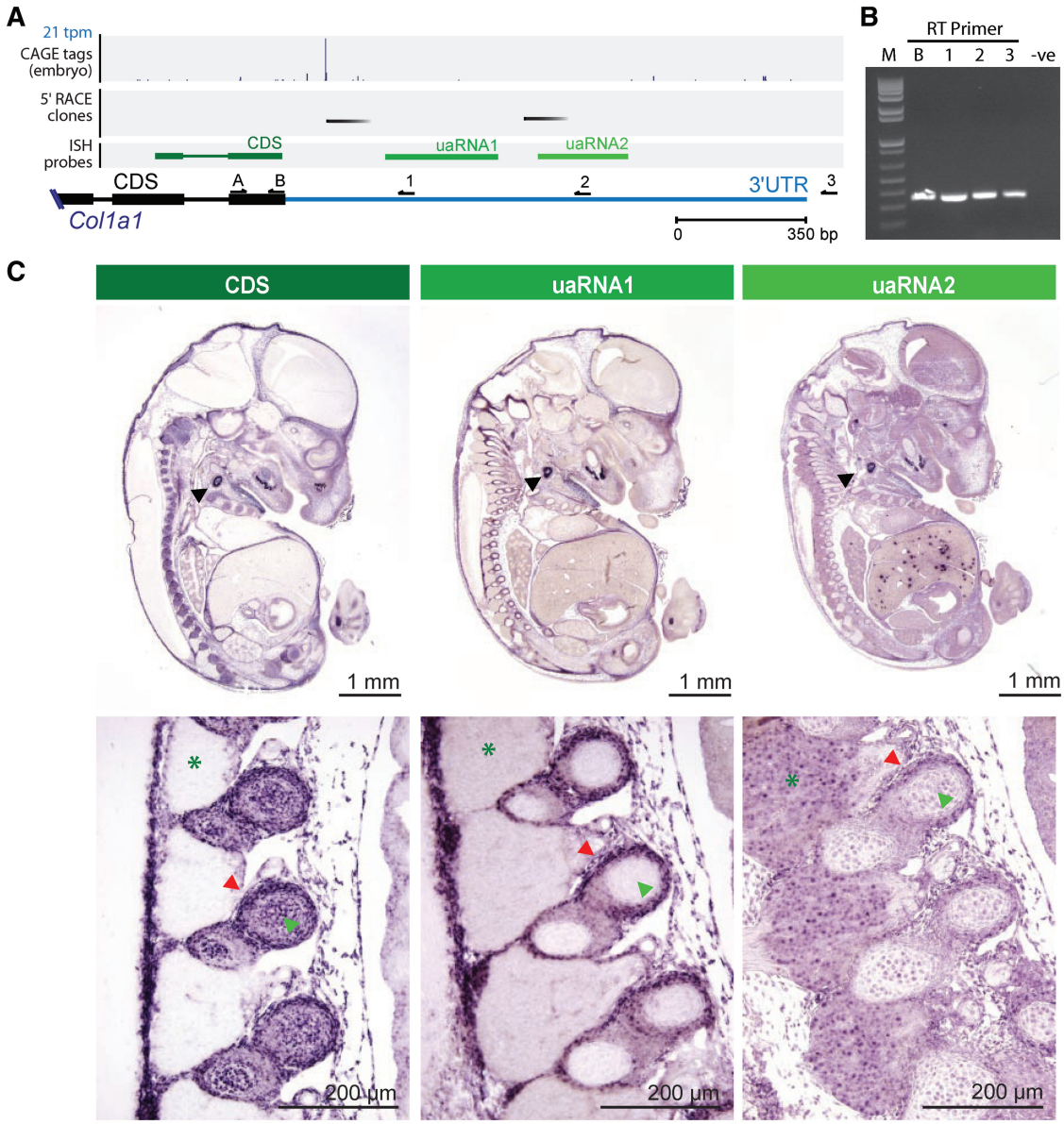

Figure 4. uaRNAs within *Col1a1* 3′UTR. (**A**) Genome browser view of *Col1a1* 3′UTR showing; histogram of CAGE distribution and density (top panel; tpm, tags per million), 5′ ends inferred from high confidence 5′ RACE products (second panel, black bars); riboprobes used in *in situ* hybridization (third panel; green bars); *Col1a1* annotated coding sequence (CDS; black bar) and 3′UTR (blue bar); reverse transcriptase (RT, 1–3) and PCR primers (A and B) used in (B) (black arrows). (**B**) Confirmation of 3′UTR annotation by RT using primers B, 1, 2 or 3 followed by PCR using primers A (forward) and B (reverse). Lanes contain (from left to right) 1-kb plus ladder (M), positive CDS control primer (B), RT primers (1–3) and no RT negative control (-ve). (**C**) ISH using one riboprobe in the terminal constitutive coding exon (CDS, dark green) of the *Col1a1* gene and two probes (uaRNA1 and uaRNA2, light green) corresponding to 3′UTR sequences, one downstream of each of the two 5′ RACE clusters. All three probes exhibit expression at sites of chondrogenesis, such as the otic vesicle (upper panel, black arrowheads) and the developing ribs (lower panel, red arrowheads). Expression of the coding region is apparent during ossification of vertebrae, in contrast to both uaRNAs whose expression is absent (lower panel, green arrowheads). The uaRNA2 probe detects expression in the the dorsal root ganglia (lower panel, green asterisks); this expression is not detected by the CDS or uaRNA1 probes. Higher magnification of the dorsal root ganglia shows that the uaRNA expression is localized to the nucleus (lower panel, green asterisk).

region. In all cases, we observed the same expression pattern consistent with the first coding probe, thereby discounting alternative splicing as a cause for the discordant expression between 3′UTRs and coding exons (Supplementary Figure S8).

Nfia encodes a transcription factor required for brain development (34), and has a highly conserved 7.5-kb 3′UTR. Comparison of section ISH using a probe targeting the terminal exon of the CDS and a probe targeting the 3′UTR (uaRNA) showed discordant expression in the developing forebrain (Figure 5A), a region where *Nfia* plays important roles in regulating gene expression (35). We also detected *Nfia* CDS expression in interstitial cells of developing testes (Figure 5A). The function of *Nfia* during testis differentiation is unknown, but it has been reported that homozygous *Nfia*-null male mice are sterile (34). In contrast to the CDS, we detected expression of the uaRNA within the testis cords (Figure 5A), rather than

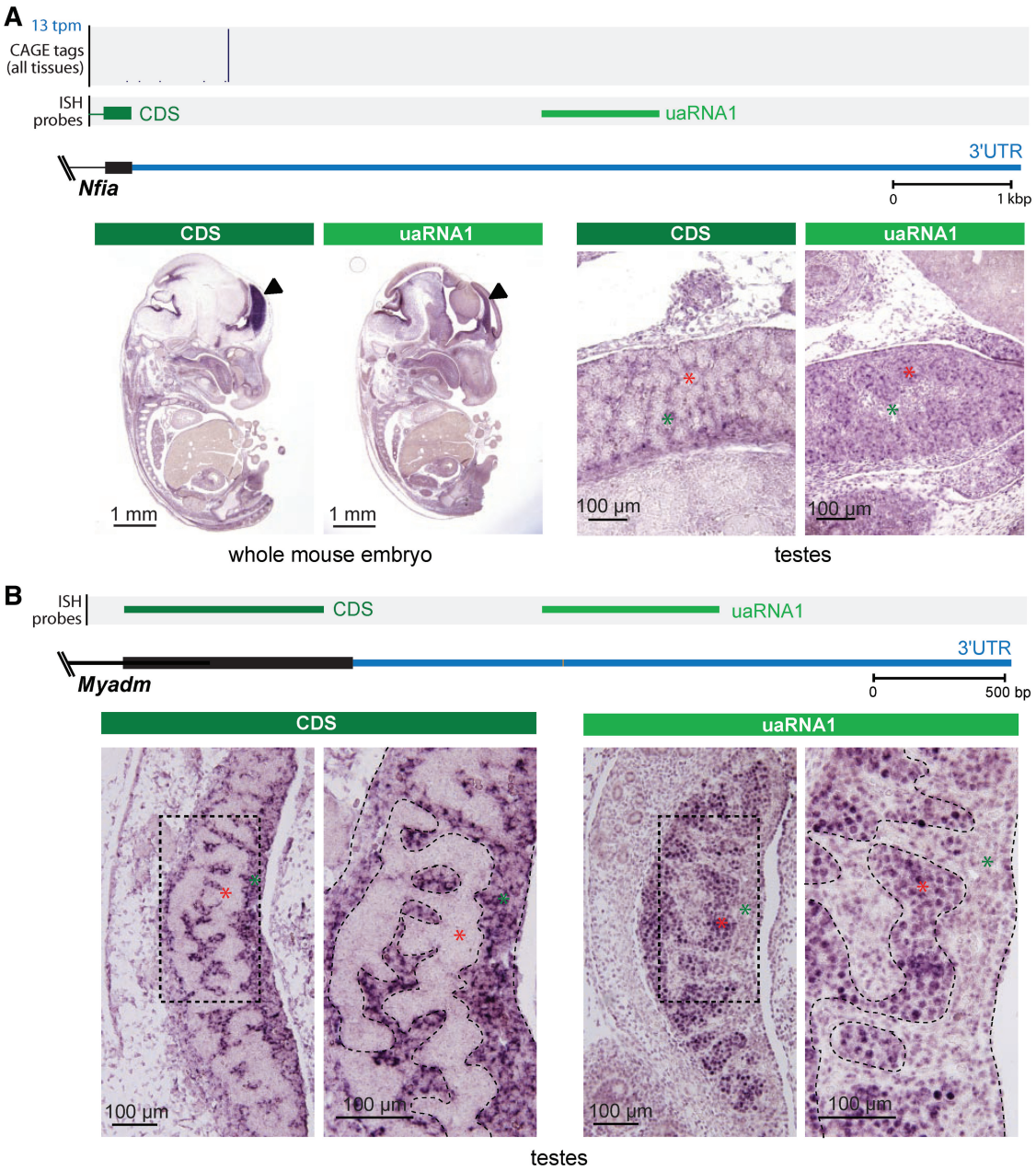

Figure 5. uaRNAs within the *Nfia* and *Myadm* 3′UTRs. (**A**) (Top panel) Genomic context of *Nfia* 3′UTR showing annotated coding sequence (CDS, black) and 3′UTR (blue), histogram of CAGE distribution and density (tpm, tags per million), and riboprobes used in ISH that target the terminal constitutive coding exons (dark green; CDS) and 3′UTR (light green; uaRNA1). (Bottom panel) Section ISH showing the *Nfia* CDS probe detecting expression in interstitial cells (green asterisks) of 12.5 dpc testes (marker) and the uaRNA probe detecting expression within the testis cords (red asterisks). (**B**) (Top panel) Genomic context of *Myadm* 3′UTR showing annotated coding sequence (CDS, black) and 3′UTR (blue), riboprobes used in ISH that target the terminal constitutive coding exon (dark green; CDS) and 3′UTR (light green; uaRNA1). (Bottom panel) Section ISH showing the *Myadm* CDS probe detecting expression in the cytoplasm of interstitial cells (green asterisks) in the developing testis at 12.5 dpc and the uaRNA probe detecting expression in the nuclei of Sertoli cells and germ cells in the testis cords (red asterisks). High *Myadm* expression was not detected by either the CDS or the uaRNA probe elsewhere in the embryo (see Supplementary Figure S8).

the interstitium, suggesting that the *Nfia* uaRNA plays a role independent from the NFIA protein.

The gene encoding myeloid-associated differentiation marker *Myadm* (36) showed decoupled expression of the CDS and uaRNA in the developing gonad (Figure 5B). ISH using independent CDS and uaRNA probes showed the terminal coding exon of *Myadm* was expressed in interstitial cells of the developing testis, whereas the uaRNA was detected in Sertoli cells and germ cells within testis cords. Moreover, high magnification examination of the ISH images revealed different subcellular localizations, with the coding exon being detected in the cytoplasm of interstitial cells, and the uaRNA localized in the nuclei of germ and Sertoli cells (Figure 5B).

Interestingly, we detected only low levels of *Myadm* uaRNA expression in testes before 12.5 dpc or in ovaries at any stage investigated, suggesting that the expression of this uaRNA was highly tissue- and developmentally specific. Furthermore, northern analysis identified a testis-specific small uaRNA transcript of ~140 nt (Supplementary Figure S7), suggesting the uaRNA is processed into a smaller stable RNA.

One general observation derived from the ISH data was that for each of the candidate genes, specific tissues could be identified where the 3′UTR probe detected expression, but the CDS probe did not. As already discussed above, the absence of indicators of active transcription suggest that uaRNAs are not transcribed independently of the associated mRNA, but rather are a cleaved product of the full-length transcript. Therefore, the presence of 3′UTR expression in the absence of CDS expression can be explained by rapid degradation of the upstream transcript containing the CDS. To examine this possibility, we performed quantitative RT-PCR (qRT-PCR) on the CDS and 3′UTR regions of *Col1a1*, *Nfia* and *Myadm* on total RNA extracted from the tissues and cell types identified by ISH as showing exclusive 3′UTR expression. Unfortunately, we were unable to extract sufficient RNA from the developing spine for analysis of *Col1a1*. However, for *Nfia* and *Myadm*, where the ISH showed elevated 3′UTR expression relative to the CDS, the qRT-PCR revealed a similar tissue-specific increase in 3′UTR expression in brain and testis, respectively (Supplementary Figure S7). In addition to independently validating the ISH data, this result supports a model where the 3′UTR is cleaved to give rise to a uaRNA, while the upstream coding region is degraded.

DISCUSSION

The conventional understanding of 3′UTRs is that they exclusively operate in *cis*, via regulatory proteins and microRNAs, to control the translation, stability and localization of mRNAs. This concept has remained unquestioned largely due to the lack of evidence to the contrary and the inherent difficulties of discerning *trans*-acting roles of 3′UTRs. Our data show that the distinct expression of 3′UTR sequences is a widespread phenomenon that is developmentally regulated, with stage-, tissue- and subcellular-specific expression. Together, these results suggest that 3′UTR sequences can fulfill biological roles different from their normally associated mRNA.

The independent function we propose for uaRNAs is supported by a number of published observations. As previously noted, the *oskar* 3′UTR in *Drosophila* has been shown to rescue an oogenesis defect in *oskar* null-mutants, independently of the oskar protein (33). Our SAGE tag analysis supported the independent expression of the *oskar* 3′UTR, consistent with the idea that the *oskar* 3′UTR is expressed and functions independently *in vivo*. Similarly, a number of additional reports have shown that the 3′UTRs of troponin I, tropomyosin, alpha-cardiac actin, ribonucleotide reductase, DM protein kinase and prohibitin genes can act *in trans* to control cell proliferation and differentiation in the absence of associated coding-regions (37–41). Such roles may also contribute toward disease etiology. For example, the 3′UTR of the DM protein kinase gene, which is involved in myotonic dystrophy, inhibits the differentiation of C2C12 myoblasts (41) and the ectopic expression of the prohibitin 3′UTR has been shown to block cell cycle progression and contains characteristic mutations in breast cancer-derived cells (42).

Despite such examples showing 3′UTRs acting in *trans*, the challenge remains to understand the mechanism by which they are generated and by which they act. The mechanism of post-transcriptional cleavage is unknown, but could potentially involve specific RNA-binding proteins or *trans*-acting RNAs that can target cleavage enzymes in a sequence-specific manner similar to the targeting of chromatin-modifying complexes by long ncRNAs and targeting of RISC enzymes by miRNAs (43). In addition, a non-transcriptionally linked capping enzyme was recently discovered in the cytoplasm of several human cell lines that may be responsible for the post-transcriptional capping of cleaved transcripts (44). The function of uaRNAs may, at least in part, be inferred from known 3′UTR function. For example, uaRNAs may act as decoys to titrate *trans*-acting factors and thereby fine-tune their regulatory function, similar to the ncRNA *IpsPS1*, which was recently shown to sequester the microRNA *miR-399* in *Arabidopsis thaliana* (45). Alternatively, they may act as a scaffold to localize proteins into regulatory complexes that are required even in the absence of the associated mRNA, as suggested by the observation that the *oskar* 3′UTR is necessary for trafficking and accumulation of Staufen from nurse cells to the oocyte during oogenesis (33). Therefore, it remains to be determined what specific information uaRNAs contribute toward gene regulation and the advantage gained by having such information either linked to mRNA or decoupled in other contexts.

Regardless of novel and unexpected roles of uaRNA transcripts, the existence and distinct expression of uaRNAs should be considered in future studies that assume the 3′UTR expression to be coincident with the associated mRNAs. Within this study, we demonstrated a number of genes that exhibited discordant expression patterns between their CDSs and 3′UTRs in both microarray and ISH (46). Given these examples, results gained using probes targeted to 3′UTRs should be interpreted with care.

In summary, our findings enhance the current understanding of 3′UTR sequences and their role in regulating differentiation and developmental processes. This study and recent observations that 3′UTRs have undergone rapid expansion in eukaryotic evolution suggest more sophisticated functionality inherent in 3′UTR sequences than previously suspected. Moreover, the finding that 3′UTRs can be independently expressed as developmentally regulated ncRNAs further blurs the distinction between coding and noncoding RNAs (47) and serves as a reminder that the traditional concept of the gene is becoming increasingly outmoded (48,49), requiring

a reassessment of our understanding of the genomic programming of complex organisms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Australian Research Council/University of Queensland co-sponsored Federation Fellowship (FF0561986; to J.S.M.); an Australian Research Council Discovery Project Grant (DP0879913; to D.W.); National Health and Medical Research Council of Australia Career Development Awards (CDA631542; to M.E.D., CDA519937; to D.W.); a Queensland Government Department of Employment, Economic Development and Innovation Smart Futures Fellowship (to M.E.D.); the University of Milan (to G.S.). Funding for open access charge: The University of Queensland.

Conflict of interest statement. None declared.

REFERENCES

1. Kuersten,S. and Goodwin,E.B. (2003) The power of the 3′ UTR: translational control and development. *Nat. Rev. Genet.*, **4**, 626–637.
2. Frith,M.C., Pheasant,M. and Mattick,J.S. (2005) The amazing complexity of the human transcriptome. *Eur. J. Hum. Genet.*, **13**, 894–897.
3. Mazumder,B., Seshadri,V. and Fox,P.L. (2003) Translational control by the 3′-UTR: the ends specify the means. *Trends Biochem. Sci.*, **28**, 91–98.
4. Siepel,A., Bejerano,G., Pedersen,J.S., Hinrichs,A.S., Hou,M., Rosenbloom,K., Clawson,H., Spieth,J., Hillier,L.W., Richards,S. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
5. 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.
6. Furuno,M., Pang,K.C., Ninomiya,N., Fukuda,S., Frith,M.C., Bult,C., Kai,C., Kawai,J., Carninci,P., Hayashizaki,Y. *et al.* (2006) Clusters of internally primed transcripts reveal novel long noncoding RNAs. *PLoS Genet.*, **2**, e37.
7. Pruitt,K.D., Tatusova,T. and Maglott,D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
8. Kuhn,R.M., Karolchik,D., Zweig,A.S., Wang,T., Smith,K.E., Rosenbloom,K.R., Rhead,B., Raney,B.J., Pohl,A., Pheasant,M. *et al.* (2009) The UCSC Genome Browser Database: update 2009. *Nucleic Acids Res.*, **37**, D755–D761.
9. Ahsan,B., Saito,T.L., Hashimoto,S., Muramatsu,K., Tsuda,M., Sasaki,A., Matsushima,K., Aigaki,T. and Morishita,S. (2009) MachiBase: a Drosophila melanogaster 5′-end mRNA transcription database. *Nucleic Acids Res.*, **37**, D49–D53.
10. Lin,H., Zhang,Z., Zhang,M.Q., Ma,B. and Li,M. (2008) ZOOM! Zillions of oligos mapped. *Bioinformatics*, **24**, 2431–2437.
11. Badger,J.H. and Olsen,G.J. (1999) CRITICA: coding region identification tool invoking comparative analysis. *Mol. Biol. Evol.*, **16**, 512–524.
12. Mercer,T.R., Dinger,M.E., Sunkin,S.M., Mehler,M.F. and Mattick,J.S. (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA*, **105**, 716–721.
13. Wang,Z., Zang,C., Rosenfeld,J.A., Schones,D.E., Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Peng,W., Zhang,M.Q. *et al.* (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.*, **40**, 897–903.
14. Valen,E., Pascarella,G., Chalk,A., Maeda,N., Kojima,M., Kawazu,C., Murata,M., Nishiyori,H., Lazarevic,D., Motti,D. *et al.* (2009) Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Res.*, **19**, 255–265.
15. Suzuki,H., Forrest,A.R., van Nimwegen,E., Daub,C.O., Balwierz,P.J., Irvine,K.M., Lassmann,T., Ravasi,T., Hasegawa,Y., de Hoon,M.J. *et al.* (2009) The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.*, **41**, 553–562.
16. Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.
17. Wilhelm,D., Hiramatsu,R., Mizusaki,H., Widjaja,L., Combes,A.N., Kanai,Y. and Koopman,P. (2007) SOX9 regulates prostaglandin D synthase gene transcription in vivo to ensure testis development. *J. Biol. Chem.*, **282**, 10553–10560.
18. Kawaji,H., Kasukawa,T., Fukuda,S., Katayama,S., Kai,C., Kawai,J., Carninci,P. and Hayashizaki,Y. (2006) CAGE Basic/Analysis Databases: the CAGE resource for comprehensive promoter analysis. *Nucleic Acids Res.*, **34**, D632–D636.
19. Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C., Maeda,N., Oyama,R., Ravasi,T., Lenhard,B., Wells,C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
20. Carninci,P., Sandelin,A., Lenhard,B., Katayama,S., Shimokawa,K., Ponjavic,J., Semple,C.A., Taylor,M.S., Engstrom,P.G., Frith,M.C. *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.*, **38**, 626–635.
21. Deutsch,E.W., Lam,H. and Aebersold,R. (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.*, **9**, 429–434.
22. Trinklein,N.D., Karaoz,U., Wu,J., Halees,A., Force Aldred,S., Collins,P.J., Zheng,D., Zhang,Z.D., Gerstein,M.B., Snyder,M. *et al.* (2007) Integrated analysis of experimental data sets reveals many novel promoters in 1% of the human genome. *Genome Res.*, **17**, 720–731.
23. Bernstein,B.E., Mikkelsen,T.S., Xie,X., Kamal,M., Huebert,D.J., Cuff,J., Fry,B., Meissner,A., Wernig,M., Plath,K. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
24. Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
25. Fejes-Toth,K., Sotirova,V., Sachidanandam,R., Assaf,G., Hannon,G.J., Kapranov,P., Foissac,S., Willingham,A.T., Duttagupta,R., Dumais,E. *et al.* (2009) Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature*, **457**, 1028–1032.
26. Ni,T., Corcoran,D.L., Rach,E.A., Song,S., Spana,E.P., Gao,Y., Ohler,U. and Zhu,J. (2010) A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods*, **7**, 521–527.
27. Plessy,C., Bertin,N., Takahashi,H., Simone,R., Salimullah,M., Lassmann,T., Vitezic,M., Severin,J., Olivarius,S., Lazarevic,D. *et al.* (2010) Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. *Nat. Methods*, **7**, 528–534.
28. Mercer,T.R., Dinger,M.E., Bracken,C.P., Kolle,G., Szubert,J.M., Korbie,D.J., Askarian-Amiri,M.E., Gardiner,B.B., Goodall,G.J., Grimmond,S.M. *et al.* (2010) Regulated post-transcriptional RNA cleavage diversifies the eukaryotic transcriptome. *Genome Res.*, **20**, doi:10.1101/gr.112128.110.
29. Hong,X., Scofield,D.G. and Lynch,M. (2006) Intron size, abundance, and distribution within untranslated regions of genes. *Mol. Biol. Evol.*, **23**, 2392–2404.
30. Dorner,D., Vlcek,S., Foeger,N., Gajewski,A., Makolm,C., Gotzmann,J., Hutchison,C.J. and Foisner,R. (2006) Lamina-associated polypeptide 2alpha regulates cell cycle

progression and differentiation via the retinoblastoma-E2F pathway. *J. Cell. Biol.*, **173**, 83–93.
31. Wayman,G.A., Lee,Y.S., Tokumitsu,H., Silva,A. and Soderling,T.R. (2008) Calmodulin-kinases: modulators of neuronal development and plasticity. *Neuron*, **59**, 914–931.
32. Niemann,C.U., Kjeldsen,L., Ralfkiaer,E., Jensen,M.K. and Borregaard,N. (2007) Serglycin proteoglycan in hematologic malignancies: a marker of acute myeloid leukemia. *Leukemia*, **21**, 2406–2410.
33. Jenny,A., Hachet,O., Zavorszky,P., Cyrklaff,A., Weston,M.D., Johnston,D.S., Erdelyi,M. and Ephrussi,A. (2006) A translation-independent role of oskar RNA in early Drosophila oogenesis. *Development*, **133**, 2827–2833.
34. das Neves,L., Duchala,C.S., Tolentino-Silva,F., Haxhiu,M.A., Colmenares,C., Macklin,W.B., Campbell,C.E., Butz,K.G. and Gronostajski,R.M. (1999) Disruption of the murine nuclear factor I-A gene (Nfia) results in perinatal lethality, hydrocephalus, and agenesis of the corpus callosum. *Proc. Natl Acad. Sci. USA*, **96**, 11946–11951.
35. Shu,T., Butz,K.G., Plachez,C., Gronostajski,R.M. and Richards,L.J. (2003) Abnormal development of forebrain midline glia and commissural projections in Nfia knock-out mice. *J. Neurosci.*, **23**, 203–212.
36. Pettersson,M., Dannaeus,K., Nilsson,K. and Jonsson,J.I. (2000) Isolation of MYADM, a novel hematopoietic-associated marker gene expressed in multipotent progenitor cells and up-regulated during myeloid differentiation. *J. Leukoc. Biol.*, **67**, 423–431.
37. Fan,H., Villegas,C., Huang,A. and Wright,J.A. (1996) Suppression of malignancy by the 3′ untranslated regions of ribonucleotide reductase R1 and R2 messenger RNAs. *Cancer Res.*, **56**, 4366–4369.
38. Rastinejad,F. and Blau,H.M. (1993) Genetic complementation reveals a novel regulatory role for 3′ untranslated regions in growth and differentiation. *Cell*, **72**, 903–917.
39. Rastinejad,F., Conboy,M.J., Rando,T.A. and Blau,H.M. (1993) Tumor suppression by RNA from the 3′ untranslated region of alpha-tropomyosin. *Cell*, **75**, 1107–1117.
40. Jupe,E.R., Liu,X.T., Kiehlbauch,J.L., McClung,J.K. and Dell'Orco,R.T. (1996) The 3′ untranslated region of prohibitin and cellular immortalization. *Exp. Cell Res.*, **224**, 128–135.
41. Amack,J.D., Paguio,A.P. and Mahadevan,M.S. (1999) *Cis* and *trans* effects of the myotonic dystrophy (DM) mutation in a cell culture model. *Hum. Mol. Genet.*, **8**, 1975–1984.
42. Jupe,E.R., Liu,X.T., Kiehlbauch,J.L., McClung,J.K. and Dell'Orco,R.T. (1996) Prohibitin in breast cancer cell lines: loss of antiproliferative activity is linked to 3′ untranslated region mutations. *Cell Growth Differ.*, **7**, 871–878.
43. Karginov,F.V., Cheloufi,S., Chong,M.M., Stark,A., Smith,A.D. and Hannon,G.J. (2010) Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases. *Mol. Cell*, **38**, 781–788.
44. Otsuka,Y., Kedersha,N.L. and Schoenberg,D.R. (2009) Identification of a cytoplasmic complex that adds a cap onto 5′-monophosphate RNA. *Mol. Cell. Biol.*, **29**, 2155–2167.
45. Franco-Zorrilla,J.M., Valli,A., Todesco,M., Mateos,I., Puga,M.I., Rubio-Somoza,I., Leyva,A., Weigel,D., Garcia,J.A. and Paz-Ares,J. (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.*, **39**, 1033–1037.
46. Dinger,M.E., Amaral,P.P., Mercer,T.R., Pang,K.C., Bruce,S.J., Gardiner,B.B., Askarian-Amiri,M.E., Ru,K., Solda,G., Simons,C. *et al.* (2008) Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.*, **18**, 1433–1445.
47. Dinger,M.E., Pang,K.C., Mercer,T.R. and Mattick,J.S. (2008) Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput. Biol.*, **4**, e1000176.
48. Dinger,M.E., Amaral,P.P., Mercer,T.R. and Mattick,J.S. (2009) Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. *Brief Funct. Genomic Proteomic*, **8**, 407–423.
49. Gerstein,M.B., Bruce,C., Rozowsky,J.S., Zheng,D., Du,J., Korbel,J.O., Emanuelsson,O., Zhang,Z.D., Weissman,S. and Snyder,M. (2007) What is a gene, post-ENCODE? History and updated definition. *Genome Res.*, **17**, 669–681.