RESOURCE

nature structural & molecular biology

Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans

Ryan J Taft[1,8], Cas Simons[1,2,8], Satu Nahkuri[1], Harald Oey[1], Darren J Korbie[1], Timothy R Mercer[1], Jeff Holst[3,4], William Ritchie[3,4], Justin J-L Wong[3], John E J Rasko[3–5], Daniel S Rokhsar[6], Bernard M Degnan[7] & John S Mattick[1]

We have recently shown that transcription initiation RNAs (tiRNAs) are derived from sequences immediately downstream of transcription start sites. Here, using cytoplasmic and nuclear small RNA high-throughput sequencing datasets, we report the identification of a second class of nuclear-specific ~17- to 18-nucleotide small RNAs whose 3′ ends map precisely to the splice donor site of internal exons in animals. These splice-site RNAs (spliRNAs) are associated with highly expressed genes and show evidence of developmental stage– and region–specific expression. We also show that tiRNAs are localized to the nucleus, are enriched at chromatin marks associated with transcription initiation and possess a 3′-nucleotide bias. Additionally, we find that microRNA-offset RNAs (moRNAs), the *miR-15/16* cluster previously linked to oncosuppression and most small nucleolar RNA (snoRNA)-derived small RNAs (sdRNAs) are enriched in the nucleus, whereas most miRNAs and two H/ACA sdRNAs are cytoplasmically enriched. We propose that nuclear-localized tiny RNAs are involved in the epigenetic regulation of gene expression.

Recent advances in high-throughput RNA sequencing have led to the detection of new members of established classes of small RNAs[1,2] and to the discovery and characterization of at least three classes of promoter-proximal species, including 5′-capped promoter-associated small RNAs (PASRs)[3], transcription start site (TSS)-associated RNAs (TSSa RNAs)[4] and transcription initiation RNAs (tiRNAs)[5]. The latter are ~18 nucleotides (nt) in length, are generated from sequences just downstream of transcription start sites in animals and are generally associated with highly expressed genes, transcription-factor binding and GC-rich promoters[5,6]. To determine the subcellular location of tiRNAs and to investigate the possibility that there might be other classes of nuclear-enriched small RNAs, we performed targeted deep sequencing of small RNAs from the nuclear and cytoplasmic fractions of a human monocytic leukemia cell line (THP-1) and the nuclei of primary mouse granulocytes.

RESULTS

tiRNAs are localized in the nucleus

We assessed the relative nuclear enrichment of tiRNAs by analysis of THP-1 nuclear, cytoplasmic and total small-RNA deep-sequencing datasets, which were designed to specifically include very small RNA species (~15–30 nt) and whose purity was validated by quantitative PCR and northern blotting (**Supplementary Fig. 1**). Using synthetic RNA spike-ins to normalize between libraries (**Supplementary Tables 1** and **2**), we found that tiRNAs are >40-fold enriched in the nucleus (**Fig. 1**). Indeed, nuclear and cytoplasmic small-RNA fractions intersect with 7,014 and 914 Refgene TSSs, respectively, suggesting that tiRNAs may be expressed at the majority of genes in any given genetic system. Consistent with previous analyses, genes with tiRNAs derived from both RNA fractions are significantly more highly expressed than those without ($P < 10^{-16}$, **Supplementary Fig. 2a**). The increased density of tiRNAs in the nuclear fraction reveals that 35% of genes have extensive sense and antisense clusters proximal to the TSSs (**Fig. 2a**) and shows that peak tiRNA and RNA polymerase II (RNAPII) density lies, on average, at the same position upstream of the +1 nucleosome, consistent with a model of tiRNA biogenesis dependent on RNAPII backtracking and transcription factor IIS (TFIIS) cleavage[5,6] (**Fig. 2b**). Evidence for a regulated biogenesis pathway is further provided by the fact that tiRNAs show a terminal-nucleotide bias. Although tiRNA 5′ ends are randomly distributed, they are enriched for 3′-terminal guanines in THP-1 cells (**Fig. 2c**), chicken and *Drosophila* (**Supplementary Fig. 3**).

tiRNAs are enriched at transcription-initiation chromatin marks

Genes with highly expressed tiRNAs (abundance >8) show a pronounced peak of tiRNA density ~30 nt downstream of the TSS (**Supplementary Fig. 4**), are three-fold enriched for RNAPII binding compared to genes with a single tiRNA deep-sequencing read and are at least two-fold enriched for chromatin marks (derived from $CD4^+$ cells[7,8]) indicative of active transcription or transcription

[1]Institute for Molecular Bioscience, University of Queensland, St. Lucia, Australia. [2]Queensland Facility for Advanced Bioinformatics, St. Lucia, Australia. [3]Gene & Stem Cell Therapy Program, Centenary Institute of Cancer Medicine and Cell Biology, Camperdown, Australia. [4]Sydney Medical School, University of Sydney, Australia. [5]Cell and Molecular Therapies, Sydney Cancer Centre, Royal Prince Alfred Hospital, Camperdown, Australia. [6]Department of Molecular and Cell Biology and Center for Integrative Genomics, University of California Berkeley, Berkeley, California, USA. [7]School of Integrative Biology, University of Queensland, St. Lucia, Australia. [8]These authors contributed equally to this work. Correspondence should be addressed to J.S.M. (j.mattick@imb.uq.edu.au).

Received 29 January; accepted 27 April; published online 11 July 2010; corrected online 21 July 2010; doi:10.1038/nsmb.1841

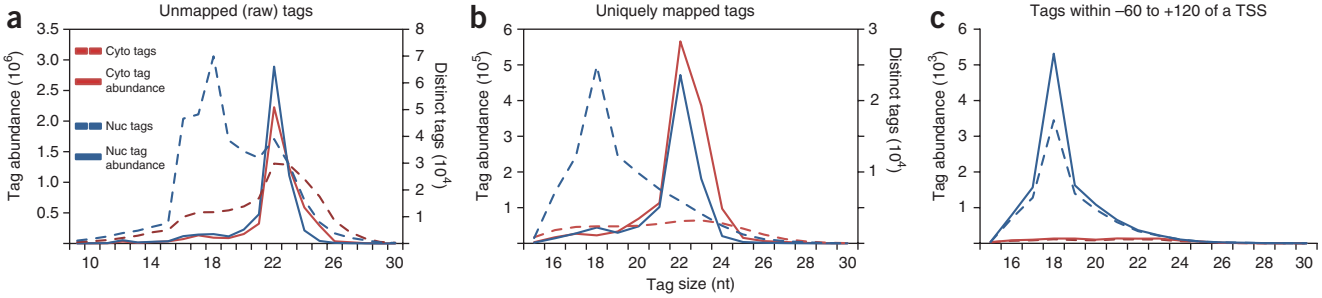

Figure 1 Human THP-1 nuclear and cytoplasmic small-RNA library characteristics. All panels: left vertical axis, total tag abundance; right vertical axis, number of distinct tags. (**a**) The size distribution of all tags and their total abundance after adaptor trimming before mapping. Note the peak of abundance at 22 nt, consistent with miRNA expression, and the peak of distinct but weakly expressed tags at 18 nt, consistent with tiRNA expression. Nuc, nucleus; cyto, cytoplasm. (**b**) The size distribution of all uniquely mapped tags and their total abundance, with peaks of abundance and unique tags at 22 and 18 nt, respectively. (**c**) The size distribution and abundance of tags proximal to Refgene TSSs. Note that in **b** and **c** normalized abundance is given. Please see Online Methods, **Supplementary Figure 1**, **Supplementary Tables 1, 2** and **6** and **Supplementary Methods** for additional details.

initiation, including histone H3 Lys4 trimethylation (H3K4me3), histone H2B Lys5 acetylation (H2BK5ac), H3K9ac, H3K27ac, H3K18ac, H4K91ac, H2BK120ac, H3K4ac, H4K20ac, H4K5ac and H3K79me3 (**Supplementary Fig. 5**). Transcription initiation marks are also enriched at over 65,000 loci with tiRNA-like clusters composed of predominantly 18-nt small RNAs. These clusters are >20-fold enriched at RNAPII binding sites and are also enriched at chromatin marks indicative of active transcription (**Fig. 2d**) but not at marks associated with silenced chromatin (for example, H3K9me2) or RNAPII elongation (for example, H3K79me3) (**Fig. 2d**). Notably, these clusters are also >11-fold enriched at CpG islands that lie outside UCSC knownGene boundaries, some of which show chromatin-modification profiles that mirror the set of marks at the nearest upstream gene, despite distances of 2 kilobases or more (for example, see **Supplementary Fig. 6**). These results suggest that ~18-nt nuclear small RNAs are generally associated with widespread transcription initiation across the genome and may mark the location of unannotated sites of RNAPII activity.

Splice-site RNAs are associated with splice donor sites

We also explored the possibility that small RNAs are associated with other genic features. Investigation of human Refseq exon boundaries

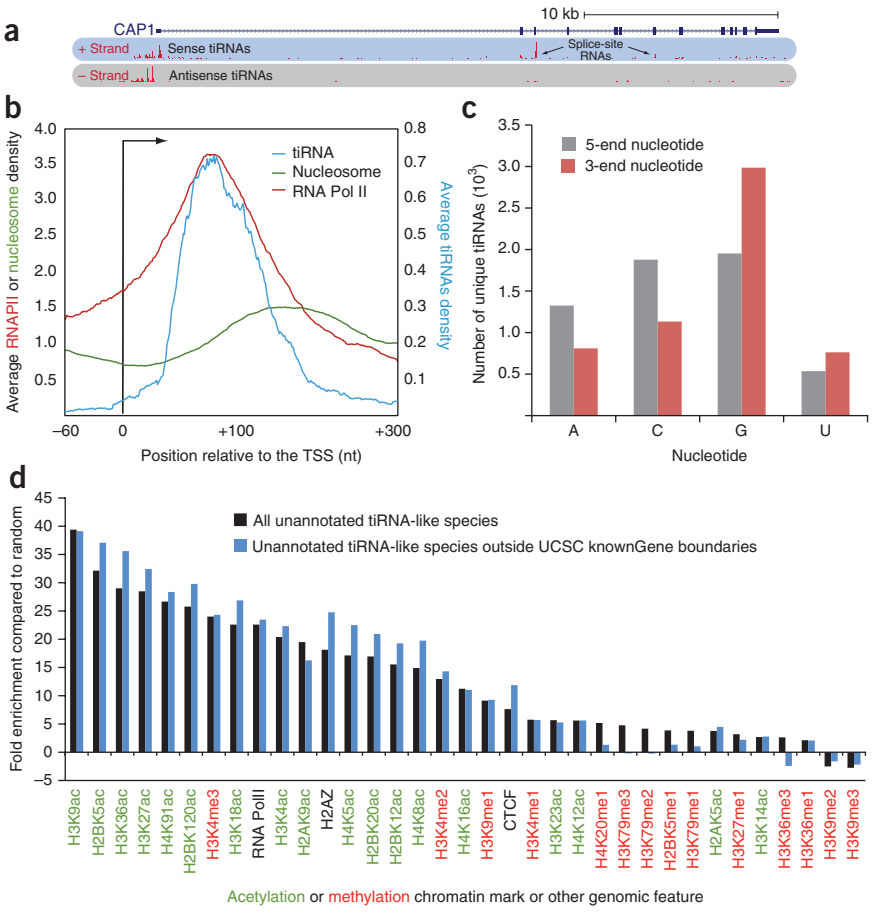

Figure 2 Features of nuclear-localized very small RNAs. (**a**) A schematic showing an example of nuclear small-RNA density at the human CAP1 locus. tiRNAs are present downstream of the CAP1 TSS as well as antisense and upstream, consistent with bidirectional promoter activity. spliRNAs, small RNAs whose 3′ ends map to the exon 3′ end (that is, the 5′ splice site), are expressed at exons 4 and 7. (**b**) Average tiRNA and RNAPII density peaks just downstream of Refgene TSSs and immediately upstream of the +1 nucleosome. RNAPII and nucleosome data are derived from CD4$^+$ T cells. Black bar and arrow denote the TSS and direction of transcription, respectively. Note that, to allow accurate comparisons, all values are averages, and tiRNA densities are smoothed and shifted to the right due to averaging across their entire length. See **Supplementary Figure 4** for cumulative tiRNA 5′ abundance density distributions, which show a pronounced peak 20–50 nt downstream of the TSS, consistent with earlier reports. (**c**) THP-1 tiRNA 3′ ends are enriched for guanines, consistent with nucleotide enrichments in *Drosophila* and chicken (**Supplementary Fig. 3**). (**d**) Nuclear THP-1 small RNAs are dominantly ~18 nt (**Fig. 1**) and are generally enriched at sites of RNAPII binding and regions with chromatin marks associated with active but not elongating transcription. In **b** and **c**, the RNAPII and chromatin-mark data are derived from CD4$^+$ T cells[7,8], as described in the main text and **Supplementary Methods**.

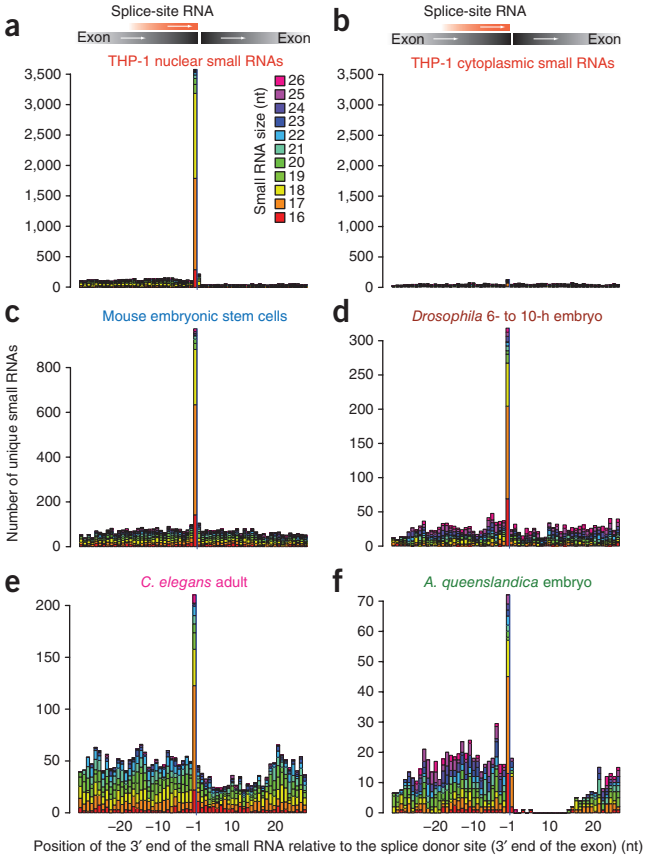

Figure 3 Splice-site RNAs are conserved across metazoa. The position of small-RNA 3′ ends is plotted with respect to the splice donor site (that is, the 3′ end of the exon). Top, schematics depict the position of spliRNAs and their strand orientation with respect to exon-exon junctions. (**a**,**b**) Small RNAs, dominantly ~17 or 18 nt, are >35-fold enriched at the 5′ splice site in THP-1 nuclei compared to either the background (**a**) or cytoplasmic THP-1 small RNAs (**b**). (**c**–**f**) Splice-site RNA expression is conserved in species representative of all major metazoan lineages, including mouse (**c**), the fruit fly *D. melanogaster* (**d**), the nematode *C. elegans* (**e**) and the marine sponge *A. queenslandica* (**f**). The data presented in **c**, **d**, **e** and **f** are derived from the publicly available National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) series GSE12521, GSE11624, GSE11738 and GSE12578, respectively. Please see **Supplementary Methods** and **Supplementary Table 7** for further details.

revealed more than 5,000 THP-1 genes with small RNAs whose 3′ termini map precisely to the splice donor site (that is, the 3′ end of the exon), are ~35-fold enriched in the nuclear deep-sequencing library (**Fig. 3a**,**b**) and are present at internal exons regardless of gene length or exon number (**Supplementary Table 3**). These splice-site RNAs (spliRNAs) are detectable using mapping strategies that consider exon-exon or exon-intron boundaries and multimapping deep-sequencing reads (**Supplementary Fig. 7a–f**).

We also found that spliRNAs are expressed in a wide range of evolutionarily distant metazoans (**Fig. 3** and **Supplementary Fig. 7g–z**). Indeed, spliRNAs are nuclear localized in primary mouse granulocyte nuclei (**Fig. 4a** and **Supplementary Fig. 8**), and a small but statistically significant subset ($n = 109$, $P < 10^{-20}$) is conserved with nuclear THP-1 spliRNAs. SpliRNAs are also detectable in mouse embryonic stem cells[9] (**Fig. 3c**), a wide range of *Drosophila melanogaster*[10] (**Fig. 3d**) and *Caenorhabditis elegans*[11] (**Fig. 3e**) tissues and in one of the most basal multicellular animals, the marine sponge *Amphimedon queenslandica*[12] (**Fig. 3f**). They have a modal length of 17 or 18 nt in human THP-1 cells and mouse primary granulocyte nuclei and a modal length of 17 nt in all other libraries and species examined. Their expression is not affected by the loss of Dicer or DGCR8 in mouse embryonic styem cells (**Fig. 4b**,**c**), nor is expression altered in *C. elegans* germline mutants (**Supplementary Fig. 7g–i**), indicating that spliRNA biogenesis is not intimately connected with the pathways that produce miRNAs or siRNAs. Indeed, with few exceptions, spliRNAs are expressed in most tissues and developmental stages in *Drosophila* and *C. elegans* (**Supplementary Fig. 7j–r**).

SpliRNAs, however, are more enriched compared to background in *Drosophila* heads compared to bodies, are almost undetectable in imaginal discs and are less abundant in adult sponge compared to embryo (**Fig. 3f** and **Supplementary Fig. 7s–x**), suggesting that spliRNAs may be connected with high gene expression in actively proliferating or undifferentiated tissues. Indeed, THP-1 and *Drosophila* genes with spliRNAs are more highly expressed than those without (**Supplementary Fig. 2b**,**c**). To determine whether spliRNAs are present outside the animal kingdom, we investigated small-RNA distributions at splice donor sites in the flowering plant *Arabidopsis thaliana* and the budding yeasts *Saccharomyces castellii* and *Saccharomyces cerevisiae*. As with tiRNAs[6], we detected no evidence of spliRNAs in yeast or plants (data not shown).

Overall, spliRNAs are weakly expressed (the median abundance in THP-1 nuclei is 1) and, similar to tiRNAs, show a strong enrichment for 3′-terminal guanines, which is likely driven by the consensus splice-site sequence (data not shown). Additionally, although spliRNAs are statistically more common at constitutive splice sites, we also observed a mild but statistically significant enrichment of spliRNAs at alternative first exons (**Supplementary Table 4**). To query the relationship between RNAPII activity and spliRNAs, we examined the recently described GRO-seq[13] dataset, which captures the position, amount and orientation of transcriptionally engaged RNA polymerases. We found a local GRO-seq minimum at the splice donor site (**Supplementary Fig. 9**), which aligns with the position of spliRNAs and may be consistent with a model of spliRNA biogenesis dependent on cleavage of the 3′ end of the nascent transcript.

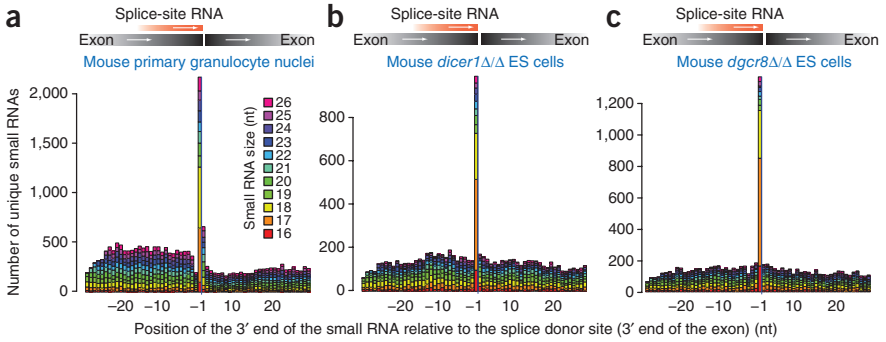

Figure 4 Mouse spliRNAs from primary granulocyte nuclei and embryonic stem cells. (**a**) The density of small RNAs in primary mouse granulocyte nuclei at exon-exon junctions. (**b**,**c**) The density of small RNAs in mouse embryonic stem (ES) cells lacking the miRNA-processing enzymes Dicer or Dgcr8 (GSE12521). Splice-site RNA biogenesis is not affected by the loss of these RNAi components in mouse ES cells.

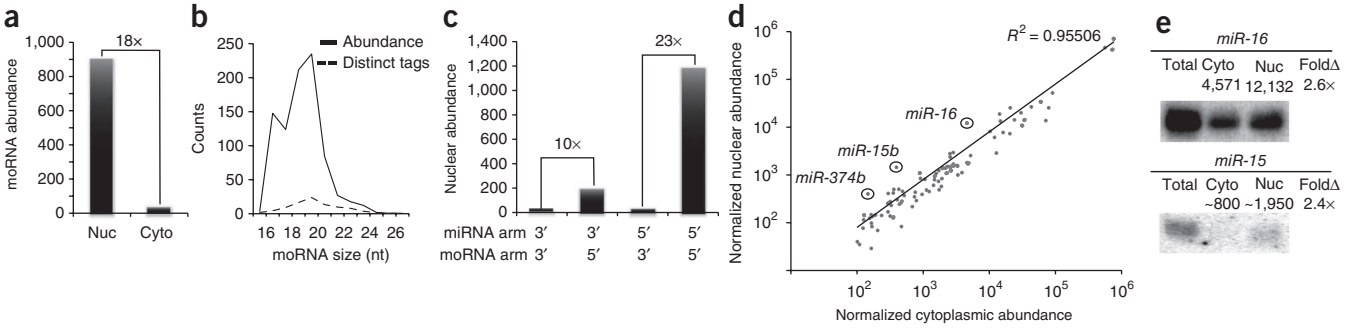

Figure 5 moRNAs and a subset of miRNAs are enriched in the nucleus of human THP-1 cells. Nuc, nucleus; cyto, cytoplasm. (**a**) The normalized abundance of moRNAs in nuclear and cytoplasmic THP-1 small-RNA libraries. (**b**) The size distribution of nuclear-localized moRNAs. (**c**) moRNAs are dominantly derived from the 5′ arm of the pre-miRNA, independent of the location of the mature miRNA (see also **Supplementary Fig. 11**). (**d**) Normalized nuclear and cytoplasmic abundance of THP-1 miRNAs. *miR-16*, *miR-15b* and *miR-374b* are all >2-fold enriched in the THP-1 nucleus (see also **Supplementary Table 5**). (**e**) Northern blot validation of the nuclear localization of the *miR-15/16* cluster. Normalized expression values are shown above each image. The expression values for *miR-15* include data from both *miR-15a* and *miR-15b* (see **Supplementary Methods**).

Indeed, we and others have recently shown that nucleosomes are preferentially positioned at exons[14–17] and that nucleosomes containing H3K36me3, which are associated with expressed exons, are positioned just downstream of the exon boundary[15,18], raising the possibility that a RNAPII-dependent mechanism, like that proposed to generate tiRNAs[5,6], also leads to the biogenesis of spliRNAs. Consistent with this hypothesis, short introns are two-fold enriched downstream of exons expressing spliRNAs, which could promote RNAPII pausing and backtracking due the proximity of the downstream exon-associated nucleosome (**Supplementary Fig. 10a**), and spliRNAs are ~2× less frequent at exons <60 nt in length (**Supplementary Fig. 10b**), which generally lack positioned nucleosomes[14].

MicroRNA-offset RNAs are nuclear enriched

In addition to tiRNAs and spliRNAs, we identified several other nuclear-enriched small-RNA species in THP-1 cells. Micro-RNA offset RNAs (moRNAs) are conserved small RNAs derived from the ends of pre-miRNAs[19]. Consistent with recent analysis that these species are present in humans[20] and are processed by the nuclear-localized RNase Drosha[19,20], we find that moRNAs are 18-fold enriched in the THP-1 nucleus and tend to be ~19 or 20 nt in length (**Fig. 5a**,**b**). Previous reports have suggested that moRNAs can be derived from either end of the pre-miRNA; however, our data indicate that moRNAs from 60 pre-miRNAs are almost exclusively derived from the 5′ arm, regardless of the position of the processed mature miRNA (**Fig. 5c** and **Supplementary Fig. 11**), suggesting that moRNA and miRNA biogenesis may be linked but are not interdependent. Indeed, consistent with reports of moRNAs in *Ciona intestinalis*[19], the abundance of some THP-1 moRNAs exceeds that of the mature miRNA derived from the same locus (**Supplementary Fig. 11**).

Select miRNAs are nuclear enriched in THP-1 cells

Expression profiling revealed that a small subset of miRNAs is also nuclear enriched (**Fig. 5d**,**e** and **Supplementary Table 5**). Although most miRNAs are cytoplasmically localized, some are present in equal concentrations in both compartments (for example, *let-7a* and *let-7i*; **Supplementary Table 5**) or are somewhat nuclear enriched (for example, *miR-15a*; **Supplementary Table 5**). Three other miRNAs, *miR-374b*, *miR-15b* and *miR-16*, are more than two-fold enriched in the nucleus (**Fig. 5d**,**e**). Downregulation of the *miR-15/16* cluster has been associated with chronic lymphocytic lymphoma, pituitary adenomas and prostate carcinoma and is known to target multiple oncogenes, including BCL2, MCL1, CCND1 and WNT3A[21]. The results presented here suggest that *miR-15/16* might have additional nuclear functions or might interact with targets within the nucleus.

sdRNAs show differential subcellular localization

There is emerging evidence that miRNAs and small nucleolar RNAs (snoRNAs) are evolutionarily related[22–25], and it has recently been shown that snoRNAs, which are classified as either C/D or H/ACA, can be processed into snoRNA-derived RNAs (sdRNAs) with distinct size distributions[22]. Consistent with snoRNA localization to the nucleolus, C/D sdRNAs are 3- to 200-fold enriched in the THP-1 nuclear fraction (**Supplementary Table 5** and **Supplementary Fig. 12**). However, sdRNAs from two H/ACA snoRNAs, SNORA36B and SNORA63 (also known as E3), which are miRNA-like and are predominantly ~22 nt in length (**Supplementary Fig. 12**), are approximately three-fold enriched in the cytoplasm (**Supplementary Table 5**), consistent with previous reports[22,23]. These data indicate that the boundary between miRNAs and other small RNAs, particularly H/ACA sdRNAs, may be blurry. Indeed, like *miR-15/16*, sdRNAs from three additional H/ACA snoRNAs are miRNA-sized but nonetheless are nuclear enriched (**Supplementary Table 5**).

DISCUSSION

Taken together, these data suggest that there is a wide range of small RNAs localized to, and abundant in, the metazoan nucleus. We propose that many of these species are involved in regulating epigenomic modifications and transcription. Transcription initiation RNAs and spliRNAs may have a common origin and function, possibly associated with the positioning of nucleosomes. If this is so, our preferred hypothesis is that this is an evolved capacity of RNAPII backtracking[6] that allows efferent signals to be produced in parallel with transcription elongation to mark the position for future reference. Indeed, 31% of THP-1 genes with spliRNAs also have tiRNAs. However, two alternative, but not mutually exclusive, possibilities are that spliRNAs are linked to, or are byproducts of, splicing or result from post-transcriptional cleavage of longer capped RNAs[3]. The absence of tiRNAs and spliRNAs in yeast and plants may reflect different systems of nucleosome positioning, chromatin marking or the criteria used to define these small-RNA classes. For example, small RNAs derived from wild-type *S. cerevisiae*[26], which lacks RNAi, are predominantly ~17 or 18 nt, have a 3′-terminal nucleotide purine (that is, adenine) bias and are phased such that local

small-RNA maxima coincide with minima of nucleosome density (**Supplementary Fig. 13**). Therefore, although these small RNAs do not meet the criteria we have used to define tiRNAs and spliRNAs in metazoans, they show many similar characteristics, suggesting that very small RNAs are a basal feature within the eukaryotic lineage that may have been co-opted to specific genomic positions and into specific roles in animals.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Gene Expression Omnibus: THP-1 and mouse granulocyte nuclei small-RNA datasets can be retrieved with accession numbers GSE20664 and GSE20683, respectively. Additional files and information can be found at http://matticklab.com/index.php?title=NuclearTinyRNAs.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank GeneWorks for assistance modifying the Illumina protocol to facilitate detection of very small RNA species and for deep sequencing the THP-1 and primary mouse granulocyte nuclei small-RNA samples and M.E. Dinger for bioinformatic assistance with the analysis of wiggle format tracks. J.S.M. and R.J.T. are supported by a Federation Fellowship grant (FF0561986) and a Discovery Project grant (DP0988851) from the Australian Research Council. J.E.J.R. received project support from the Australian National Health and Medical Research Council (358300) and the Sydney Cancer Centre Foundation. J.E.J.R. and J.H. received project and equipment support from Cancer Institute NSW and NSW Cancer Council. J.E.J.R. and J.J.-L.W. received support from Cure The Future Foundation. W.R. received salary support from an Australian National Health and Medical Research Council Training Fellowship.

AUTHOR CONTRIBUTIONS

R.J.T. designed the THP-1 deep sequencing and bioinformatic experiments, led the analysis and wrote the manuscript; C.S. made the initial spliRNA observation, designed the bioinformatic analysis of spliRNAs with R.J.T. and helped to write the manuscript; S.N. performed the analysis of spliRNA expression with respect to exon position and exon and intron size and helped to write the manuscript. H.O. and D.J.K. isolated the THP-1 nuclear and cytoplasmic RNA and performed the northern blots, respectively; T.R.M. performed the initial GRO-seq analysis; J.H., W.R., J.J.-L.W. and J.E.J.R. isolated and sequenced the mouse primary granulocyte nuclei small RNAs; D.S.R. and B.M.D. provided *A. queenslandica* genome sequences; J.S.M. helped to design the study and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/nsmb/.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Ghildiyal, M. & Zamore, P.D. Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**, 94–108 (2009).
2. Malone, C.D. & Hannon, G. Small RNAs as guardians of the genome. *Cell* **136**, 656–668 (2009).
3. ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs. *Nature* **457**, 1028–1032 (2009).
4. Seila, A.C. *et al.* Divergent transcription from active promoters. *Science* **322**, 1849–1851 (2008).
5. Taft, R.J. *et al.* Tiny RNAs associated with transcription start sites in animals. *Nat. Genet.* **41**, 572–578 (2009).
6. Taft, R.J., Kaplan, C.D., Simons, C. & Mattick, J.S. Evolution, biogenesis and function of promoter-associated RNAs. *Cell Cycle* **8**, 2332–2338 (2009).
7. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
8. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
9. Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P. & Blelloch, R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* **22**, 2773–2785 (2008).
10. Chung, W.J., Okamura, K., Martin, R. & Lai, E.C. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr. Biol.* **18**, 795–802 (2008).
11. Batista, P.J. *et al.* PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol. Cell* **31**, 67–78 (2008).
12. Grimson, A. *et al.* Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193–1197 (2008).
13. Core, L.J., Waterfall, J.J. & Lis, J.T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
14. Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* **19**, 1732–1741 (2009).
15. Nahkuri, S., Taft, R.J. & Mattick, J.S. Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle* **8**, 3420–3424 (2009).
16. Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon-intron structure. *Nat. Struct. Mol. Biol.* **16**, 990–995 (2009).
17. Tilgner, H. *et al.* Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* **16**, 996–1001 (2009).
18. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41**, 376–381 (2009).
19. Shi, W., Hendrix, D., Levine, M. & Haley, B. A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. *Nat. Struct. Mol. Biol.* **16**, 183–189 (2009).
20. Langenberger, D. *et al.* Evidence for human microRNA-offset RNAs in small RNA sequencing data. *Bioinformatics* **25**, 2298–2301 (2009).
21. Aqeilan, R.I., Calin, G.A. & Croce, C.M. miR-15a and miR-16–1 in cancer: discovery, function and future perspectives. *Cell Death Differ.* **17**, 215–220 (2010).
22. Taft, R.J. *et al.* Small RNAs derived from snoRNAs. *RNA* **15**, 1233–1240 (2009).
23. Ender, C. *et al.* A human snoRNA with microRNA-like functions. *Mol. Cell* **32**, 519–528 (2008).
24. Saraiya, A.A. & Wang, C.C. snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog.* **4**, e1000224 (2008).
25. Scott, M.S., Avolio, F., Ono, M., Lamond, A.I. & Barton, G.J. Human miRNA precursors with box H/ACA snoRNA features. *PLOS Comput. Biol.* **5**, e1000507 (2009).
26. Drinnenberg, I.A. *et al.* RNAi in budding yeast. *Science* **326**, 544–550 (2009).

ONLINE METHODS

THP-1 RNA isolation and validation. THP-1 cells were grown in suspension culture[5,27]. Nuclear and cytoplasmic RNA was isolated as previously described[28], except that washes were carried out using 1 ml of wash buffer, and Tween-40 was substituted for Tween-20 in the final wash. RNA was extracted using the TRIzol protocol (Invitrogen), and the resulting RNA pellets were resuspended in equal volumes to obtain cell-equivalent concentrations. Seven RNA species were assessed by quantitative PCR and/or northern blotting to validate nuclear and cytoplasmic RNA fractionation. (**Supplementary Fig. 1** and **Supplementary Table 6**). For more details, see **Supplementary Methods**.

Mouse granulocyte nuclei preparation and RNA isolation. Bone marrow from C57BL/6J was harvested from the femur, tibia and spine using a mortar and pestle in PBS supplemented with 2% (v/v) FCS and mature granulocytes purified by flow cytometry as described previously[29,30]. Purification was validated by reanalysis by flow cytometry and May-Grünwald Giemsa staining. Nuclear purification was carried out using the PARIS kit (Ambion). RNA was extracted from the nuclear fraction using Trizol before deep sequencing. For more details, see **Supplementary Methods**.

Small RNA deep sequencing. Deep sequencing of cytoplasmic and nuclear small RNAs from THP-1 cells and mouse granulocyte nuclei was performed by GeneWorks on the Illumina GAII. For THP-1 small RNA sequencing, sample isolation from the PAGE gel after adaptor ligation was performed with a modified set of size markers to facilitate sequencing of small RNAs ≥15 nt. For more details, please see **Supplementary Methods**.

Other small-RNA deep sequencing datasets. Small-RNA datasets from mouse[9], *Drosophila*[10], *C. elegans*[11], *A. queenslandica*[12] and budding yeast[26] were obtained from the NCBI GEO (**Supplementary Table 7**). Human GRO-seq data[13] was also obtained from NCBI GEO (GSE13518).

Reference genome and annotation sources. Human (hg18, NCBI build 36.1), mouse (mm9, NCBI build 37), *Drosophila* (dm3, BDGP release 5), *C. elegans* (ce6, WS190) and *S. cerevisiae* (sacCer2, SGD June 2008) genome sequences and gene and genome feature annotations were obtained from a local mirror of the UCSC Genome Browser[31]. Human and mouse Refseq genes were obtained from the respective refGene databases. *Drosophila* Flybase, *C. elegans* Sanger and *S. cerevisiae* SGD gene annotations were obtained from the dm3.flyBaseGene, ce6.sangerGene and sacCer2.sgdGene databases, respectively. *A. queenslandica* sequences and annotations were obtained from the University of Queensland sponge genome sequence database. The *S. castellii* genome sequence and annotations were obtained from the Yeast Gene Order Database[32]. We used the *Arabidopsis* TAIR8 genome sequence and the TAIR8 Ensemble gene annotations[33].

$CD4^+$ T cell nucleosome modification data were downloaded from the authors' website[7,8]. Control $CD4^+$ T cell nucleosome datasets were obtained from the NCBI Sequence Read Archive (SRR000711–SRR000720). *S. cerevisiae* combined H3 and H4 nucleosome data was obtained from the authors' website[34].

Bioinformatic analyses. Bioinformatic analyses were performed on a local high-performance computer that houses a mirror of the UCSC Genome Browser[31]. We used a suite of in-house AWK, C, Perl, and Python scripts and UCSC backend tools. Small-RNA datasets, raw $CD4^+$ nucleosome data and *S. cerevisiae* H3 and H4 nucleosome data were mapped to the appropriate genome using ZOOM[35]. Small RNA, GRO-seq, chromatin modification and nucleosome density distributions were generated by converting mapped tag positions to genome-wide wiggle density plots and averaging these densities across all loci of interest. For $CD4^+$ T cell nucleosome data[7,8], we extended the genomic matches of all uniquely mapping tags in the 3′ direction so that they reached a total length of 150 nt, consistent with the expected length of nucleosome-associated DNA, as described previously[15,36].

The abundance of THP-1 nuclear and cytoplasmic small-RNA datasets was normalized by the relative expression of spike-ins 2 and 6 (**Supplementary Table 2**). Bioinformatic queries against spike-ins were performed without mismatches to ensure accurate quantification and normalization. Identification and analysis of THP-1 nuclear tiRNAs was performed as previously described[5]. Analysis of the expression of genes with tiRNAs was accomplished using gene-expression data from undifferentiated THP-1 cells[27], as described previously[5]. Refgenes with high tiRNA abundance (>8) or low tiRNA abundance (1) were obtained, and regions −60 to +300 relative to the TSS were assessed for chromatin-mark densities. Unannotated 18-mers were identified after eliminating all canonical tiRNAs and then further filtering to exclude those proximal to any knownGene TSS or within a knownGene boundary (that is, within the bounds defined by the transcription start and stop sites). Enrichments at chromatin marks used loci with chromatin-mark tag densities two standard deviations higher than the mean for that mark across the genome. Loci located near TSSs (within 200 nt) or that mapped to known small-RNA annotations were excluded. The relative enrichment of nuclear small RNAs at each chromatin mark or protein-binding site was assessed using an in-house (Perl) bootstrapping program over 1,000 iterations.

Splice-site RNAs are defined as small RNAs, dominantly 17 or 18 nt, whose 3′ ends map to the 3′ end of internal exons. spliRNAs were mapped to both the genome and a library of splice-site junctions for each organism. Dips of small RNAs just across the splice site in some organisms may reflect poor gene annotations (that is, missed exons). To examine the conservation of spliRNAs between human THP-1 and mouse granulocyte nuclei, a Fisher's exact test was used to examine the significance of the association between syntenic splice donor sites ($N = 162,807$) that have one or more spliRNAs in only the human THP-1 nuclear dataset ($N = 3,044$), only the mouse granulocyte nuclei dataset ($N = 2,037$) or both ($N = 109$). Analysis of the expression of genes with spliRNAs in human and *Drosophila* was accomplished using gene-expression data from undifferentiated THP-1 cells[27] and a *Drosophila* developmental time course[37], as described previously[5]. To examine the association of spliRNAs with alternative and constitutive exons, UCSC knownGene exon annotations were used to derive the splicing status of exons with spliRNAs versus exons without spliRNAs in the same genes. The prevalence of four different alternative splicing events (**Supplementary Table 4**) in both datasets was assessed, and the statistical significance of the observed difference was calculated using the Fisher's exact test.

We used annotations from miRbase version 12 (ref. 38) to assess THP-1 nuclear and cytoplasmic miRNA expression. To ensure accurate expression values, we included all uniquely mapping and multimapping tags that mapped exclusively to miRNA loci. Relative microRNA expression was calculated as the sum of the normalized abundance of all tags that mapped to any particular pre-miRNA. We defined moRNAs as any RNA tag that covered the most 5′ or 3′ ends of a pre-miRNA annotation. Using EDC northern blots[39], the nuclear enrichment of *miR-15* was assessed using a probe spanning the 5′ 16 nucleotides, and *miR-16* was detected using a probe spanning its entire length (**Supplementary Table 6**).

27. Suzuki, H. *et al.* The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nat. Genet.* **41**, 553–562 (2009).
28. Hwang, H.W., Wentzel, E.A. & Mendell, J.T. A hexanucleotide element directs microRNA nuclear import. *Science* **315**, 97–100 (2007).
29. Holst, J. *et al.* Generation of T-cell receptor retrogenic mice. *Nat. Protoc.* **1**, 406–417 (2006).
30. Guibal, F.C. *et al.* Identification of a myeloid committed progenitor as the cancer-initiating cell in acute promyelocytic leukemia. *Blood* **114**, 5415–5425 (2009).
31. Kuhn, R.M. *et al.* The UCSC Genome Browser database: update 2009. *Nucleic Acids Res.* **37**, D755–D761 (2009).
32. Byrne, K.P. & Wolfe, K.H. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* **15**, 1456–1461 (2005).
33. Poole, R.L. The TAIR database. *Methods Mol. Biol.* **406**, 179–212 (2007).
34. Mavrich, T.N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18**, 1073–1083 (2008).
35. Lin, H., Zhang, Z., Zhang, M., Ma, B. & Li, M. ZOOM! Zillions of oligos mapped. *Bioinformatics* **24**, 2431–2437 (2008).
36. Schmid, C.D. & Bucher, P. ChIP-Seq data reveal nucleosome architecture of human promoters. *Cell* **131**, 831–832 author reply 832–833 (2007).
37. Arbeitman, M.N. *et al.* Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**, 2270–2275 (2002).
38. Griffiths-Jones, S., Saini, H.K., van Dongen, S. & Enright, A.J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2008).
39. Pall, G.S., Codony-Servat, C., Byrne, J., Ritchie, L. & Hamilton, A. Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Res.* **35**, e60 (2007).