



HAL
open science

Short intron-derived ncRNAs

Florent Hubé, Damien Ulveling, Alain Sureau, Sabrina Forveille, Claire Francastel

► **To cite this version:**

Florent Hubé, Damien Ulveling, Alain Sureau, Sabrina Forveille, Claire Francastel. Short intron-derived ncRNAs. *Nucleic Acids Research*, 2016, 45 (8), pp.4768-4781. 10.1093/nar/gkw1341 . hal-02127265

HAL Id: hal-02127265

<https://u-paris.hal.science/hal-02127265>

Submitted on 13 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Short intron-derived ncRNAs

Florent Hubé^{1,2,*}, Damien Ulveling^{1,2}, Alain Sureau^{1,2}, Sabrina Forveille^{1,2} and Claire Francastel^{1,2,*}

¹Université Paris Diderot, Sorbonne Paris Cité, Paris, France and ²Epigénétique et Destin Cellulaire, CNRS UMR 7216, Paris, France

Received July 12, 2016; Revised December 19, 2016; Editorial Decision December 20, 2016; Accepted December 21, 2016

ABSTRACT

Introns represent almost half of the human genome, although they are eliminated from transcripts through RNA splicing. Yet, different classes of non-canonical miRNAs have been proposed to originate directly from intron splicing. Here, we considered the alternative splicing of introns as an interesting source of miRNAs, compatible with a developmental switch. We report computational prediction of new Short Intron-Derived ncRNAs (SID), defined as precursors of smaller ncRNAs like miRNAs and snoRNAs produced directly by splicing, and tested their dependence on each key factor in canonical or alternative miRNAs biogenesis (Drosha, DGCR8, DBR1, snRNP70, U2AF65, PRP8, Dicer, Ago2). We found that about half of predicted SID rely on debranching of the excised intron-lariat by the enzyme DBR1, as proposed for mirtrons. However, we identified new classes of SID for which miRNAs biogenesis may rely on intermingling between canonical and alternative pathways. We validated selected SID as putative miRNAs precursors and identified new endogenous miRNAs produced by non-canonical pathways, including one hosted in the first intron of SRA (Steroid Receptor RNA activator). Consistent with increased SRA intron retention during myogenic differentiation, release of SRA intron and its associated mature miRNA decreased in cells from healthy subjects but not from myotonic dystrophy patients with splicing defects.

INTRODUCTION

The discovery of non-coding RNAs (ncRNA) and the variety of molecular processes in which they have been implicated suggest that they increase and diversify the amount of regulatory molecules available in the cell. In contrast to housekeeping or infrastructural ncRNAs, which are normally constitutively expressed and required for normal

function and viability of the cell, regulatory ncRNAs are expressed in response to external stimuli or at particular stages of development and cell differentiation, and can affect the expression of other genes at the level of transcription or translation (1). Among those, microRNAs (miRNA) are a class of naturally occurring small ncRNAs, about 20–25 nucleotides (nt) in length, which have been identified in almost all eukaryotic cells. miRNAs are post-transcriptional regulators that bind to complementary sequences on target mRNA, usually resulting in translational repression or target degradation and thus, gene silencing (2). The human genome may encode over 1000 miRNAs, targeting ~60% of gene products in mammals. Therefore, because they affect gene regulation and are often deregulated in human diseases, their systematic identification has been the focus of many experimental and computational analyses [for reviews see (3,4)].

Canonical miRNAs are generated in a two-step processing pathway, mediated by two major enzymatic complexes containing the RNase III-family of endonucleases Drosha and Dicer. Drosha, together with DGCR8, is part of the microprocessor multiprotein complex that mediates nuclear processing of the primary miRNA into stem-loop precursors of ~60–70 nt (pre-miRNA). Exportin-5 (XPO5) mediates the nuclear export of correctly processed miRNA precursors. In the cytoplasm, the pre-miRNA is cleaved by Dicer into the mature 20–25 nt miRNA, which is then incorporated as single-stranded RNA into a ribonucleoprotein complex containing Argonaute 2 (Ago2) protein, known as the RNA-induced silencing complex (RISC). This RISC complex directs the miRNA to its target mRNA leading to its translational repression or its degradation [for a review see (5)].

It has also been postulated that regulatory RNA molecules could originate from the introns of protein-coding genes as functional by-products (6–8). Several recent studies indeed uncovered an atypical pathway to generate miRNA precursors in a way that bypasses the Drosha/DGCR8 complex (9–16). Instead, the pre-miRNA-like hairpins are produced by the action of the splicing machinery followed by lariat-debranching by the

*To whom correspondence should be addressed. Tel: +33 1 57 27 89 32; Fax: +33 1 57 27 89 10; Email: florent.hube@univ-paris-diderot.fr
Correspondence may also be addressed to Claire Francastel. Tel: +33 1 57 27 89 18; Fax: +33 1 57 27 89 10; Email: claire.francastel@univ-paris-diderot.fr
Present address: Sabrina Forveille, Metabolomics and Cell Biology Platforms, Gustave Roussy Cancer Campus, Villejuif, France.

enzyme DBR1. The 5'- and/or the 3'-tails around the hairpin are then trimmed by the RNA exosome (12). The mirtron pathway merges with the canonical miRNA pathway at hairpin export by XPO5, and subsequent processing of hairpins by Dicer. These unusual miRNA precursors are called mirtrons owing to their embedding into introns of coding and non-protein coding genes (9,14,16). Only a few mirtrons have been described to date although they have been shown to exist from drosophila to humans (9–16). More recently, two other unconventional pathways have been described; the simtron pathway involves a small nuclear ribonucleoprotein (snRNP), which is part of the spliceosome complex snRNP70 that directly recruits Drosha on the stem-loop hairpin formed by the pre-mRNA, independently of DGCR8 (17,18). After cleavage by Drosha, pre-miRNAs are exported to the cytoplasm. The agotron pathway is more intricate given that the spliced introns are processed by Ago2 directly in the nucleus. Hence, miRNAs generated by the agotron pathway are processed independently of Drosha and Dicer (19). Interestingly, several miRNAs have been described to be processed from snoRNAs (20) or tRNAs (21) [for a review see (22)] and the description of non-canonical pathways has accumulated over the years (15,17,19,23–26).

As we just discussed above, many new unconventional miRNAs, and most if not all small nucleolar RNAs (snoRNAs), are produced from introns through splicing mechanisms. Although introns, which represent about half of the genome in mammals, have long been considered as ‘junk’ and ‘dark matter’ since they are quickly degraded within minutes following their excision, an increasing number of studies have shown the importance of introns in producing small ncRNAs. The possibility that there are more to be discovered, expressed at lower levels or in more specialized cellular contexts, calls for the exploitation of genome sequencing information to accelerate their discovery and ease their structural characterization.

We propose herein the bioinformatic identification of human candidates as intron-derived small ncRNAs precursors through data mining of publically available information provided by high-throughput sequencing projects. We used the term SID as an acronym for Short Intron-Derived ncRNAs precursors of smaller ncRNAs like miRNAs and snoRNAs regardless of their biogenesis pathway but produced directly from splicing. Hence, this term may group precursors of snoRNAs that are already known to originate from splicing of introns (27,28), and non-canonical precursors of miRNAs like agotrons and mirtrons (15,19) or other classes of splicing-dependent miRNA precursors yet to be identified.

We focused on small introns (shorter than 200 nt) since they are more prone to alternative splicing (29). We provide a list of 56 new human SID, extracted from databases of alternative splicing events and matching entries in databases of human small RNAs. We further experimentally validated that these SID candidates were indeed detectable, in a panel of six cell lines and primary cells, as discrete and persistent entities. Among the highest detectable candidates in human cells, we validated six of them, which indeed produced a miRNA with variable dependence on known fac-

tors of miRNAs pathways. We employed several independent approaches to characterize the miRNA of one of these new SID that is hosted in the first intron of SRA (Steroid Receptor RNA activator). Of particular interest, we already showed that alternative splicing serves to increase the transcriptional output of the SRA1 gene, by producing a protein-coding gene when this intron is excised or a functional non-coding RNA when this intron is retained, disrupting the open reading frame (ORF), concomitant with myogenic differentiation (30,31). We now show that splicing of intron 1 of SRA generates a new miRNA with high levels in undifferentiated myogenic progenitors.

As a whole, this work represents the first experimental validation of human endogenous SID and identified new classes of small ncRNAs of intron origin. It also puts forward that alternative splicing might represent an attractive source of small regulatory RNAs compatible with a tissue-specific or differentiation-stage specific role, which will provide new actors or biomarkers deregulated in disease.

MATERIALS AND METHODS

Data mining and data sets

Datasets of ‘Human Introns’, ‘Alternative Events’, ‘CpG island’, ‘SNP’ and ‘Genes’ were retrieved from the UCSC table browser, using ‘Genes and gene prediction’, ‘Alt events’, ‘Regulation’ and ‘RefSeq Genes’ tracks from the hg19 assembly covering the whole human genome. Using Galaxy and ‘Operate on Genomic Intervals’ tool, we selected introns shorter than 200 nucleotides and 100% covered in ‘Alternative Events’ datasets. Location of intron retention relative to the start and stop codons was obtained by genomic coordinate comparison (from ‘Genes’ dataset). Palindrome formation of selected introns was assessed using ‘palindrome search’ tool (parameters: minimum length of palindrome: 18; maximum length of palindrome: 28; Maximum gap between repeated regions: >100; number of mismatches allowed: <6). Finally, putative SID (Short Intron-Derived ncRNA) and non-SID introns were manually curated to remove potential redundancy. All excluded introns (1713), i.e. which did not present the ability to form a palindrome, were used to construct the NI (normal introns) dataset. The miRNA dataset (1426) was downloaded from miR-Base (<http://www.mirbase.org/>; release 17), the snoRNA dataset (402) was downloaded from snoRNABase (<https://www-snoRNA.biotoul.fr/>; v3) and a random dataset (10000) was generated using the Random DNA Sequence Generator software. Data for small RNA-seq was obtained from ENCODE/Cold Spring Harbor Lab track from the UCSC table browser (GEO DataSet number GSE24565; part of the BioProject 30709) and consist in NextGen sequencing information for RNAs between 20–200 nt in size, isolated from RNA samples from tissues or sub-cellular compartments of cell lines (K562, GM12878 and prostate cells).

Datasets analysis

All tools were from The European Molecular Biology Open Software Suite EMBOSS. GC content was calculated using ‘infoseq’; CpG islands were mapped using ‘CpG islands’ track from UCSC browser. Single Nucleotide Poly-

morphisms were searched using the dbSNP132 track from UCSC browser. RNALfold, which calculates local stable secondary structures of RNAs, was used with default parameters. Minimum free energies (MFE, kcal/mol) were corrected relative to the size of the transcripts.

Software and databases

Software and web servers used were as follows: The Mfold web server v2.3, <http://mfold.rna.albany.edu>; UCSC genome Browser, <http://genome.ucsc.edu>; Galaxy, <http://main.g2.bx.psu.edu>; Genomatix Software Suite v2.1, <http://www.genomatix.de>; Venny, <http://bioinfogp.cnb.csic.es>; Random DNA sequence generator, <http://users-birc.au.dk>; miRBase v17; <http://www.mirbase.org>; snoRNABase v3, <https://www-snoRNA.biotoul.fr>; miRNA Primer Design Tool, <http://genomics.dote.hu:8080/mirnadesigntool>; RNALfold, <http://www.tbi.univie.ac.at>.

Cell culture

Primary human satellite cells (LHCN-M2, myoblast, MB) and their *in vitro* differentiated myotubes (MT) counterpart, HEK-293, K562, MCF-7 and MDA-MB-231 cells were grown as previously described (30,31). RNA from primary human satellite cells isolated from muscle biopsies of three fetuses showing clinical symptoms of the congenital Myotonic Dystrophy type 1 (DM1), and from one immortalized cell line (DM11), all carrying more than 2000 CTG repeats, were a gift from Denis Furling (Myology Institute, Paris, France) and were used previously (31,32).

Antibodies

Primary antibodies used were directed against Drosha (ab12286, Abcam), DGCR8 (ab90579, Abcam), DBR1 (sc-99369, Santa Cruz), snRNP70 (sc-9571 c-18, Santa Cruz), PRP8 (sc-55534 F-6, Santa Cruz), U2AF65 (sc-48804 H-300, Santa Cruz), Dicer (sc-56651), Ago2 (sc-32877 H-300, Santa Cruz) and γ -Tubulin (T6557, Sigma).

Plasmids and constructs

pSicoR human Dicer1 (ID 14763), pSicoR human Drosha1 (ID 14766) and pSicoR human DGCR8-1 (ID 14769) were requested from Addgene (www.addgene.org). Short hairpin RNAs (shRNA) directed against human DBR1 (DBR1-1, DBR1-2 and DBR1-3), snRNP70+PRP8, U2AF65+PRP8, Ago2 (Ago2-1 and Ago2-2) and Luciferase were produced using MessageMuter™ shRNA production kit (Epicentre Biotechnologies) and *in vitro* transcribed using T7 Ribomax large-scale production system (Promega) following manufacturer's instructions. Sequences are depicted in Supplemental Table S2.

Knockdown of snRNP70 or U2AF65 was combined to that of the major splicing factor PRP8, since interference of snRNP70 or U2AF65 alone was insufficient to reduce splicing significantly in another study (18).

RNA preparation

Total RNA was isolated using Trizol reagent (Life Technologies) according to the manufacturer's instructions and as previously described (31,33). Short and long RNAs were purified using mirVana™ miRNA Isolation Kit (Life Technologies) according to instructions.

Oligonucleotide array

Oligonucleotides, 45–55 nt in length and complementary to candidate SID, at a final concentration of 10 μ M, were spotted on GeneScreen Plus (PerkinElmer) membranes using a 96 well manifold system (BioRad). To confirm the specificity of hybridization, we included a series of control oligonucleotides on the array. Briefly, RNU6-1 (U6) oligonucleotide served as loading control, two randomly-generated oligonucleotides were used to set the baseline signal and two oligonucleotides matching constitutive exon 4 of GAPDH and exon 4 of SRA mRNA were used as size fractionation controls. All oligonucleotides were first phosphorylated using 1U Polynucleotide kinase (New England Biolabs) per microgram of oligonucleotide for 2h at 37°C followed by enzyme inactivation by addition of 1 vol. of 0.5 M EDTA.

Membranes were briefly washed once with 1 \times MOPS–NaOH pH 7 and once with 0.1 \times SSC, 0.125 N NaOH. Oligonucleotides were denaturated in 500 mM NaOH before being diluted in 0.1 \times SSC, 0.125 N NaOH and spotted onto membranes. Membranes were further washed again twice in 1 \times MOPS–NaOH pH 7. Spotted oligonucleotides were cross-linked to the membrane using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) method described in (34). Arrays were not stored but used directly after spotting.

Five micrograms of small RNAs freshly isolated from cell lines, transfected or not with shRNA, were preheated at 80°C for 3 min, cooled on ice, and labelled overnight at 37°C using 1 U poly (A) polymerase and 50 μ Ci ³²P- α -ATP in the presence of 40 U RNase Out (Life Technologies).

Membranes were pre-hybridized in Ultrahyb[®]-oligo (Life Technologies) at 42°C for at least 30 min followed by an overnight hybridization in the same solution containing the RNA probe. Following hybridization, membranes were washed twice with 2 \times SSC/0.5% SDS at 42°C and once at 1 \times SSC/0.5% SDS. Membranes were exposed to a phosphor storage screen, scanned using Phosphor Imager (Fujifilm), and hybridization signals were quantified using Multi Gauge v3 (Fujifilm). Membranes were never stripped and used only once.

Hybridization signals for each spot of the array and background values at eight empty spots were measured. Averaged hybridization signal at empty spots was subtracted from each other spot signal and normalized using the U6 signal. Signals from spots with random oligonucleotides were averaged and served as baseline value. Signals from candidate SID from among duplicate spots and triplicate arrays were averaged and considered positive when being over the baseline value. Hybridization with RNA from the different cell types was performed three times ($n = 3$), or four times ($n = 4$) following RNA interference in K562 cells.

Northern blot

RNA northern blot using short RNA fractions were performed as described by Pall and Hamilton (34) using 5 μ g of small RNAs purified as described above and probed with the 32 P- γ -ATP radio-labelled oligonucleotides used in the macro-array.

Pocket-sized RNA-seq

The method was fully described in reference (35). Sequences used to produce human SRA intron 1 or luciferase baits are provided in Supplemental Table S2.

RT-PCR, RT-qPCR and stem-loop RT-qPCR

RNA was isolated as described above and reverse transcribed as described previously (30,31,33,36,37). Primers used for amplification are described in Supplemental Table S2. Radio-labeled PCR products were separated on denaturing poly-acrylamide/urea gels as previously described (38). Following electrophoresis, the gels were dried and exposed to a phosphor storage screen, and scanned as above ($n = 2$). Quantitative PCR was carried out using the LightCycler[®] 480 SYBR Green I master on the Light cycler 480 real-time PCR system (Roche Diagnostics). For stem-loop RT-qPCR, reverse transcriptase reactions were performed replacing random hexamer primers by stem-loop primers. All reactions, including no-template controls and RT-minus controls, were run in duplicate ($n = 3$). Relative expression of genes was calculated by using the comparative CT method (30,31), i.e. $2^{-((\text{shMock sample} - \text{shMock control}) - (\text{shTarget sample} - \text{shTarget control}))}$. Expression of immunoprecipitated RNAs was obtained by calculating the 'relative to the irrelevant antibody' method, i.e. $2^{-((\text{Ct specific Ab} - \text{Ct irrelevant Ab}))}$ and was expressed as fold enrichment ($n = 2$). Mature miRNA-specific primers were designed using miRNA Primer Design Tool Website and normalization was performed relative to the signal of RNU6-1 (U6) amplification. All primers used in qPCR experiments had efficiencies $>98\%$, much more above the MIQE requirement.

RNA and protein immunoprecipitation

Cells were lysed, RNA and proteins were extracted and quantified as described previously (31) and used for native RNA immunoprecipitation (RIP) or protein immunoprecipitation experiments, respectively. Protein extracts (1 mg) were incubated with the appropriate antibody (1 μ g/mg of proteins) for 2 h at 4°C as described earlier (33,36). Total proteins were analyzed by SDS-PAGE and immunoblotting as previously described (36) or when appropriate, co-precipitated RNA was extracted using Trizol method, reverse transcribed and PCR amplified as described above ($n = 2$).

Transfection experiments

Plasmid constructs and shRNA were transiently transfected ($n = 3$) using jetPEI[™] and interferin reagents (PolyPlus transfection), respectively, following the manufacturer's instructions and as previously described (31).

RESULTS

Bioinformatic prediction of candidate Short Intron-Derived ncRNA (SID)

In order to systematically identify SID in the human genome database we used the workflow outlined in the Material and Methods section and in Supplementary Figure S1, starting from 326, 537 introns extracted from UCSC table browser. Since 100% of pre-miRNA hairpins (calculated from miRBase) and 95% of snoRNA (calculated from snoRNABase) are not >200 nt in length, our computational strategy retained only introns <200 nt but >50 nt so they can form a hairpin compatible with the size of a pre-miRNA or a snoRNA. This restriction has also been chosen for further experimental validation after RNA size fractionation that imposes a cut-off around 150–200 nt. Although splicing of introns is a prerequisite to produce mirtrons and snoRNAs, we reasoned that alternative splicing events would be of prime interest to account for their regulatory functions during cellular differentiation or to integrate external or environmental signals. Therefore, the 28 118 extracted introns of 50–200 nt in size were further confronted with 92 222 alternative splicing events from UCSC (39), of which 6450 represent retained introns. This first step led to the identification of 2147 short introns subjected to alternative splicing. In addition to effective splicing and intron length, structural characteristics are also important to define a SID. We thus selected 218 unique introns shorter than 200 nucleotides, which contained inverted repeats necessary for the formation of a stem-loop structure. These short hairpin introns, unlike typical introns that are rapidly degraded following splicing (40), are much more stable in the cell. Thus, out of these 218 introns, we retained introns that matched collections of small RNAs of 20–200 nt in length isolated from tissues or sub-cellular compartments from ENCODE cell lines (<http://genome.ucsc.edu/ENCODE/>) and deep sequencing of small RNA-seq samples from ENCODE project (GEO Series accession number GSE24565), potentially containing both discrete and stable introns or mature miRNAs.

Overall, we identified 56 introns that are compatible in length, structure and sequence, with the formation of SID hairpins and for which small RNAs exist in small RNAs databases (Figure 1 and Supplementary Table S1).

Structural features of candidate Short Intron-Derived ncRNA (SID)

We then analyzed structural features of the putative 56 new human SID compared to the excluded normal introns (NI) and to annotated pre-miRNAs and snoRNAs extracted from miRBase v17 and snoRNABase v3, respectively. We showed that human SID exhibit a higher G+C content (Figure 2A) and preferentially reside within CpG island regions (Figure 2B). Furthermore, the stability of their secondary structure is higher than that of NI, snoRNAs (41) or random sequences as determined using RNALfold, and is comparable to that of annotated pre-miRNAs (Figure 2C). Altogether, these findings suggest that formation of a pre-miRNA-like structure is not a common feature of introns but is a structural characteristic of both canonical

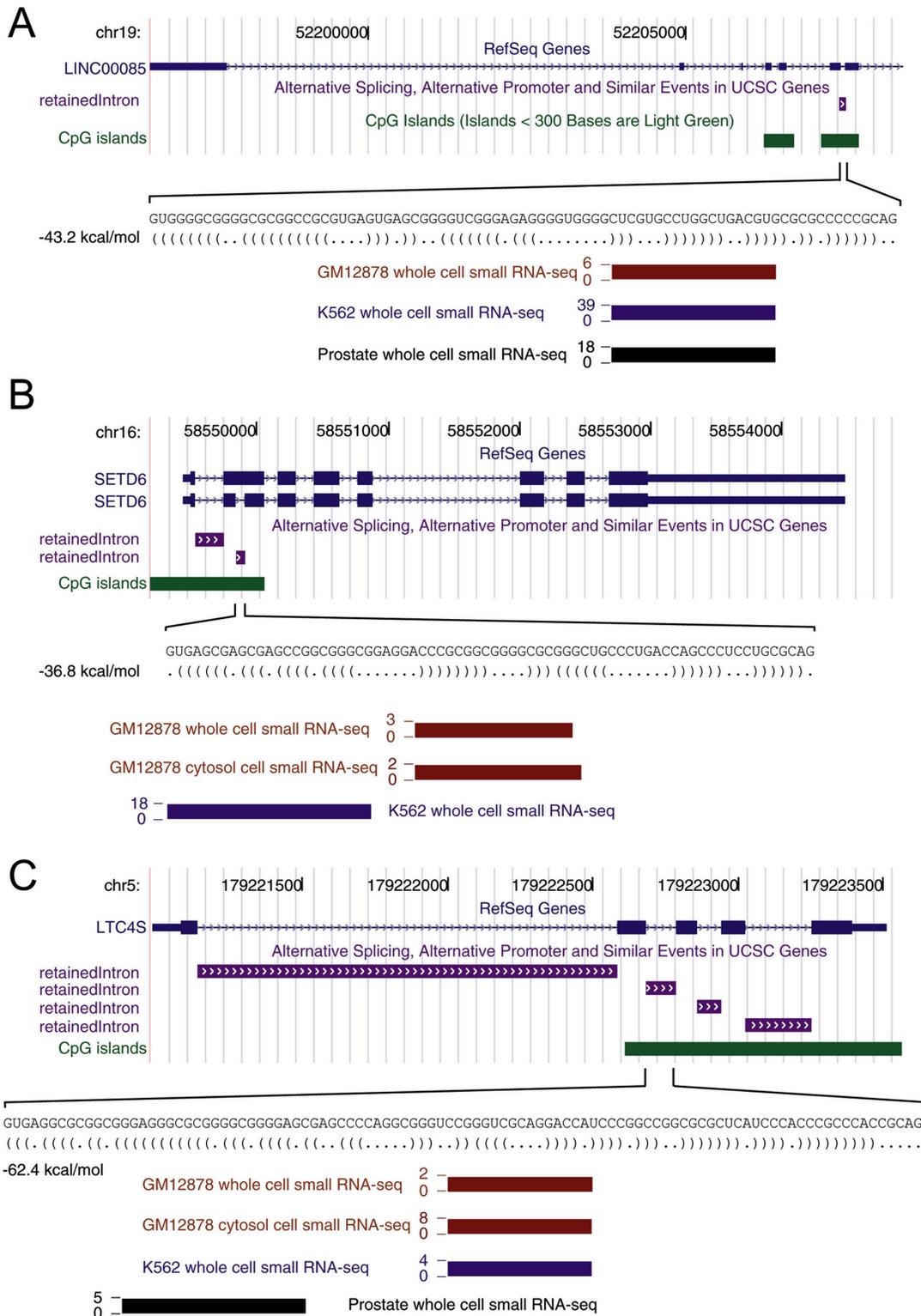


Figure 1. typical example of SID candidates. (A) Schematic representation of the human LINC00085 gene located on chromosome 19 (chr19:52, 196, 593-52, 208, 443 described as SID #39 in Supplementary Table S1). The sixth intron was identified as a retained intron in database. In this particular example, the gene considered was described as a non-coding RNA (NR_024330). (B) Schematic representation of the human SETD6 gene located on chromosome 16 (chr16:58, 549, 383-58, 554, 456 described as SID #8 in Supplementary Table S1). The second intron was identified as a retained intron in database. In this particular example, the two transcript variants 1 and 2 have been already described as NM_001160305 (containing the intron in frame with the ORF) and NM_024860, respectively. (C) Schematic representation of the human LTC4S gene located on chromosome 5 (chr5:179, 220, 986-179, 223, 512 described as SID #47 in Supplementary Table S1). Although its four introns were described as retained introns in database, only the second intron was compatible in size. (A–C) A zoom capture is presented below each gene map including small RNA-seq data obtained from ENCODE in three cell lines (GM12878, K562 and prostate cells and fractionated data when available). Sequence and structure determined using RNALfold are also indicated.

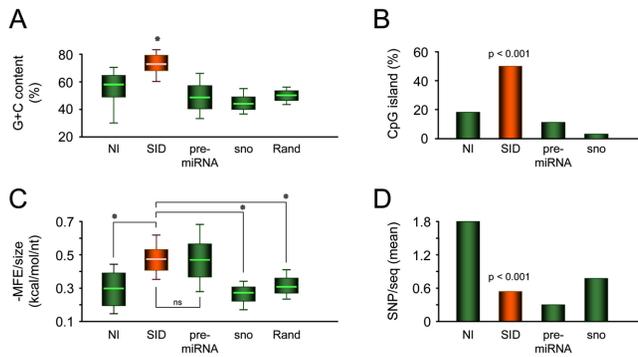


Figure 2. Structural and sequence features of newly identified SID. (A) G+C content of NI (introns; $n = 1713$), SID ($n = 56$), pre-miRNAs (from miRBase v17; $n = 1426$), snoRNAs (sno, from snoRNABase v3; $n = 402$), and random sequences (generated as described in Materials and Methods; $n = 10\,000$), all with similar nucleotide length, were analyzed and plotted as box-and-whisker diagrams. (B) Percentage of sequences that mapped in an annotated CpG island region. (C) Corrected MFE (minimum free energy) was calculated for each sequence by dividing the absolute value obtained from RNALfold program (see Materials and Methods) by the corresponding sequence length. Data were plotted as box-and-whisker diagrams. (D) Total number of SNP was calculated for each dataset using Galaxy tools and the dbSNP132 track and a mean of SNP per sequence was then plotted. (A and B) * mean $P < 0.001$ (Student's t -test). (C and D) Statistical analyses were performed using Chi-squared test.

pre-miRNAs and SID. Interestingly, searching for genomic common single-nucleotide polymorphism (SNP), we found that the mean number of SNP *per* intron is almost three-times higher in NI than in SID (Figure 2D). These results are consistent with a higher selective pressure on SID than on NI, equivalent to that on pre-miRNA (42) and snoRNAs (43), and suggestive of the functional importance of these sequences.

Our approach was further validated with the finding that previously identified human mirtrons (44) like pre-miR-1234 (SID #44) and pre-miR-6743 (SID #26) were present in our data set (Supplementary Table S1). Moreover, two of the identified host genes, TCIRG1 (T-Cell, Immune Regulator 1, ATPase, H⁺ Transporting, Lysosomal V0 Subunit A3) hosting SID #12 and CPSF1 (Cleavage and polyadenylation specificity factor 1) hosting SID #44 and #46, were also host genes for agoRNAs (19).

Experimental validation of SID candidates

To test whether the 56 potentially new human SID could be detected as discrete entities, we developed low-density oligonucleotide arrays hybridized with RNA extracted from cultured cells and fractionated in size (Supplementary Figure S2). Oligonucleotides matching the 56 SID were spotted as duplicates. Positive controls for hybridization were designed to detect highly expressed RNU6-1 snRNA (U6) and the hsa-miR-21 expressed in most tumour cell lines. Negative controls consisted of randomly chosen oligonucleotides. Oligonucleotides matching exons of housekeeping mRNAs served as controls for size fractionation since exon-containing mRNAs should not be found in the small RNA fraction. Arrays were hybridized using radio-labelled RNA probes extracted from six different human cell types and fractionated to retain RNAs shorter than 200 nt in

length. Hybridization signals were quantified and normalized as described in the Material and Methods section. Out of the 56 candidates, 36 showed signals above baseline levels in at least one cell type, representing our best candidates as new human SID (Supplementary Table S1). Although SID candidates were selected using a specific workflow that integrates alternative splicing events, we first validated expression of their host gene as well as alternative splicing of their RNA precursors as measured by detection of intron-retaining or -spliced isoforms, using radioactive RT-PCR (Supplementary Figures S3 and S4). All but 3 (#1, #27, #47) of the 36 SID candidates showed retention and excision events, even weak, in at least one of the six chosen cell lines. Candidates #6 and #8 showed intermediate amplification products in all cell lines studied and #36 showed intermediate amplification products in MDA-MB-231 cells (Supplementary Figures S3), consistent with alternative acceptor or donor splicing sites already reported for candidates #6 and #8, respectively (<http://fasterdb.ens-lyon.fr>). Cloning and sequencing of candidate #36 intermediate product also identified an alternative non-canonical (GT→GA) donor splicing site (Supplementary Figures S5).

As a whole, 33 SID (Supplementary Table S1) were detectable in at least one of the six cell lines, and showed variable levels consistent with introns originating from alternative splicing events being cell-type- or differentiation-stage-specific. For example, SID #10, #24 and #39 showed higher levels in ERalpha-negative, highly invasive, fibroblast-like MDA-MB-231 than in ERalpha-positive, weakly invasive, luminal epithelial-like MCF-7. In contrast, SID #42 showed higher levels in MCF-7 cells than in MDA-MB-231. Similarly, SID #39 and #43 showed higher levels in undifferentiated human myoblasts than in their differentiated counterpart, whereas SID #8 and #53 increased during myoblast differentiation. Some SID, such as #10, #16, #24 and #39, were detected in all six cell types at high levels. In contrast, high levels of #42 and #46 were a hallmark of cancer cells whatever the origin of the tumour, whereas levels of #21 were low in cancer cell lines compared to normal myogenic cells.

Biogenesis of miRNAs from SID candidates

In order to decipher the pathways involved in the biogenesis of miRNAs from SID, we used the 19 candidates that showed signals above baseline levels in the ENCODE common cell line K562, in which we knocked down all major players involved in miRNAs, mirtrons, simtrons and snoRNAs biogenesis, using RNA interference (see Material and Methods section and Supplementary Table S2). Size fractionated RNAs extracted from K562 were radio-labelled and then used as probes on low-density oligonucleotide arrays to assess the impact of knock down assays on individual SID levels. RT-PCR and western blots controlling efficiency of RNA interference are shown in Supplementary Figure S6A and B, respectively. We used shRNA directed against luciferase gene as a mock control (Supplementary Table S2).

We first performed experiments to knock down Dicer and Ago2 (Figure 3). In addition to controls for size fractionation already mentioned earlier for the experimental val-

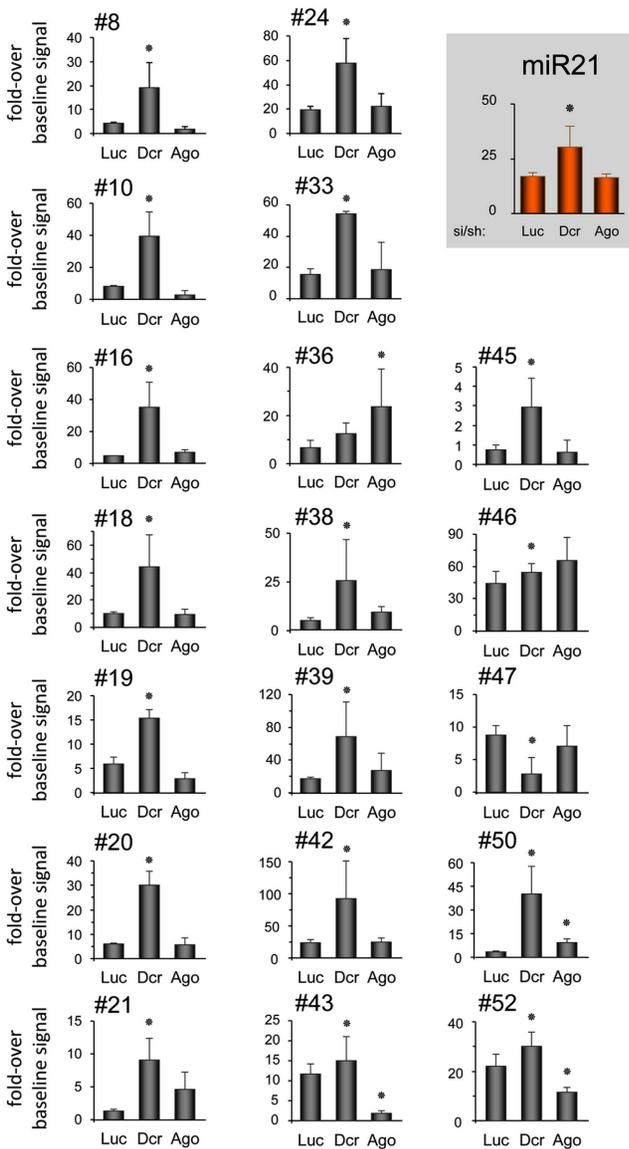


Figure 3. Role of Dicer and Ago2 in biogenesis of SID. Oligonucleotide arrays developed in Supplementary Figure S2, and spotted with the SID that shows a signal above baseline in K562 cells (Supplementary Table S1) were hybridized with radio-labeled small RNA isolated from K562 cells in which Dicer, Ago2 or Luc were knocked down ($n = 4$). Hybridization signals were plotted as a fold-over baseline signal as described in Material and Methods section. Luc, control luciferase knockdown; Dcr, Dicer knockdown; Ago, Ago2 knockdown. The hsa-miR-21 (miR21) was used as positive control. Significant differences were assessed using Student's *t*-test ($*P < 0.05$). Error bars represent standard error at the mean (SEM).

idation of SID candidates, the increased signal following Dicer knockdown suggests that only SID, *i.e.* precursors of smaller ncRNAs that accumulate as the result of their impaired processing into smaller ncRNAs, are detectable by the method. For all SID candidates but 2 (#36 and #47), we observed an accumulation of SID in conditions of reduced levels of Dicer protein (Figure 3), suggesting that Dicer is indeed involved in SID maturation into smaller RNAs as shown for canonical miRNAs. As expected, decreased expression of Ago2 did not affect levels of most SID, although

increased levels of SID #36 might be explained by the previously described dicing activity of Ago2 that can replace Dicer in pre-miRNA processing (24,25).

We then used shRNAs directed against Drosha and DGCR8 that compose the microprocessor (45), snRNA70+PRP8 and U2AF65+PRP8 that are components of the splicing machinery (18), and DBR1 that debranches intron lariats (46) (Figure 4). The 19 SID showed distinct dependency for one to several of these proteins, allowing to classify them in five different categories: (i) the members of the first category correspond to classical pre-miRNAs and showed dependency on the microprocessor only. It included pre-miR-21, used here as a control, and SID #16, #18, #19 and #20; (ii) the second category was composed of one mirtron candidate (#8), as it showed dependency on the debranching enzyme DBR1 but not on the microprocessor components or snRNP70+PRP8 and U2AF65+PRP8; (iii) SID #45 was dependent on Drosha and snRNP proteins, as shown previously for simtrons (17,18); (iv) a new category of candidates, including #24, #39, #42, #43, #46, #47 and #52, showed dependency on both the microprocessor (Drosha and DGCR8) and DBR1; (v) unclassified SID for which no dependency could clearly and/or significantly be determined, corresponding to SID #10, #21, #33, #36, #38 and #50.

Physical interaction of SID candidates with key factors involved in biogenesis of miRNAs

Because knock down experiments could lead to indirect effects, and to test whether proteins involved in these pathways act directly on the processing of SID into miRNAs (Figure 5), we performed native RIP to detect interactions of a given SID with proteins implicated in the canonical miRNA pathway (microprocessor proteins Drosha and DGCR8), in intron-lariat debranching (DBR1) or in splicing (snRNP70, component of the U1 complex; U2AF65, component of the U2 complex). We also assessed the interaction of SID with Ago2, which is a prerequisite to suggest loading into the RISC complex and hence, to suggest a regulatory function for the resulting miRNA. Western blots controlling efficiency of immunoprecipitation assays are shown in Figure 5A. Results shown in Figure 5B are expressed as a fold enrichment relative to immunoprecipitation using an irrelevant antibody.

As anticipated from data described above, SID #16, #18, #19 and #20 were significantly associated with Drosha and DGCR8, but not with DBR1 nor snRNP70 or U2AF65. Conversely, SID #45, which we classified as a simtron-like, showed a significant association with Drosha and snRNP70 as expected (17,18). SID #43, which we have demonstrated the dependence on Drosha and DBR1 by RNA interference (Figure 4), was indeed associated with these two proteins. However, SID #8 that we classified as a mirtron according to its dependency on DBR1 but not on the microprocessor (14), unexpectedly co-precipitated with Drosha and snRNP70 as well as with DBR1.

Because of the direct interactions described above, SID #8, that we defined as a mirtron, should be referred to as a mirtron-like RNA since it also interacts with snRNP70, which has never been described in the case of mirtrons. Sim-

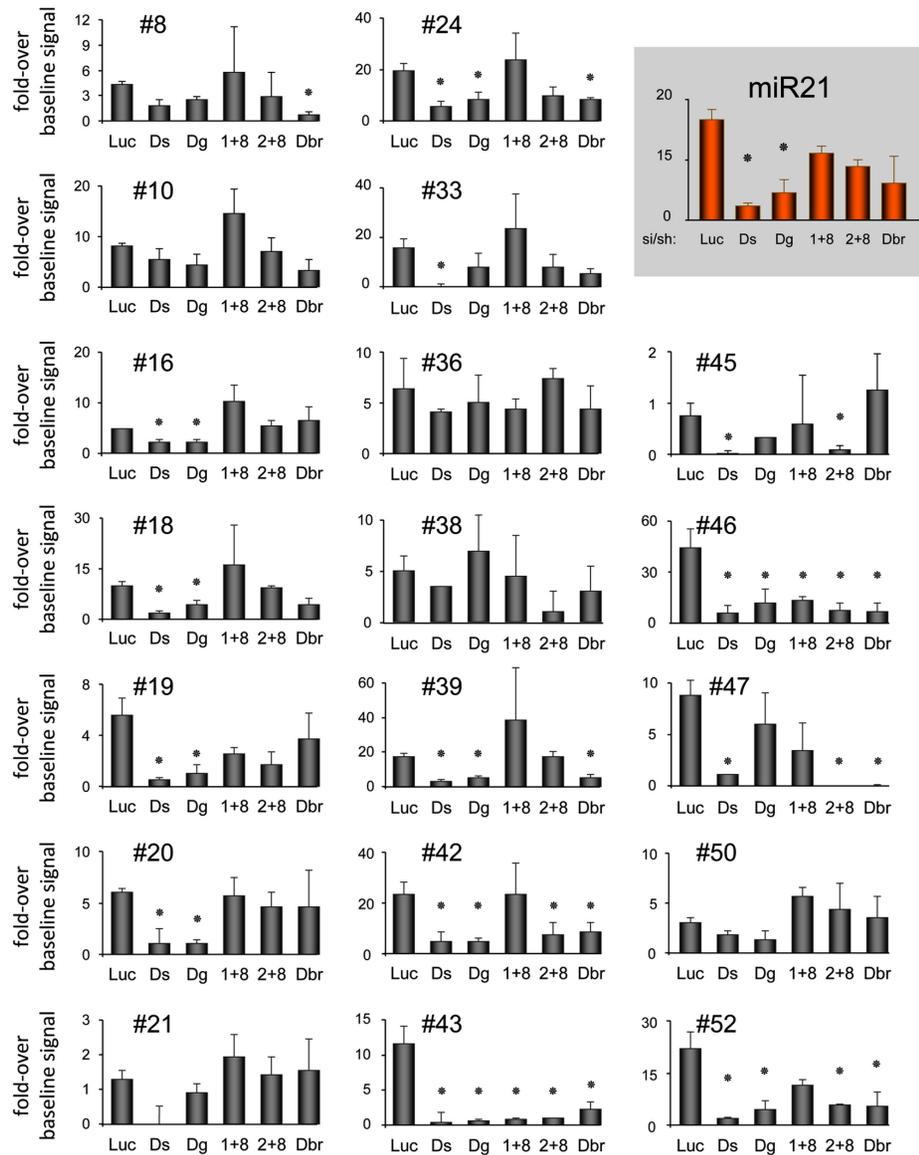


Figure 4. Role of Drosha, DGCR8, snRNP70+PRP8, U2AF65+PRP8 and DBR1 in biogenesis of SID. Oligonucleotide arrays developed in Supplementary Figure S2, and spotted with the SID that shows a signal above baseline in K562 cells (Supplementary Table S1) were hybridized with radio-labelled small RNA isolated from K562 cells in which Drosha, DGCR8, snRNP70+PRP8, U2AF65+PRP8, DBR1 or Luc were knocked down ($n = 4$). Hybridization signals were plotted as a fold-over baseline signal as described in Material and Methods section. Luc, control luciferase knockdown; Ds, Drosha knockdown; Dg, DGCR8 knockdown; 1+8, snRNP70+PRP8 knockdown; 2+8, U2AF65+PRP8 knockdown; Dbr, DBR1 knockdown. The hsa-miR-21 (miR21) was used as positive control. Significant differences were assessed using Student's *t*-test ($*P < 0.05$). Error bars represent standard error at the mean (SEM).

ilarly, SID #45, which we classified as a simtron, should be referred to as a simtron-like because of its interaction with Ago2 that is dispensable in the simtron pathway (17). Of note, it is not so surprising to find the full SID associated with the Ago2 complex since Dicer, Ago2 and TAR RNA-binding protein 2 (TRBP2) that compose the RISC-loading complex (RLC) are first loaded together on the stem-loop before cleavage of the loop by Dicer and incorporation of the guide-strand into the RISC complex and repression of the target RNAs (47–50).

Data from Figures 3, 4 and 5 relative to the biogenesis of SID candidates (interference experiments and native RIP

assays) are summarized in Supplementary Table S3 in which we propose a categorization of SID candidates.

Mature miRNAs originating from SID candidates

In order to confirm that SID candidates are genuine precursors of smaller ncRNAs, in particular mature miRNAs, we developed northern blot dedicated to the detection of endogenous small RNAs (see Material and Methods section). We have arbitrarily selected SID candidates from each category described above, i.e. #8, #10, #18, #43, #45 and #46. As shown in Figure 6, northern blot analyses further revealed that endogenous miRNAs originating from these six

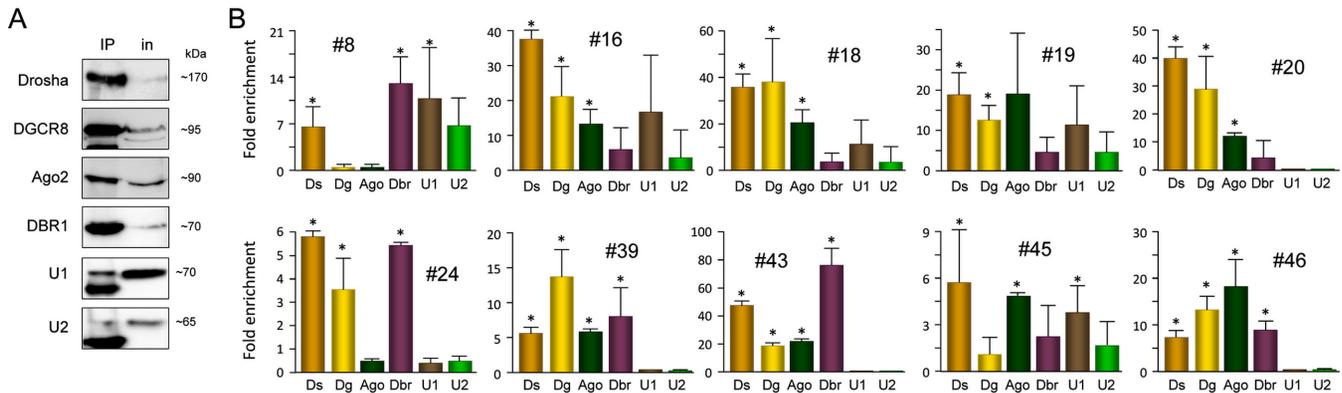


Figure 5. SID interactions with miRNA pathway key factors. (A) Native RIP was performed using the indicated antibodies and the effective protein precipitation was confirmed by western blotting. IP, immunoprecipitation; in, 5% input. (B) Co-precipitated RNAs were extracted as described in Material and Methods section ($n = 2$). Quantitative PCR were performed using specific primer pairs (Supplementary Table S2) to amplify SID #8, #16, #18, #19, #20, #24, #39, #43, #45 and #46. Amount of immunoprecipitated material is expressed as fold enrichment compared to irrelevant antibody. Ds, Drossha; Dg, DGCR8; Ago, Ago2; Dbr, Dbr1; U1, snRNP70; U2, U2AF65. Significant differences with irrelevant antibody were assessed using Student's t -test ($*P < 0.05$). Error bars represent standard error at the mean (SEM).

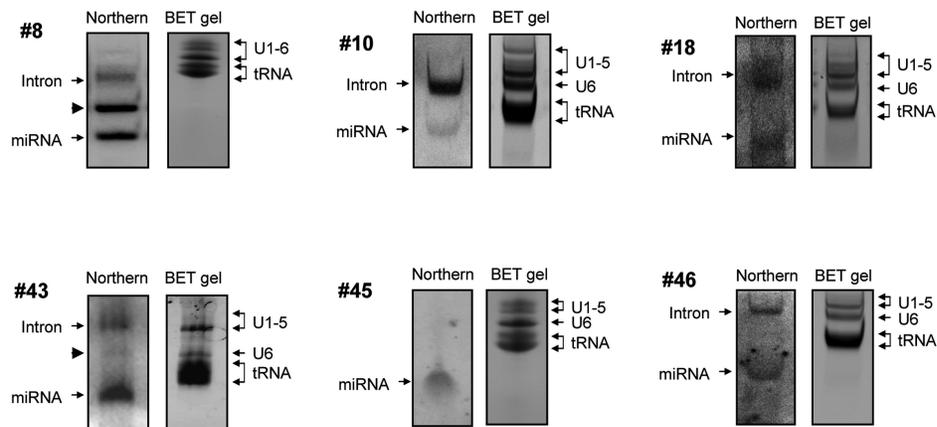


Figure 6. Small RNA isolated from HEK-293 cells were analysed by northern blot using specific radio-labelled probes complementary to SID #8, #10, #18, #43, #45 and #46. The corresponding acrylamide gels stained with ethidium bromide are shown. The upper and lower bands corresponding to the SID and its miRNA product, respectively, are indicated by small arrows, and the intermediate product for #8 and #43 is indicated by an arrow-head.

SID were indeed detectable in the small fraction of RNAs. These results strongly suggest that these introns meet the requirements to be classified as new human SID. For reasons that are not clear yet, the intermediate product corresponding to the hairpin (pre-miRNA) has never been detected in this study, except for SID #8 and SID #43. Likewise, many pre-miRNAs have not been detected in other studies either, like for hsa-miR-148a (51), hsa-miR-17 (52) or mmu-miR-103 (53). It must also be emphasized that only the mature small RNA with the expected size for miRNAs was detected using a probe against SID #45 that we classified as a simtron-like. Indeed, in the case of simtrons, the primary sequence (pri-miRNA) does not correspond to the intron alone but to the pre-messenger RNA (which is larger than 200 nt and therefore eliminated from the size fractionation).

Characterization of a new mature miRNA processed from SID #43

SID #43 corresponds to intron 1 of *SRA1* (Steroid Receptor RNA Activator 1) gene for which we already reported the importance of alternative splicing in other contexts (30,31,38,54). In order to clone and sequence the exact mature sequence of the miRNA processed from intron 1 of *SRA1*, we used the method that we recently developed (35) based on *in vitro* transcription, RNA pull down and adapted RACE-PCR techniques. As shown in Figure 7A, we were able to characterize the mature SRA miRNA, allowing to assess its expression profile in myoblasts and myotubes using stem-loop RT-qPCR. We used hsa-miR-1224 and hsa-miR-199a2 as positive controls. During myogenic differentiation of human healthy myoblasts to myotubes, we observed the previously reported increase in miR-199a2 during myogenic differentiation (55). Consistent with the previously described decrease in SID #43 levels during the course of myoblast differentiation (Supplementary Table S1), levels of the mature miRNA derived from SID #43

tion (9,11,13,14,16,44,64–70) but never on splicing events. Despite these approaches being different from the one proposed herein, we also computationally identified SID producing hsa-miR-6743 (#26) and hsa-miR-1234 (#44), previously described as mirtrons (66). Although levels of these SID were too weak for further experimental validation, we now show that they originate from an alternative splicing event of the transcripts RIC8A (Resistance to Inhibitors of Cholinesterase 8 homolog A) and CPSF1 (Cleavage and Polyadenylation Specificity Factor), respectively. *CPSF1* gene encodes the largest subunit of the CPSF complex, a multi-subunit complex that plays a central role in 3' processing of pre-mRNA (71). The gene locus appears to be very complex in that it contains around 50 introns, of which, interestingly, 11 are alternatively spliced (72). We identified an additional SID produced by alternative splicing of CPSF1 (#46), which is distinct from the previously reported SID producing hsa-miR-1234 (9), hsa-miR-939 (16), and an agotron (19) also hosted in this gene. It is not surprising to find multiple SID in genes containing high numbers of introns, which in addition are subject to alternative splicing. Along the same lines, it is not surprising to find a higher density of SID on chromosome 19 (12/56) that contains the highest density of genes (8,73), but more importantly, the highest number of genes with short introns (<http://www.ncbi.nlm.nih.gov/genome>) compatible with their processing and formation of stable hairpin structures (29). Like *CPSF1*, chromosome 19 appears to host both SID and canonical miRNAs precursors as it contains the largest cluster of pre-miRNAs identified in mammals, where at least 46 different pre-miRNAs were found within a 100-kb region (74).

We did not restrict our search to introns of protein-coding genes since the possibility that non-coding RNAs might also produce shorter RNAs has already been proposed (59,75,76) as exemplified by *SNHG* (Small Nucleolar RNA Host Gene) family of genes and *GAS5* (Growth Arrest Specific 5) that are non-coding multi-snoRNAs hosting genes (77), or by *MIR17HG* also known as the miR-17/92 cluster (78). Consistent with this, we identified SID candidates #39, #16 and #43 respectively hosted in *LINC00085* (long intergenic non-protein coding RNA 85, NR_024330), *RHBG* (Rh family, B glycoprotein, which variant 2 is a non-coding isoform, NR_026549) and *SRA1* [which we previously described as both non-coding RNAs and mRNA isoforms, see ref. (30,31,38,54)]. Interestingly, *RHBG*, similarly to what was shown for *SRA*, could represent a new bifunctional RNA since retention of its intron disrupts the ORF whereas, as for classical alternatively spliced mRNA, splicing preserves the ORF (30,31,58). In both cases, we have now shown that alternative splicing of introns not only favours the production of a protein product, but also promotes the formation of a miRNA. Adding a degree of complexity, alternative donor and acceptor splicing sites might also contribute to the diversification of the RNA forms produced, *i.e.* mRNA, long ncRNA and/or short ncRNA, as it might be the case for SID #6, #8 and #36. Whether the shorter SID generated by alternative donor or acceptor sites can also produce a miRNA in certain cellular contexts remains to be tested.

Although bioinformatic prediction of human SID eased their identification based on conservation, splicing origin or structural features [see refs. (9,11,13,14,16,44,64,65,67) and the present study], computational analysis is likely to generate large amounts of false positive predictions. Indeed, out of the first 19 mirtrons previously identified (9), 3 were experimentally tested and one of them (hsa-miR-1233 precursor) was suggested not to be a mirtron or a simtron (68). Thus, it stresses the need to experimentally test these SID candidates and characterize their biogenesis pathways.

Our primary goal was to identify SID originating from alternative splicing events being cell-type- or differentiation-stage-specific. However, we identified intronic miRNAs, which biogenesis should not rely in theory on splicing mechanisms. Indeed, intronic miRNAs are either transcribed independently of, or share promoter with, their host gene (79–82), but in both cases transcripts that correspond to pri-miRNAs should be much longer than the intron alone. However, SID #18, which we defined as a miRNA, had the exact size of the spliced intron, suggesting that, at least for this candidate, splicing mechanism was a prerequisite to produce the corresponding intronic miRNA. Apparent discrepancy also exists if we consider the levels of SID-hosting transcripts isoforms (Supplementary Figures S3 and S4) and that of the SID candidates (summarized in Supplementary Table S1). However, the fact that only SID and not mature miRNAs are detected on the oligonucleotide arrays could explain this discrepancy for SID that are rapidly processed in a given cellular context. In addition, the stability and fate, and hence the detectable levels, of the spliced intron are independent of that of the longer transcripts from which they derived. Such example already exists with *EGFL7* (epidermal growth factor domain 7) that hosts the hsa-miR-126 in intron 6. *EGFL7* was shown to be highly expressed in lung and to a lesser extent in heart whereas the opposite was observed for hsa-miR-126 [(83) and <http://www.microrna.org>].

In agreement with splicing proposed to be required for the production of non-canonical precursors of miRNAs, we report the experimentally-tested dependence of several SID on components of the splicing machinery (snRNP70+PRP8, U2AF65+PRP8) and of the microprocessor (Drosha, DGCR8), as exemplified by SID #8, #43, #45 and #46. Several other studies have reported a crosstalk between spliceosome and microprocessor. One striking example comes from plants, along which production of ath-miR-163 in *Arabidopsis thaliana* is favoured by components of the snRNP70 that potentiate miRNA biogenesis by hiding the proximal polyA signal and thus prevent polyadenylation and premature cleavage/degradation of the pri-miRNA (84). Other examples of such interplay have been reported in mammals, but almost invariably represent a competition between components of the spliceosome and those of the microprocessor (61,84–86). Interestingly, Drosha and DGCR8 were found in association with the spliceosome machinery in a supraspliceosome context (87), where knockdown of Drosha resulted in increased splicing whereas inhibition of the spliceosome enhanced miRNA production (87).

To date, the canonical pathway (Drosha/DGCR8–XPO5–Dicer–Ago2) is considered as the main biogenesis

pathway for the majority of miRNAs (88). In the last few years, alternative pathways for miRNA biogenesis have also been reported. The first identified, the mirtron pathway, bypasses the cleavage by the microprocessor (9,15) since spliced-introns, first debranched by DBR1, are directly used as pre-miRNAs and exported in the cytoplasm by XPO5. The simtron pathway (17,18) uses the snRNP70 small nuclear ribonucleoprotein to recruit Drosha that will directly process the stem-loop structure from the pre-mRNA to produce the pre-miRNA. This mechanism is independent of DGCR8 that usually recognizes pri-miRNAs and directs Drosha RNase III activity to release the pre-miRNA hairpin (89). We now report that these pathways are likely to be more complicated and inter-connected than previously expected, at least for miRNAs produced from short introns (SID). We described four intronic pre-miRNAs, 1 simtron-like and 1 mirtron-like, but also several other SID which biogenesis relies on both microprocessor (Drosha and DGCR8) and debranching enzyme (DBR1). Nonetheless, we did not exclude the possibility that miRNA originating from SID could be produced through multiple pathways, either depending of the host-gene expression/regulation, or relying on the intron excision/retention balance.

Among the new SID that we experimentally validated, we focused on the interesting case of the miRNA contained in the first intron of SRA RNA. SRA RNA was first described as a non-coding RNA (90) that acts as a transcriptional co-activator of nuclear receptors [For recent reviews see refs. (91,92)]. Since then, we described new isoforms produced by alternative splicing (30,31,54). SRA RNA was then proposed as the first member of a new class of bifunctional RNAs since both a protein and a non-coding RNA can be produced from the same genetic entity (30,31,54). We now report that intron 1 of SRA is detectable as a discrete entity in all 6 cell types that we have tested, although less abundantly in differentiated primary myotubes. Besides, we have recently found that coding (spliced intron 1) SRA isoforms were more abundant in differentiated human myogenic cells whereas non-coding (intron 1 retaining) SRA isoforms increased during myogenic differentiation and enhanced myogenic differentiation and myogenic conversion of non-muscle cells through the co-activation of MyoD activity (31). Interestingly, the amount of intron 1 and derived-mature miRNA were also higher in undifferentiated myoblasts suggesting that during myogenic differentiation, splicing of SRA intron 1 parallels the production of a SID. We previously showed that induction of muscle differentiation was not accompanied by a change in the ratio between non-coding and coding SRA isoforms in DM1 cells compared to non-affected muscle cells (31). We showed here that the levels of intron 1 of SRA were not affected during the course of differentiation of DM1 cells. While we highlighted a role of non-coding and coding SRA isoforms in the differentiation of normal cells (31), a role of the SID (intron 1) and more likely of the mature miRNA derived from it, cannot be excluded. Indeed, the impact of the SRA-mature miRNA could be direct (by inhibiting the non coding SRA isoforms retaining intron 1) or indirect (through other targets remaining to be identified). Interestingly, among the top five predicted target genes, and although not much data is available on *FBXO41* (F box

protein 41) or *PRX* (Periaxin) in the literature, the three other candidates have a potential link with myoblast differentiation. *DAGLA* (Diacylglycerol Lipase Alpha) expression is significantly increased upon myogenic differentiation *in vitro* (93). *NECTINI* (Nectin Cell Adhesion Molecule 1) belongs to the family of cellular adhesion molecules that play central roles in cell adhesion, cell motility, proliferation and survival, and contributes to the morphogenesis and differentiation of many cell types and tissues (94,95). Increased levels of *MECP2* (MEthyl CpG binding Protein 2) during myogenic differentiation have been linked to global heterochromatin rearrangements that occur during and are a prerequisite for myogenic terminal differentiation (96). Of course, the whole range of miRNA targets originating from the SID identified in this study still remains to be identified in a given cellular context.

As a whole, we present evidence for the existence of new classes of Short Intron-Derived ncRNAs that we termed SID, that regroup agotron, mirtron and simtron RNAs. In addition, we uncovered new types of miRNAs which biogenesis appears to be splicing-dependant but nevertheless closely connected to the classical miRNA pathway.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Jean-François Ouimette and H el ene Quach for valuable comments on the manuscript, members of the team for critical reading and Drs Vincent Mouly and Denis Furling (Myology Institute, Paris, France) for providing primary human myogenic cells and RNAs.

FUNDING

AFM (Association Fran aise contre les Myopathies) [16055]; ANR (Agence Nationale pour la Recherche), LNCC (Ligue Nationale Contre le Cancer) and Gefluc (to C.F. team). Funding for open access charge: AFM (Association Fran aise contre les Myopathies).

Conflict of interest statement. None declared.

REFERENCES

- Clark, M.B. and Mattick, J.S. (2011) Long noncoding RNAs in cell biology. *Semin. Cell Dev. Biol.*, **22**, 366–376.
- He, L. and Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.*, **5**, 522–531.
- Berezikov, E. (2011) Evolution of microRNA diversity and regulation in animals. *Nat. Rev. Genet.*, **12**, 846–860.
- Pais, H., Moxon, S., Dalmay, T. and Moulton, V. (2011) Small RNA discovery and characterisation in eukaryotes using high-throughput approaches. *Adv. Exp. Med. Biol.*, **722**, 239–254.
- Kawamata, T. and Tomari, Y. (2010) Making RISC. *Trends Biochem. Sci.*, **35**, 368–376.
- Mattick, J.S. and Gagen, M.J. (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.*, **18**, 1611–1630.
- Mattick, J.S. (2001) Non-coding RNAs: the architects of eukaryotic complexity. *EMBO Rep.*, **2**, 986–991.

8. Hube,F and Francastel,C. (2015) Mammalian introns: when the junk generates molecular diversity. *Int. J. Mol. Sci.*, **16**, 4429–4452.
9. Berezikov,E., Chung,W.J., Willis,J., Cuppen,E. and Lai,E.C. (2007) Mammalian mirtron genes. *Mol. Cell*, **28**, 328–336.
10. Berezikov,E., Liu,N., Flynt,A.S., Hodges,E., Rooks,M., Hannon,G.J. and Lai,E.C. (2010) Evolutionary flux of canonical microRNAs and mirtrons in *Drosophila*. *Nat. Genet.*, **42**, 6–9.
11. Chung,W.J., Agius,P., Westholm,J.O., Chen,M., Okamura,K., Robine,N., Leslie,C.S. and Lai,E.C. (2011) Computational and experimental identification of mirtrons in *Drosophila melanogaster* and *Caenorhabditis elegans*. *Genome Res.*, **21**, 286–300.
12. Flynt,A.S., Greimann,J.C., Chung,W.J., Lima,C.D. and Lai,E.C. (2010) MicroRNA biogenesis via splicing and exosome-mediated trimming in *Drosophila*. *Mol. Cell*, **38**, 900–907.
13. Glazov,E.A., Kongsuwan,K., Assavalapsakul,W., Horwood,P.F., Mitter,N. and Mahony,T.J. (2009) Repertoire of bovine miRNA and miRNA-like small regulatory RNAs expressed upon viral infection. *PLoS One*, **4**, e6349.
14. Okamura,K., Hagen,J.W., Duan,H., Tyler,D.M. and Lai,E.C. (2007) The mirtron pathway generates microRNA-class regulatory RNAs in *Drosophila*. *Cell*, **130**, 89–100.
15. Ruby,J.G., Jan,C.H. and Bartel,D.P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature*, **448**, 83–86.
16. Valen,E., Preker,P., Andersen,P.R., Zhao,X., Chen,Y., Ender,C., Dueck,A., Meister,G., Sandelin,A. and Jensen,T.H. (2011) Biogenic mechanisms and utilization of small RNAs derived from human protein-coding genes. *Nat. Struct. Mol. Biol.*, **18**, 1075–1082.
17. Havens,M.A., Reich,A.A., Duelli,D.M. and Hastings,M.L. (2012) Biogenesis of mammalian microRNAs by a non-canonical processing pathway. *Nucleic Acids Res.*, **40**, 4626–4640.
18. Janas,M.M., Khaled,M., Schubert,S., Bernstein,J.G., Golan,D., Veguilla,R.A., Fisher,D.E., Shomron,N., Levy,C. and Novina,C.D. (2011) Feed-forward microprocessing and splicing activities at a microRNA-containing intron. *PLoS Genet.*, **7**, e1002330.
19. Hansen,T.B., Veno,M.T., Jensen,T.I., Schaefer,A., Damgaard,C.K. and Kjems,J. (2016) Argonaute-associated short introns are a novel class of gene regulators. *Nat. Commun.*, **7**, 11538.
20. Ender,C., Krek,A., Friedlander,M.R., Beitzinger,M., Weinmann,L., Chen,W., Pfeffer,S., Rajewsky,N. and Meister,G. (2008) A human snoRNA with microRNA-like functions. *Mol. Cell*, **32**, 519–528.
21. Lee,Y.S., Shibata,Y., Malhotra,A. and Dutta,A. (2009) A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.*, **23**, 2639–2649.
22. Abdelfattah,A.M., Park,C. and Choi,M.Y. (2014) Update on non-canonical microRNAs. *Biomol. Concepts*, **5**, 275–287.
23. Jha,A., Panzade,G., Pandey,R. and Shankar,R. (2015) A legion of potential regulatory sRNAs exists beyond the typical microRNAs microcosm. *Nucleic Acids Res.*, **43**, 8713–8724.
24. Cifuentes,D., Xue,H., Taylor,D.W., Patnode,H., Mishima,Y., Cheloufi,S., Ma,E., Mane,S., Hannon,G.J., Lawson,N.D. *et al.* (2010) A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. *Science*, **328**, 1694–1698.
25. Cheloufi,S., Dos Santos,C.O., Chong,M.M. and Hannon,G.J. (2010) A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature*, **465**, 584–589.
26. Havens,M.A., Reich,A.A. and Hastings,M.L. (2014) Drosha promotes splicing of a pre-microRNA-like alternative exon. *PLoS Genet.*, **10**, e1004312.
27. Liu,J. and Maxwell,E.S. (1990) Mouse U14 snRNA is encoded in an intron of the mouse cognate hsc70 heat shock gene. *Nucleic Acids Res.*, **18**, 6565–6571.
28. Mattick,J.S. and Makunin,I.V. (2005) Small regulatory RNAs in mammals. *Hum. Mol. Genet.*, **14 Spec No 1**, R121–R132.
29. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
30. Hube,F., Guo,J., Chooniedass-Kothari,S., Cooper,C., Hamedani,M.K., Dibrov,A.A., Blanchard,A.A., Wang,X., Deng,G., Myal,Y. *et al.* (2006) Alternative splicing of the first intron of the steroid receptor RNA activator (SRA) participates in the generation of coding and noncoding RNA isoforms in breast cancer cell lines. *DNA Cell Biol.*, **25**, 418–428.
31. Hube,F., Velasco,G., Rollin,J., Furling,D. and Francastel,C. (2011) Steroid receptor RNA activator protein binds to and counteracts SRA RNA-mediated activation of MyoD and muscle differentiation. *Nucleic Acids Res.*, **39**, 513–525.
32. Furling,D., Coiffier,L., Mouly,V., Barbet,J.P., St Guily,J.L., Taneja,K., Gourdon,G., Junien,C. and Butler-Browne,G.S. (2001) Defective satellite cells in congenital myotonic dystrophy. *Hum. Mol. Genet.*, **10**, 2079–2087.
33. Velasco,G., Hube,F., Rollin,J., Neuillet,D., Philippe,C., Bouzinba-Segard,H., Galvani,A., Viegas-Pequignot,E. and Francastel,C. (2010) Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl. Acad. Sci. U.S.A.*
34. Pall,G.S. and Hamilton,A.J. (2008) Improved northern blot method for enhanced detection of small RNA. *Nat. Protoc.*, **3**, 1077–1084.
35. Hube,F. and Francastel,C. (2015) "Pocket-sized RNA-Seq": a method to capture new mature microRNA produced from a genomic region of interest. *Non-Coding RNA*, **1**, 127–138.
36. Ferri,F., Bouzinba-Segard,H., Velasco,G., Hube,F. and Francastel,C. (2009) Non-coding murine centromeric transcripts associate with and potentiate Aurora B kinase. *Nucleic Acids Res.*, **37**, 5071–5080.
37. Varkonyi-Gasic,E. and Hellens,R.P. (2011) Quantitative stem-loop RT-PCR for detection of microRNAs. *Methods Mol. Biol.*, **744**, 145–157.
38. Cooper,C., Guo,J., Yan,Y., Chooniedass-Kothari,S., Hube,F., Hamedani,M.K., Murphy,L.C., Myal,Y. and Leygue,E. (2009) Increasing the relative expression of endogenous non-coding Steroid Receptor RNA Activator (SRA) in human breast cancer cells using modified oligonucleotides. *Nucleic Acids Res.*, **37**, 4518–4531.
39. Sugnet,C.W., Kent,W.J., Ares,M. Jr. and Haussler,D. (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.*, 66–77.
40. Moore,M.J. (2002) Nuclear RNA turnover. *Cell*, **108**, 431–434.
41. Zhang,B., Han,D., Korostelev,Y., Yan,Z., Shao,N., Khrameeva,E., Velichkovsky,B.M., Chen,Y.P., Gelfand,M.S. and Khaitovich,P. (2016) Changes in snoRNA and snRNA abundance in the human, chimpanzee, macaque, and mouse brain. *Genome Biol. Evol.*, **8**, 840–850.
42. Quach,H., Barreiro,L.B., Laval,G., Zidane,N., Patin,E., Kidd,K.K., Kidd,J.R., Bouchier,C., Veuille,M., Antoniewski,C. *et al.* (2009) Signatures of purifying and local positive selection in human miRNAs. *Am. J. Hum. Genet.*, **84**, 316–327.
43. Bhartiya,D., Talwar,J., Hasija,Y. and Scaria,V. (2012) Systematic curation and analysis of genomic variations and their potential functional consequences in snoRNA loci. *Hum. Mutat.*, **33**, E2367–E2374.
44. Wen,J., Ladewig,E., Shenker,S., Mohammed,J. and Lai,E.C. (2015) Analysis of nearly one thousand mammalian mirtrons reveals novel features of dicer substrates. *PLoS Comput. Biol.*, **11**, e1004441.
45. Gregory,R.I., Yan,K.P., Amuthan,G., Chendrimada,T., Doratotaj,B., Cooch,N. and Shiekhattar,R. (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature*, **432**, 235–240.
46. Kim,H.C., Kim,G.M., Yang,J.M. and Ki,J.W. (2001) Cloning, expression, and complementation test of the RNA lariat debranching enzyme cDNA from mouse. *Mol. Cell*, **11**, 198–203.
47. MacRae,I.J., Ma,E., Zhou,M., Robinson,C.V. and Doudna,J.A. (2008) In vitro reconstitution of the human RISC-loading complex. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 512–517.
48. Chendrimada,T.P., Gregory,R.I., Kumaraswamy,E., Norman,J., Cooch,N., Nishikura,K. and Shiekhattar,R. (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature*, **436**, 740–744.
49. Tang,G. (2005) siRNA and miRNA: an insight into RISCs. *Trends Biochem. Sci.*, **30**, 106–114.
50. Maniataki,E. and Mourelatos,Z. (2005) A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev.*, **19**, 2979–2990.
51. Kozłowska,E., Krzyżosiak,W.J. and Koscińska,E. (2013) Regulation of huntingtin gene expression by miRNA-137, -214, -148a, and their respective isomiRs. *Int. J. Mol. Sci.*, **14**, 16999–17016.
52. Jin,H.Y., Gonzalez-Martin,A., Miletic,A.V., Lai,M., Knight,S., Sabouri-Ghomi,M., Head,S.R., Macauley,M.S., Rickert,R.C. and

- Xiao, C. (2015) Transfection of microRNA mimics should be used with caution. *Front. Genet.*, **6**, 340.
53. Ohno, M., Shibata, C., Kishikawa, T., Yoshikawa, T., Takata, A., Kojima, K., Akanuma, M., Kang, Y.J., Yoshida, H., Otsuka, M. *et al.* (2013) The flavonoid apigenin improves glucose tolerance through inhibition of microRNA maturation in miRNA103 transgenic mice. *Sci. Rep.*, **3**, 2553.
54. Chooniedass-Kothari, S., Emberley, E., Hamedani, M.K., Troup, S., Wang, X., Czosnek, A., Hube, F., Mutawe, M., Watson, P.H. and Leygue, E. (2004) The steroid receptor RNA activator is the first functional RNA encoding a protein. *FEBS Lett.*, **566**, 43–47.
55. Jung, S.Y., Malovannaya, A., Wei, J., O'Malley, B.W. and Qin, J. (2005) Proteomic analysis of steady-state nuclear hormone receptor coactivator complexes. *Mol. Endocrinol.*, **19**, 2451–2465.
56. Kapusta, A. and Feschotte, C. (2014) Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends Genet.*, **30**, 439–452.
57. Ulveling, D., Francastel, C. and Hube, F. (2011) When one is better than two: RNA with dual functions. *Biochimie*, **93**, 633–644.
58. Ulveling, D., Francastel, C. and Hube, F. (2011) Identification of potentially new bifunctional RNA based on genome-wide data-mining of alternative splicing events. *Biochimie*, **93**, 2024–2027.
59. Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Hum. Mol. Genet.*, **15** Spec No 1, R17–R29.
60. Rearick, D., Prakash, A., McSweeney, A., Shepard, S.S., Fedorova, L. and Fedorov, A. (2011) Critical association of ncRNA with introns. *Nucleic Acids Res.*, **39**, 2357–2366.
61. Brown, J.W., Marshall, D.F. and Echeverria, M. (2008) Intronic noncoding RNAs and splicing. *Trends Plant Sci.*, **13**, 335–342.
62. Jung, C.H., Hansen, M.A., Makunin, I.V., Korbie, D.J. and Mattick, J.S. (2010) Identification of novel non-coding RNAs using profiles of short sequence reads from next generation sequencing data. *BMC Genomics*, **11**, 77.
63. Zhou, H. and Lin, K. (2008) Excess of microRNAs in large and very 5' biased introns. *Biochem. Biophys. Res. Commun.*, **368**, 709–715.
64. Chen, D., Meng, Y., Yuan, C., Bai, L., Huang, D., Lv, S., Wu, P., Chen, L.L. and Chen, M. (2011) Plant siRNAs from introns mediate DNA methylation of host genes. *RNA*, **17**, 1012–1024.
65. Joshi, P.K., Gupta, D., Nandal, U.K., Khan, Y., Mukherjee, S.K. and Sanan-Mishra, N. (2012) Identification of mirtrons in rice using MirtronPred: a tool for predicting plant mirtrons. *Genomics*, **99**, 370–375.
66. Ladewig, E., Okamura, K., Flynt, A.S., Westholm, J.O. and Lai, E.C. (2012) Discovery of hundreds of mirtrons in mouse and human small RNA data. *Genome Res.*, **22**, 1634–1645.
67. Meng, Y. and Shao, C. (2012) Large-scale identification of mirtrons in Arabidopsis and rice. *PLoS One*, **7**, e31163.
68. Schamberger, A., Sarkadi, B. and Orban, T.I. (2012) Human mirtrons can express functional microRNAs simultaneously from both arms in a flanking exon-independent manner. *RNA Biol.*, **9**, 1177–1185.
69. St, L.G., Shtokalo, D., Tackett, M.R., Yang, Z., Eremina, T., Wahlestedt, C., Urcuqui-Inc Seilheimer, B., McCaffrey, T.A. and Kapranov, P. (2012) Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. *BMC Genomics*, **13**, 504.
70. Zhu, Q.H., Spriggs, A., Matthew, L., Fan, L., Kennedy, G., Gubler, F. and Helliwell, C. (2008) A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Res.*, **18**, 1456–1465.
71. Mandel, C.R., Bai, Y. and Tong, L. (2008) Protein factors in pre-mRNA 3'-end processing. *Cell Mol. Life Sci.*, **65**, 1099–1122.
72. Thierry-Mieg, D. and Thierry-Mieg, J. (2006) AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol.*, **7** (Suppl 1), S12–S14.
73. Grimwood, J., Gordon, L.A., Olsen, A., Terry, A., Schmutz, J., Lamerdin, J., Hellsten, U., Goodstein, D., Couronne, O., Tran-Gyamfi, M. *et al.* (2004) The DNA sequence and biology of human chromosome 19. *Nature*, **428**, 529–535.
74. Bortolin-Cavaille, M.L., Dance, M., Weber, M. and Cavaille, J. (2009) C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. *Nucleic Acids Res.*, **37**, 3464–3473.
75. Mercer, T.R., Dinger, M.E., Bracken, C.P., Kolle, G., Szubert, J.M., Korbie, D.J., skarian-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**, 1639–1650.
76. Mercer, T.R., Gerhardt, D.J., Dinger, M.E., Crawford, J., Trapnell, C., Jeddeloh, J.A., Mattick, J.S. and Rinn, J.L. (2011) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.*
77. Williams, G.T., Mourtada-Maarabouni, M. and Farzaneh, F. (2011) A critical role for non-coding RNA GAS5 in growth arrest and rapamycin inhibition in human T-lymphocytes. *Biochem. Soc. Trans.*, **39**, 482–486.
78. de, P.L., Yao, E., Callier, P., Faivre, L., Drouin, V., Cariou, S., Van, H.A., Genevieve, D., Goldenberg, A., Oufadem, M. *et al.* (2011) Germline deletion of the miR-17 approximately 92 cluster causes skeletal and growth defects in humans. *Nat. Genet.*, **43**, 1026–1030.
79. Bartel, D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
80. Kim, Y.K. and Kim, V.N. (2007) Processing of intronic microRNAs. *EMBO J.*, **26**, 775–783.
81. Monteys, A.M., Spengler, R.M., Wan, J., Tecedor, L., Lennox, K.A., Xing, Y. and Davidson, B.L. (2010) Structure and activity of putative intronic miRNA promoters. *RNA*, **16**, 495–505.
82. Ramalingam, P., Palanichamy, J.K., Singh, A., Das, P., Bhagat, M., Kassab, M.A., Sinha, S. and Chattopadhyay, P. (2014) Biogenesis of intronic miRNAs located in clusters by independent transcription and alternative splicing. *RNA*, **20**, 76–87.
83. Fitch, M.J., Campagnolo, L., Kuhnert, F. and Stuhlmann, H. (2004) Egl7, a novel epidermal growth factor-domain gene expressed in endothelial cells. *Dev. Dyn.*, **230**, 316–324.
84. Szweykowska-Kulinska, Z., Jarmolowski, A. and Vazquez, F. (2013) The crosstalk between plant microRNA biogenesis factors and the spliceosome. *Plant Signal. Behav.*, **8**, e26955.
85. Mattioli, C., Pianigiani, G. and Pagani, F. (2014) Cross talk between spliceosome and microprocessor defines the fate of pre-mRNA. *Wiley. Interdiscip. Rev. RNA*, **5**, 647–658.
86. Melamed, Z., Levy, A., shwal-Fluss, R., Lev-Maor, G., Mekahel, K., Atias, N., Gilad, S., Sharan, R., Levy, C., Kadener, S. *et al.* (2013) Alternative splicing regulates biogenesis of miRNAs located across exon-intron junctions. *Mol. Cell*, **50**, 869–881.
87. Agranat-Tamir, L., Shomron, N., Sperling, J. and Sperling, R. (2014) Interplay between pre-mRNA splicing and microRNA biogenesis within the supraspliceosome. *Nucleic Acids Res.*, **42**, 4640–4651.
88. Ha, M. and Kim, V.N. (2014) Regulation of microRNA biogenesis. *Nat. Rev. Mol Cell Biol.*, **15**, 509–524.
89. Seitz, H. and Zamore, P.D. (2006) Rethinking the microprocessor. *Cell*, **125**, 827–829.
90. Lanz, R.B., McKenna, N.J., Onate, S.A., Albrecht, U., Wong, J., Tsai, S.Y., Tsai, M.J. and O'Malley, B.W. (1999) A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell*, **97**, 17–27.
91. Colley, S.M. and Leedman, P.J. (2011) Steroid receptor RNA activator - a nuclear receptor coregulator with multiple partners: Insights and challenges. *Biochimie*, **93**, 1966–1972.
92. Cooper, C., Vincett, D., Yan, Y., Hamedani, M.K., Myal, Y. and Leygue, E. (2011) Steroid receptor RNA activator bi-faceted genetic system: heads or tails? *Biochimie*, **93**, 1973–1980.
93. Zhao, D., Pond, A., Watkins, B., Gerrard, D., Wen, Y., Kuang, S. and Hannon, K. (2010) Peripheral endocannabinoids regulate skeletal muscle development and maintenance. *Eur. J. Transl. Myol.*, **1**, 167–179.
94. Takai, Y., Ikeda, W., Ogita, H. and Rikitake, Y. (2008) The immunoglobulin-like cell adhesion molecule nectin and its associated protein afadin. *Annu. Rev. Cell Dev. Biol.*, **24**, 309–342.
95. Takai, Y., Miyoshi, J., Ikeda, W. and Ogita, H. (2008) Nectins and nectin-like molecules: roles in contact inhibition of cell movement and proliferation. *Nat. Rev. Mol Cell Biol.*, **9**, 603–615.
96. Brero, A., Easwaran, H.P., Nowak, D., Grunewald, I., Cremer, T., Leonhardt, H. and Cardoso, M.C. (2005) Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J. Cell Biol.*, **169**, 733–743.