

# Identification of differentially expressed non-coding RNAs in embryonic stem cell neural differentiation

Konstantinia Skreka<sup>1</sup>, Simon Schafferer<sup>1</sup>, Irina-Roxanna Nat<sup>2</sup>, Marek Zywicki<sup>1</sup>, Ahmad Salti<sup>2</sup>, Galina Apostolova<sup>2</sup>, Matthias Griehl<sup>1</sup>, Mathieu Rederstorff<sup>3</sup>, Georg Dechant<sup>2</sup> and Alexander Hüttenhofer<sup>1,\*</sup>

<sup>1</sup>Division of Genomics and RNomics, Biocenter, Medical University Innsbruck, Innrain 80/82, A-6020 Innsbruck,

<sup>2</sup>Institute for Neuroscience, Medical University Innsbruck, Anichstrasse 35/MZA, A-6020 Innsbruck and

<sup>3</sup>Lorraine University, Biopôle, UMR 7214 AREMS, 9 Avenue de la Forêt de Haye, BP 184, 54506 Vandoeuvre-lès-Nancy, France

Received January 24, 2012; Revised February 27, 2012; Accepted March 24, 2012

## ABSTRACT

Protein-coding genes, guiding differentiation of ES cells into neural cells, have extensively been studied in the past. However, for the class of ncRNAs only the involvement of some specific microRNAs (miRNAs) has been described. Thus, to characterize the entire small non-coding RNA (ncRNA) transcriptome, involved in the differentiation of mouse ES cells into neural cells, we have generated three specialized ribonucleo-protein particle (RNP)-derived cDNA libraries, i.e. from pluripotent ES cells, neural progenitors and differentiated neural cells, respectively. By high-throughput sequencing and transcriptional profiling we identified several novel miRNAs to be involved in ES cell differentiation, as well as seven small nucleolar RNAs. In addition, expression of 7SL, 7SK and vault-2 RNAs was significantly up-regulated during ES cell differentiation. About half of ncRNA sequences from the three cDNA libraries mapped to intergenic or intragenic regions, designated as interRNAs and intrRNAs, respectively. Thereby, novel ncRNA candidates exhibited a predominant size of 18–30 nt, thus resembling miRNA species, but, with few exceptions, lacking canonical miRNA features. Additionally, these novel intrRNAs and interRNAs were not only found to be differentially expressed in stem-cell derivatives, but also in primary cultures of hippocampal neurons and

astrocytes, strengthening their potential function in neural ES cell differentiation.

## INTRODUCTION

In recent years, the number of proposed non-coding RNA (ncRNA) transcripts has dramatically been rising. For example, in the human genome there is an estimated number of up to 450 000 ncRNA transcripts predicted (1). In agreement, a recent study designated as ENCODE project (2), which focused on 1% of the human genome in high resolution, revealed that up to 90% of the human genome might be transcribed, with only 1.5% of RNA transcripts encoding for proteins. Thus, it has been proposed that the remaining 88.5% of RNA transcripts might serve as a source for regulatory ncRNAs (3). However, it is currently still unclear which of the 450 000 predicted ncRNA candidates, encoded on the human genome, are functional and which ones represent spurious transcription products or degradation intermediates (4). Therefore, it is important to clearly identify the functional portion of the ncRNA transcriptome in model organisms. Several features might be employed to filter out and preselect functional, regulatory ncRNAs from a background of spurious transcription/degradation intermediates such as (i) analysis of differential expression of ncRNAs during cell differentiation and development, (ii) ncRNA expression in disease or (iii) ncRNA expression during development. In addition, since most functional ncRNAs are known to bind to proteins forming ribonucleo-protein particles (RNPs), isolation by RNPs

\*To whom correspondence should be addressed. Tel: +43 512 9003 70250; Fax: +43 512 9003 73100; Email: alexander.huetttenhofer@i-med.ac.at  
Present address:

Marek Zywicki, Laboratory of Computational Genomics, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznan, Poland.

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

might increase the likelihood for identifying functional ncRNAs (5).

For embryonic stem (ES) cell maintenance and pluripotency, non-coding RNAs have recently emerged as important regulators of gene expression (6–8). Up till now, specific microRNAs, a class of small regulatory ncRNAs, sized 21–24 nt (9,10), have been investigated in neural development during ES cell differentiation. In particular, expression of the ES cell specific miR-290 cluster, harboring miRNAs-290, –291, –293, –294 and –295, respectively, has been shown to be significantly down-regulated upon differentiation (11), while regulating *de novo* methylation in ES cells by repressing the transcriptional repressor Rb2 (12,13). Inhibition of mir-145 in human ES cells has been shown to reduce their capacity for differentiation (14), while maturation repression of the pre-let7 miRNA precursor by the Lin28 protein is a mechanism blocking commitment to neural fate (15). In addition, microRNA-array analysis in ES cells versus differentiated cells reveals specific microRNA expression signatures (16,17), thus implying transcriptome changes during differentiation related to microRNA function. Notably, lack of expression of pre-microRNA-processing proteins Dicer (18,19) and DGCR8 (20) was shown to result in severe differentiation defects.

In order to identify the complete set of small ncRNAs involved in neural differentiation of mouse ES cells *in vitro*, we have generated specialized, RNP-derived cDNA libraries, as previously described (21,22), for three differentiation stages. Transcriptional profiling by high-throughput sequencing revealed the presence of numerous differentially expressed known and novel ncRNAs, which predominantly locate to intergenic and intronic regions of the mouse genome. The majority of the novel ncRNAs exhibited a sequence length bias of 18–30 nt. For selected, newly identified ncRNA transcripts differential expression in primary hippocampal neurons and astrocytes was confirmed by real-time PCR.

## MATERIALS AND METHODS

### RNP library generation

RNP libraries were generated as previously described (21,22). Briefly, cells were lysed and cell extracts were size-fractionated on 10–30% glycerol gradients. Subsequently, glycerol gradients were fractionated and phenol–chloroform precipitated. The extracted RNA was 3'-C-tailed using poly (A) polymerase from yeast (Epicenter, Madison, USA). A 19-mer 5'-adaptor (–GTCAGCAATCCCTAACCAG, bold and underlined are ribonucleotides) was ligated by T4 RNA ligase (Fermentas, St. Leon-Rot, Germany) to the C-tailed RNA. The RNAs were reverse transcribed by using an anchor primer (5'-AGGAGCCATCGTATGTCTCGGGGGGGH) and amplified by PCR at 53°C annealing, for 25 cycles using the following primers: 5'-libPCR GTCAGCAATCCC TAACGAG, 3'-libPCR AGGAGCCATCGTATGTCTCG. The cDNA was PAGE purified and size selected from 20 to 400 bp. cDNA was cloned into the pGEM-T vector, (Promega, Mannheim, Germany) for diagnostic Sanger sequencing before high-throughput sequencing. For Solexa

sequencing, additional bar-coded forward primers were added to cDNAs by PCR:

5'-AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTATACGTCAGCAATCCCTAACGAG-3' for the ES library, 5'-AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTCATCGTCAGCAATCCCTAACGAG for the NP library 5'ATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCCGATCTGACAGTCAGCAATCCCTAACGAG for the N/G library. The following reverse primer was employed for all libraries: 5'-CAAGCAGAAGACGGCATAACGAGCTCTTCCGATCTAGGAGCCATCGTATGTCTCG 3'. The eluted cDNAs were analyzed by high-throughput sequencing employing the Solexa (Illumina) platform. Sequencing was performed with the Genome Analyzer GAII at FASTERIS SA, Plan-les-Quates, (Switzerland). Reads were generated as single reads with a maximum of 76 bp in length.

### ES cell cultures

The mouse ES cell line used was ES–E14TG2a (passage 10–14) (23). ES cells were cultured as previously described (24), with some modifications, on a feeder layer of mouse embryonic fibroblasts; feeder cells were inactivated with 10 µg/ml mitomycin C for 2.5 h. ES cell culture medium (ESCM) consisted on Knockout Dulbecco's Modified Eagle's medium with Knockout serum replacement (15%), 2 mM Glutamax, 0.1 mM MEM non-essential amino acids, 0.5 mM 2-mercaptoethanol (all from Gibco Invitrogen Corporation, Paisley, UK) supplemented with 1000 U/ml leukaemia inhibitory factor (LIF) (Chemicon, Temecula, CA). Cell cultures were maintained in a humidified atmosphere with 5% CO<sub>2</sub>, at 37°C. Medium was changed every day. ES cell colonies were passaged with 0.25% Trypsin–EDTA (Gibco) at a 1:3–1:5 split ratio every other day.

### Neural differentiation cultures

Neural differentiation of ES cells occurred in discrete steps, which include neural induction, neural proliferation and neural specification, modified from previously published protocols (25–27).

Neural induction medium (NIM) consisted of Dulbecco's Modified Eagle Medium DMEM/F12 with Glutamax (Gibco) containing 25 µg/ml insulin, 100 µg/ml transferrin, 5 ng/ml sodium selenite, 2.5 µg/ml fibronectin (all from Sigma-Aldrich St. Louis, MO, USA) and 0.1 mM MEM nonessential amino acids solution.

Neural proliferation medium (NPM) consisted of DMEM/F12 with Glutamax, 1% N2 supplement and 0.1 mM MEM nonessential amino acids, supplemented with Fibroblast Growth Factor 2 (FGF2) and Epidermal Growth Factor (EGF) (10 ng/ml each).

Neuronal differentiation medium (NDM) medium consisted of Neurobasal medium with 2% (vol/vol) B27 supplement without vitamin A (Gibco), 2 mM glutamax, 1 µM cAMP, 200 µM ascorbic acid and 20 µg/ml laminin (all from Sigma). Media were changed every 48 h.

ES cells, at Day 0 of neural differentiation (D0), were separated from feeder cells, dissociated with Accutase (Sigma), and cultured as cell suspension ( $2 \times 10^5$  cells/ml) in ESCM without LIF in non-adhesive bacterial-grade dish. Cells that became floating cellular aggregates were cultured in the same medium for 2 days. The adherent colony culture was initiated by plating cellular aggregates on plastic or glass surface coated with polyornithine (Sigma,  $15 \mu\text{g/ml}$  for plastic and  $50 \mu\text{g/ml}$  for glass, 1 h RT) and laminin ( $20 \mu\text{g/ml}$  ON  $4^\circ\text{C}$ ), and further cultured in NIM. Attached aggregates flattened over 1–2 days and columnar primitive neuroepithelial (NE) cells developed and formed neural rosettes at approximately D6. The neural tube-like rosettes formed during neural induction were used as a selection criterion; neural colonies were mechanically separated, collected and dissociated with Accutase, and cultured as cell suspension or in adherent culture onto polyornithine/laminin-coated plates in NPM ( $2 \times 10^5$  cells/ml) for 4 days, for the proliferation of the neural progenitor cells. For neural differentiation/specification, neural progenitors were dissociated with Accutase and plated onto polyornithine/laminin-coated plates or coverslips, followed by culturing in NDM.

Colonies and cellular morphologies were monitored by phase contrast microscopy. ES or neural colonies were mechanically detached or split under a stereomicroscope. Three independent experiments were performed. Samples for RNP libraries generation, RNA extraction or immunocytochemistry were taken at the following time-points: D0 (stage designated as ES), D10 (stage designated as NP) and D20 (stage designated as N/G).

#### Initial mapping and analysis of the cDNA libraries

Prior to analysis, the pool of sequencing reads has been divided into three libraries according to sequence barcodes: ATAC for ES library, CATC for NP library and GACA for N/G library. The analysis of the individual libraries has been performed employing the APART pipeline (automated pipeline for annotation of RNA transcripts) (28). All APART modules were used. For removal of adapter sequences, we employed a 20-nt window of the adaptor sequences with three allowed mismatches and included removal of the C-tail, located between the insert and 3'-adaptor. For downstream analysis, only reads with a minimal length of 18 nt by which both adaptors had been detected, were used. For sequence reads, containing only a partial 3'-adaptor sequence, at least six terminal bases were required to match the C-tail, or the beginning of the 3'-adaptor sequence. Next, sequences were aligned to the mouse genome (mm9), allowing for a single mismatch. Assembly and annotation was performed employing the APART pipeline, employing default settings.

#### Transcriptional profiling of three cDNA libraries from ES cells

For differential expression analysis the non-clustered list of contigs generated by APART, has been employed. The identification of overlapping contigs between libraries has been performed through BEDTools package (29) with 50% overlap threshold. Contigs appearing in at least two libraries have been selected for subsequent differential

expression analysis. Next, the contig redundancy caused by allowing for multiple read matches has been removed by using in-house perl scripts employing APART clustering data. Normalization of the read number has been carried out utilizing the package edgeR (30) from Bioconductor (31). In order to test for differences in expression edgeR and several in-house scripts have been used.

#### miRNA prediction

For identification of novel miRNAs, present in our dataset, the software miRDeep2 (32) was employed, in particular scripts mapper.pl and miRDeep2.pl (33). Mapping was performed by using the default parameter set and the FASTQ files from each library, except that the adaptor sequences were already removed prior to analysis. The miRDeep2 script was then applied to find potential miRNAs in the data set based on the suggested parameters. Results were compared to data assembled in the differential expression analysis by using the package rtracklayer (34) of the Bioconductor project.

#### Real-time PCR

Total RNA was isolated from mouse ES cells, primary hippocampal neurons and primary astrocytes with TRI Reagent (Sigma-Aldrich, Vienna, Austria) according to the manufacturer's protocol. An amount of 500 ng of total RNA were poly-A tailed and reverse transcribed to cDNA using the microRNA first strand synthesis kit (Agilent Technologies, Böblingen, Germany) technologies, following the manufacturer's protocol.

The cDNA was used as template for the real-time PCR on ES cells, primary hippocampal neurons and astrocytes. The universal reverse primer provided with the kit was used together with the following forward primers in 5'–3' orientation:

```
mmu_interRNA30: CATCCCACTTCTGACACCACAAA
mmu_interRNA32: GCCTACCAAACCCTGTCAA
mmu_intraRNA16: CCATAAGGTAGGCATTGCA
mmu_interRNA6: GAACGTGAGCTGGGTTTAGACCGTC
mmu_interRNA36: GATCCCACTTCTGACACCA
mmu_interRNA26: GGGGAATCTGACTGTCTA
mmu_intraRNA11: CAAGATAAGATTTCCCGG
mmu_intraRNA140: AGTCCCTGCCCTTTGTACA
mmu_interRNA143: TCTAAATTTCCACCTTTTTTCAGTTT
mmu_interRNA105: GCGTTGGTGGTATAGTGGTGAGCATAGCT
mmu_interRNA140: GGGCCGGCGGCGGCGGCGG
mmu_interRNA35: ATCCCACTTCTGACACCA
mmu_interRNA68: TCCTCGTTAGTATAGTGGTTAGTA
```

Primers were ordered from Sigma-Aldrich. Real-time PCR was performed using Power SYBR<sup>®</sup> Green PCR Master Mix (Applied Biosystems, Darmstadt, Germany). Reactions were performed at  $60^\circ\text{C}$  annealing for 1 min and for 40 cycles. Normalization was performed with U6.

## RESULTS AND DISCUSSION

### Generation of three specialized RNP-derived cDNA libraries from mouse ES cell neural differentiation

In order to identify novel ncRNAs that regulate ES cell neural differentiation we generated three RNP-derived

cDNA libraries from three stages of neural differentiation of mouse ES cells: (i) pluripotent stem cells (designated as ES), (ii) neural progenitor cells (designated as NP) and (iii) a differentiated heterogeneous population of neurons and glial cells (designated as N/G), as previously reported (21,22). Briefly, mouse ES cells (expressing the pluripotent markers *Nanog* and *Oct4*) were differentiated into NP (expressing the neural markers *nestin* and *Sox1*), which subsequently spontaneously differentiated to neurons (expressing *Tau*) and glial cells (expressing *Gfap* and *Osp*) by withdrawing growth factors FGF2 and EGF (see ‘Materials and Methods’ section and Supplementary Figure S1A).

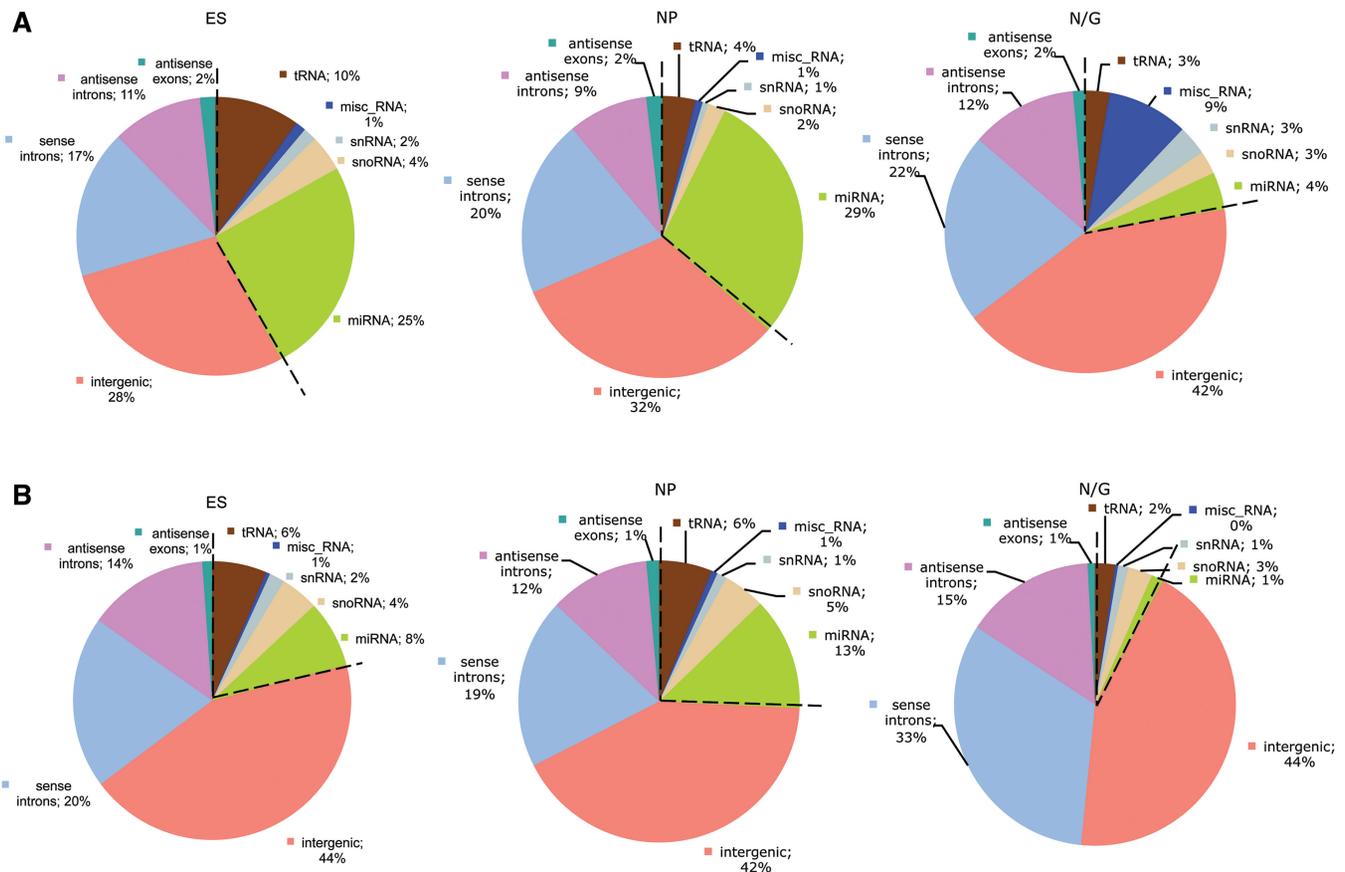
Real-time PCR analysis of marker genes revealed high levels of *Oct4* and *Nanog* in ES cells, while their levels were significantly down-regulated at the NP and not detected at the N/G stage. The neural marker genes *Nestin* and *Sox1*, were significantly up-regulated at NP and N/G stages, while the neuronal (*Tau*), astrocytic (*Gfap*) and oligodendrocytic (*Osp*) gene expression was highly up-regulated in the last stage (as compared to ES stage; see Supplementary Figure S1B).

From each of the three stages, specialized RNP libraries encoding small ncRNAs were generated (see ‘Materials and Methods’ section). Subsequently, cDNA libraries were analyzed by high-throughput sequencing employing

the Solexa platform and ~26 Mio. sequence reads for the three libraries were obtained [sequences have been deposited in the Sequence Read Archive (NCBI) with the accession number: SRP008250].

### Transcriptional profiling of three cDNA libraries from ES cell neural differentiation

In order to determine differentially expressed RNA transcripts within the ES, NP and N/G stages, we bioinformatically analyzed sequence reads from the respective cDNA libraries by APART (automated pipeline for annotation of RNA transcripts), a bioinformatical algorithm recently developed in our lab (28). Thereby, differential expression analysis was based on barcoding of cDNAs from the three libraries prior to cloning and sequencing. Subsequently, two bioinformatical analyses were performed by employing APART: (i) we first annotated all sequence reads to known or unknown ncRNA species, e.g. miRNAs, snoRNAs, tRNAs, intranRNAs (i.e. intragenic RNAs) or interRNAs (intergenic RNAs; Figure 1A); (ii) we next grouped all identical sequences into ‘contigs’ and determined the distribution of these unique RNA transcripts in our libraries (Figure 1B). By this approach, 706 contigs with expression changes >2-fold between any of the three cDNA libraries and a mean



**Figure 1.** Distribution of ncRNA sequences between the three cDNA libraries. (A) Distribution based on sequence reads. (B) Distribution of contigs. The most abundant known ncRNAs are shown in addition to reads mapping in intergenic and intronic regions (i.e. intergenic, sense/antisense intron). ES = embryonic stem cells, NP = neural progenitors, N/G = neurons and glial cells.

expression value of at least 28 were identified (Supplementary File S1). Only those sequences were included which appeared at least in two of the three cDNA libraries in order to assess expression differences.

In this analysis, miRNAs represented the majority of differentially expressed known ncRNA sequences: while in the N/G library, miRNAs accounted for only 4% of all sequences, in ES or NP libraries 25–29% of sequences corresponded to miRNAs (Figure 1A). This is due to significantly fewer miRNA species being present at the N/G stage, as demonstrated by the number of unique contigs (1% versus 8–13%; Figure 1B), which hints to a more prominent participation of miRNAs in regulation of gene expression during the early stages of ES cell differentiation. Interestingly, a 9-fold increase in the abundance of sequences from the class of miscellaneous ncRNAs is observed in total sequence reads at the N/G stage, mainly due to an increase of 7SK and 7SL ncRNA expression (Figure 1A; see below).

The majority of novel ncRNA transcripts in all three cDNA libraries mapped to either intergenic or intronic regions of the mouse genome, which were designated as interRNAs (i.e. intergenic RNAs) or intraRNAs (i.e. intragenic RNAs such as intronically encoded RNAs), respectively (Figure 1B). Sequence reads representing intraRNAs or interRNAs showed a rather similar distribution in all cell stages (60% in ES and NP stages or 75% in the N/G stage, respectively). We have previously shown that intergenic regions between and intragenic regions within protein-coding genes harbour the majority of functional known ncRNAs (21), in particular snoRNAs and miRNAs, suggesting the presence of novel representatives of these ncRNA classes also in our cDNA libraries (see below).

From intraRNA species, about one-third of sequences mapped to introns in antisense orientation while the remaining sequences derived from the sense orientation of introns (Figure 1B). In addition, a small number of sequences were annotated to exons of mRNAs, in antisense orientation (Figure 1B). Previously, by analysis of a mouse brain cDNA library encoding small ncRNAs we have presented evidence that ncRNAs mapping in antisense orientation to introns or exon/intron boundaries might be involved in regulation of alternative splicing of respective genes (21).

### Validation of ES cell neural differentiation by transcriptional profiling of miRNAs

To validate the transcriptional profiling analysis, we investigated expression of known miRNA species, previously described to be involved in neural differentiation. Indeed, we verified expression of stem cell specific microRNAs in the ES library, in particular abundant expression of the mir-290 (11) and the miR-17-92 clusters (35), as well as miRNAs miR-199, miR-106, miR-299 and miR-214 (Figure 2 and Supplementary File S2), as previously reported (11,36). In addition, we observed expression of these miRNA species being significantly down-regulated upon ES cell differentiation in the NP and N/G population

of cells (see above and Figure 2), in agreement with earlier studies (35,36).

We were also able to verify expression of miRNAs miR-21, miR-15b, miR-669, miR-329, miR-335, miR-16, miR-411 and miR-541 in mouse ES cells (11,35) and, in addition, we observed that their expression was significantly down-regulated in the later stages of ES cell differentiation (i.e. NP, N/G stages; Figure 2). Notably, we find, as reported, the stem-cell-specific miR-302 (16,37–41) exclusively expressed in ES cells while completely absent at the NP and N/G stage (Supplementary File S2).

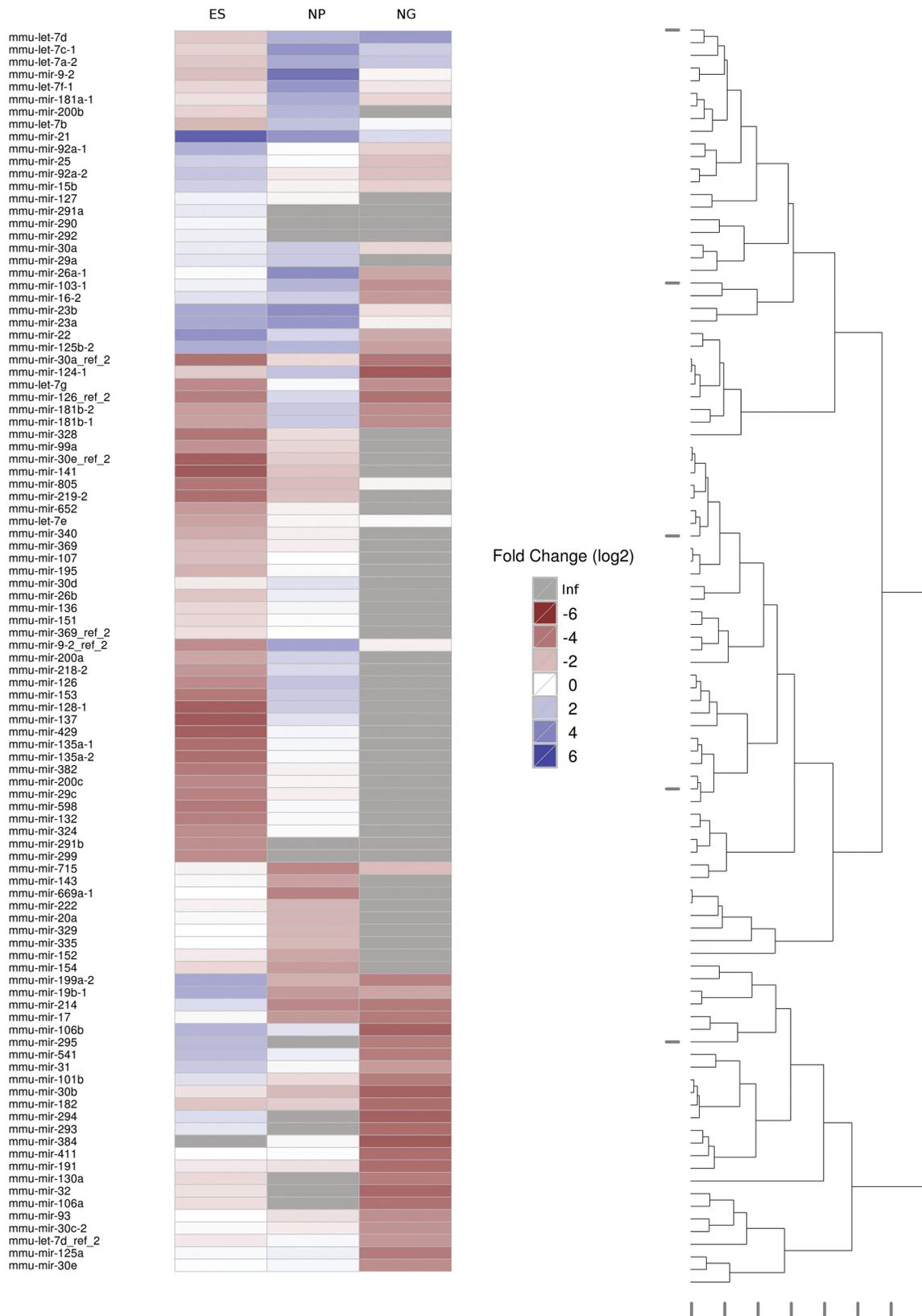
Conversely, expression of neural-specific microRNAs mir-9 and miR-124, as well as the miRNA let-7 family, which have been reported to modulate stem cell derived neurogenesis (42), was found to be significantly up-regulated at the NP stage in our analysis (Figure 2). For human ES cells, several microRNAs have been reported to be up-regulated upon spontaneous differentiation, such as mir-181a, mir-181b-2, mir-26a, mir-23b, mir-137 (39), consistent with their mouse homologues from this study.

Interestingly, we also identified microRNAs exhibiting expression patterns differing from previous studies. Thereby, expression of microRNAs miR-29a (16,36), miR-135a (39), miR-141 (40,43), miR-340 (41), miR-200c (16,37,43), miR-328 (36) and miR-30e (40) has been reported to be up-regulated in ES cells. In contrast, we find an increase in their expression at the NP stage (Figure 2). As for miR-135a, miR-141, miR-340, and miR-200c this is probably due to the fact that previous studies were carried out in human ES cells (16,39–41), which derive from a different developmental stage than mouse ES cells. In addition, expression analyses carried out in mouse ES cells (36) were only extended to embryoid bodies (EBs) and not to neural progenitor cells, which might explain the observed up-regulation of expression of miR-29a and miR-328 in NPs with respect to ES cells in our study.

Finally, we have identified expression of specific miRNAs, not previously being reported in ES cells, such as miR-715. Also, we observed microRNAs whose expression has been reported in mouse ES cells (35), without further analysis on their differential expression, to be significantly up-regulated in NPs and NGs (Figure 2). From these, miR-200a, miR-200b, miR-218-2, miR-153, miR-103-1, miR-26b, miR-30d, miR-324, miR-429 and miR-382, were found to be the highest expressed at the NP stage, while miR-805 and miR-674 were the highest expressed at the N/G stage (Figure 2 and Supplementary File S3). We also observed expression of miR-598 being up-regulated at the NP stage while miR-877, miR-682, miR-678 and miR-217 are expressed at the N/G stage, only. Taken together, our data demonstrate that the regulation of gene expression during neural differentiation by a large number of miRNAs might be even more complex than previously anticipated.

### Differentially expressed known ncRNAs from the three stages of ES cell neural differentiation

In addition to miRNAs, also several C/D box snoRNAs (small nucleolar RNAs) were found among the most



**Figure 2.** The 100 most differentially expressed miRNAs. Expression values were normalized against the mean expression value (257.6). Fold changes are depicted in a log<sub>2</sub> scale. Inf: infinite fold-change because expression was not detectable in the respective library. The dendrogram to the right indicates the Euclidean distance of the expression values.

prominent differentially expressed known ncRNA species (Figure 3), designated as SNORD12, 29, 31, 35, 74, 101, 104 and 115 (44). The majority of snoRNAs has been reported to guide covalent modifications of ribosomal RNAs (rRNAs) or small nuclear RNAs (snRNAs), respectively (45,46). Thereby, most snoRNAs, with few exceptions, are encoded within intronic sequences of protein-coding genes (45,47,48). Up till now, two classes of snoRNAs have been described, i.e. C/D box snoRNAs and H/ACA box snoRNAs. Unlike canonical snoRNAs, so-called 'orphan' snoRNAs lack any complementarity to rRNA or snRNA targets and might target other RNA molecules such as mRNAs (45,46,49).

We show that snoRNAs SNORD12, 29, 31, 74, 101 and 104 are abundantly expressed in ES cells, while their expression significantly decreases upon ES cell differentiation by ~10- to 30-fold (Figure 3). In contrast, expression of SNORD35 and 115 is low in ES cells while their expression is significantly up-regulated in NP cells or N/G cells (Figure 3). Interestingly, SNORD31, 101 and 115 have no validated rRNA targets (49,50) and might be involved in regulation of other RNA species such as mRNAs (see below). In addition, unlike canonical snoRNAs, SNORD12, 29, 31, 74, 104 and 115 are reported to be encoded within intronic sequences of non-protein coding transcripts or hypothetical open reading frames, which might not code for proteins (44,49,51,52); up till now, the role of these non-protein-coding snoRNA host gene transcripts has been elusive; this implies the possibility that, in addition to intronically encoded snoRNAs, also their respective host gene transcripts might be involved in the regulation of neural differentiation.

Interestingly, earlier reports have implicated SNORD115 (also designated as HBII-52) to be involved in brain development and disease (49,53). Thereby, SNORD115 has been proposed to target the brain-specific serotonin receptor 2C mRNA (49), thus regulating alternative splicing and/or editing of the serotonin receptor 2C pre-mRNA (54,55). We observed an ~100-fold increase in SNORD115 expression between mouse ES cells and differentiating cells, which was confirmed by northern blotting (Supplementary Figure S3). At the N/G stage, expression of SNORD115 is reduced by ~50-fold in comparison to the NP stage. However, the N/G library mainly consists of astrocytes and oligodendrocytes with fewer neurons (see above); since SNORD115 has previously been shown to be expressed exclusively in neurons, this might explain the lower abundance at the N/G stage.

Previously, it has been demonstrated that SNORD115 maps to the Prader-Willi Syndrome (PWS) locus on chromosome 15 (49). PWS is a neurodevelopmental disease with patients showing severe obesity and varying degrees of mental retardation (56). Notably, two snoRNAs from that locus have directly been implicated in the etiology of the disease, namely SNORD115 and SNORD116 (also designated as HBII-85), respectively (57-60). It is thus interesting to note that expression of SNORD115 is significantly up-regulated at a very early stage of development, i.e. already during neural differentiation.

In addition, expression of three other known ncRNAs, i.e. 7SL, 7SK and vault-2 RNA, respectively, is up-regulated

at the N/G stage compared to NP and ES cells (Figure 3). Thereby, 7SL RNA is an abundant ncRNA species, which serves as an integral part of the signal recognition particle (SRP) (61,62). The SRP has been shown to promote the insertion of proteins into the cellular membrane or to regulate protein secretion (63). These two processes might especially be required at the N/G stage where neural cells interconnect to form neural networks guided by receptors on the surface of cells, such as serotonin receptor 2C (see above); concomitantly, an increase in protein secretion might be required by secretion of neurotransmitter peptides.

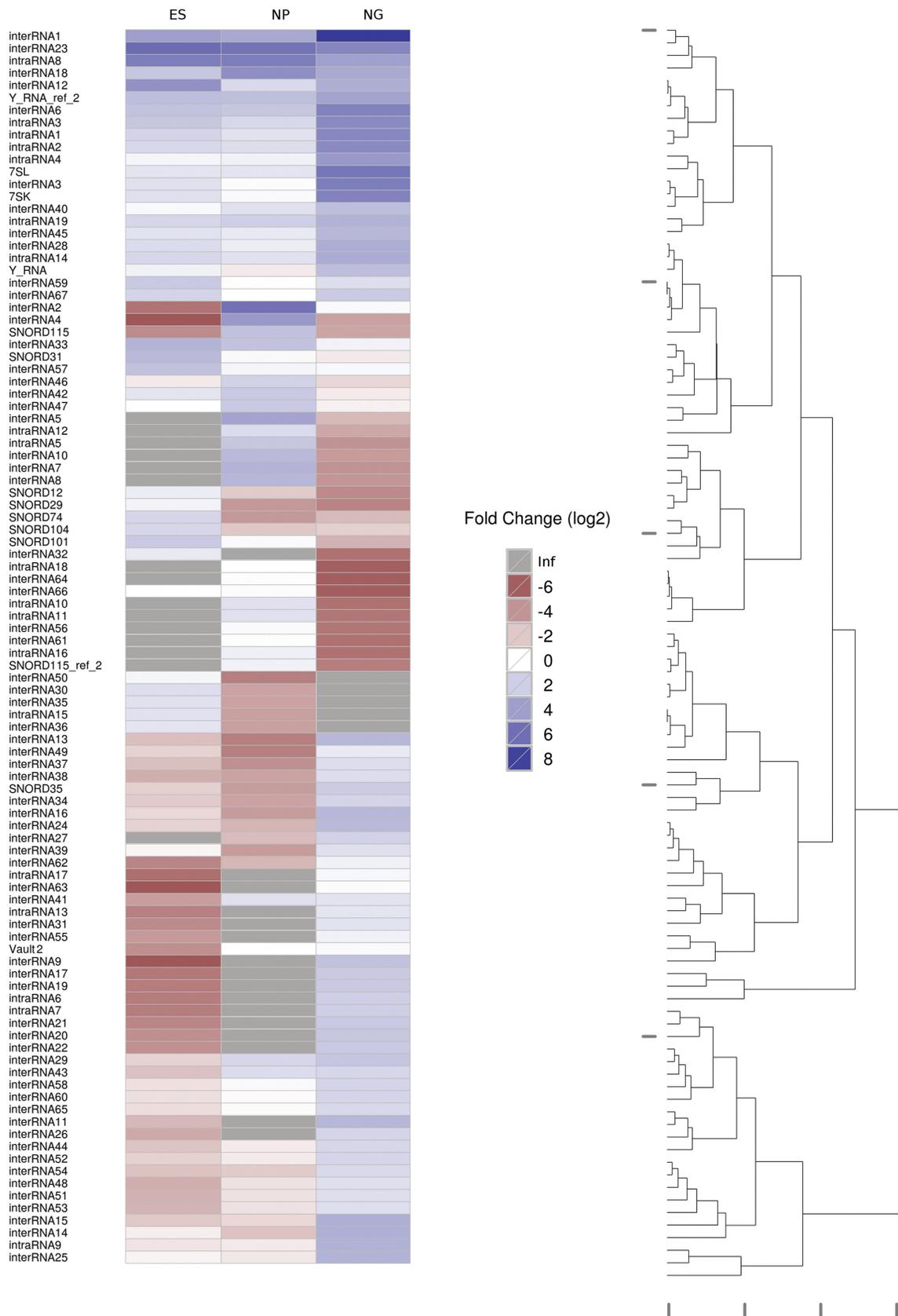
Previously, 7SK RNA has been reported to negatively regulate RNA Polymerase II transcription by inactivating the positive transcription elongation factor b (P-TEFb) as well as affecting the function of the chromatin regulator HMGA1 (64,65). By this mechanism, an important transcriptional regulatory role of 7SK RNA in HMGA1-dependent cell differentiation regulation has been described (64). Hence, 7SK RNA might also be involved in the regulation of the transition from mouse ES cells into neural/glial population of cells.

Lastly, among differentially expressed known ncRNAs we have identified expression of vault-2 ncRNA (66), an ncRNA component of the vault RNP, to be upregulated at NP and N/G stages, respectively (Figure 3). Recently, by a subtractive hybridization approach we have observed expression of vault RNAs 1-3 to be highly up-regulated upon EBV (Epstein-Barr virus) infection of human B cells (67,68).

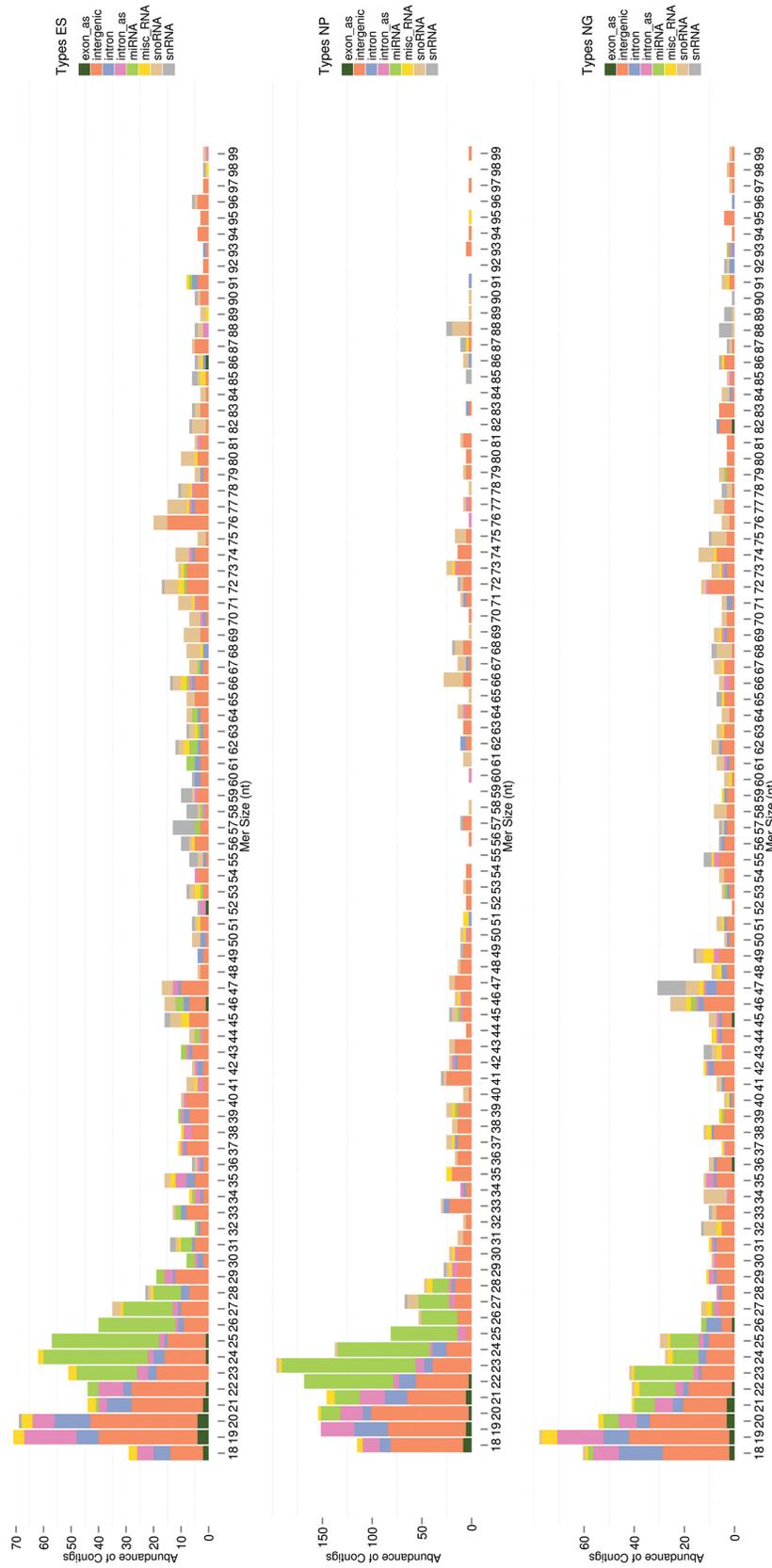
#### Size distribution of novel ncRNAs from the three stages of ES cell neural differentiation

Analysis of the length distribution of novel and known ncRNA candidates from the three cDNA libraries showed a bias towards 18-30 nt sized ncRNA species in all three cDNA libraries (Figure 4). Thereby, in the NP library the 18-30 nt peak is ~2-fold larger compared to ES and N/G libraries. A second, smaller peak of ncRNA sequences appears at RNAs sized 45-47 nt which increases at the N/G stage; a third potential peak, sized 71-78 nt, is also observed (Figure 4).

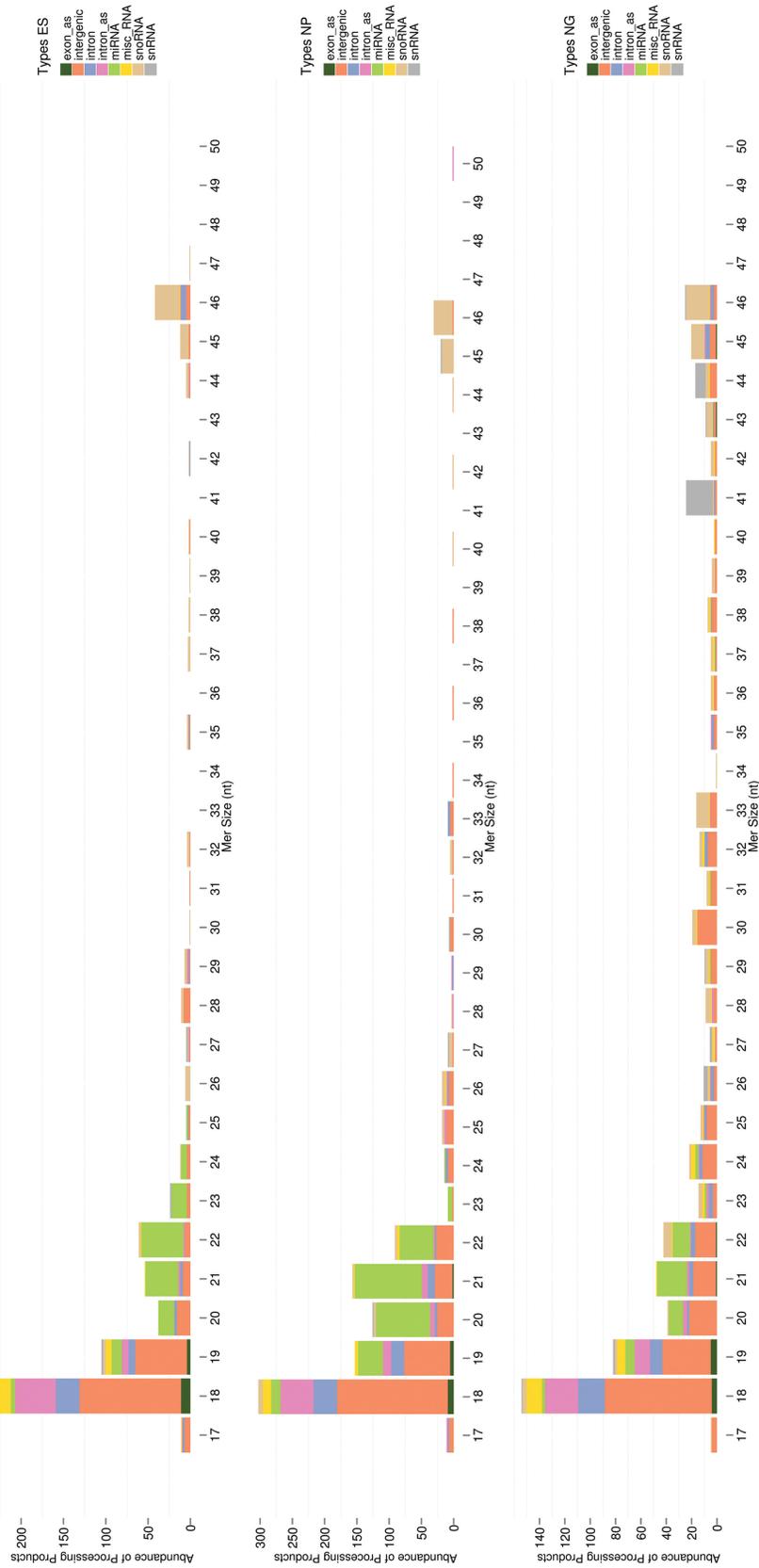
The fraction of 18-30 nt sized ncRNA candidates is comprised of miRNA species in addition to inter- and intraRNAs, i.e. ncRNAs located in intergenic or intronic regions (Figure 4). Although we tried to sub-classify the latter RNA species, including motif finding algorithms provided by the MEME suite (69) and searching for structural conservation employing the Vienna RNA package (70) and RNaz (71), as of now, we could not identify any common motifs. In addition, a correlation of features such as RNA size did not result in satisfying findings. However, due to their size and high abundance, similar to the abundance of miRNAs, some intra- and interRNAs might represent, yet undiscovered, novel microRNAs. Therefore, we employed the miRDeep algorithm (32,33) to search for novel microRNAs in our dataset. Thereby, we investigated folding of the pre-miRNAs by mfold, as well as the conservation of seed sequences and the abundance of the miR\* sequences in the cDNA libraries. In some cases, novel identified microRNA candidates were predicted to derive



**Figure 3.** The 100 most differentially expressed novel ncRNA candidates. Further legend analogous to Figure 2.



**Figure 4.** Length distribution and abundance of contigs from the three RNP libraries. ncRNAs were clustered in unique contigs; the consensus sequence was employed to determine the absolute abundance versus the length of respective ncRNAs (contigs exhibiting a size >99 nt are rare and are therefore not shown).



**Figure 5.** Length distribution and abundance of processed contigs from the three RNP libraries. ncRNAs were clustered in unique contigs; the consensus sequence was employed to determine the absolute abundance of the processed ncRNAs, as assessed by APART, versus the length of the respective processed ncRNAs (processed contigs exhibiting a size >50 nt are rare and are therefore not shown).

from snoRNA sequences. Since it has previously been reported that some snoRNAs can indeed serve as precursor RNAs for miRNAs (72–74), we included these candidates in our analysis. Based on above criteria, we characterized 17 novel microRNA candidates (Supplementary File S4). Thereby, the highest scoring miRNA candidate was mmu-interRNA26, whose expression was also validated by a mouse ESTs (expressed sequence tags) database screen; mmu-interRNA26 is abundantly expressed in the N/G library, unlike most other miRNAs (Figure 3), as well as in primary hippocampal neurons and astrocytes (see below).

The remaining inter- and intraRNAs in the size range of 18–30 nt scored very low as potential miRNAs suggesting that these might represent a distinct class of entirely novel neural ncRNA candidates. Interestingly, we note that sequences of numerous inter- and intraRNAs are included in reported ESTs (Supplementary File S1), thus validating their expression. Several inter- and intraRNAs were found also to cover binding sites for transcription factor p300 and the transcriptional repressor CTCF, which are reported to regulate gene expression in ES cells and in the cortex and cerebellum of the mouse brain. We further identified a number of inter- and intraRNAs overlapping regions of chromatin acetylation, which has been reported in total brain and cerebellum (Supplementary File S1). Taken together, these results indicate that most inter- and intraRNAs are actively transcribed in the brain, while at least some of these might be involved in regulating brain function and development on a transcriptional level.

#### Processing analysis of known and novel ncRNA candidates by APART

In Eukarya, many functional known ncRNA species are processed from larger RNA transcripts (75). In particular, rRNA, miRNA-, snoRNA- or tRNA-precursor transcripts are processed by distinct processing pathways, which include site-specific exo- and endonucleases. By employing APART we have identified a large number of small stable RNA species, potentially derived from larger precursor transcripts (for selected examples see Supplementary Figure S2). Detection of RNA processing by APART is achieved by scanning of a contig-coverage plot to search for significant changes (28). APART considers a position as a putative processing site when the coverage shift between 2 nt is larger than one-third of the maximum coverage of the contig (Supplementary Figure S2).

For proof of principle we investigated miRNA processing by the APART algorithm for the class of miRNAs, which are reported to be processed by Drosha/Pasha and Dicer from larger primary and precursor miRNA transcripts (76,77). Indeed, we were able to identify this processing event for miRNAs mmu-let-7c-1 as well as for mmi-mir-21 (Supplementary Figure S2). In addition, in particular for the class of 18–30 nt sized ncRNA candidates, we were able to identify in many cases potential precursor transcripts (Figure 5) which might strengthen the functionality of these novel ncRNA candidates.

#### Novel ncRNA candidates are expressed in primary hippocampal neurons and astrocytes

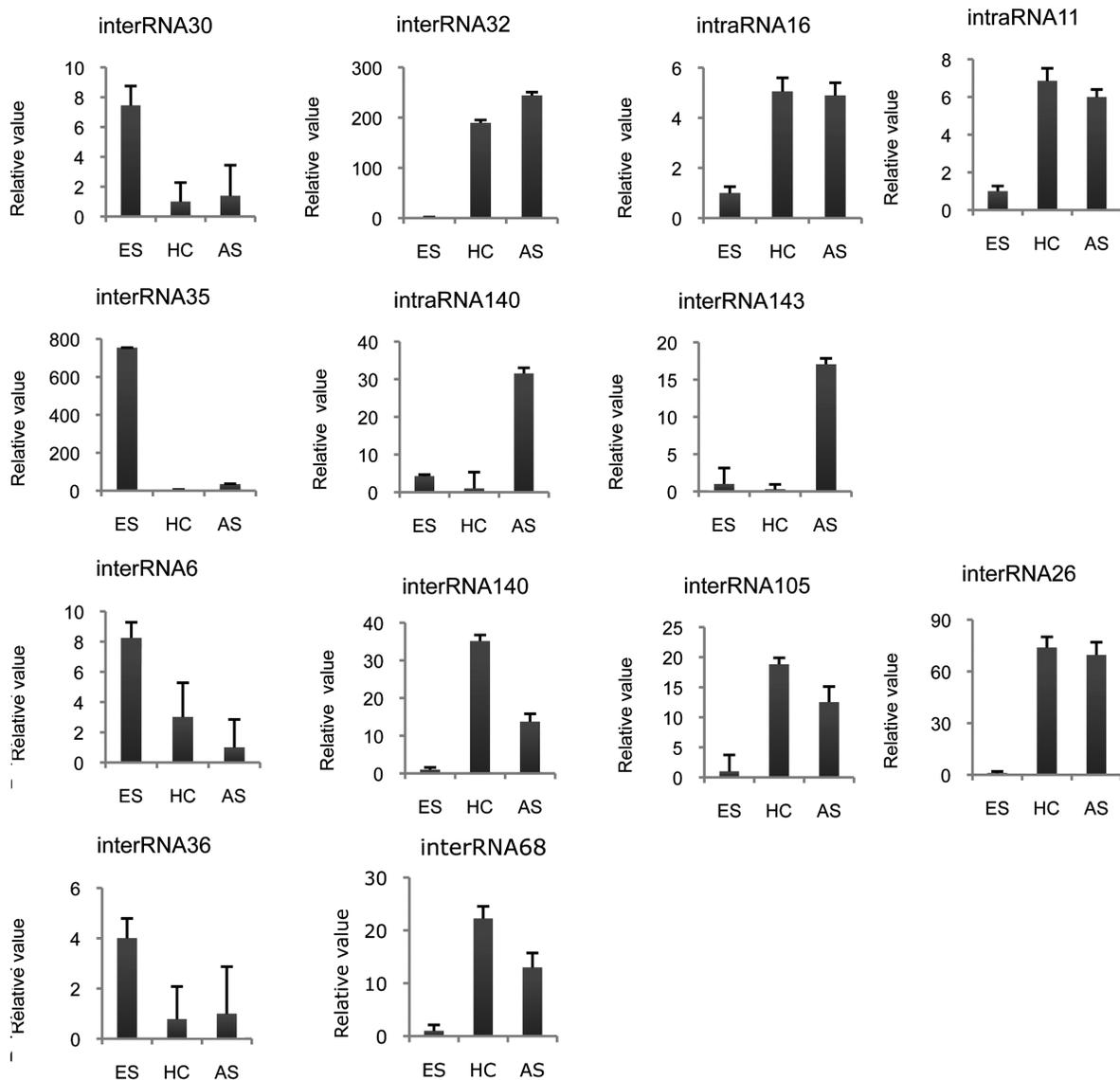
Due to the heterogeneity of the N/G stage, comprised of neurons and glial cells (astrocytes and oligodendrocytes) (Supplementary Figure S1), we additionally analyzed in which type of neural cells selected inter- and intraRNAs are expressed. To that aim, we analyzed primary cell cultures of neuronal and glial cells for the presence of selected novel ncRNA candidates by real-time PCR. Based on the library profiling by protein markers (see above), we employed astrocytes (AS) as well as primary hippocampal neurons (HC) (see Supplementary Methods section) to investigate ncRNA expression.

Indeed, we observed neural-specific as well as ES cell-specific expression of selected candidates (Figure 6). For example, interRNA32 and interRNA26 were highly expressed in HC and AS but showed very low expression in ES cells; thereby, up-regulation of expression was observed between 70- and 200-fold (Figure 6). Other ncRNAs, such as interRNA30, interRNA35 and interRNA36, which exhibited a reduced expression level during differentiation (i.e. in the NP and N/G stages), also show reduced expression in HC and AS cell cultures comprised mainly of a single cell type (Figure 6). Interestingly, interRNA143 and interRNA140 exhibited a highly AS-specific expression when compared to HC and ES cells, thereby being potentially involved in AS differentiation. Conversely, interRNA105, interRNA140 and interRNA68 are higher expressed in HC than in AS cells, with interRNA140 being ~3-fold higher expressed in HC compared to AS (Figure 6).

#### Abundance versus differential expression of ncRNA candidates

To further increase the likelihood for the identification of novel functional ncRNA candidates, we assessed their relative abundance in the libraries and compared it to their differential expression within the three differentiation stages of ES cells. The differentially expressed candidates were assigned scores, calculated by their mean expression value versus their absolute mean fold change in all three libraries. We rationalized that there should be a correlation between the abundance of a ncRNA, its differential expression, and the likelihood for representing a functional ncRNA species. This does not rule out, that lowly expressed ncRNAs with an even distribution between the three ES cell stages might not be functional. However, differential expression is one of the key indications of regulatory function of the gene products. Moreover, for future functional analysis such as the identification of protein binding partners, a robust, abundant and differential expression would be desirable.

The results of the abundance/differential expression plot show that ncRNA candidates with the highest scores (i.e. highest abundance/highest differential expression) are predominantly interRNAs or miRNAs, sized 18–30 nt next to miRNAs (Figure 7, upper right corner). We also identified other known ncRNAs such as 7SL RNA and 7SK RNA, vault-2 RNA and snoRNAs, in particular SNORD115 (see above), which are among the



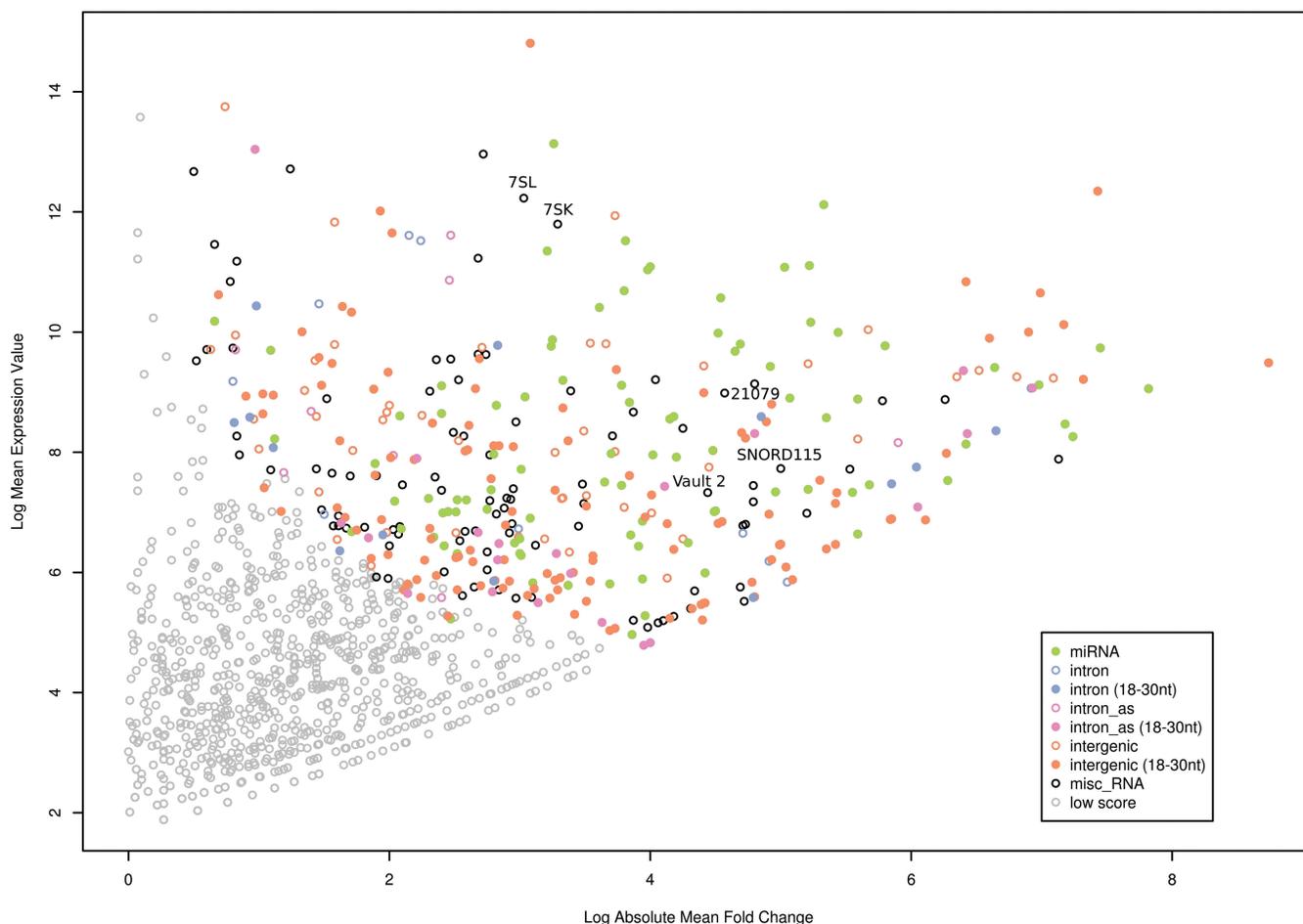
**Figure 6.** Expression analysis of selected inter- and intraRNAs in primary hippocampal neurons and astrocytes by real-time PCR. Expression values were normalized against U6.

top-ranked candidates. The potentially novel miRNA candidate mmu-interRNA26 is indicated in Figure 7 (designated as 21 079, corresponding to the ID number indicated in Supplementary File S4). Differential expression of selected ncRNA candidates from this analysis was also verified by real-time PCR (Supplementary Figure S4). Further functional analysis of selected ncRNA candidates will thus focus on highest-ranked novel and known ncRNAs from our screen.

## CONCLUSION

In this study, we defined the small ncRNA transcriptome involved in regulation of mouse ES cell differentiation into neural cells in mouse by a deep-sequencing approach. In general, high-throughput deep-sequencing approaches have the potential to identify a large number of potential ncRNA candidates, however, these studies usually lack

functional evidence of novel RNA transcripts. Functional analysis of potential ncRNA candidates, however, is highly time consuming, with currently no high-throughput methods available. Hence, it is important to preselect or enrich for functional ncRNA candidates in deep-sequencing screens for further analysis. In order to increase the likelihood for identification of such functional ncRNA species, we applied three novel selection filters: (i) cDNA libraries were generated from RNPs rather than from protein-devoid RNAs, since most functional eukaryal RNAs are known to form RNA–protein complexes; (ii) we analyzed differential expression of novel ncRNA candidates by generation of three RNP libraries from three stages of ES cell differentiation taking into consideration that many functional ncRNAs are regulated during development; (iii) since most functional ncRNAs have been shown to be processed from larger precursor RNAs, we also analyzed potential RNA processing products.



**Figure 7.** Plot of the mean fold-change versus the mean expression of ncRNAs expressed in all three libraries. The ncRNAs are assigned scores (see Supplementary Material section) based on their mean expression value ( $y$ -axis) versus their absolute mean fold change ( $x$ -axis). Expression values and fold changes represent the mean in all three libraries. Fold changes are shown in a  $\log_2$  scale. The grey dots (low score) refer to candidates having a mean expression value of  $<50$  and a fold-change of  $<2$ .

By these analyses, we identified several known ncRNAs from two abundant RNA classes, i.e. miRNAs and snoRNAs, respectively, which were differentially expressed. In addition, expression of three other known ncRNAs, i.e. 7SL, 7SK and vault-2 RNA, respectively, was shown to be significantly up-regulated during neural differentiation.

While regulation of gene expression by 7SK RNA is exerted on the transcriptional level, miRNAs and some selected snoRNAs have been reported to regulate gene expression on the post-transcriptional level; finally, 7SL RNA was shown to act on a post-translational level. Hence, regulation of gene expression during neural differentiation might be exerted by these ncRNAs on transcriptional, translational and post-translational levels.

The majority of differentially expressed ncRNAs were represented by novel ncRNA candidates and mapped to intergenic and intronic regions, designated as interRNAs or intraRNA, respectively. These RNA species predominantly exhibited sizes between 18 and 30 nt. Thereby, 17 of these interRNAs and intraRNAs might represent novel members from the class of miRNAs, as assessed by the miRDeep algorithm. However, the majority of interRNAs or intraRNAs could not be assigned as miRNAs and thus

might represent novel candidates for regulatory ncRNAs in neural differentiation. For future analyses, it will be interesting to study the effects of known and novel ncRNAs on ES cell differentiation by overexpression or inactivation of selected ncRNA candidates.

#### ACCESSION NUMBERS

Sequence Read Archive (NCBI) accession number: SRP008250.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Files 1–4, Supplementary Figures 1–4, Supplementary Methods and Supplementary References [78–80].

#### FUNDING

The Austrian Science Funds FWF (SFB F4411); the Austrian Ministry of Science and Research grant Gen-AU (D 110420-011-015); the PhD program

W 1206-B18 (funding of K.S.) to A.H.; the Austrian Ministry of Science and Research grant Gen-AU (D110420-012-012) (funding of M. Z). Funding for open access charge: Austrian Ministry of Science and Research grant (SFB F4411).

*Conflict of interest statement.* None declared.

## REFERENCES

- Willingham, A.T. and Gingeras, T.R. (2006) TUF love for “junk” DNA. *Cell*, **125**, 1215–1220.
- Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
- Mattick, J.S., Taft, R.J. and Faulkner, G.J. (2009) A global view of genomic information—moving beyond the gene and the master regulator. *Trends Genet.*, **26**, 21–28.
- van Bakel, H., Nislow, C., Blencowe, B.J. and Hughes, T.R. (2010) Most ‘dark matter’ transcripts are associated with known genes. *PLoS Biol.*, **8**, e1000371.
- Hüttenhofer, A. and Schattner, P. (2006) The principles of guiding by RNA: chimeric RNA-protein enzymes. *Nat. Rev. Genet.*, **7**, 475–482.
- Stadler, B.M. and Ruohola-Baker, H. (2008) Small RNAs: keeping stem cells in line. *Cell*, **132**, 563–566.
- Mallanna, S.K. and Rizzino, A. (2010) Emerging roles of microRNAs in the control of embryonic stem cells and the generation of induced pluripotent stem cells. *Dev. Biol.*, **344**, 16–25.
- Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P. and Blelloch, R. (2008) Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.*, **22**, 2773–2785.
- Ambros, V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.
- Plasterk, R.H. (2006) Micro RNAs in animal development. *Cell*, **124**, 877–881.
- Houbavii, H.B., Murray, M.F. and Sharp, P.A. (2003) Embryonic stem cell-specific microRNAs. *Dev. Cell*, **5**, 351–358.
- Sinkkonen, L., Hugenschmidt, T., Berninger, P., Gaidatzis, D., Mohn, F., Artus-Revel, C.G., Zavolan, M., Svoboda, P. and Filipowicz, W. (2008) MicroRNAs control de novo DNA methylation through regulation of transcriptional repressors in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.*, **15**, 259–267.
- Benetti, R., Gonzalo, S., Jaco, I., Munoz, P., Gonzalez, S., Schoeftner, S., Murchison, E., Andl, T., Chen, T., Klatt, P. *et al.* (2008) A mammalian microRNA cluster controls DNA methylation and telomere recombination via Rbl2-dependent regulation of DNA methyltransferases. *Nat. Struct. Mol. Biol.*, **15**, 998.
- Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.A. and Kosik, K.S. (2009) MicroRNA-145 regulates OCT4, SOX2, and KLF4 and represses pluripotency in human embryonic stem cells. *Cell*, **137**, 647–658.
- Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E.E., Nitsch, R. and Wulczyn, F.G. (2008) A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell. Biol.*, **10**, 987–993.
- Lakshmi, U., Love, B., Goff, L.A., Jörnsten, R., Graichen, R., Hart, R.P. and Chesnut, J.D. (2007) MicroRNA expression pattern of undifferentiated and differentiated human embryonic stem cells. *Stem Cells Dev.*, **16**, 1003–1016.
- Chen, C., Ridzon, D., Lee, C.T., Blake, J., Sun, Y. and Strauss, W.M. (2007) Defining embryonic stem cell identity using differentiation-related microRNAs and their potential targets. *Mamm. Genome*, **18**, 316–327.
- Kanellopoulou, C., Muljo, S.A., Kung, A.L., Ganesan, S., Drapkin, R., Jenuwein, T., Livingston, D.M. and Rajewsky, K. (2005) Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.*, **19**, 489–501.
- Murchison, E.P., Partridge, J.F., Tam, O.H., Cheloufi, S. and Hannon, G.J. (2005) Characterization of Dicer-deficient murine embryonic stem cells. *Proc. Natl Acad. Sci. USA*, **102**, 12135–12140.
- Wang, Y., Medvid, R., Melton, C., Jaenisch, R. and Blelloch, R. (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat. Genet.*, **39**, 380–385.
- Rederstorff, M., Bernhart, S.H., Tanzer, A., Zywicki, M., Perfler, K., Lukasser, M., Hofacker, I.L. and Huttenhofer, A. (2010) RNPomics: defining the ncRNA transcriptome by cDNA library generation from ribonucleo-protein particles. *Nucleic Acids Res.*, **38**, e113.
- Rederstorff, M. and Huttenhofer, A. (2011) cDNA library generation from ribonucleoprotein particles. *Nat. Protoc.*, **6**, 166–174.
- Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S. and Smithies, O. (1987) Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, **330**, 576–578.
- Shimizukawa, R., Sakata, A., Hirose, M., Takahashi, A., Iseki, H., Liu, Y., Kunita, S., Sugiyama, F. and Yagami, K. (2005) Establishment of a new embryonic stem cell line derived from C57BL/6 mouse expressing EGFP ubiquitously. *Genesis*, **42**, 47–52.
- Lee, S.H., Lumelsky, N., Studer, L., Auerbach, J.M. and McKay, R.D. (2000) Efficient generation of midbrain and hindbrain neurons from mouse embryonic stem cells. *Nat. Biotechnol.*, **18**, 675–679.
- Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. and Smith, A. (2003) Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.*, **21**, 183–186.
- Nat, R., Salti, A., Suci, L., Ström, S. and Dechant, G. (2012) Pharmacological modulation of the hedgehog pathway differentially affects dorsal/ventral patterning in mouse and human embryonic stem cell models of telencephalic development. *Stem Cells Dev* [Epub ahead of print].
- Zywicki, M., Bakowska-Zywicka, K. and Polacek, N. (2012) Revealing stable processing products from ribosome-associated small RNAs by deep-sequencing data analysis. *Nucleic Acids Res.*, **40**, 4013–4024.
- Quinlan, A.R. and Hall, I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
- Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
- Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
- Friedländer, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S. and Rajewsky, N. (2008) Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.*, **26**, 407–415.
- Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W. and Rajewsky, N. (2011) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.*, **40**, 37–52.
- Lawrence, M., Gentleman, R. and Carey, V. (2009) rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*, **25**, 1841–1842.
- Calabrese, J.M., Seila, A.C., Yeo, G.W. and Sharp, P.A. (2007) RNA sequence analysis defines Dicer’s role in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA*, **104**, 18097–18102.
- Tripathi, R., Saini, H.K., Rad, R., Abreu-Goodger, C., van Dongen, S. and Enright, A.J. (2011) Messenger RNA and microRNA profiling during early mouse EB formation. *Gene Expr. Patterns*, **11**, 334–344.
- Tzur, G., Levy, A., Meiri, E., Barad, O., Spector, Y., Bentwich, Z., Mizrahi, L., Katzenellenbogen, M., Ben-Shushan, E., Reubinoff, B.E. *et al.* (2008) MicroRNA expression patterns and function in

- endodermal differentiation of human embryonic stem cells. *PLoS One*, **3**, e3726.
38. Laurent, L.C., Chen, J., Ulitsky, I., Mueller, F.J., Lu, C., Shamir, R., Fan, J.B. and Loring, J.F. (2008) Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence. *Stem Cells*, **26**, 1506–1516.
  39. Bar, M., Wyman, S.K., Fritz, B.R., Qi, J., Garg, K.S., Parkin, R.K., Kroh, E.M., Bendoraite, A., Mitchell, P.S., Nelson, A.M. *et al.* (2008) MicroRNA discovery and profiling in human embryonic stem cells by deep sequencing of small RNA libraries. *Stem Cells*, **26**, 2496–2505.
  40. Stadler, B., Ivanovska, I., Mehta, K., Song, S., Nelson, A., Tan, Y., Mathieu, J., Darby, C., Blau, C.A., Ware, C. *et al.* (2010) Characterization of microRNAs involved in embryonic stem cell states. *Stem Cells Dev.*, **19**, 935–950.
  41. Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M. *et al.* (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.*, **18**, 610–621.
  42. Krichevsky, A.M., Sonntag, K.C., Isacson, O. and Kosik, K.S. (2006) Specific microRNAs modulate embryonic stem cell-derived neurogenesis. *Stem Cells*, **24**, 857–864.
  43. Ren, J., Jin, P., Wang, E., Marincola, F.M. and Stronck, D.F. (2009) MicroRNA and gene expression patterns in the differentiation of human embryonic stem cells. *J. Transl. Med.*, **7**, 20.
  44. Lestrade, L. and Weber, M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.
  45. Bachellerie, J.P., Cavallé, J. and Huttenhofer, A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
  46. Kiss, T. (2002) Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell*, **109**, 145–148.
  47. Filipowicz, W. and Pogacic, V. (2002) Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.*, **14**, 319–327.
  48. Kiss, T. (2004) Biogenesis of small nuclear RNPs. *J. Cell. Sci.*, **117**, 5949–5951.
  49. Cavallé, J., Buiting, K., Kiefmann, M., Lalande, M., Brannan, C.I., Horsthemke, B., Bachellerie, J.P., Brosius, J. and Huttenhofer, A. (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc. Natl Acad. Sci. USA*, **97**, 14311–14316.
  50. Vitali, P., Royo, H., Seitz, H., Bachellerie, J.P., Huttenhofer, A. and Cavaille, J. (2003) Identification of 13 novel human modification guide RNAs. *Nucleic Acids Res.*, **31**, 6543–6551.
  51. Tycowski, K.T., Shu, M.D. and Steitz, J.A. (1996) A mammalian gene with introns instead of exons generating stable RNA products. *Nature*, **379**, 464–466.
  52. Smith, C.M. and Steitz, J.A. (1998) Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol. Cell Biol.*, **18**, 6897–6909.
  53. Runte, M., Huttenhofer, A., Gross, S., Kiefmann, M., Horsthemke, B. and Buiting, K. (2001) The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum. Mol. Genet.*, **10**, 2687–2700.
  54. Kishore, S. and Stamm, S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 230–232.
  55. Vitali, P., Basyuk, E., Le Meur, E., Bertrand, E., Muscatelli, F., Cavallé, J. and Huttenhofer, A. (2005) ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J. Cell Biol.*, **169**, 745–753.
  56. Butler, M.G. (2011) Prader-Willi syndrome: obesity due to genomic imprinting. *Curr. Genomics*, **12**, 204–215.
  57. Ding, F., Prints, Y., Dhar, M.S., Johnson, D.K., Garnacho-Montero, C., Nicholls, R.D. and Francke, U. (2005) Lack of Pwcr1/MBII-85 snoRNA is critical for neonatal lethality in Prader-Willi syndrome mouse models. *Mamm. Genome*, **16**, 424–431.
  58. Skryabin, B.V., Gubar, L.V., Seeger, B., Pfeiffer, J., Handel, S., Robeck, T., Karpova, E., Rozhdestvensky, T.S. and Brosius, J. (2007) Deletion of the MBII-85 snoRNA gene cluster in mice results in postnatal growth retardation. *PLoS Genet.*, **3**, e235.
  59. Sahoo, T., del Gaudio, D., German, J.R., Shinawi, M., Peters, S.U., Person, R.E., Garnica, A., Cheung, S.W. and Beaudet, A.L. (2008) Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat. Genet.*, **40**, 719–721.
  60. Hogart, A., Leung, K.N., Wang, N.J., Wu, D.J., Driscoll, J., Vallero, R.O., Schanen, N.C. and LaSalle, J.M. (2009) Chromosome 15q11-13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. *J. Med. Genet.*, **46**, 86–93.
  61. Keenan, R.J., Freymann, D.M., Stroud, R.M. and Walter, P. (2001) The signal recognition particle. *Annu. Rev. Biochem.*, **70**, 755–75.
  62. Wild, K., Weichenrieder, O., Strub, K., Sinning, I. and Cusack, S. (2002) Towards the structure of the mammalian signal recognition particle. *Curr. Opin. Struct. Biol.*, **12**, 72–81.
  63. Lutcke, H. (1995) Signal recognition particle (SRP), a ubiquitous initiator of protein translocation. *Eur. J. Biochem.*, **228**, 531–550.
  64. Eilebrecht, S., Benecke, B.J. and Benecke, A. (2011) 7SK snRNA-mediated, gene-specific cooperativity of HMGAI and P-TEFb. *RNA Biol.*, **8**, 1084–93.
  65. Eilebrecht, S., Brysbaert, G., Wegert, T., Urlaub, H., Benecke, B.-J. and Benecke, A. (2011) 7SK small nuclear RNA directly affects HMGAI function in transcription regulation. *Nucleic Acids Res.*, **39**, 2057–2072.
  66. van Zon, A., Mossink, M.H., Scheper, R.J., Sonneveld, P. and Wiemer, E.A.C. (2003) The vault complex. *Cell. Mol. Life Sci.*, **60**, 1828–1837.
  67. Mrázek, J., Kreutmayer, S.B., Grässer, F.A., Polacek, N. and Huttenhofer, A. (2007) Subtractive hybridization identifies novel differentially expressed ncRNA species in EBV-infected human B cells. *Nucleic Acids Res.*, **35**, e73.
  68. Nandy, C., Mrázek, J., Stoiber, H., Grasser, F.A., Huttenhofer, A. and Polacek, N. (2009) Epstein-barr virus-induced expression of a novel human vault RNA. *J. Mol. Biol.*, **388**, 776–784.
  69. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
  70. Hofacker, I.L. (2009) RNA secondary structure analysis using the Vienna RNA package. *Curr. Protoc. Bioinformatics*, Chapter 12, Unit12.12.
  71. Gruber, A.R., Findeiß, S., Washietl, S., Hofacker, I.L. and Stadler, P.F. (2010) Rnaz 2.0: improved noncoding Rna detection. *Pac. Symp. Biocomput.*, **15**, 69–79.
  72. Ender, C., Krek, A., Friedländer, 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.
  73. Saraiya, A.A. and Wang, C.C. (2008) snoRNA, a novel precursor of microRNA in *Giardia lamblia*. *PLoS Pathog.*, **4**, e1000224.
  74. Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P. and Mattick, J.S. (2009) Small RNAs derived from snoRNAs. *RNA*, **15**, 1233–1240.
  75. Tollervey, D. and Caceres, J.F. (2000) RNA processing marches on. *Cell*, **103**, 703–709.
  76. Conrad, R., Barrier, M. and Ford, L.P. (2006) Role of miRNA and miRNA processing factors in development and disease. *Birth Defects Res. C Embryo Today*, **78**, 107–117.
  77. Lehrbach, N.J. and Miska, E.A. (2011) Regulation of pre-miRNA Processing. *Adv. Exp. Med. Biol.*, **700**, 67–75.
  78. Pfaffl, M.W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.*, **29**, e45.
  79. Polleux, F. and Ghosh, A. (2002) The slice overlay assay: a versatile tool to study the influence of extracellular signals on neuronal development. *Sci. STKE*, **2002**, pl9.
  80. Zeitelhofer, M., Vessey, J.P., Xie, Y., Tübing, F., Thomas, S., Kiebler, M. and Dahm, R. (2007) High-efficiency transfection of mammalian neurons via nucleofection. *Nat. Protoc.*, **2**, 1692–1704.