11014–11029 *Nucleic Acids Research, 2018, Vol. 46, No. 20* *Published online 8 September 2018*
doi: 10.1093/nar/gky791

Small RNA sequences derived from pre-microRNAs in the supraspliceosome

Shelly Mahlab-Aviv[1], Ayub Boulos[2], Ayelet R. Peretz[2], Tsiona Eliyahu[3], Liran Carmel[2], Ruth Sperling[2,*] and Michal Linial[3,*]

[1]The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel, [2]Department of Genetics, Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel and [3]Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

Received July 07, 2017; Revised August 11, 2018; Editorial Decision August 21, 2018; Accepted September 06, 2018

ABSTRACT

MicroRNAs (miRNAs) are short non-coding RNAs that negatively regulate the expression and translation of genes in healthy and diseased tissues. Herein, we characterize short RNAs from human HeLa cells found in the supraspliceosome, a nuclear dynamic machine in which pre-mRNA processing occurs. We sequenced small RNAs (<200 nt) extracted from the supraspliceosome, and identified sequences that are derived from 200 miRNAs genes. About three quarters of them are mature miRNAs, whereas the rest account for various defined regions of the pre-miRNA, and its hairpin-loop precursor. Out of these aligned sequences, 53 were undetected in cellular extract, and the abundance of additional 48 strongly differed from that in cellular extract. Notably, we describe seven abundant miRNA-derived sequences that overlap non-coding exons of their host gene. The rich collection of sequences identical to pre-miRNAs at the supraspliceosome suggests overlooked nuclear functions. Specifically, the abundant hsa-mir-99b may affect splicing of LINC01129 primary transcript through base-pairing with its exon-intron junction. Using suppression and overexpression experiments, we show that hsa-mir-7704 negatively regulates the level of the lncRNA HAGLR. We claim that in cases of extended base-pairing complementarity, such supraspliceosomal pre-miRNA sequences might have a role in transcription attenuation, maturation and processing.

INTRODUCTION

MicroRNAs (miRNAs) are small ∼22 nt long molecules that have been implicated in regulating all cellular signaling pathways, and alterations in their expression level are associated with the development and progression of diseases, including cancer (1). Their main role is thought to be the negative regulation of gene expression and translation, mostly by base-pairing to the 3′-UTR of target mRNA transcripts in the cytoplasm (2,3).

The canonical biogenesis of miRNAs from RNA Polymerase II (Pol II) transcripts involves a number of steps. The first step occurs in the nucleus by the microprocessor, whose key proteins are DGCR8 and Drosha. DGCR8 binds the RNA molecule, while Drosha, an RNase III type enzyme, cleaves the primary (pri) miRNA transcript into a precursor (pre) miRNA stem-loop molecule of 70–80 bases (4–6). Next, the pre-miRNA is exported by Exportin-5 to the cytoplasm, where it is cleaved by the RNase III Dicer, yielding double stranded RNA that comprises the mature miRNA and its complement. The leading strand of the miRNA is then loaded on Argonaute protein (AGO), the key component of the RNA-induced silencing complex (RISC), which stabilizes the miRNA and directs its binding to its target gene (Reviewed in (3)).

Recent studies revealed the presence of mature miRNAs in the nucleus (7–9), suggesting nuclear functions in addition to the known cytoplasmic ones (10,11). Several studies provide evidence for active shuttling of miRNAs from the cytoplasm to the nucleus (12,13), involving Exportins (14). Although the function(s) of miRNAs in the nucleus are not yet well understood, recent reports demonstrated their involvement in a number of processes, such as the regulation of non-coding RNAs (15–18), and transcriptional silencing (19). Search for nuclear miRNAs complementarity in human promoters revealed additional targets, suggesting that they may function in transcription inhibition (20). Nuclear miRNAs were also shown to affect transcriptional activation, a process that requires the involvement of AGO and proximal promoter of non-coding transcripts (21). Anal-

*To whom correspondence should be addressed. Tel: +972 2 6585162; Fax: +972 2 6586975; Email: r.sperling@mail.huji.ac.il
Correspondence may also be addressed to Michal Linial. Tel: +972 2 6585425; Fax: +972 2 6523429; Email: michall@cc.huji.ac.il

ysis of miRNA–mRNA–AGO interactions, revealed substantial AGO–miRNA mapping to intronic sequences (22).

A large proportion of miRNA genes are located in introns of coding genes, albeit many are likely transcribed from their own Pol II dependent promoters (3,23,24). Multiple lines of evidence show strong links between splicing and miRNA processing. For example, splicing components were reported to co-sediment with the microprocessor and pre-miRNAs (25); processing of miRNA-211 was shown to promote the splicing of its hosting intron (26); knockdown of AGO1, AGO2 and Dicer affect alternative splicing (AS) (13,27); splicing factors were found within chromatin-associated AGO proteins (13); nuclear small RNAs were shown to affect AS, and altered splicing requires AGO2 expression.

Most mammalian Pol II transcripts undergo pre-mRNA splicing (28). Notably, splicing and AS play a major role in the regulation of gene expression, and changes in splicing are common in many human diseases including cancer (29,30). Splicing, as well as additional processing events of Pol II transcripts, occurs in the cell nucleus within a huge (21 MDa) and highly dynamic machine known as the supraspliceosome (31–33). It is a ribonucleoprotein (RNP) complex, composed of four native spliceosomes, each similar to the *in-vitro* assembled spliceosome, that are connected by the pre-mRNA (34,35). The entire repertoire of nuclear pre-mRNAs, independent of their length and number of introns, is individually assembled in supraspliceosomes that include the five spliceosomal U snRNPs, and all known splicing factors (32,33). Supraspliceosome offers coordination and regulation for pre-mRNA processing events including AS (36,37) and other processing steps, such as 5′-end and 3′-end processing, and RNA editing (38). Thus, the supraspliceosome—the endogenous spliceosome—is involved in all nuclear processing activities of pre-mRNAs (31–33).

We hypothesized that there is a cross-talk between intronic miRNA processing and pre-mRNA splicing within the supraspliceosome, primarily for miRNAs residing on the same transcriptional unit. The main microprocessor components, Drosha and DGCR8 are found in supraspliceosomes (39). In addition, the presence in the supraspliceosome of several pre-miRNAs from intronic sequences have been previously reported (39,40).

Here, we focus on supraspliceosomal small RNA sequences derived from the hairpin precursors that include pre-miRNA sequences of HeLa cells. We refer to these small RNA sequences as miRNA gene aligned sequences (miR-GAS). Using small ncRNA (<200 nt) deep sequencing, we identified hundreds of miR-GAS sequences in the supraspliceosome. Three quarters are mature miRNAs and the rest represent other, typically well-defined, regions of the hairpin precursor miRNAs (HP-miRNAs) sequences. We focus on miR-GAS that are exclusively expressed in the supraspliceosomal fraction (SF), point out potential functions for such set of sequences, and demonstrate a negative effect of miR-7704 on the expression of lncRNA HAGLR. Our results imply an unexplored function for sequences identical to miRNAs and pre-miRNAs at the supraspliceosome in post-transcriptional gene regulation, splicing and nuclear pre-mRNA processing.

MATERIALS AND METHODS

Plasmids

For overexpression of hsa-mir-7704, the pre-miR-7704 sequence (5′-CGGGGTCGGCGGCGACGTGCTCAGC TTGGCACCCAAGTTCTGCCGCTCCGACGCCCG GC-3′) was cloned into the pPRIME-CMV-GFP-FF3 vector at the Xho1 and EcoR1 sites by GENEWIZ (South Plainfield, NJ, USA), generating pPRIME-miR-7704-CMV-GFP-FF3 plasmid. A pPRIME-CMV-GFP-FF3 empty vector was used as a control.

Isolation of supraspliceosomes

All isolation steps were conducted at 4°C. Supraspliceosomes were prepared from nuclear supernatants enriched in supraspliceosomes as described previously (34). Briefly, nuclear supernatants were prepared from purified cell nuclei by microsonication of the nuclei and precipitation of the chromatin in the presence of excess of tRNAs. The nuclear supernatant was fractionated on 10–45% (v/v) glycerol gradients. Centrifugations were carried out at 4°C in an SW41 rotor run at 41 krpm for 90 min [or an equivalent $w^2t = 2500$ (ω is in krpm; t is in h)]. The gradients were calibrated using the tobacco mosaic virus as a 200S sedimentation marker. Supraspliceosome peak fractions were confirmed by western blot and by electron microscopy visualization.

Protein detection

Western blotting (WB) analyses were performed as previously described (41). We used anti-hnRNP G (kindly provided by Prof. Stefan Stamm, University of Kentucky, Lexington), visualized with horseradish peroxidase conjugated to affinity-pure Goat anti-Rabbit IgG (H+L; Jackson Immunoreaserch, 1:5000). The MAb-104 that is directed against the phosphorylated epitopes of the SR proteins (42) was used as previously described (43).

RNA isolation from supraspliceosomes and deep sequencing

RNA was extracted from supraspliceosomes prepared from HeLa cells, as previously described (34). The integrity of the RNA was evaluated by an Agilent 2100 bioAnalyzer. For small RNA library construction, ∼1 μg of RNA was used. After phosphatase and T4 polynucleotide kinase (PNK) treatments, the RNA was ethanol precipitated to enrich for small RNA, and small RNA libraries (in triplicates) were prepared according to NEBNext Small RNA Library Prep Set for Illumina (Multiplex Compatible) Library Preparation Manual. Adaptors were then ligated to the 5′ and 3′ ends of the RNA, and cDNA was prepared from the ligated RNA and amplified to prepare the sequencing library. The amplified sequences were purified on E-Gel® EX 4% Agarose gels (ThermoFisher # G401004), and sequences representing RNA smaller than 200 nt were extracted from the gel. The library was sequenced using the Illumina NextSeq 500 Analyzer. The sequencing data, after removal the adaptors and filtering out low quality sequences, were aligned to miRBase (Release 21). In addition,

the filtered high-quality fragments were mapped to the human transcriptome of hg19 gtf file from UCSC provided by Galaxy. The hg19 transcriptome contains 963 559 exons from 45 314 transcripts. Sequences aligned to the miRNA genes as compiled in miRBase are reported.

Small RNA preparation from HeLa cell extract

Total RNA from HeLa cells was purified using QIAzol Lysis Reagent RNeasy Plus Universal Kit (QIAGEN, Germany). Sample has been transferred up to an RNeasy Minispin column and centrifuged for 15 s at >8000 g at room temperature, and the supernatant was processed according to the manufacturer's standard protocol. Samples with an RNA Integrity Number (RIN) of >8.5, as measured by Agilent 2100 Bioanalyzer, were considered for further analysis. Small RNA libraries (<200 nt) were generated using the TruSeq RNA V2 (Illumina) library protocol. The RNA was tested for quality and subjected to a library preparation of miRNA and mRNA according to TruSeq RNA standard protocol. Read length was 100 nt. Data used are derived from two biological duplicates.

Next generation sequencing (NGS) analysis

RNA extracted from the supraspliceosome fractions (SFs) in HeLa cells was taken from six independent library preparations. HeLa cells supraspliceosomes were taken from three independent biological preparations. Each preparation was split, and used to produce two libraries. NGS was performed for each sample on small RNA (<200 nt) molecules using standard Illumina Protocol. Each library consisted of ∼25 million reads of maximum length 76 (see supplementary Table S1).

Raw data of the sequenced small RNA were trimmed using Cutadapt version 1.13. Low-quality reads were filtered out using FASTX toolkit. Reads from the SF were aligned against human genome hg19 and miRBase database (version 21) using TopHat 2.1.1, allowing 90% sequence identity and a maximum of two mismatches. Reads whose start or end position were mapped to miRNA genes were considered. High quality reads from the six SF preparations were combined. The correlation between the expression of all miRNAs from HeLa-1 to HeLa-6 ranges from 0.58 to 0.82. Each sample is associated with 100–170 pre-miRNA genes (Supplementary Figure S1). Out of the mapped reads, only reads of length ≥17 were considered. miRNA gene aligned sequences (miR-GAS) refer to all mapped, high quality reads that are aligned to any of the pre-miRNA as defined by miRBase.

A significant correlation is calculated between the mapping schemes of high-quality reads against the sequences in miRBase and the human transcriptome (Supplementary Table S2). Furthermore, the correctly mapped reads between the two mapping methods are comparable (138.0k and 138.8k reads from mapping applied on miRBase, and the whole transcriptome, respectively).

We exclude the sequences identical to HP-miRNAs that are poorly supported (i.e. ∼300 miRNAs that were detected by 1–9 reads, see Supplementary Table S2). For the rest of the analyses, only miR-GAS with ≥10 reads were considered. This threshold was used to ensure reliable support in view of the limited total reads assigned to miR-GAS in the SF (total mapped reads from all cell samples are 139 k, Supplementary Table S2). Notably, we confirm that using this threshold, high consistency is observed among the multiple preparations (see Supplementary Figure S1D). Sequencing of the cellular extract (CE) used read length of 100 nt, that were mapped to miRNA genes using miRExpress 2.0. The total number of reads of the CE for miRNA library is 29.8 M reads, among them 3.25 M reads are mapped to miRNAs (Supplementary Table S1).

Validation of gene expression

RNA was extracted from HeLa and MCF-10A cells. The MCF-10A, human mammary epithelial cell-line, was purchased from ATCC (CRL-10317). RT-PCR was performed on RNA extracted from HeLa and MCF-10A cells and from nuclear supernatants of the above cells as described (59). The following sets of primers for HAGLR was used: Forward (exon 1) 5′-CGTCGGAGCGGCA GAACTT-3′ and Reverse 5′-AAGGGCCCATTTTCAGG CCA-3′ (exon 2). The primers for HOXD1 are: Forward (exon 1) 5′-ATTTACCTCCGGCTCACTCG-3′ and Reverse 5′-AGGTGCAAGCAGTTGGCTAT-3′ (exon 2). The identity of all PCR products was confirmed by sequencing. Each experiment was repeated at least 3 times. The relative abundance was quantified in view of the intensity of the β-actin that was used as a control. The β-actin forward and reverse primers, for an amplicon of 140 nt, are: 5′-CTGGAACGGTGAAGGTGACA-3′ and 5′-AAGGG ACTTCCTGTAACAATGCA-3′, respectively.

Transfection and RNA isolation

HeLa cells were grown in six-well plates. For downregulation of hsa-mir-7704, the cells were transfected with Anti-hsa-mir-7704 inhibitor AM29132 (ThermoFisher) according to manufacturer's instructions at 100 nM for 48 h. As controls we used HeLa cells transfected with mirVana miRNA inhibitor, negative control #1 (Ambion), and non-treated HeLa cells. Similarly, transfections with Anti-hsa-mir-7704 inhibitor AM29132 (ThermoFisher) were performed in MCF-10A cells.

For overexpression of hsa-miR-7704, HeLa cells grown in six-well plates were transfected for 48 h with 2 μg/well of the pPRIME-miR-7704-CMV-GFP-FF3 plasmid, using Lipofectamine-2000 reagent (ThermoFisher). For control, we used transfection with 2 μg/well of the empty vector. Transfection efficiency was determined using GFP expression as detected by fluorescence microscopy.

Nuclear RNA isolation was performed as previously described (39). Briefly, 48 hr post transfection, the six-well plates were washed with PBS followed by the addition of 175 μl of cold RLN buffer (50 mM Tris pH 8, 140 mM NaCl, 1.5 mM $MgCl_2$ and 0.5% NP40). The cells were then scraped and moved to an Eppendorf tube on ice for 5 min. Centrifugation for 2 min at 300 g, at 4°C, was then performed. The supernatant was transferred to a new tube and the pellet (nuclei) was centrifuged again. RNA was then extracted from nuclei with miRNeasy mini kit (Qiagen), following the manufacturer's instructions. All experiments were performed with at least three biological replicates.

Quantitative PCR

TaqMan microRNA assay. For RT-PCR of miR-7704, the TaqMan Advanced miRNA cDNA synthesis kit (ThermoFisher) was used according to the manufacturer's instructions**,** which included polyA tailing, adaptor ligation and miR-Amp reaction.

RT of mRNA. RT of nuclear RNA was performed using the High Capacity cDNA Reverse Transcription Kit (ThermoFisher) according to the manufacturer's instructions, using RT Random Primers, and MultiScribeTM Reverse Transcriptase.

Quantitative PCR reaction. mRNA and miR-7704 levels were measured using the TaqMan Fast Advanced Mix (ThermoFisher) and the following TaqMan Assays with FAM/MGB-NFQ primers/probe: TaqMan Advance MiR Assay hsa-mir-7704, 480576_mir (AB-25576 ThermoFisher); TaqMan Gene Expression Assays MTO, XS/PC: beta actin: Hs99999903_m1 (AB-4453320); TaqMan Gene Expression Assays /PC: HOXD1: Hs04334671_g1 (AB-4448892); Custom TaqMan Copy Number Assays, SM/PC: HAGLR_NR_033979.2 (AB-4400294). Fw primer: 5' -TGCCAAGCTGAGCACGTC-3', Rev: 5'-TACTCCAGATCTGGGGAC-3', FAM Probe: 5'-ACGTACTCCAGATCTG-3'. Assays were performed according the manufacturer's instructions. Amplification was carried out using a QuantStudio 12K Flex Real-Time PCR System (for downregulation), and StepOnePlus Real-Time PCR System (for overexpression), for 40 cycles at annealing temperature of 60°C. Analysis was performed using the delta delta C_T, $2^{-\Delta\Delta C_T}$ method. All experiments were performed in at least three biological repeats.

RNA folding and Bioinformatics tools

We have used the RNAfold software as part of a set of web-tools for analyzing the thermodynamics of folding (44). We applied the program by default parameters. The set of miR-GAS that are longer than 22 nt were analyzed. The minimum free energy (MFE) that is calculated is reported along other quantities such as the number of predictions and the average free energy for the folded structure.

Ethics approval

The authors declare that the study did not raise ethical issues in the context of the current legislation.

RESULTS

Hundreds of small RNA types are associated with the supraspliceosome

A large proportion of miRNA genes are harbored within introns of genes transcribed by Pol II. As processing of Pol II transcripts occur in supraspliceosomes, we hypothesized that key microprocessor components should be present in supraspliceosomes. Here, we wished to characterize the composition of supraspliceosome-associated small RNA

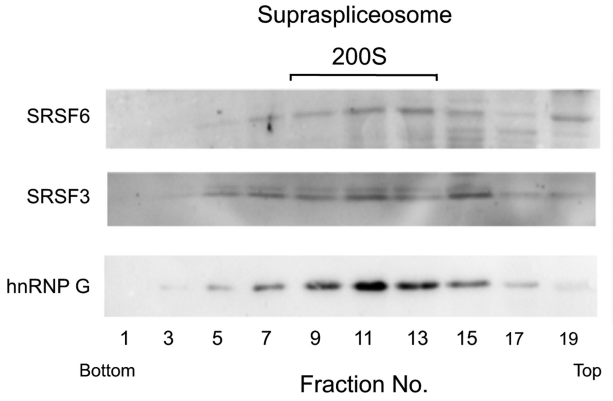

Figure 1. Splicing factors mark the supraspliceosome fraction (SF). WB analysis of the distribution across the glycerol gradient of splicing factors, previously shown to be predominantly found and associated with supraspliceosomes. Nuclear supernatants enriched for supraspliceosomes were prepared from HeLa cells and were fractionated in 10–45% glycerol gradients. Aliquots from odd gradient fractions were analyzed by WB using anti-hnPNP G and anti-SR antibodies. Supraspliceosomes peak in fractions 9–13. The distribution of hnRNP G (42 kDa), SRSF3 (SRp20, 20 kDa) and SRSF6 (SRp55, 55 kDa) across the gradient is presented. TMV was used as a size marker for the sedimentation (200 S).

sequences that align to the hairpin precursor, which includes pre-miRNA, as presented by the miRbase (hereinafter miR-GAS). To this end, we prepared nuclear supernatants enriched with supraspliceosomes under native salt conditions from HeLa cells, and fractionated them on glycerol gradients as previously described [(34), see Materials and Methods]. The extraction protocol preserves the higher order splicing complexes as shown by electron microscopy (34,45). Supraspliceosomes sediment at 200 S; and the splicing factors SR proteins and hnRNP G, predominantly found in supraspliceosomes and associated with them, mark their position in these gradients (36,46). Figure 1 shows the results of the WB analyses across the gradient using antibodies directed against these splicing factors.

Next, we extracted small RNA (<200 nt) from the SF (fractions 9–12, Figure 1). The RNA was used to build a barcoded library for sequencing by NGS (see Materials and Methods). Altogether, we analyzed six sequencing datasets, each library yielded 20 M to 30 M raw reads (Supplementary Table S1). Alignment of these supraspliceosomal sequences to the human transcriptome revealed a diverse collection of small RNA species, including pre-miRNAs, small nucleolar RNAs (SNORDs) (47), intronic sequences and more. In this study, we only consider reads that are aligned to the hairpin precursor miRNA (HP-miRNA) sequences that includes the stem-loop portion from the primary miRNA as determined by miRBase. Alignment was performed using the miRBase catalogue as well as the full human hg19 transcriptome (See Materials and Methods, Supplementary Table S1).

Overall, 503 miR-GAS were identified in the combined data from supraspliceosomes in HeLa cells (Supplementary Table S2). Only the 200 SF miR-GAS that are ≥17 nt long, and are supported by ≥10 reads will be further discussed (see Materials and Methods, Supplementary Table S2). To test that we do not introduce a bias by restricting mapping

of the NGS reads to miRBase sequences, we have repeated the analysis while mapping the reads to the entire human transcriptome (hg19 genome, see Materials and Methods). The two mappings gave consistent results (Spearman correlation $r = 0.78$, P-value = 3.3e–113, Supplementary Table S2). Additionally, independent libraries were prepared from total small RNA (<200 nt) from HeLa cell extract (CE). The data from CE is used as a reference, representing the mature, unfractionated, cellular miRNAs in HeLa cells.

Supraspliceosome-associated miRNA profile differs from the cellular one

A large population of human miRNAs are located in introns. These are either transcribed together with their host gene on the same transcriptional unit or at the opposite strand. Among them, many are transcribed by their own promoters, others are coupled to splicing as mirtrons or tailed mirtrons (48). The rest of the miRNAs are associated with intergenic regions and are single exon genes transcribed by Pol II as autonomous transcriptional unit. To test whether the collection of miRNAs identified at the SF are limited to intronic miRNAs, and thus are naturally involved with the splicing machinery during miRNA biogenesis, we compared the genomic location of the miRNAs that were identified in SF versus CE (Supplementary Figure S2). The SF is significantly enriched with miRNA sequences of intronic localization (72%) relative to the CE (62%; P-value = 0.00076). Note that the fraction of reads associated with miRNAs in the SF is only 4.2% of their number in the CE (Supplementary Table S1), suggesting that the impact of the SF on the CE (which also includes the nucleus and the supraspliceosome within it) is negligible. The substantial fraction (28%) of intergenic miR-GAS in the supraspliceosome indicates that the presence of these sequences in SF are not a mere reflection of pre-miRNA biogenesis but rather suggest a broader role in gene expression.

We assessed the overlap between the miRNA sequences identified in HeLa SF and CE, the latter being dominated by cytoplasmic miRNAs. Out of the 200 supraspliceosomal miR-GAS, 147 appear in both preparations (SF and CE), and 53 could only be identified in SF (Figure 2A). Focusing on mature miRNAs in the two fractions, we identified 107 mature miRNAs that are shared between SF and CE, and 13 that are exclusive to the SF (Figure 2B). In addition, there are many mature miRNAs (245) that appear in the CE but not in the SF (see Supplementary Table S3).

Table 1 lists the SF miR-GAS that are supported by ≥20 reads. The most striking observation is the extremely high abundance of miR-6087 that accounts for 17% of the total reads associated with miR-GAS from the SF, yet it is virtually absent in the CE. Note that most sequences identical to miRNAs that are exclusive to the SF were recently identified, and are considered novel miRNAs (48,49). Other highly abundant SF miR-GAS that are undetected in CE include miR-633a, miR-7704 and miR-622. For a complete list of miR-GAS in SF and CE, and for the subset of mature miRNAs in the SF, see Supplementary Table S3.

As the total number of HP-miRNA mapped reads from the SF is substantially smaller relative to the CE (4.2%), we compared the ranking of pre-miRNAs from both preparations rather than their absolute amounts. For this analysis, we considered the 200 miRNAs from SF that were mapped to all human miRBase sequences with a predetermined threshold of ≥10 reads (Figure 3A), and identified those that exhibit substantial ranking difference in the SF compared to the CE (where 1 is the rank of the miRNA with the highest number of reads). The ranking of most miRNAs is roughly preserved. However, the ranking of 38 miRNAs (∼25%) substantially differ between the SF and the CE (Supplementary Table S3). From them, 12 miRNAs show over-representation in the CE (e.g. miR-148a, let-7a-1, let-7a-2 and let-7a-3). However, the great majority show an opposite trend, and are over-represented in the SF. For example, miR-1246 is ranked 7 in the SF, but only 96 in the CE. Other extreme examples include miR-378C, miR-663b and miR-7704 (Figure 3A). We suggest that these miR-GAS are of a special interest, and their prevalence at the supraspliceosome indicates potentially overlooked functions in the nucleus.

For testing whether the observed phenomenon is dominated by mature miRNAs, we repeated the above analysis for mature miRNAs only (Figure 3B, Supplementary Table S3). For SF miRNA-GAS that are aligned mostly to mature mRNAs we observed, as expected, no change between Figure 3A and B (e.g., miR-148a, let-7a-1, let-7a-2 and let-7a-3, Supplementary Table S3). It is evident that the ranking bias (Figure 3A) applies also to miR-GAS that are not aligned to mature miRNA sequences (e.g. miR-19b-1, miR-622). These non-canonical sequences strongly point at an additional miRNA processing mechanism, presumably within the supraspliceosome.

Table 2 compares the most abundant miR-GAS in the SF to their cognate miRNAs in the CE. Note that the top 10 miR-GAS account for over 70% of all miR-GAS reads from the SF. Among the most prevalent miR-GAS are variants of let-7, miR-21 and miR-100, which are also extremely abundant in CE.

Table 1. Top miR-GAS that are detected in the SF but not in the CE

miRNA	# reads in SF	% in the SF
miR-6087	23686	17.04
miR-663a	2853	2.05
miR-7704	2798	2.01
miR-622	2436	1.75
miR-3960	226	0.16
miR-663b	210	0.15
miR-2861	142	0.10
miR-7110	94	0.07
miR-3648	81	0.06
miR-7705	65	0.05
miR-492	60	0.04
miR-639	55	0.04
miR-7641-1	54	0.04
miR-7641-2	48	0.03
miR-3654	48	0.03
miR-8086	40	0.03
miR-1244-3	38	0.03
miR-3652	35	0.03
miR-4792	30	0.02
miR-7706	27	0.02
miR-5047	23	0.02
miR-4444-1	22	0.02
miR-4709	21	0.02

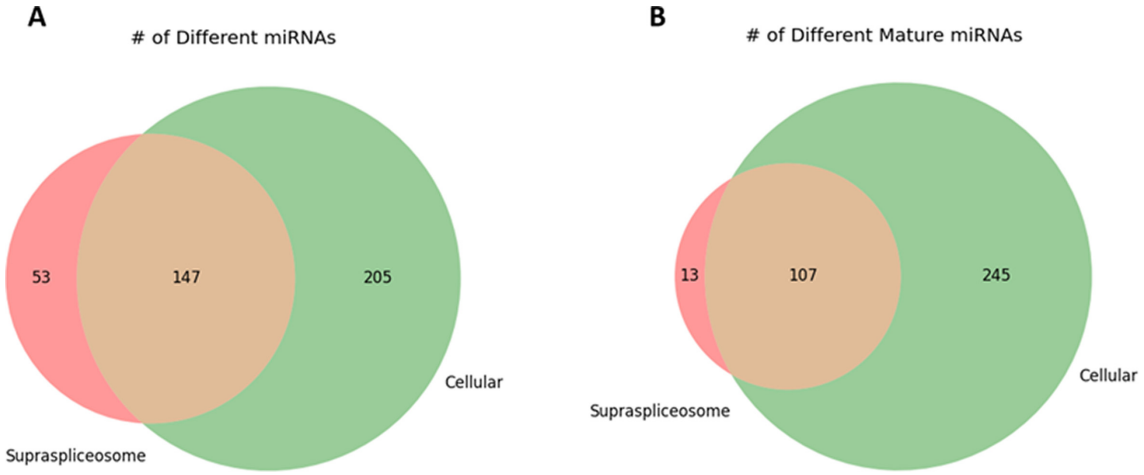

Figure 2. The collection of sequences identical to miRNAs in supraspliceosomes from HeLa cells. (**A**) Intersection of HP-miRNAs aligned sequences found in SF and CE. (**B**) Intersection of mature miRNAs aligned sequences found in the SF and the CE. Counting of miRNAs is a unified collection from the six independent libraries. Each miRNA is associated with ≥10 reads according to mapping to the entire miRBase sequence collection. The evidence for the 200 miR-GAS are supported by multiple HeLa libraries (see Supplementary Figure S1D).

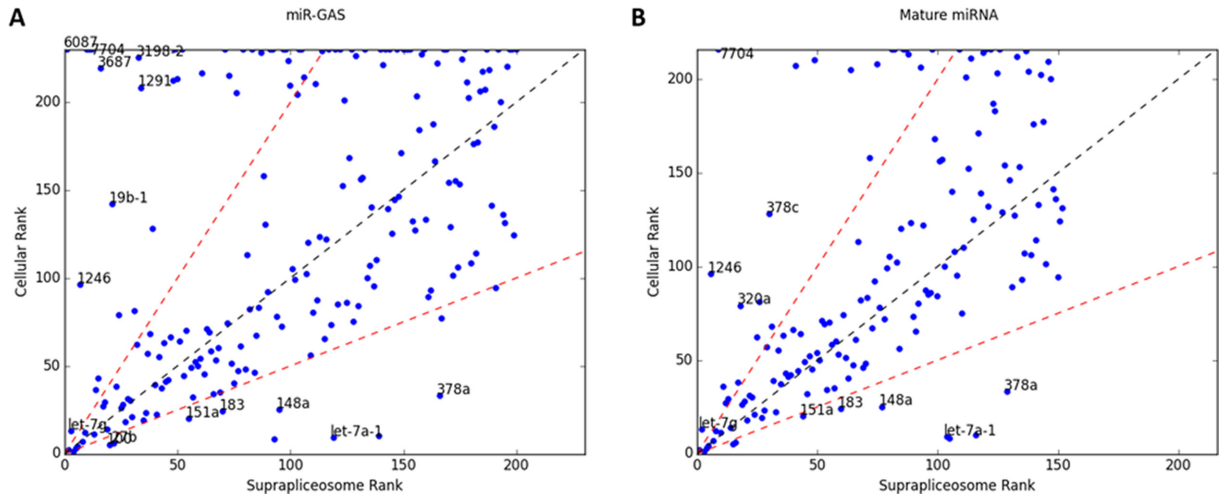

Figure 3. Comparison of miR-GAS abundance ranking in SF and CE. (**A**) Comparison of the rank order for miR-GAS with a threshold of ≥10 reads in each, in view of their rank in CE. Note that rank 1 is associated with the most abundant miRNA sequence and rank marked as 230 is associated with all cases in which no reads were identified in the CE. (**B**) Comparison of ranks when only mature miRNAs are compared in both preparations. The dashed lines show arbitrary boundaries for defining a miRNA that show a significant deviation in ranking by at least 4-fold. The annotated miRNAs are representatives of those showing a significant bias in their rank order. Those that are biased towards a higher abundance in SF relative to CE are shown in the top-left section and the miRNAs with an opposite trend are in the lower-right section of the figure. Source data are in Supplementary Table S3 and S4.

Table 2 also indicates the percentage of mature miRNAs that are associated with each of the identified miRNA-GAS (see Supplementary Table S4).

A quarter of the supraspliceosome-associated miR-GAS are not mature miRNAs

We classified the miR-GAS found in the SF according to the regions to which they align. These regions are based on HP-miRNAs (as determined by miRBase): mature miRNA (combining mature sequences at the 5p and 3p); undefined complement (the complementary sequence according to miRBase, where there is no evidence for a second miRNA); complete HP-miRNA that includes sequences identical to pre-miRNA; and miRNA extension (including the tails of <50 nt of genomic extended sequences from the 5′ mature miRNA and the 3′ of the complementary miRNA). In addition, reads that do not reside within a single region, but rather overlap two or more regions, are categorized as overlapping region (Figure 4A).

Figure 4B and C show the distribution of reads among these disjoined categories, as well as the read length distribution within each category. Mature miRNAs accounts for the majority of miR-GAS (76% of normalized counts), and their length is, as expected, centered around 22 nt. The remaining 24% miR-GAS mostly belong to overlapping regions and miRNA extension categories. Notably, the length distribution in these two categories is wide (Figure 4B).

Most miR-GAS associated with a high number of reads are assigned entirely to mature miRNAs (Table 2). How-

Table 2. Top 20 miR-GAS ranked by their abundance in the SF relative to the CE

miRNA	SF read count	CE read count	Rank SF (re-mapping)[a]	Rank CE[b]	% in the SF	% in the CE	% SF mature miRNA
miR-6087	23686	0	1	230	17.04	0.00	0.19
hsa-let-7i	16229	243109	2	2	11.67	7.49	99.75
hsa-let-7g	11704	68575	3	13	8.42	2.11	99.97
miR-21	11085	1003460	4	1	7.97	30.91	98.70
hsa-let-7f-2	10257	137382	5	3	7.38	4.23	99.91
hsa-let-7f-1	10169	133159	6	4	7.31	4.10	99.18
miR-1246	3365	1025	7 (8)	96	2.42	0.03	100.00
miR-92a-1	3364	87008	8 (7)	7	2.42	2.68	99.11
miR-92a-2	3173	78660	9	12	2.28	2.42	99.97
miR-663a	2853	0	10 (11)	230	2.05	0.00	0.88
miR-7704	2798	0	11 (10)	230	2.01	0.00	100.00
miR-622	2436	0	12 (14)	230	1.75	0.00	0.08
miR-30a	2311	79239	13 (12)	11	1.66	2.44	98.92
hsa-let-7b	2063	17074	14 (13)	36	1.48	0.53	99.47
miR-20a	1446	8216	15	43	1.04	0.25	18.95
miR-3687	1415	4	16	219	1.02	0.00	4.31
miR-24-2	1410	21682	17	27	1.01	0.67	98.94
miR-24-1	1370	20909	18	29	0.99	0.64	99.93
miR-99a	1368	63965	19	14	0.98	1.97	100.00
miR-100	1342	107344	20	5	0.97	3.31	99.40

[a]Ranking according to mapping vs human transcriptome (in parenthesis). The rank for all other miRNAs remained identical in both mappings schemes.
[b]Rank marked as 230 is associated with all cases in which no reads were identified in the CE.

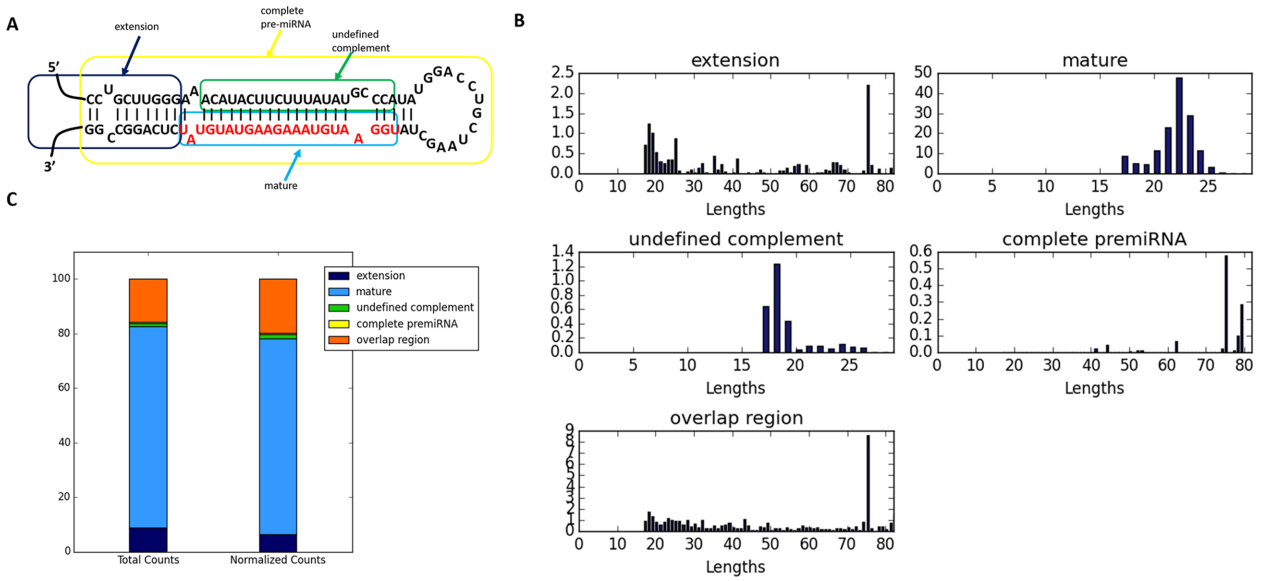

Figure 4. Partition of the miR-GAS to predefined regions of HP-miRNAs. (**A**) A schematic view on an HP-miRNA prototype (marked yellow, example of miR-663a) where the different regions are marked according to their positions on the HP-miRNA (see text). (**B**) Length histogram of reads mapped to each pre-defined region according to A. A flexible assignment of = <4 nt are used for the predefined regions. (**C**) The relative counts of reads to each of the predefined regions. Total counts of reads (left) and normalized counts (right; the relative abundance of reads aligned to a specific miRNA is summed to 1) are shown. All miRNAs that are analyzed have ≥10 reads, a minimal read length of ≥17.

ever, there are some exceptions. For miR-6087, miR-663a and miR-622, mature miRNA sequences account for <1% of the aligned reads, suggesting an abundant yet-uncharacterized HP-miRNAs fragments. Additionally, for miR-20a and miR-3687, mature miRNAs account for only 4.3% and 19% of the reads, respectively. In both instances, we identified a large number of overlapping reads, suggesting their non-conventional processing.

We analyzed all miR-GAS with ≥10 reads. For only 94 of them (47%) we observed that the mature miRNA is supported by >99% of the reads (Figure 5A, Supplementary Table S4). These mature miRNAs might be explained as shuttled back from the cytoplasm. In 77 of the miR-GAS (38.5%), at least 10% of the reads are not assigned to mature miRNAs. Among them, for 33 miRNAs (16.5%), the amount of mature miRNAs is <1%. For example, 216 out of 233 reads of miR-4436 are assigned to overlapping regions (Supplementary Table S4). The results in Figure 5A strongly suggest an overlooked processing mode that differs from the canonical one, that generates the bulk of cytoplasmic miRNAs.

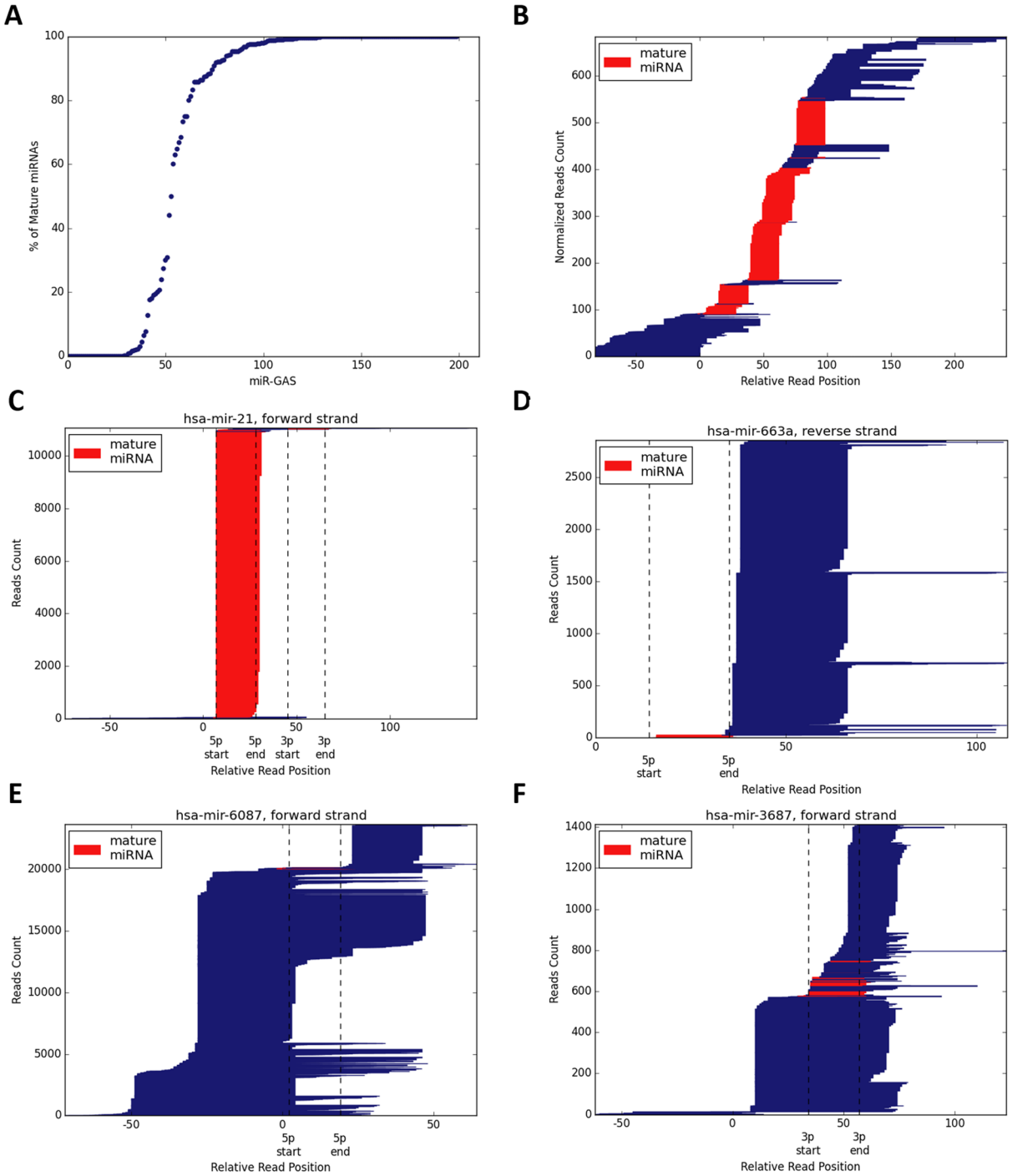

Figure 5. Read-length and relative position of aligned fragments. (**A**) Analysis of miR-GAS that are supported by ≥10 reads and mapped to miRBase sequences, sorted by their fraction of mature miRNA versus the total reads. (**B**) Sum of reads mapped to all analyzed miRNAs. Read-length of the different miRNAs are aligned relative to the HP-miRNA start position. On the x-axis, coordinate zero is identical to the 5′ of HP-miRNAs as defined by miRBase. (**C**) Sum of reads mapped to miR-21. (**D**) Sum of reads mapped to miR-663a (located on the reverse strand). (**E**) Sum of reads mapped to miR-6087. (**F**) Sum of reads mapped to miR-3687. Blocks of reads from the same starting points reflect non-randomized cleavage sites. The mature miRNAs are colored in red, and the width of the bars is proportional to the number of reads. Dashed block marks the position of the mature miRNAs as annotated by miRBase. Coordinate marked as 0 refers to the 5′-end of the HP-miRNA as reported by miRBase. Reads that originate from miRNA extension (<50 nt from each end of the HP-miRNA) are reported.

Figure 5B displays the overall position and counts of reads along the HP-miRNA, showing the unified, normalized data from 200 miR-GAS from SF (with ≥10 reads mapped to miRBase, Supplementary Table S4). It is clear that many reads are associated with miRNA extension in both ends of their HP-miRNAs. Some proximal expressed miRNA transcripts (separated along the genome by less than a read length, 76 nt) may share reads of neighboring miRNA genes that are in a cluster (e.g. miR-92a1 and miR-92b1 are only 36 nt apart, and mir-20A are only 57 nt apart from miR-92b1).

A varied distribution of reads along the pre-miRNA characterizes the SF miR-GAS collection

Figure 5C shows the profile and length distribution of miR-21. This miRNA is among the most abundant miRNAs identified in the SF. In this case, the mature miRNA dominates, and its fraction is >98.8% of all reads. It is likely that miR-21 was shuttled back from the cytosol and as such it is a prototype for other miRNAs that are processed in the cytosol but are still identified at the SF. Furthermore, most reads in Figure 5C have the same sharp and well-defined edges, supporting a well-defined processing pathway. In miR-6087, the most abundant miR-GAS expressed in SF (17%, Table 1, Figure 5E), 44% of the reads are aligned to the extension region of miRNA that includes <50 nt around the HP-miRNA in the genome. The rest are aligned to overlapping regions, with only minute fraction aligned to the mature miRNA sequence. Another illuminating example is miR-663a (2853 reads, Figure 5D), where most of its reads map to overlapping regions and not to the classical mature 5p. Figure 5F illustrates miR-3687 that exhibits a complex, non-conventional mapping pattern and provides evidence for a full miRNA hairpin. The set of examples described above shows the unique characteristics that specify the small RNA sequences derived from HP-miRNAs in the SF (Supplementary Table S4). This unique pattern supports a non-conventional processing, presumably at the supraspliceosome.

We tested the possibility that miR-GAS that are longer than classical mature miRNAs (22 nt) are characterized by stable secondary structures. A structure-based blockage in the processivity of a nuclease at the supraspliceosome may lead to accumulation of such reads. The folding energy that is associated with non-canonical miR-GAS fragments was calculated. Among the long miR-GAS that are abundant in the SF, we observed miR-6087, miR-3687, miR-622 and miR-639, as candidates for structural-dependent processing at the supraspliceosome (Supplementary Table S5).

miRNAs 5p and 3p strand preference in the SF in comparison to the CE

To further search for characteristics of supraspliceosomal miRNA sequences, we tested the relative abundance of the strands identified by the mature miRNAs (5p and 3p). As both strands may originate from a single pre-miRNA, we defined the percentage of the 5p population using the fraction of 5p/(5p+3p). Supplementary Figure S3 presents these values for 127 mature miRNAs that are found in both the SF and the CE (Supplementary Table S6, a threshold ≥10 reads). We note that as expected from a canonical processing of the duplex miRNAs, mature miRNAs originate from either the 5p arm or the 3p arm, and rarely from both. Indeed, mature miRNAs from the CE have a balanced representation, on average (Supplementary Figure S3). Mature miRNAs from the SF show a balanced representation as well, with a slight (insignificant) preference to the 5p strand.

While the overall distribution of the 5p versus 3p is similar between the SF and CE, some individual miRNAs exhibit marked differences. Supplementary Figure S3 shows miRNAs where the difference in 5p percentage between the SF and CE is >20% (e.g. miR-30e and miR-27a). A complete list is provided in Supplementary Table S6.

Table 3 lists the sequences identical to miRNA, with the most significant difference between the prevalence of the 5p and 3p strands between the SF and the CE (only miRNAs with >20 reads in the SF are considered, to increase the reliability of detection). Table 3 also includes miR-GAS that seem to have preference towards a strand that is not documented in miRBase (50), thus labeled by us as 'undefined complementary'. For example, miR-622 is considered to be produced entirely of the 3p arm. However, we identified 1513 reads that presumably emerge from the complementary 5p arm.

Potential functions of the pre-miRNAs from SF

To obtain insights regarding possible functions of miR-GAS, we focused on the 23 miR-GAS that were observed exclusively at the SF (Table 1). The number of reads per miRNA varies among them by two orders of magnitudes (Supplementary Figure S4A). We focused on seven of these miRNA sequences that reside within exons, as their possible function may be more readily understood (Supplementary Figure S4A, orange bars).

Supplementary Figure S4B shows a genome browser view for these miRNAs. Whereas miR-7705 and miR-4709 are located at the 3′UTR of their host gene (framed by light blue), the other five are located at the first 5′UTR exon of their host gene (framed by light red). Their location near the start of the transcription hints at a possible role of these five miRNAs in recruiting components of the transcriptional machinery and consequently are likely to attenuate the expression of their host genes.

Nuclear miR-7704 sequences in regulation of gene expression

We identified 2798 reads for miR-7704-5p (Supplementary Table S4) that resides within the HOXD1 gene, a key nuclear transcription factor that is involved in limb differentiation, covering the coding region of the first exon. This high number of reads suggests that this miRNA has the potential to indirectly attenuate the expression of genes in its vicinity. It is possible that this mature miRNA prevents transcription by base-pairing with the HAGLR gene (called HOXD Antisense Growth-Associated Long Non-Coding RNA). As its name implies, it is complementary to miR-7704-5p and to HOXD1 (Figure 6A).

To validate potential function of SF miR-GAS in gene expression we tested the effect of changes in expression of SF

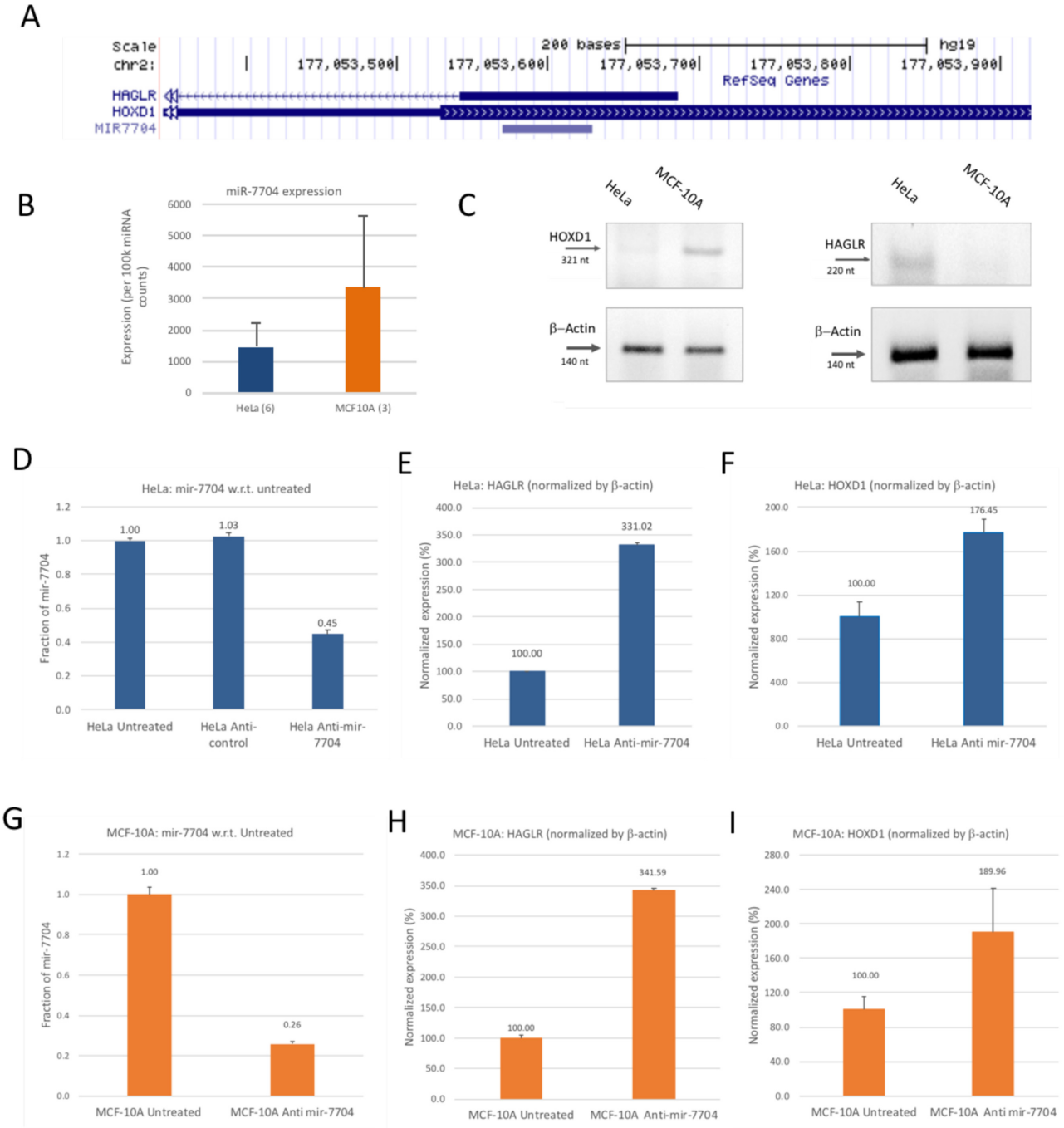

Figure 6. Regulation of gene expression by nuclear miR-7704. (**A**) Genome browser view for miR-7704 and the overlap with HOXD1 and HAGLR, a long non-coding RNA gene. (**B**) Average expression levels of miR-7704 from the SF based on HeLa (six samples, blue) and MCF-10A cells (three samples, orange). The expression values are based on normalizing each sample to miRNA read counts (100k per sample). The statistical bars are the calculated standard errors. (**C**) Comparison of the results of RT-PCR assays of HOXD1 and HAGLR expression in HeLa and mammary epithelial MCF-10A cells. Results represent RT-PCR from three biological repeats. The identity of the extracted bands was confirmed by DNA sequencing. (D–F) Effect of down-regulation of miR-7704 on HAGLR expression in HeLa cells. Transfection with Anti-miR-7704 inhibitor resulted in down-regulation of miR-7704 (**D**), increase in the expression level of HAGLR mRNA (**E**), and mild elevation of HOXD1 (**F**). (G–I) Effect of down-regulation of miR-7704 on HAGLR expression in MCF-10A cells. Transfection with Anti-miR-7704 inhibitor resulted in down-regulation of miR-7704 (**G**), increase in the expression level of HAGLR mRNA (**H**) and mild elevation of HOXD1 (**I**). D–I, each, show results of qPCR of three independent biological preparations. The expression levels of HAGLR and HOXD1 were normalized to β-actin expression from the same preparation.

Table 3. Percentage of mature miRNA sequences originated from the 5p arm (Top). Reads assigned to 'undefined complement' according to miRBase (Bottom)

miRNA intersect	SF reads	SF 5p (%)	CE reads	CE 5p (%)	Difference (%)
miR-374a	36	8.3	2070	49.9	41.6
miR-27a	431	2.1	24594	43.6	41.5
miR-365a	23	56.5	762	20.3	36.2
miR-181a-1	81	100.0	6513	69.0	31.0
miR-30e	173	8.7	7258	34.6	25.9

miRNA	**SF reads**	**SF 5p (%)**	**miRBase annotation**	**miRBase 5p (%)**	**Difference (%)**
miR-492	26	0	only 5p	100	100
miR-4449	52	100	only 3p	0	100
miR-622	1513	100	only 3p	0	100
miR-6087	110	41.8	only 5p	100	58.2

miR-7704 on gene expression of HOXD1 and HAGLR. To this end, we first compared the expression of HOXD1 and HAGLR transcripts in HeLa and in mammary epithelial cells MCF-10A with the level of SF miR-7704 in these two cell types (Figure 6B-C). The same protocols for isolating SF, sequencing the small RNA (<200 nt) and NGS analysis were applied for the MCF-10A cells. We find an inverse correlation between the expression of SF miR-7704 and that of HAGLR that is expressed in the antisense direction. The level of SF miR-7704 seems to be lower in HeLa cells relative to MCF-10A cells. The RT-PCR analysis revealed that the expression of HAGLR is higher in the HeLa cells compared to MCF-10A cells (Figure 6C). HOXD1 is likely to be indirectly affected by the abundance of miR-7704, presumably via the attenuation of HAGLR transcript (an antisense of HOXD1).

To validate the effect of supraspliceosomal miR-7704 sequences on the expression of HAGLR, we transfected HeLa cells with anti-miR-7704 inhibitor. Nuclear RNA was extracted from the transfected cells and the expression levels of miR-7704, HAGLR and HOXD1 was measured by qPCR, using β-actin for normalization. Cells transfected with non-relevant anti-miR and untreated cells were used as controls (See Materials and Methods). As can be seen in Figure 6D, transfection of HeLa cells with anti-miR-7704 inhibitor reduced the level of nuclear miR-7704 to 45% compared to non-treated cells, while non-relevant Anti-miR had hardly any effect. As result of the decrease in the level of nuclear miR-7704, the level of lincRNA HAGLR was upregulated >3-fold (Figure 6E). The expression level of HOXD1 in untreated cells was low, and was mildly upregulated by decreasing the level of miR-7704 (Figure 6F). A similar experiment performed on MCF10-A cells (Figure 6G–I) revealed that following anti-miR-7704 transfection, the level of nuclear miR-7704 in the cells was down regulated to 24% relative to non-treated cells. Consequently, the level of HAGLR was upregulated over 3-fold (Figure 6H), while the expression level of HOXD1 was mildly upregulated.

To further validate the effect of nuclear miR-7704 on the expression of HAGLR, we overexpressed miR-7704, by transfecting HeLa cells with the pPRIME-miR-7704-CMV-GFP-FF3 plasmid (see Materials and Methods), using the empty vector as control. Nuclear RNA was isolated and the level of miR-7704 and mRNAs was determined by qPCR (See Materials and Methods). Figure 7A shows a strong overexpression of miR-7704. This upregulation of miR-7704 resulted in a decrease in the level of HAGLR mRNA by over two folds, while the level of HOXD1 mRNA was slightly upregulated (Figure 7B, C). We thus conclude that miR-7704 negatively regulates the expression of the lncRNA HAGLR.

Supraspliceosomal miRNA sequences may affect gene expression and splicing

The presence of miR-GAS in supraspliceosomes suggests that at least some of these sequences might compete with splicing. An interesting example is miR-99b (Figure 8A). Almost all of the reads (96%) originates from the 5p arm of the miRNA (miR-99b-5p, Figure 8C) and the rest of the reads are from the 3p arm. Figure 8B shows that miR-99b is harbored in intron 1 of an isoform of the SPACA6 gene (Sperm Acrosome Associated 6), which has an extended upstream 5′ non-coding exon. This isoform is one of several experimentally validated variants of the SPACA6 gene. Interestingly, miR-99b is fully complementary to the 5′ splice junction of intron 1 of SPACA6P-AS, which is transcribed in the antisense direction of SPACA6 (Figure 8B), spanning positions –6 to +16 with respect to the splice junction. SPACA6P-AS, also termed LINC01129, is a long non-coding RNA for which very little is known. The abundant miR-99b-5p (Figure 8C) is complementary to the splice junction of SPACA6P-AS, 5′-AAG/gtcggt-3′ (exon sequence is in capital letters) (Figure 8D), possibly forming a duplex of 22 nt.

DISCUSSION

miR-GAS from the SF differs from miRNAs in CE

Although the most studied role of miRNAs is the down-regulation of transcription in the cytoplasm, information on nuclear functions of miRNAs is accumulating (reviewed in (7–9)). In this study, we focus on sequences aligned to HP-miRNAs found in supraspliceosomes, the nuclear pre-mRNA processing machine. We identified a large number of sequences aligned to 200 HP-miRNAs in supraspliceosomes, supporting our earlier finding of pre-miRNAs of in-

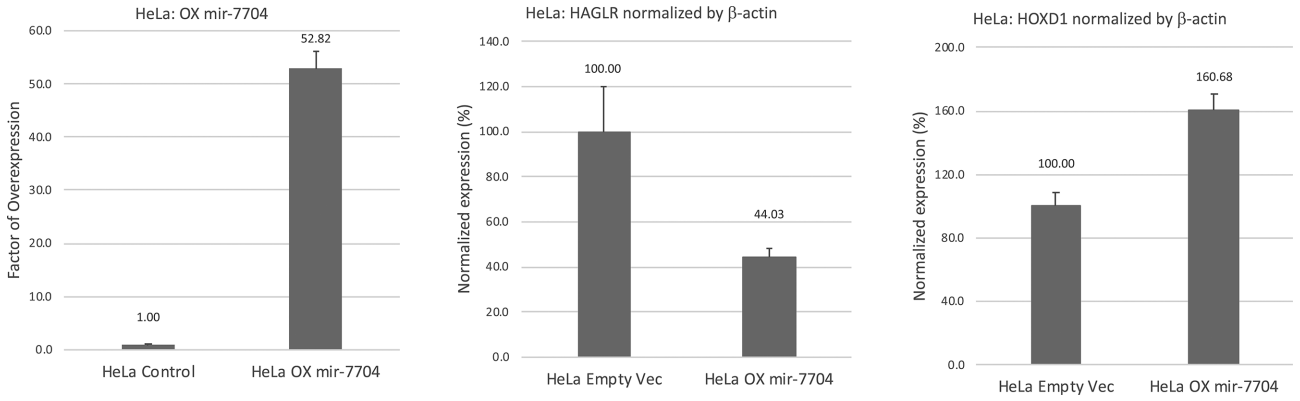

Figure 7. Negative regulation of HAGLR by miR-7704 – overexpression experiments. Transfection of HeLa cells with pPRIME-miR-7704-CMV-GFP-FF3 plasmid resulted in over 50-fold increase in the level of miR-7704, as measured by qPCR (**A**). This increase in the level of miR-7704 is accompanied by significant decrease in the level of HAGLR mRNA as measured by qPCR of HAGLR (**B**). The expression level of HOXD1, as measured by qPCR is presented (**C**). The results of qPCR are extracted from 3 independent biological preparations. Transfection efficiency ranged between 35% and 40% as determined by GFP. These obtained transfection efficiencies indicates that the actual effect of miR-7704 on gene expression is likely to be higher than the measured one. The expression levels of HAGLR and HOXD1 were normalized to β-actin expression from the same preparation.

tron 2 of HTR2C (40), and intron 13 of MCM7 in supraspliceosomes (39). It is also in accordance with our finding of the presence of DGCR8 and Drosha in supraspliceosomes, and the cross-talk between the microprocessor and the splicing machine (39). Taken together, it is possible that miRNA sequences also play a nuclear role in the supraspliceosome, possibly regulating gene expression and processing, through a yet undetermined mechanism.

Moreover, comparing HeLa miR-GAS profile in SF with their profile in CE, we identified sequences that are exclusive to the SF (Table 1, Figure 2). Some of the most abundant sequences in the SF (e.g. miR-6087, miR-663a, miR-7704, miR-622, Table 1) are hardly detected in the CE even though the cellular extract coverage is 23.5-fold larger. This suggests the presence of a unique repertoire of pre-miRNA derived sequences at the supraspliceosome.

Many supraspliceosomal pre-miRNA derived sequences result from a non-canonical RNA processing

As many miRNAs are harbored in introns, we expected to find sequences derived from intronic miRNAs in SF. Whereas the SF is indeed enriched with intronic miRNAs, we nevertheless find many sequences derived from miRNA sequences that are not embedded in introns (e.g. miR-1246, let-7i and miR-92a-2, Supplementary Figure S2). Furthermore, the canonical pathway of miRNA biogenesis includes cleavage of the pri-miRNA in the nucleus by the microprocessor to generate the pre-miRNA stem-loop molecule of about 65–80 nt long. The pre-miRNA is then exported to the cytoplasm, where it is further processed by Dicer to generate the mature miRNAs (2,3,51,52). Thus, the expectation is to find mostly complete pre-miRNA in the SF. Yet, we observe sequences identical to various parts of the HP-miRNAs. Namely, sequences that align to the complete pre-miRNA sequences, to mature miRNAs, and to other regions along the miRNA hairpin (Figure 4 and Supplementary Table S4). Atypical biogenesis pathways of pri-miRNAs in the nucleus have been documented (3), including the mirtron pathway which is DGCR8 and Drosha independent (53,54), the simtron pathway, which is DGCR8 independent (55), and the agotron pathway that is both DGCR8/Drosha and Dicer independent (3,56). Nevertheless, all these alternative processing pathways involve transport of a 65–80 nt stem-loop pre-miRNA from the nucleus to the cytoplasm. Thus, these pathways cannot explain the discovery of predominantly short fragments from the SF miRNA sequences.

Many of the fragments of miRNA sequences that we identified in supraspliceosomes are of a defined length and are characterized by sharp edges (Figure 5), suggesting that they are not byproducts of unspecific degradation. Furthermore, these defined fragments cannot be a result of the failure of the DNA polymerase to proceed due to modifications in the RNA along the process of library preparation. The library was prepared directly from the extracted small RNA fraction, and the 5′ and 3′ adaptors were ligated prior to any amplification step (see Materials and Methods).

A key finding in our study concerns the high percentage of mature miRNAs in the SFs, representing 76% of all normalized reads (Figure 4, Supplementary Table S4). The presence of mature miRNAs in nuclei has been reported as a result of an active shuttling from the cytoplasm to the nucleus (10,11). Notably, many of the exclusively identified miRNAs in the SF belong to the novel set of miRNAs that were recently identified from large-scale bioinformatics surveys [e.g. (57)].

Finding of reads which are aligned to overlap regions (Figure 4), and the fact that many of the reads are short and cover regions other than that of the mature miRNAs (Figures 4 and 5), suggest a local, nuclear processing through an unknown mechanism. We can therefore conclude that the defined sequences that we find represent genuine sequences, whose biogenesis and localization in the supraspliceosome calls for further research.

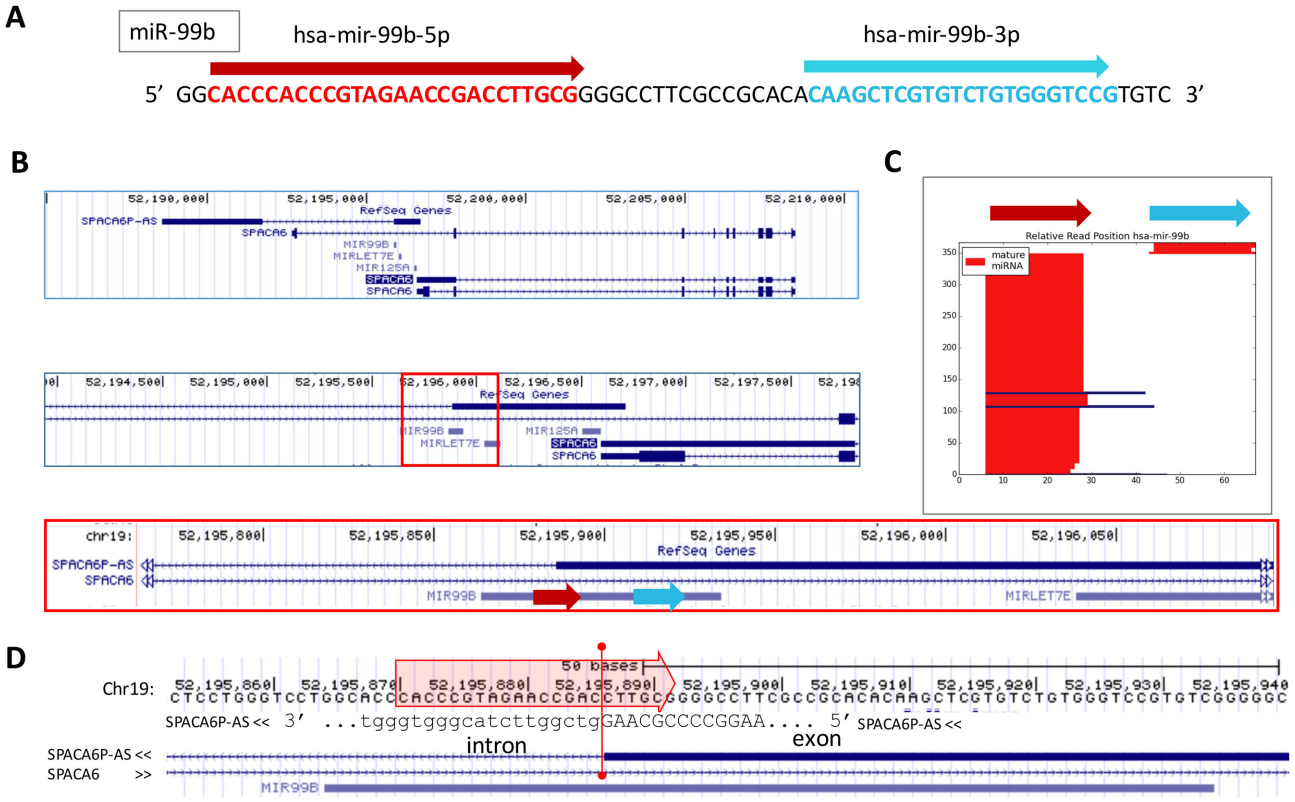

Figure 8. The chromosomal context of miR-99b portrays potential function in splicing. (**A**) The sequence of the pre-miRNA is shown and the 5p and the 3p are colored in red and blue, respectively. The image focuses on isoforms of SPACA6 gene (Sperm Acrosome Associated 6). The isoform with an extended upstream 5′ non-coding exon. (**B**) Several resolutions illustrating the ability of miR-99b-5p to completely complement, by base-pairing, to the 5′ splice junction of the intron of SPACA6P-AS, which is transcribed in the antisense direction of SPACA6. (**C**) Sum of reads mapped to miR-99b. (**D**) Detailed base-paring at the intron-exon junction. See text for details.

Supraspliceosomal pre-miRNA-like sequences originate from genomic strategic locations

Currently miRBase (version 21) includes 1881 miRNA precursors (2588 mature miRNAs) (50). Many of the newly reported miRNAs listed in miRBase may not comply with the canonical biogenesis and a genuine cytoplasmic cellular location [e.g. (48)]. Tools such as novoMIRank (58) are testing the confidence of novel candidate miRNAs from NGS experiments. Revised definitions for authentic miRNAs versus other miRNA-like sequences are proposed [see discussion in (59)].

A careful analysis of sequences that are abundant in SF and are not detected in CE (or detected in minute amounts, Supplementary Table S3) reveals about 10 cases where the sequences might also come from highly transcribed regions such as the rRNA gene clusters (60). Specifically, several of the exclusive miR-GAS are identical to sequences derived from rRNA clusters, and are often ignored at the NGS mapping step. Yet, they are also identical to sequences of the relevant pre-miRNA. For example, miR-6087 (within 28S), miR-3687 and miR-663a (within 5′-external transcribed spacers of rRNA, 5′-ETS) (61). These fragments were renamed as rRNA-hosted miRNA analogs (rmiRNAs), and were shown to be involved in cancer (61). Whether these stable miRNA-like fragments are processing byproducts of abundant RNA species calls for future validations.

Potential functions of supraspliceosomal miRNA derived sequences in gene expression and splicing

The supraspliceosome is a dynamic complex that packages pre-mRNA transcripts of different sizes and number of introns into complexes of a unique structure, that enables coordination of all pre-mRNA processing in the nucleus, including splicing, AS, 5′/3′ end processing, and RNA editing. It is not clear yet where and how the processing of the SF miR-GAS occurs, and further studies are required to elucidate this point. However, our results reveal that miR-GAS represent a defined population that varies from the cellular population in several aspects. These results suggest that at least some of the SF miR-GAS may function in regulating gene expression.

Recent studies revealed the involvement of nuclear miRNAs in a number of functions including transcription silencing and activation (7–11). An interesting instance is miR-320, a tumor suppressor that targets the BCR/ABL oncogene (62). Kim *et al.* (19) have studied miR-320 that is encoded within the promoter region of POLR3D (RNA Polymerase III Subunit D) in the anti-sense orientation, and thus fully complementary to it. An inverse correlation between the expression of miR-320 and POLR3D in a number of cell lines was shown. Furthermore, down-regulation of miR-320 resulted in an increased expression of POLR3D,

and overexpression of miR-320 led to down-regulation of POLR3D (19).

It should be noted that we identified 898 reads of miRNA-320a in supraspliceosomes, 99.5% of them were identical to mature miRNA-3p that likely affects expression of POLR3D (Supplementary Tables S4, S6). Kim *et al.* (19) also point at additional miRNAs that show full complementarity to promoter regions. Interestingly, among these miRNAs is miR-639, also identified in our study, which is expressed from the same strand as the protein coding gene TECR. This miRNA overlaps with the 5′-UTR of TECR (Supplementary Figure 4B), and is thus likely to affect TECR transcription by sequestering factors. We identified four additional supraspliceosomal miRNA-derived sequences that overlap the 5′-UTR regions of genes. Similarly, we noted the abundance of miRNAs that overlap genes at their 3′-UTR as well. These include miR-7705 that overlaps the gene PABPC1, and miR-4709, located at the 3′-UTR of the NPC2 gene (Supplementary Figure S4B). Most of these genes play a role in key cellular processes, and are thought to function in the nucleus. It is likely that such miRNAs may affect the stability and/or localization of their respective pre-mRNAs. Three of the SF miRNAs, miR-7704, miR-7705 and miR-7706 (Supplementary Figure S4), were recently identified together with a small set of unknown miRNAs that are activated in macrophages by interferon-27 (63).

We validated the potential of miR-7704 to alter the expression of the gene products at its vicinity. This miRNA overlaps with the 5′ end of the HOXD1 gene, and also displays a perfect complementary to the 5′-UTR of the HAGLR gene (Figure 6A). Therefore, it can play a role in coordinating their expression patterns. Our results show an inverse correlation between the expression of HAGLR and miR-7704 in two different cell lines, as expected from a direct competition on transcription. In a set of experiments to supress and overexpress the nuclear level of miR-7704, we found that the lowly expressed HOXD1 is mildly affected. Notably, in these experiments, we demonstrated that downregulation of miR-7704 led to upregulation HAGLR, while overexpression of miR-7704 resulted in suppression of HAGLR expression. We can thus conclude that nuclear miR-7704 negatively regulates the expression of the lincRNA HAGLR (Figures 6 and 7).

Transcriptional gene silencing and other nuclear functions were proposed to occur by mature miRNA (19,20). Importantly, in the case of SF miR-GAS with potential nuclear function in gene expression, only the sequences derived from miR-7704 are mature miRNA sequences. In five of the other six cases, the majority of the reads are derived from overlap regions (miR-7705, miR-4709, miR-639, miR-3652, miR-4444-1). The question how sequences derived from overlap-regions of HP-miRNA were produced, and how they remain associated with the supraspliceosome is beyond the scope of this study. Irrespective of that, the identification of SF miR-GAS that can target or compete transcripts by base pairing, suggests that sequences beyond the mature miRNAs should be considered in playing a role in the nucleus.

The interaction of small RNAs with the supraspliceosome is not limited to miRNAs. Indeed, supraspliceosome-associated SNORD27, lacking the methylase fibrillarin, was demonstrated as regulating the alternative splicing (AS) of E2F7 transcription factor, through base-pairing, and also affecting AS of several other genes (47). Among the miR-GAS found in supraspliceosomes we identified miR-99b that might play a similar role in splicing. This miRNA (Figure 8), which is harbored in the intron of the SPACA6 gene, is fully complementary to the 5′ splice junction of the first intron of SPACA6P-AS, transcribed in the antisense direction of SPACA6 (Figure 8B). This complementarity suggests that SF miR-99b can compete with the binding of U1 and U6 snRNPs required for splicing of LINC01129 (Figure 8D). Thus, miR-99b can play a role in determining the ratio between spliced and unspliced LINC01129 RNA. Furthermore, the supraspliceosomal miR-99b might be a new player in the balance of expression of SPACA6 and LINC01129.

CONCLUSIONS

The supraspliceosome is a dynamic machine involved in all pre-mRNA processing activities. Because both the canonical and atypical biogenesis steps of miRNAs involve the export of pre-miRNA to the cytoplasm, we do not expect to find mature miRNAs, or processed fragments that originate from pre-miRNAs in the supraspliceosome. Our findings illustrate the rich spectrum of miRNAs-aligned sequences in the supraspliceosome, suggesting a novel potential function for some of these miRNA sequences within the supraspliceosome, in addition to their known function of attenuating gene expression and suppressing translation in the cytoplasm. Here we show a novel function for nuclear miR-7704. Specifically, miR-7704 negatively regulates the expression of lncRNA HAGLR, which is overlapping and resides on the complementary strand. In addition, we provided a body of evidence that strongly suggest a role for specific sequences identical to miRNA genes in regulating transcription, RNA processing and splicing.

DATA AVAILABILITY

The sequencing data of this manuscript have been deposited in GEO and the accession number is GSE100803.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Aviva Pecho for excellent technical assistance. We thank Prof. Stefan Stamm (University of Kentucky) for the anti-hnRNP G antibodies, and the CS system at the Hebrew University for the support in processing and storage the data.

FUNDING

Israel Cancer Research Fund (ICRF) Acceleration Grant (to R.S and L.C) (in part); ELIXIR Excelerate [676559 to M.L.]. Funding for open access charge: ICRF.

Conflict of interest statement. None declared.

REFERENCES

1. Di Leva,G., Garofalo,M. and Croce,C.M. (2014) microRNA in cancer. *Annu. Rev. Pathol.*, **9**, 287–314.
2. Krol,J., Loedige,I. and Filipowicz,W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, **11**, 597–610.
3. Ha,M. and Kim,V.N. (2014) Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.*, **15**, 509–524.
4. 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.
5. Han,J., Lee,Y., Yeom,K.H., Nam,J.W., Heo,I., Rhee,J.K., Sohn,S.Y., Cho,Y., Zhang,B.T. and Kim,V.N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901.
6. Landthaler,M., Yalcin,A. and Tuschl,T. (2004) The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. *Curr. Biol.*, **14**, 2162–2167.
7. Roberts,T.C. (2014) The MicroRNA biology of the mammalian nucleus. *Mol. Ther. Nucleic Acids*, **3**, e188.
8. Guil,S. and Esteller,M. (2015) RNA-RNA interactions in gene regulation: The coding and noncoding players. *Trends Biochem. Sci.*, **40**, 248–256.
9. Huang,V. and Li,L.C. (2012) miRNA goes nuclear. *RNA Biol.*, **9**, 269–273.
10. Liao,J.Y., Ma,L.M., Guo,Y.H., Zhang,Y.C., Zhou,H., Shao,P., Chen,Y.Q. and Qu,L.H. (2010) Deep sequencing of human nuclear and cytoplasmic small RNAs reveals an unexpectedly complex subcellular distribution of miRNAs and tRNA 3′ trailers. *PLoS one*, **5**, e10563.
11. Jeffries,C.D., Fried,H.M. and Perkins,D.O. (2011) Nuclear and cytoplasmic localization of neural stem cell microRNAs. *RNA*, **17**, 675–686.
12. Ohrt,T., Mutze,J., Staroske,W., Weinmann,L., Hock,J., Crell,K., Meister,G. and Schwille,P. (2008) Fluorescence correlation spectroscopy and fluorescence cross-correlation spectroscopy reveal the cytoplasmic origination of loaded nuclear RISC in vivo in human cells. *Nucleic Acids Res.*, **36**, 6439–6449.
13. Ameyar-Zazoua,M., Rachez,C., Souidi,M., Robin,P., Fritsch,L., Young,R., Morozova,N., Fenouil,R., Descostes,N., Andrau,J.C. *et al.* (2012) Argonaute proteins couple chromatin silencing to alternative splicing. *Nat. Struct. Mol. Biol.*, **19**, 998–1004.
14. Weinmann,L., Hock,J., Ivacevic,T., Ohrt,T., Mutze,J., Schwille,P., Kremmer,E., Benes,V., Urlaub,H. and Meister,G. (2009) Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell*, **136**, 496–507.
15. Leucci,E., Patella,F., Waage,J., Holmstrom,K., Lindow,M., Porse,B., Kauppinen,S. and Lund,A.H. (2013) microRNA-9 targets the long non-coding RNA MALAT1 for degradation in the nucleus. *Sci. Rep.*, **3**, 2535.
16. Hansen,T.B., Wiklund,E.D., Bramsen,J.B., Villadsen,S.B., Statham,A.L., Clark,S.J. and Kjems,J. (2011) miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.*, **30**, 4414–4422.
17. Tang,R., Li,L., Zhu,D., Hou,D., Cao,T., Gu,H., Zhang,J., Chen,J., Zhang,C.Y. and Zen,K. (2012) Mouse miRNA-709 directly regulates miRNA-15a/16-1 biogenesis at the posttranscriptional level in the nucleus: Evidence for a microRNA hierarchy system. *Cell Res.*, **22**, 504–515.
18. Nishi,K., Nishi,A., Nagasawa,T. and Ui-Tei,K. (2013) Human TNRC6A is an Argonaute-navigator protein for microRNA-mediated gene silencing in the nucleus. *RNA*, **19**, 17–35.
19. Kim,D.H., Saetrom,P., Snove,O. Jr and Rossi,J.J. (2008) MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 16230–16235.
20. Younger,S.T., Pertsemlidis,A. and Corey,D.R. (2009) Predicting potential miRNA target sites within gene promoters. *Bioorg. Med. Chem. Lett.*, **19**, 3791–3794.
21. Schwartz,J.C., Younger,S.T., Nguyen,N.B., Hardy,D.B., Monia,B.P., Corey,D.R. and Janowski,B.A. (2008) Antisense transcripts are targets for activating small RNAs. *Nat. Struct. Mol. Biol.*, **15**, 842–848.
22. Chi,S.W., Zang,J.B., Mele,A. and Darnell,R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.
23. Shomron,N. and Levy,C. (2009) MicroRNA-biogenesis and pre-mRNA splicing crosstalk. *J. Biomed. Biotechnol.*, **2009**, 594678.
24. Mattioli,C., Pianigiani,G. and Pagani,F. (2014) Cross talk between spliceosome and microprocessor defines the fate of pre-mRNA. *Wiley Interdiscipl. Rev. RNA*, **5**, 647–658.
25. Kataoka,N., Fujita,M. and Ohno,M. (2009) Functional association of the microprocessor complex with the spliceosome. *Mol. Cell. Biol.*, **29**, 3243–3254.
26. 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.
27. Allo,M., Buggiano,V., Fededa,J.P., Petrillo,E., Schor,I., de la Mata,M., Agirre,E., Plass,M., Eyras,E., Elela,S.A. *et al.* (2009) Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.*, **16**, 717–724.
28. Papasaikas,P. and Valcarcel,J. (2016) The Spliceosome: The ultimate RNA chaperone and sculptor. *Trends Biochem. Sci.*, **41**, 33–45.
29. Kelemen,O., Convertini,P., Zhang,Z., Wen,Y., Shen,M., Falaleeva,M. and Stamm,S. (2013) Function of alternative splicing. *Gene*, **514**, 1–30.
30. Lee,Y. and Rio,D.C. (2015) Mechanisms and Regulation of Alternative Pre-mRNA Splicing. *Annu. Rev. Biochem.*, **84**, 291–323.
31. Sperling,J., Azubel,M. and Sperling,R. (2008) Structure and function of the Pre-mRNA splicing machine. *Structure*, **16**, 1605–1615.
32. Shefer,K., Sperling,J. and Sperling,R. (2014) The supraspliceosome-a multi-task-machine for regulated pre-mRNA processing in the cell nucleus. *Comput. Struct. Biotechnol. J.*, **11**, 113–122.
33. Sperling,R. (2017) The nuts and bolts of the endogenous spliceosome. *Wiley Interdiscip. Rev. RNA*, **8**, e1377.
34. Azubel,M., Habib,N., Sperling,J. and Sperling,R. (2006) Native spliceosomes assemble with pre-mRNA to form supraspliceosomes. *J. Mol. Biol.*, **356**, 955–966.
35. Cohen-Krausz,S., Sperling,R. and Sperling,J. (2007) Exploring the architecture of the intact supraspliceosome using electron microscopy. *J. Mol. Biol.*, **368**, 319–327.
36. Heinrich,B., Zhang,Z., Raitskin,O., Hiller,M., Benderska,N., Hartmann,A.M., Bracco,L., Elliott,D., Ben-Ari,S., Soreq,H. *et al.* (2009) Heterogeneous nuclear ribonucleoprotein G regulates splice site selection by binding to CC(A/C)-rich regions in Pre-mRNA. *J. Biol. Chem.*, **284**, 14303–14315.
37. Sebbag-Sznajder,N., Raitskin,O., Angenitzki,M., Sato,T.A., Sperling,J. and Sperling,R. (2012) Regulation of alternative splicing within the supraspliceosome. *J. Struct. Biol.*, **177**, 152–159.
38. Raitskin,O., Angenitzki,M., Sperling,J. and Sperling,R. (2002) Large nuclear RNP particles-the nuclear pre-mRNA processing machine. *J. Struct. Biol.*, **140**, 123–130.
39. 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.
40. Zhang,Z., Falaleeva,M., Agranat-Tamur,L., Pages,A.P., Eyras,E., Sperling,J. and Stamm,S. (2013) The 5′ untranslated region of the serotonin receptor 2C pre-mRNA generates miRNAs and is expressed in non-neuronal cells. *Exp. Brain Res.*, **230**, 387–394.
41. Raitskin,O., Cho,D.S., Sperling,J., Nishikura,K. and Sperling,R. (2001) RNA editing activity is associated with splicing factors in lnRNP particles: The nuclear pre-mRNA processing machinery. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 6571–6576.
42. Roth,M.B., Murphy,C. and Gall,J.G. (1990) A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *JCB*, **111**, 2217–2223.
43. Yitzhaki,S., Miriami,E., Sperling,J. and Sperling,R. (1996) Phosphorylated Ser/Arg-rich proteins: Limiting factors in the assembly of 200S large nuclear ribonucleoprotein particles. *Proc. Natl. Acad. Sci. U.S.A.*, **93**, 8830–8835.
44. Gruber,A.R., Lorenz,R., Bernhart,S.H., Neubock,R. and Hofacker,I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.***36**, W70–W74.

45. Spann,P., Feinerman,M., Sperling,J. and Sperling,R. (1989) Isolation and visualization of large compact ribonucleoprotein particles of specific nuclear RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **86**, 466–470.
46. Kotzer-Nevo,H., de Lima Alves,F., Rappsilber,J., Sperling,J. and Sperling,R. (2014) Supraspliceosomes at defined functional states present portray the Pre-Assembled nature of the pre-mRNA processing machine in the cell nucleus. *Int. J. Mol. Sci.*, **15**, 11637–11664.
47. Falaleeva,M., Pages,A., Matsuzek,Z., Hidmi,S., Agranat-Tamir,L., Korotkov,K., Nevo,Y., Eyras,E., Sperling,R. and Stamm,S. (2016) Dual function of C/D box snoRNAs in rRNA modification and alternative pre-mRNA splicing. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E1625–E1634.
48. 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.
49. Friedlander,M.R., Lizano,E., Houben,A.J., Bezdan,D., Banez-Coronel,M., Kudla,G., Mateu-Huertas,E., Kagerbauer,B., Gonzalez,J., Chen,K.C. *et al.* (2014) Evidence for the biogenesis of more than 1,000 novel human microRNAs. *Genome Biol.*, **15**, R57.
50. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
51. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
52. Fabian,M.R. and Sonenberg,N. (2012) The mechanics of miRNA-mediated gene silencing: A look under the hood of miRISC. *Nat. Struct. Mol. Biol.*, **19**, 586–593.
53. 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.
54. Ruby,J.G., Jan,C.H. and Bartel,D.P. (2007) Intronic microRNA precursors that bypass Drosha processing. *Nature*, **448**, 83–86.
55. 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.
56. 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.
57. Vitsios,D.M., Davis,M.P., van Dongen,S. and Enright,A.J. (2017) Large-scale analysis of microRNA expression, epi-transcriptomic features and biogenesis. *Nucleic Acids Res.*, **45**, 1079–1090.
58. Backes,C., Meder,B., Hart,M., Ludwig,N., Leidinger,P., Vogel,B., Galata,V., Roth,P., Menegatti,J., Grasser,F. *et al.* (2016) Prioritizing and selecting likely novel miRNAs from NGS data. *Nucleic Acids Res.*, **44**, e53.
59. Desvignes,T., Batzel,P., Berezikov,E., Eilbeck,K., Eppig,J.T., McAndrews,M.S., Singer,A. and Postlethwait,J.H. (2015) miRNA Nomenclature: a view incorporating genetic origins, biosynthetic pathways, and sequence variants. *Trends Genet.*, **31**, 613–626.
60. Chak,L.L., Mohammed,J., Lai,E.C., Tucker-Kellogg,G. and Okamura,K. (2015) A deeply conserved, noncanonical miRNA hosted by ribosomal DNA. *RNA*, **21**, 375–384.
61. Yoshikawa,M. and Fujii,Y.R. (2016) Human ribosomal RNA-derived resident MicroRNAs as the transmitter of information upon the cytoplasmic cancer stress. *Biomed. Res. Int.*, **2016**, 7562085.
62. Xishan,Z., Ziying,L., Jing,D. and Gang,L. (2015) MicroRNA-320a acts as a tumor suppressor by targeting BCR/ABL oncogene in chronic myeloid leukemia. *Sci. Rep.*, **5**, 12460.
63. Swaminathan,S., Hu,X., Zheng,X., Kriga,Y., Shetty,J., Zhao,Y., Stephens,R., Tran,B., Baseler,M.W., Yang,J. *et al.* (2013) Interleukin-27 treated human macrophages induce the expression of novel microRNAs which may mediate anti-viral properties. *Biochem. Biophys. Res. Commun.*, **434**, 228–234.