

Characterization of novel primary miRNA transcription units in human cells using Bru-seq nascent RNA sequencing

Karan Bedi¹, Michelle T. Paulsen¹, Thomas E. Wilson^{2,*} and Mats Ljungman^{1,3,*}

¹Department of Radiation Oncology, Rogel Cancer Center and Center for RNA Biomedicine, University of Michigan, Ann Arbor, MI 48109, USA, ²Department of Pathology and Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109, USA and ³Department of Environmental Health Sciences, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA

Received June 20, 2019; Revised October 07, 2019; Editorial Decision October 13, 2019; Accepted October 26, 2019

ABSTRACT

MicroRNAs (miRNAs) are key contributors to gene regulatory networks. Because miRNAs are processed from RNA polymerase II transcripts, insight into miRNA regulation requires a comprehensive understanding of the regulation of primary miRNA transcripts. We used Bru-seq nascent RNA sequencing and hidden Markov model segmentation to map primary miRNA transcription units (TUs) across 32 human cell lines, allowing us to describe TUs encompassing 1443 miRNAs from miRBase and 438 from MirGeneDB. We identified TUs for 61 miRNAs with an unknown CAGE TSS signal for MirGeneDB miRNAs. Many primary transcripts containing miRNA sequences failed to generate mature miRNAs, suggesting that miRNA biosynthesis is under both transcriptional and post-transcriptional control. In addition to constitutive and cell-type specific TU expression regulated by differential promoter usage, miRNA synthesis can be regulated by transcription past polyadenylation sites (transcriptional read through) and promoter divergent transcription (PROMPTs). We identified 197 miRNA TUs with novel promoters, 97 with transcriptional read-throughs and 3 miRNA TUs that resemble PROMPTs in at least one cell line. The miRNA TU annotation data resource described here reveals a greater complexity in miRNA regulation than previously known and provides a framework for identifying cell-type specific differences in miRNA transcription in cancer and cell transition states.

INTRODUCTION

MicroRNAs (miRNAs) play critical roles in conferring robustness to cellular processes including timing of cellular development, hematopoiesis, organogenesis, apoptosis, cell proliferation, circadian rhythm and differentiation (1–4). Dysregulation of miRNA expression has been implicated in the onset and progression of many diseases, including cancer (5–9). The primary function of miRNAs is to modulate gene expression by targeting mRNAs for translational repression, deadenylation and degradation (10–12). It has been estimated that half of all protein-coding transcripts are under miRNA regulation (13).

Most miRNA genes are transcribed by RNA polymerase II generating primary transcripts containing 5'-caps and 3' poly(A) tails (14,15). These primary transcripts (pri-miRNAs) are variable in length and rapidly processed in the nucleus by the microprocessor complex consisting of DROSHA and DGCR8 into ~60–80 nucleotide precursors (pre-miRNAs) (16–19). The pre-miRNAs are exported to the cytoplasm where they are further processed into mature miRNAs by DICER (12,20–24). The mature miRNAs are then loaded along with Argonaute proteins (AGOs) into RISC complexes (RNA-induced silencing complexes) that bind primarily to the 3' UTR of mRNA targets (10,25).

The steady-state expression level of miRNAs can be regulated at many steps: initial transcription, processing into mature miRNAs, and turnover of both pri-miRNAs and mature miRNAs (19,26,27). Because pri-miRNAs are rapidly processed into pre-miRNA and subsequently into mature miRNAs, it has been difficult to identify the transcription start and end sites (TSSs and TESs) of pri-miRNA transcription units (TUs) to obtain accurate miRNA gene annotations (26). Such TU annotations of miRNAs are

*To whom correspondence should be addressed. Tel: +1 734 764 3330; Email: ljungman@umich.edu
Correspondence may also be addressed to Thomas E. Wilson. Tel: +1 734 764 2212; Email: wilsonte@umich.edu

critical for understanding the transcriptional regulation of miRNA genes. In lieu of identifying full-length transcripts, annotations of miRNA genes have been performed indirectly by assessing chromatin features suggestive of promoters upstream of the miRNA genes (28) and by analyzing data from Cap analysis gene expression sequencing (CAGE-seq) and RNA-seq (29,30). Other approaches to study pri-miRNAs have been to suppress the activity of DROSHA (31) or capture nascent RNA using *in vitro* GRO-seq or PRO-seq (32).

In this study, we used nascent RNA Bru-seq to map primary miRNA transcripts across 32 diverse human cell lines, which allowed systematic assessment of the various TUs encompassing known miRNAs. The data revealed multiple intergenic miRNA TUs initiating from their own promoters as well as miRNA genes relying on transcriptional read-through from upstream genes or divergent promoter upstream transcription (PROMPTs). About 108 TUs (21.3%) were expressed in all lines, with >68% of those falling within protein-coding genes. About 340 TUs showed variable expression patterns between cell lines indicative of different modes of regulation in different cellular contexts.

MATERIALS AND METHODS

Cell lines and cell culture

A list of cell lines used in this study is provided in Supplementary Table S1. BxPC3, UML49, UM5, GM12878, GM12891 and HCT116 were grown in RPMI (RPMI 1640, 10% FBS, 100 U/ml penicillin and 100 U/ml streptomycin). A2058, HEK293, A375, A673, MiaPaCa, panc1, U2OS, UM16, UM28 and UM59 were grown in DMEM (DMEM, 10% FBS, 100 U/ml penicillin and 100 U/ml streptomycin) and MEM growth media (Minimal Essential Medium, 10% FBS, 1× MEM amino acids, 1× non-essential amino acids, 2 mM L-glutamine, 1× antibiotic-antimycotic, 1× MEM vitamin mixture and 0.15% (w/v) sodium bicarbonate) was used for HEPB3, HepG2, MCF7, U87, UMUC9 and human fibroblasts HF, CSB and XPC. HPDE were grown in keratinocyte serum-free medium (Invitrogen Life Technologies, Inc., Carlsbad, CA) supplemented with 50 µg/ml bovine pituitary extract (Invitrogen) and 5.0 ng/ml recombinant human EGF (Invitrogen); HPNE were grown in three volumes of DMEM and one volume of medium M3, with the mixture supplemented with 5% fetal calf serum and 10 ng/ml EGF; HAP1 and K562 were grown in IMDM with 10% FBS and antibiotics; SHEP1 in 1:1 mix of MEM and F12 media with 10% FBS. iPSCs were reprogrammed from HF (33). Typically, 4 × 10⁶ cells were harvested per sample in 2–3 10cm plates (~80% confluency), yielding ~80–100 µg of total RNA.

Bru-seq nascent RNA sequencing and read mapping

For Bru-seq, cells were incubated in media containing bromouridine (Bru) (Aldrich) at a final concentration of 2 mM for 30 min at 37°C to label nascent RNA. Following labeling, cells were lysed directly in Trizol followed by isolation of total RNA, immunocapturing of Bru-labeled RNA using anti-BrdU antibodies, preparation

of strand-specific DNA libraries with the Illumina TruSeq Kit (Illumina) and deep sequencing using the Illumina sequencing platform, all as previously described (34,35). Sequenced reads were strand-specific, single-ended, and of read lengths from 51–65 bp. Reads were pre-mapped to the ribosomal RNA (rRNA) repeating unit (GenBank U13369.1) and the mitochondrial and EBV genomes (from the hg38 analysis set) using Bowtie2 (2.3.3) (36). Unaligned reads were subsequently mapped to human genome build hg38/GRCh38 using STAR (v 2.5.3a) (37) and a STAR index created from GENCODE annotation version 27 (ftp://ftp.ebi.ac.uk/pub/databases/genencode/Gencode_human/release_27/genencode.v27.basic.annotation.gtf.gz) (38). This strategy allowed mapping of reads throughout the genome, including any spliced reads found in purified nascent RNA. Uniquely mapped reads were counted in 1 kb genomic bins and bin RPKM values calculated as previously described (34,35). For cell lines where multiple replicates were available, the mapped reads were merged for more robust TU calling. Read statistics for each cell line and library are provided in Supplementary Table S1.

Genome segmentation and identification of transcription units

We previously described a Hidden Markov model (HMM) for *ab initio* identification of transcription segments, defined as contiguous spans of the genome with a similar nascent RNA read density such as occurs in a gene. Briefly, the model used 17 logarithmically distributed Bru-seq bin input states constructed individually for each sample based on its 1 kb bin RPKM values. GENCODE genes acted as putative segments to train emission probabilities for 10 logarithmically distributed bin output states of different transcription levels (34,35). Solving the HMM by the Viterbi algorithm assigned an output transcription state to every 1 kb genomic bin. Runs of adjacent bins of the same state were fused into transcription segments.

HMM output segments included both transcribed and non-transcribed regions of the genome. To identify transcribed segments with a read density statistically deviant from the random genomic background in a sample, we first estimated the background read density (reads per bp) using all reads that aligned outside of any annotated GENCODE genes. Our logic was that most intergenic regions are not highly transcribed and reads found there have a higher probability of being artifacts. The average number of such reads expected to occur randomly in any genomic segment was calculated as this intergenic read density multiplied by the segment length. Modeling genomic segments as Poisson processes allowed us to calculate a *P*-value for the number of reads actually observed in a segment using the R expression $1 - \text{ppois}(\text{observed}, \text{expected})$. HMM segments with a *P*-value <0.001 were considered to be transcribed above the genomic background.

We next parsed transcribed segments into non-overlapping transcription units (TUs), defined as genomic regions likely to be traversed by a single transcribing RNA polymerase molecule in a significant fraction of cells. We first fused all adjacent transcribed segments into

preliminary TUs, noting that such spans included segments transcribed at different levels according to the HMM. Segment fusion is necessary and appropriate due to variable efficiency of read recovery in different regions of single genes, and because transcription extends well beyond the 3' ends of annotated genes with a progressive decrease in bin RPKM (a 'peak-valley' configuration with respect to the RPKM of fused segments in the direction of transcription). Single genes can also have multiple TSSs resulting in a 'valley-peak' configuration, in which case we sought to call the longest isoform, i.e. the 5'-most TSS, as the TU. However, unattended segment fusion can lead to errors at closely adjacent genes, especially when they show similar transcription levels, because transcription past the 3' end of the upstream gene can result in a run of bins of similar read density that continues into the downstream gene. Accordingly, we computationally split preliminary TUs at the downstream peak in a 'peak-valley-peak' configuration when the downstream peak either (i) corresponded to a second known gene in the GENCODE annotation or (ii) had an RPKM value >10-fold higher than the preceding valley. Aggregate RPKM values were finally calculated for each HMM segment and each fused TU.

Importantly, the process used to call TUs was largely independent of gene annotations. GENCODE genes were only used to train the initial HMM, to establish a background read density for determining segment transcription probabilities, and to split TUs that had a high likelihood of being inappropriately fused. As a result, called TUs need not begin or end at annotated gene boundaries and can include previously unannotated genes. TU calling is imperfect due to the nature of nascent RNA sequencing data and the fact that TU endpoints were only resolved to 1 kb resolution. Nevertheless, TUs are a powerful tool for revealing the spans of RNA polymerase traversal that produce pri-miRNAs.

miRNA sources and categorization

Genomic locations of mature miRNA sequences were obtained from miRBase database (Release 22.1, $n = 1918$, <ftp://mirbase.org/pub/mirbase/22.1/genomes/hsa.gff3>) (39). A smaller subset of mature miRNAs were also obtained from MirGeneDB2.0 ($n = 557$, <http://mirgenedb.org/static/data/hsa/hsa-all.bed>) (40). These 557 MirGeneDB miRNAs overlap with 507 unique miRBase miRNAs. Both miRBase names and the corresponding MirGeneDB names (when available) are provided in figures and in Supplementary Table S2 in the format 'miRBase/MirGeneDB'. The GENCODE annotation (v27) was used to determine the biotype (<https://www.ensembl.org/info/genome/genebuild/biotypes.html>) of the transcript(s) overlapping each miRNA, which was simplified into three types: protein-coding (PC), long-non-coding (LNC) and other biotypes (OT, for anything besides PC and LNC). The relationship between each miRNA and the Bru-seq TU encompassing it were further categorized, for each cell line individually, into the following classes:

- (i) if the TU completely overlapped an annotated transcript that contained the miRNA, the TU was classi-

fied as 'abs', to signify absolute overlap of a gene, suffixed with the miRNA GENCODE type (PC, LNC, or OT);

- (ii) if the TU overlapped <80% of the annotated transcript, the TU was classified as 'partial', suffixed with the GENCODE type (PC, LNC or OT);
- (iii) if the miRNA was downstream of an annotated transcript and the TU completely overlapped that transcript, the TU was classified as 'abs' suffixed with 'IGds' to signify that the miRNA is intergenic and downstream of an annotated transcript;
- (iv) if the miRNA was upstream of an annotated transcript and the TU completely overlapped that transcript, the TU was classified as 'abs' suffixed with 'IGus' to signify that the miRNA is intergenic and upstream of an annotated transcript;
- (v) if the TU overlapped <80% of the annotated transcript and the miRNA was downstream of an annotated transcript, the TU was classified as 'partial' suffixed with 'IGds';
- (vi) if the TU overlapped <80% of the annotated transcript and the miRNA was upstream of an annotated transcript, the TU was classified as 'partial' suffixed with 'IGus';
- (vii) if the miRNA fell within an annotated transcript but the TU began in and completely overlapped a different upstream transcript (i.e. it is the transcriptional read-through of an upstream transcript that overlaps with the miRNA inside a downstream transcript), the TU was classified as 'part_features';
- (viii) if the miRNA did not overlap with an annotated transcript, the TU overlapping the miRNA was classified as 'novel'; and
- (ix) if the miRNA did not have an overlapping TU, it was considered to be 'not transcribed'.

If the start sites of TUs fell within ± 5 kb of transcription start sites (TSSs) of known transcripts, the TU were classified as having a 'known_promoter', all others were classified as 'novel_promoter'. If the TU overlapped two or more genes, and the miRNA lay within or downstream of the second gene in the direction of transcription, the promoter type could not be ascertained and was classified as 'NA_promoter'. A visual characterization of these various classifications has been provided in Supplementary Figure S1A.

miRNA-seq (small-RNA) data for GM12878, HCT116, K562 and MCF7 cell lines and H3K4Me3 data in seven cell lines were obtained from ENCODE (<https://www.encodeproject.org>). Gene Ontology analysis was performed using the topGO package in R (41). miRNA promoter names and transcription start site coordinates, as determined from CAGE-seq peaks, were obtained from file human.promoters.tsv downloaded from the FANTOM5 site http://fantom.gsc.riken.jp/5/suppl/De_Rie.et.al.2017/vis_viewer_novel/#/human (42–45). The coordinates were converted to hg38 assembly using the UCSC liftOver tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>).

RESULTS

Identification of inter- and intragenic miRNA TUs across 32 cell lines

We applied HMM-based *ab initio* segmentation to 65 high quality Bru-seq nascent RNA sequencing data sets from 32 human cell lines, including 22 cancer cell lines (A2058, A375, A673, BxPC3, HAP1, HCT116, HEPB3, HEPG2, K562, MCF7, miaPaCa, panc1, SHEP1, T47D, U2OS, U87, UM16, UM28, UM5, UM59, UML49 and UMUC9) and 10 non-cancer cell lines (HEK293, csb, GM12878, GM12891, HME, HPDE, HPNE, iPSC, HF1 and xpc) (Supplementary Table S1). In Bru-seq, cells are incubated with bromouridine for 30 min to label any RNA synthesized during that time. The nascent, Bru-labeled RNA is then immunocaptured, converted to cDNA and sequenced. Sequencing reads were mapped to the human reference genome (hg38) and contiguous transcription spans (i.e. TUs) were identified genome-wide using HMM segmentation. The process identified miRNA-associated TUs with statistically significant transcription for 1443 of the 1918 miRNAs annotated in miRBase in at least one cell line. (Supplementary Table S2). TU endpoints, lengths, transcriptional intensities in RPKM units, class, promoter types and the presence or absence of antisense transcription and PROMPTs were determined, per cell line, for all TUs that included these 1443 miRNAs (Figure 1A, red bars). However, most subsequent analyses and calculations were performed using the smaller subset of 438 TUs associated with 507 miRNAs obtained from the MirGeneDB database, which have been validated against a stricter set of criteria (40).

MirGeneDB miRNA transcription among cell lines displayed a bimodal nature, with 138 (27.2%) having a called TU in at least 30 of the 32 cell lines, and 151 (29.78%) having a called TU in two cell lines or fewer (Figure 1A, blue bars). There were 108 (21.3%) annotated miRNAs transcribed in all 32 cell lines, while 69 (13.6%) were never transcribed in any cell line. Of the 69 miRNAs, roughly half (29/69, 42%) showed transcription in the antisense direction in at least some cell lines (Figure 1A (left pie chart)). This percentage shrank to 29.6% (32/108) for miRNAs that were transcribed in all 32 cell lines, suggesting transcription on the opposite strand as a possible means of suppressing TU expression (Figure 1A (right pie chart)). Transcription in the antisense direction could result from either (i) presence of a gene (e.g. Figure 3B, ACADVL gene antisense to *MIR324*) or (ii) capturing bin(s) upstream of a downstream gene (for example, *MIR3613* as seen in our web resource <https://bruseq.org/backdoor/projects/miRNA/miRNA.html>) or (iii) transcriptional readthrough from an upstream gene (for example, *MIR616* and *MIR93/MIR25/MIR106B*).

The bimodal nature of miRNA transcription suggests potentially two classes of miRNAs, one serving housekeeping functions, the others with functions that are more specialized, as has been previously suggested (46). Potential examples of the latter class include 39 miRNAs transcribed in only one cell line with an unambiguous median TU RPKM of at least 0.3 (Supplementary Table S2 and Figure 1B). It is

possible that miRNAs that were rarely or never transcribed may represent false miRNA annotations or miRNAs expressed in cell types other than those sampled here. In this regard, we note that miRBase miRNAs were more likely to never be transcribed than MirGeneDB miRNAs (Figure 1A, blue versus red bar at $x = 0$). The TU lengths of miRNAs that were always transcribed ($n = 108$) were also similar across cell lines, with a median length around 100 kb (Figure 1C).

A majority (74/108, 68.5%) of the TUs overlapping universally transcribed miRNAs corresponded well to protein-coding transcripts (category *abs_PC*, Figure 1D). Genome-wide analysis also showed these miRNAs to be enriched in annotated genes (protein-coding or long-non coding, P -value $< 2.2e-16$). Gene ontology (GO) analysis of the genes harboring universally transcribed miRNAs revealed DNA binding and mRNA processing as highly enriched functions, as compared to a background set of 907 genes that overlap with an miRNA regardless of their transcription state (Supplementary Figure S1B–D). Nevertheless, the types, lengths and transcriptional levels (RPKM) of most TUs were similar across all cell lines, suggesting that at least the most commonly expressed miRNAs tend to share common transcriptional control mechanisms. (Supplementary Figure S1E–G). We did not observe a difference in miRNA transcription between cancer and non-cancer groups of cell lines (Supplementary Figure S1H). However, we found 30 and 24 miRNAs to be uniquely transcribed in cancer and non-cancer groups, respectively (Supplementary Figure S1I). Nineteen out of twenty-four miRNAs in the non-cancer group were exclusively transcribed in iPSCs (Supplementary Figure S1J).

TUs can be extremely long with multiple miRNAs

miRNA TUs can be very long. The longest MirGeneDB miRNA TU measured ~ 1.55 Mb, encompassing *MIR582* in the GM12878 cell line, while being less than half that length (765 kb) in the UML49 cell line (Supplementary Table S2). TUs for two intergenic miRNAs were 128 kb (*MIR138-1/MIR138-P1*, Figure 2A) and ~ 1.28 Mb (*MIR873/MIR873-v1/MIR873-v2* and *MIR876*, Figure 2B, Supplementary Figure S2A) in HPDE and HME cell lines, respectively. One large cluster of miRNAs located on chromosome 14 (Figure 2C) was transcribed as a single TU over ~ 250 kb in length, suggesting that all of these miRNAs may be processed from a single primary transcript. Indeed, BruUV-seq, whose signal accumulates immediately downstream of active promoters (47), verified the presence of only one major TSS at this locus (Figure 2C; bottom). Included in this TU are also multiple snoRNAs and a known, annotated lncRNA MEG3.

A critical importance of pri-miRNA TU length is that it, together with miRNA placement within a TU, determines the timing of miRNA expression relative to transcriptional initiation. Timing (tabulated in Supplementary Table S2) can be predicted based on the assumption that miRNA processing occurs co-transcriptionally (48) and that RNA polymerases move through DNA at an approximate average rate of 2–3 kb/min (49–53). For example,

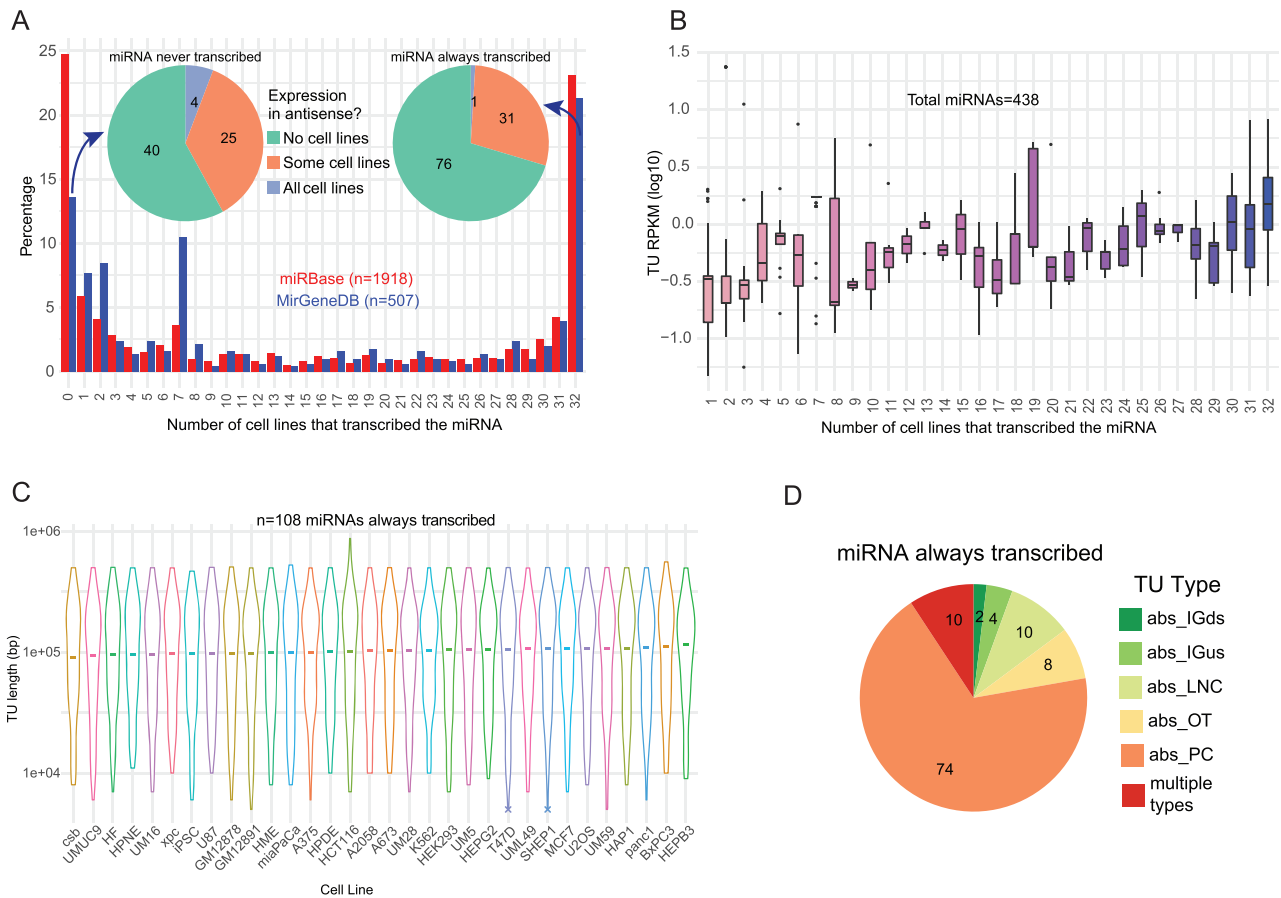


Figure 1. Identification and properties of miRNA transcription units (TUs). (A) Histogram of miRNA transcription frequency among 32 cell lines for the miRbase set ($n = 1918$, red bars) and the MirGeneDB subset ($n = 507$, blue bars). For miRNAs that were not transcribed in any cell line ($n = 69$, left pie chart), or those transcribed in all cell lines ($n = 108$, right pie chart), pie charts illustrate whether transcription through the miRNA was found in the antisense direction in all, some or no cell lines. (B) Boxplots showing Bru-seq transcription levels (RPKM) of TUs of all MirGeneDB miRNAs transcribed in 32 cell lines ($n = 438$). (C) Distribution of TU lengths (bp) for 108 miRNAs transcribed in all cell lines. (D) For miRNAs transcribed in all cell lines ($n = 108$), a pie chart shows the types of TUs observed ('abs_PC' indicates protein coding genes; 'multiple types' means different TU types were seen in different cell lines; see 'Materials and Methods' section for other definitions).

although *MIR873/MIR873-v1/MIR873-v2* and *MIR876is* encompassed by a 1.28 Mb TU in HME (Figure 2B, and an even longer TU in case of U2OS cell line (1.49 Mb, Supplementary Figure S2A)), it is located 351 kb downstream of its promoter so is expected to have an ~176-min delay from initiation to processing. Moreover, the processing of various RNA elements in the TU in Figure 2C is expected to occur in a temporal order according to their location in the TU.

miRNA transcription due to differential promoter usage

There are many reasons why miRNAs may show differential transcription between cell lines. One obvious explanation is cell type-specific activation of pri-miRNA promoters. Some miRNA TUs are initiated from multiple promoters allowing different cells to express the same miRNA-containing gene in response to distinct regulatory signals. We identified novel promoters for 197 miRNAs in the MirGeneDB data set based on a comparison with the underlying GENCODE annotation. The number of miRNAs with a novel

promoter reduced to 180 when we excluded those where a primary miRNA CAGE promoter peak from the FANTOM5 data set was found within ± 1 kb of our TU start site (42–45).

de Rie *et al.* (44) annotated TSSs of pri-miRNAs that were identified from CAGE-seq data (one per miRNA). To compare our Bru-seq TUs to these CAGE TSSs, we calculated the distance from each genomic miRNA to either the Bru-seq TU start site or the CAGE pri-miRNA TSS coordinate, and then calculated the difference between these two distances as a measure of TSS correspondence. Plotting the minimum of this difference value across all cell lines for each miRNA demonstrated that the Bru-seq TU start site was in close proximity (within ± 2 kb) to the CAGE TSS signal in at least one of our cell lines for most miRNAs (sharp peak around 0 in Supplementary Figure S2B). Thus, there is an excellent correspondence of the CAGE and Bru-seq data sets. However, plotting the median of the difference value demonstrated that our typical TU start site over multiple cell lines was frequently farther away from the miRNA than the CAGE TSS (Supplementary Figure S2C). Importantly,

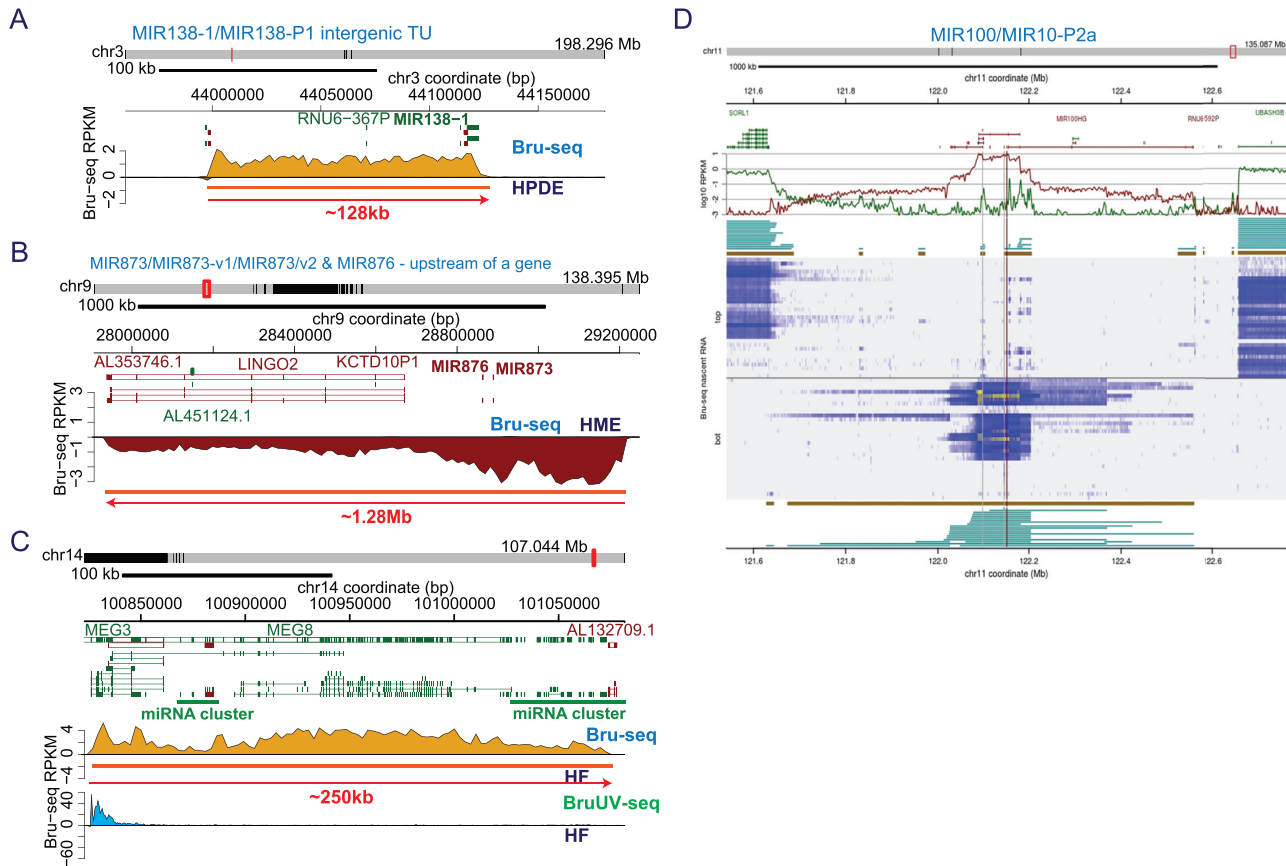


Figure 2. Differential miRNA TU lengths and promoter usage. (A) Genomic view of the ~128 kb long TU for the intergenic miRNA *MIR138-1/MIR138-P1* in HPDE cells. The orange bar below the Bru-seq trace represents the span of the called TU by the segmentation algorithm. (B) Similar view of the ~1.28 Mb long TU for the intergenic miRNAs *MIR876* and *MIR873/MIR873-v1/MIR873-v2* in HME cells. (C) The ~250 kb long TU for the miRNA cluster on chr14 in human fibroblasts, with the lower panel showing a single TSS picked up by the BruUV-seq technique. (D) Depiction of *MIR100/MIR10-P2a* transcription over all 32 cell lines. Genes represent GENCODE transcript spans. Line plots show the aggregate transcription level over all 32 lines on the top (green) and bottom (red) strands. Heat maps show one row for each cell line over all genome bins on the top (top panel) and bottom (bottom panel) strands, where darker blue indicates a higher level of transcription. Finally, cyan lines show all unique TU spans called in any cell line, and thick brown lines show the span of TU overlap groups. Images of this type can be viewed for any miRNA at <https://bruseq.org/backdoor/projects/miRNA/miRNA.html>.

Bru-seq TUs represent the longest span of contiguous transcription at a locus based on nascent RNA data, in contrast to de Rie *et al.* who used the highest expressed promoter among the set of all candidate pri-miRNA promoters (identified based on all transcripts annotated in GENCODE v19 (human) or NCBI Entrez Gene database, followed by manual curation) as the pri-miRNA promoter (44). As a result, Bru-seq data reported here expand the known diversity in the set of sometimes multiple TSSs used by individual miRNAs, by exposing alternative promoters that are farther away from the miRNA. We also identified TUs for 61 MirGeneDB miRNAs that lacked a CAGE TSS (Supplementary Table S2). For example, the intergenic miRNA *MIR219A-1/MIR219-P1* was transcribed in all 32 cell lines.

In an extreme case of multiple promoter usage, Bru-seq identified at least six distinct TU start sites for *MIR100HG/MIR10-P2a* HG (host gene) that could give rise to the mature *MIR100/MIR10-P2a* (Figure 2D, which is presented in the heat map format used by our online resource, and Supplementary Figure S2D). Using matching BruUV seq data from three cell lines, we were able to iden-

tify many more TSSs that were corroborated by H3K4Me3 signal in seven cell lines from ENCODE (Supplementary Figure S2D, bottom). Finally, a distinct group of miRNAs is regulated autonomously from their own promoters. For example, in four pancreatic cancer cell lines (BxPC3, panc1, UM5 and UM59), miRNAs *MIR182/MIR96-P2*, *MIR96/MIR96-P1* and *MIR183/MIR96-P3-v1/MIR96-P3-v2* are transcribed as part of a single TU on the negative strand with a common, novel promoter (Supplementary Figure S2E).

miRNA transcription due to 3' transcriptional read-through

Transcription termination of RNA Pol II entails synthesis past the annotated transcription end site (TES) of genes with subsequent cleavage of nascent transcripts and polyadenylation (54–56). The degree of this transcriptional read-through can vary between genes and cell lines, leading to miRNA expression diversity since 59 intergenic miRNAs are annotated within 50 kb downstream of protein-coding genes. For example, GM12878 and MCF7 cells appear to generate *MIR296* as part of read-through transcrip-

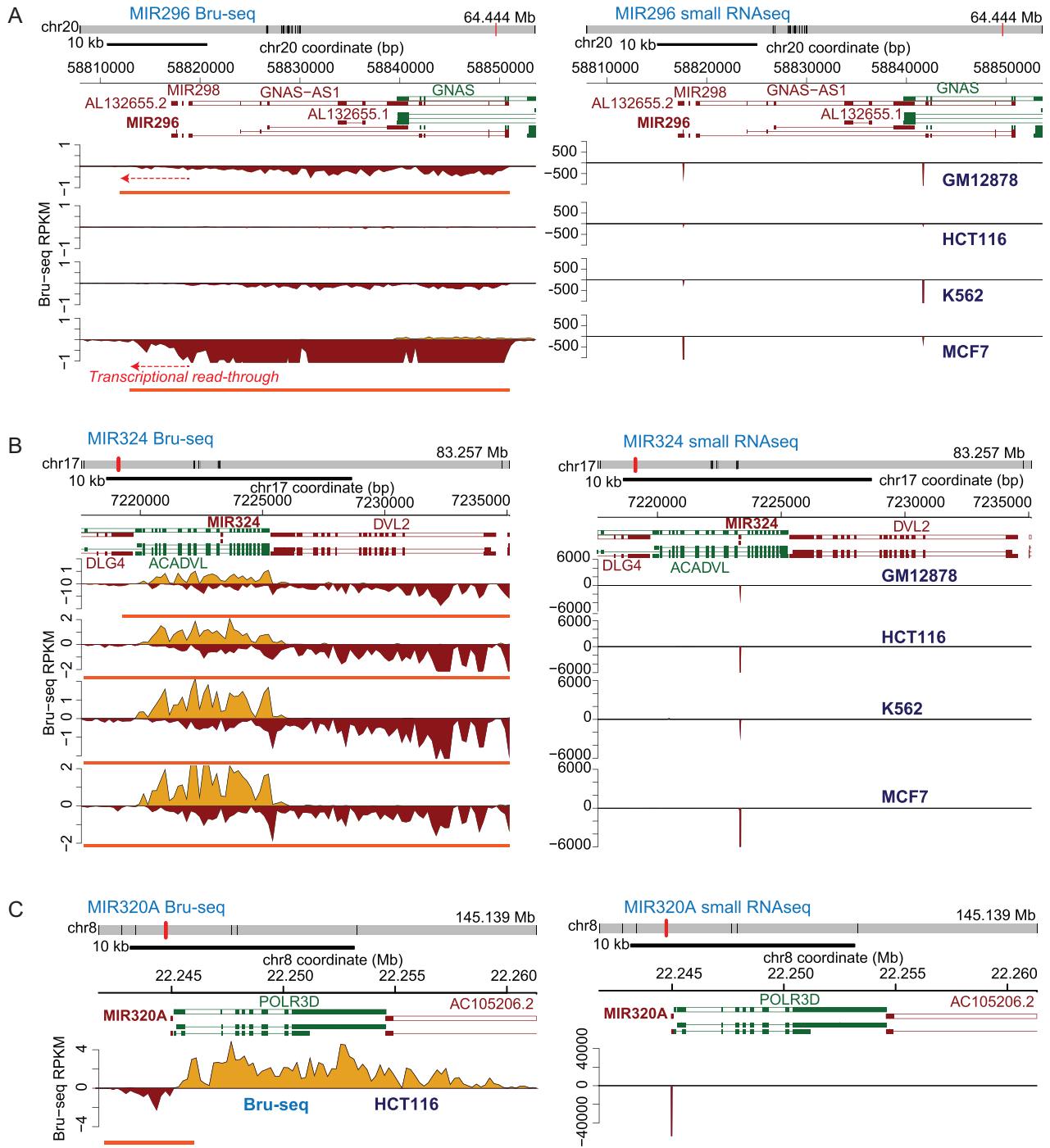


Figure 3. miRNA regulation by transcription read-through and PROMPTs. (A) Bru-seq transcription in the GNAS-AS1 gene (left panel) with transcriptional read-through crossing *MIR296* in GM12878 and MCF7 cell lines but not HCT116 and K562 cell lines. Small RNA-seq data (right panel) shows correlated expression of mature *MIR296* in GM12878 and MCF7. (B) Similar depiction of transcriptional read-through from DVL2 leading to expression of *MIR324* across multiple cell lines. (C) A PROMPT from gene POLR3D results in transcription of *MIR320A* (top panel), as verified by the small RNA-seq data (lower panel).

tion from the upstream gene *GNAS-AS1* (Figure 3A, left panel). To determine whether read-through transcription results in production of mature *MIR296* miRNA, we obtained ENCODE small RNA-seq data from these cell lines (Supplementary Table S1). Mature *MIR296* was detected in only GM12878 and MCF7 cells, matching the predictions from the called read-through TUs (Figure 3A, right panel). An additional example of apparent transcription read-through resulting in mature miRNAs is *MIR331* (Supplementary Figure S3A). Even when annotated as intergenic, some miRNAs are transcriptionally linked to upstream genes via read-through transcription and thus, its location may have been evolutionary selected for as it may share a functional relationship with the upstream gene.

Read-through transcription further exposes miRNAs to regulatory complexity due to TU overlaps. For example, *MIR324* is transcribed by read-through transcription of the gene *DVL2* (Figure 3B, left panel) and is also expressed in GM12878, K562, HCT116 and MCF7 cells (Figure 3B, right panel). However, the *ACADVL* gene on the strand opposite to *MIR324* is also transcribed in all four cell lines. Presumably, transcription of *MIR324*, via *DVL2* read-through, and *ACADVL* is mutually exclusive at any given time in a single cell due to potential conflicts between opposing RNA polymerases. In this way, miRNA read-through expression could be secondarily regulated by transcription of nearby genes in a cell type-specific manner.

Differential miRNA transcription due to upstream promoter divergent transcription

The promoters of many protein-coding genes produce divergent promoter upstream transcripts, or PROMPTs, proceeding away from the gene promoter on the opposite strand (57–59). PROMPTs are typically unstable and their functional roles remain unclear. We found 14 TU's spread across 30 cell lines that resembled PROMPTs for miRNAs in the miRBase database. For example, HCT116 cells transcribe *MIR320A* as part of a 4 kb long PROMPT that proceeds divergently away from the *POLR3D* gene, resulting in the expression of *MIR320A* (Figure 3C). *MIR320A* was present in the miRBase but not the MirGeneDB data sets, presumably due to it being classified as a non-canonical miRNA (60).

Taken together, regulation of miRNA transcription is remarkably diverse and through evolutionary selection, these miRNAs have been placed under transcriptional control of nearby genes with which they presumably share functional relationships. Many miRNAs are found within introns of protein-coding genes and transcribed as those genes are transcribed. Other miRNAs reside outside the span of protein-coding genes but are nevertheless regulated through transcription from these genes by 3' transcriptional read-through or through transcription of PROMPTs. Furthermore, the positioning of miRNAs on the opposite strand of protein-coding genes could create transcriptional interference.

Only a subset of TUs generate mature miRNAs

Primary transcription of miRNAs is necessary but not sufficient for the expression of mature miRNAs. Further

comparisons of primary TUs determined by Bru-seq and mature miRNAs determined by small RNA-seq from GM12878, HCT116, K562 and MCF7 cell lines demonstrated a notable discordance. For example, all cell lines transcribed the TP53-regulated *MIR34A* host gene (61) (Figure 4A, left panel), except K562 that has a known p53 mutation. However, only HCT116 and MCF7 cells expressed mature *MIR34A/MIR34-P1* despite strong primary transcription in GM12878 (Figure 4A, right panel). This suggests that *MIR34A/MIR34-P1* is either processed much less efficiently or turned over much more quickly in GM12878 cells. Six miRNAs (*MIR17/MIR17-P1a*, *MIR18A/MIR17-P2a*, *MIR19A/MIR19-P1*, *MIR20A/MIR17-P3a*, *MIR19B1/MIR19-P2a* and *MIR92A1/MIR92-P1a*) were transcribed together as one TU in GM12878 cells, but only *MIR17/MIR17-P1a*, *MIR18A/MIR17-P2a* and *MIR20A/MIR17-P3a* were detected as mature miRNAs (Figure 4B). Many other miRNA clusters also showed differential processing, such as *MIR30B/MIR30-P2c* and *MIR30D/MIR30-P1c* (Supplementary Figure S4A). Thus, cells can regulate the final expression of mature miRNAs with respect to the specific identity of the individual miRNAs, even when they are transcribed together. We also found rare examples where a mature miRNA was detected in small RNA-seq data even though a TU was not called by our algorithm, but in most cases low levels of transcription could be observed upon visual inspection (e.g. *MIR203A/MIR203-v1/MIR203-v2* in HCT116 cells, Supplementary Figure S4B).

To further explore the contribution of post-transcriptional regulation to miRNA biogenesis, we compared the transcription levels of TUs (RPKM) against the expression of mature miRNA (counts) using the MirGeneDB data set (Figure 4C and Supplementary Figure S4C). Primary and mature miRNA expression consistently correlated poorly across cell lines (the correlations (Spearman) were similar across the four cell lines), again suggesting that post-transcriptional processing and turnover play important roles in regulating steady-state levels of miRNAs. The correlations worsened across all four cell lines when all the miRNAs from the miRBase data set were taken into consideration (Supplementary Figure S4D). We further found that the read counts of mature miRNA forms were highly correlated across cell lines for miRNAs that were transcribed in all four lines based on Bru-seq (Figure 4D). Thus, the post-transcriptional regulation of miRNAs appears to be more linked to the miRNA itself rather than being cell line specific.

TUs and cellular reprogramming

We finally compared TU profiles of two specific cell lines in our study, human fibroblasts and the same cell line converted into induced pluripotent stem cells (iPSCs) (33). Close to 300 miRNAs, TUs were transcribed in each of the two cell lines (292 and 278 in iPSCs and fibroblasts, respectively, Figure 5A), with around one-third (34.2% and 30.9% for iPSCs and fibroblasts respectively) transcribed uniquely in the two cell states indicating cell-state specificity. iPSCs showed specific enrichment of TU types encompassing intergenic miRNAs. miRNAs upstream of annotated genes

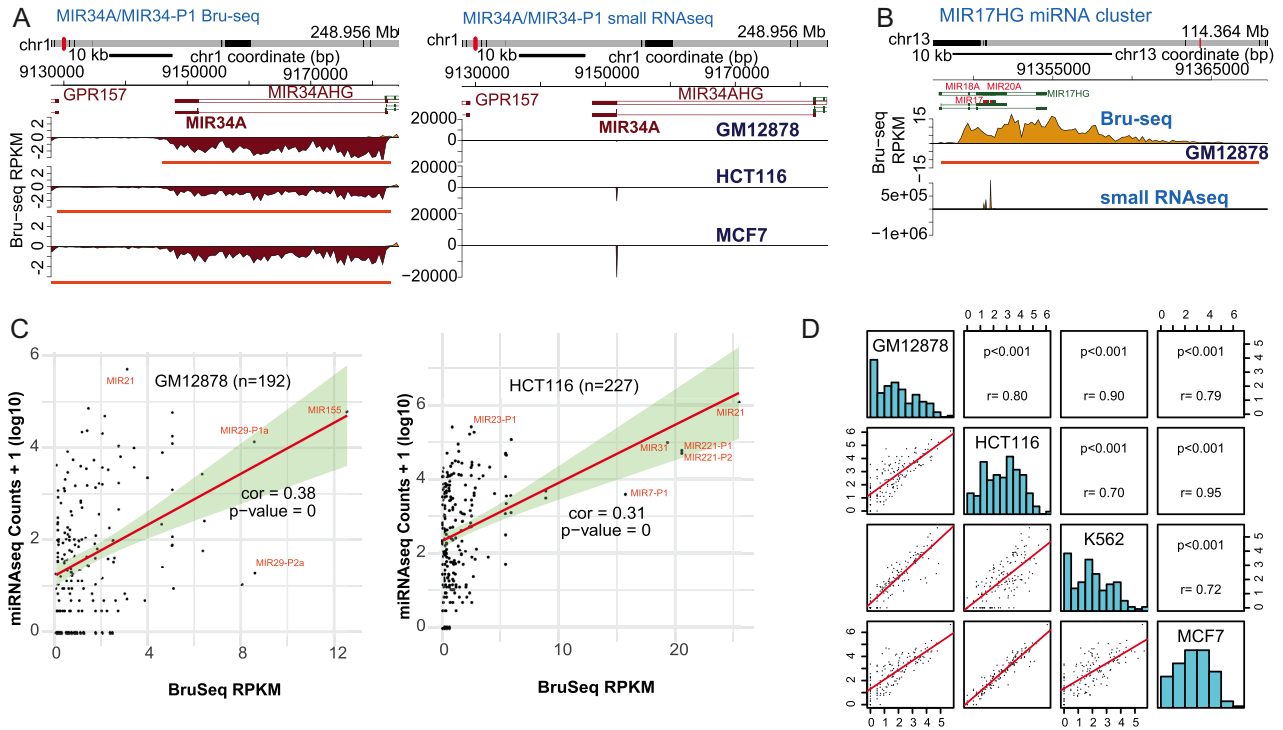


Figure 4. Only a subset of TUs generate mature miRNAs. (A) Bru-seq (left panel) reveals transcription of the MIR34A/MIR34-P1 host gene in three cell lines, but small RNA-seq data (right panel) reveal expression of mature MIR34A/MIR34-P1 in only two cell lines. (B) Bru-seq (upper panel) reveals transcription through the entire MIR17 host gene, yet only MIR17/MIR17-P1a, MIR18A/MIR17-P2a and MIR20A/MIR17-P3a are processed into mature form (indicated in red) in GM12878 (lower panel). (C) Bru-seq transcription levels (RPKM) of TUs are plotted against expression levels (counts) of mature miRNAs for GM12878 (left) and HCT116 (right) cell lines, with some included miRNAs labeled. Spearman correlation coefficients were calculated. (D) Correlation plots of mature miRNA expression for miRNA transcribed in each of GM12878, HCT116, K562 and MCF7 cell lines.

(abs_IGus, $n = 47$), novel ($n = 6$), miRNAs downstream of annotated genes (abs_IGds, $n = 2$), lncRNAs (abs_LNC, $n = 5$) and others ($n = 15$) accounted for roughly three-quarters of the uniquely transcribed TUs. Several miRNAs have been identified to be uniquely expressed in pluripotent cells both *in vivo* and *in vitro*, such as the MIR302/367 (or MIR430/MIR92-P2a according to MirGeneDB nomenclature) cluster (62) (Figure 5B). Fibroblasts, also saw an enrichment of distinct miRNAs in the lncRNA and intergenic categories, with a significant contribution from the large miRNA cluster on chromosome 14 discussed above (Figure 2C).

DISCUSSION

MicroRNAs have many important roles in regulating gene expression during normal cellular homeostasis, development and in response to cellular stresses. Dysregulation of miRNAs is implicated in pathological conditions such as cancer, so there is a need to obtain new knowledge about the regulation of miRNA biosynthesis to better understand miRNA biology and combat diseases. Due to the poor annotation of intergenic miRNAs, it has been difficult to elucidate what transcription factors are responsible for regulating their initial synthesis. Furthermore, assessment of the relative contribution of transcription and post-transcriptional regulation to miRNA expression has

not been directly explored. In this study, we identified and mapped the nascent TUs that generate miRNAs in human cells. Since pri-miRNA transcripts are rapidly processed into shorter intermediates, traditional RNA-seq techniques based on steady-state RNA are not suitable to map the full-length miRNA transcripts. Therefore, we used Bru-seq nascent RNA sequencing and a segmentation algorithm to map miRNA TUs across 32 cell lines.

Considering that the mature form of miRNAs is only ~23 nt long, it is striking that some miRNA TUs are extremely long, up to 100,000 times larger than the miRNA itself. Assuming a transcription elongation rate of 2–3 kb/min (50–53), it would take up to 5 hours to transcribe past MIR876 in U2OS cells (Supplementary Figure S2A, Supplementary Table S2). This transcriptional exercise might appear to be wasting both time and energy. However, transcriptional length can act as a biological clock spacing initiation of the gene with the completion of its product for optimal biological effectiveness (63). Thus, to delay the inhibitory effect of miRNAs on their mRNA targets may be crucial for a particular biological pathway and therefore these miRNA genes may have evolved to have very long TUs.

A large number of miRNAs are intragenic, located within protein-coding and non-coding genes. Their transcription and function may therefore be linked to their host genes. As transcription passes through an intron containing a particular miRNA, Drosha will begin processing the pri-miRNA

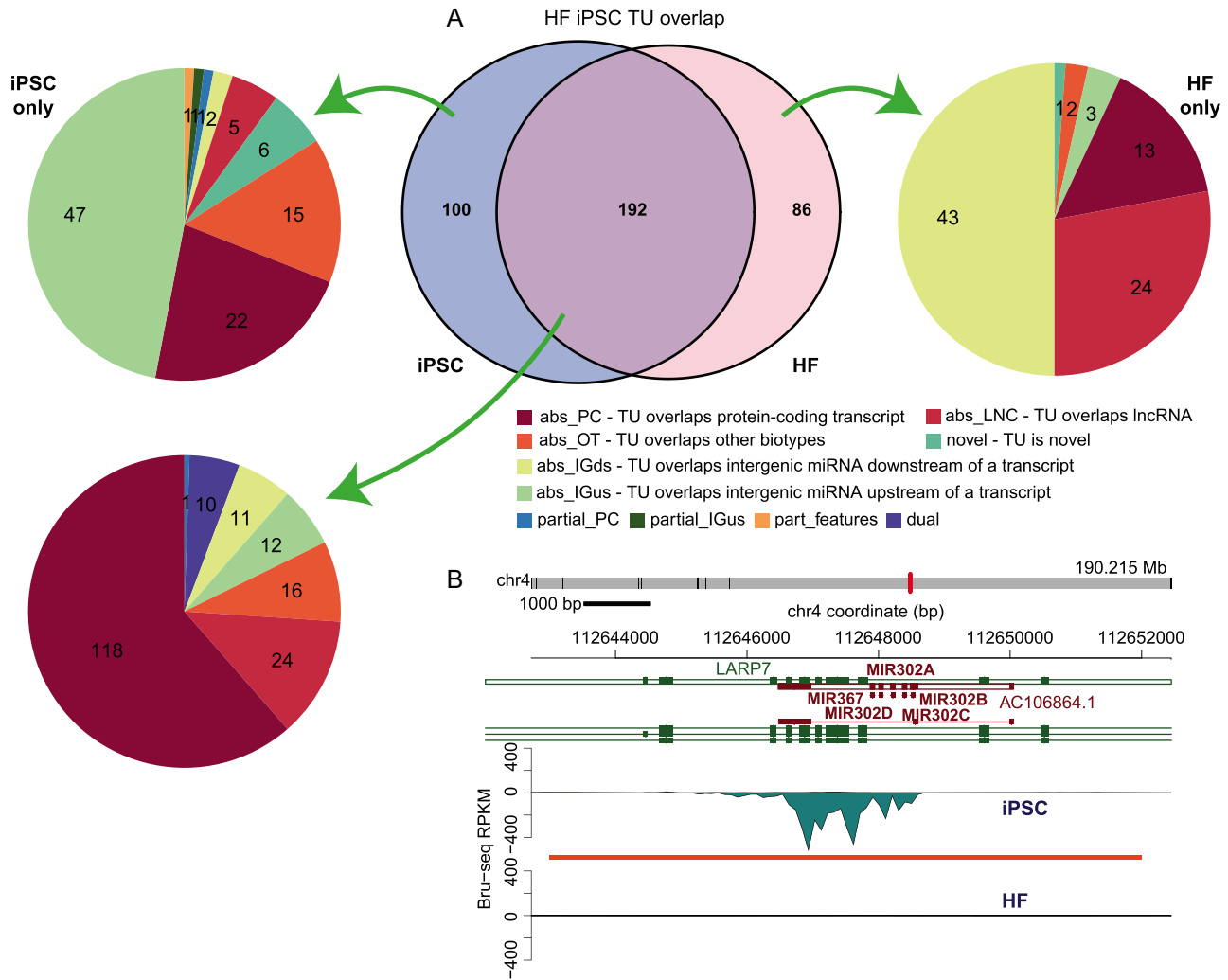


Figure 5. Altered miRNA TUs in response to cellular reprogramming. (A) Set analysis showing the overlap of TUs between human fibroblast and iPSC cell lines, plotted as a Venn diagram. Three pie charts show the TU type distributions for iPSC only TUs (100, top left), human fibroblasts only TUs (86, top right) and TUs common to both (192, bottom left). (B) TU span for the *MIR302/367* (or *MIR430/MIR92-P2a* according to MirGeneDB nomenclature) cluster in iPSCs and human fibroblasts.

(48). However, there are examples of intragenic miRNA genes that have their own promoters that are regulated independently of the host gene's promoters, which might be difficult to detect in nascent RNA samples (see below) or that could be masked if the host genes were also transcribed at a high level (29,64). For miRNAs not located within genes, we found three different transcription mechanisms. First, the miRNA gene may be transcribed following activation of its own autonomous promoter. Second, it is transcribed as a result of transcriptional read-through of an upstream gene. Third, it is transcribed as part of run-on transcription from a PROMPT (Figure 3C). It is possible that the locations of miRNAs in the genome may have been sculpted in specific ways during evolution so that miRNAs that have a functional relationship with the genes leading to their transcription can be co-regulated even when the miRNA does not reside within the gene body. Differential promoter usage can uncouple genic miRNAs from transcription of the parent gene and expose a miRNA to multiple modes of ex-

pression regulation that are integrated across all of its promoters, depending on the exact placement of the miRNA.

A previous study using metabolic labeling of nascent RNA concluded that steady-state levels of miRNAs are primarily controlled by the synthesis of the miRNAs (65). In our study, the correlations between synthesis of the primary miRNA and the expression of the mature miRNAs were surprisingly low across four cell lines for both intergenic and intragenic miRNAs (Figure 4; Supplementary Figure S4C and D). The reason for this discrepancy between the two studies is not clear, but we show clear examples where pri-miRNA synthesis occurred without expression of the mature miRNA, indicating that differences in post-transcriptional regulation play a significant role in miRNA regulation (Figure 4A). It is possible that the recruitment of the microprocessor complex to sites of transcription could be gene- and cell type-specific resulting in selective miRNA processing as pri-miRNAs are synthesized. Additionally, degradation may differ greatly across different miRNAs

both by direct modifications of the miRNAs as well as through mRNA target-directed miRNA degradation (65–67). We also found rare examples where a called pri-miRNA TU was paradoxically not necessary for mature miRNA expression (Supplementary Figure S4B), which could result from insufficient read-depth of the Bru-seq experiments. We might also miss a true TU if it was small and closely flanked the genomic miRNA, making the pri-miRNA very small, short-lived and difficult to sequence and/or computationally detect at our 1 kb bin resolution.

It is interesting that although we broadly sampled from both normal and cancer cell lines, most expressed miRNAs had similar TUs across all cell lines (Supplementary Figure S1G). Furthermore, the number of active TUs was not significantly different between the cancer and non-cancer cell line groups (Supplementary Figure S1H). Nevertheless, a limitation of this work is that all pri-miRNA TUs were established using cell lines cultured *in vitro*, representing a limited number of cell types and tissue sources, which limit the potential biological diversity of the TUs we characterized. Among the 69 miRNAs that were not expressed in any of our cell lines, we provide continued evidence that some of these may not be true miRNAs (Figure 4C; Supplementary Figure S4C and D), but others are likely to require further characterization in a wider spectrum of biological samples.

CONCLUSIONS

By using nascent RNA mapping in 32 cell lines, we have provided a data resource and public visualization tool (see ‘Materials and Methods’ section) that describes the transcription profile of 438 MirGeneDB miRNAs, including 61 with previously unknown CAGE TSS signal. We believe this approach will be very valuable for future assessments of transcriptional regulation of miRNAs following cellular exposure to stimuli or stresses, during cell state transitions and in normal and disease states.

DATA AVAILABILITY

Tables of TU endpoints and categories grouped by miRNA and cell line are provided in Supplementary Table S2. Bru-seq read data are available from the Gene Expression Omnibus (GEO) via accession number, GSE132392. Some Bru-seq data sets have been previously reported (GSE75398 for HF1 BruUV-seq data, GSE115310 for iPSC Bru-seq data). We have finally created a public web resource available at <https://bruseq.org/backdoor/projects/miRNA/miRNA.html> for visualization of Bru-seq and BruUV-seq signal surrounding all miRNA genes for all cell lines characterized in this study, with heat map views and spans of called TUs relative to GENCODE gene annotations and previously established CAGE TSSs.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGEMENTS

We thank the following collaborators for providing us with the cell lines (and Bru-seq data generated using these

cell lines) used in this study: Diane Simeone (HEP293, BxPC3, HPDE, HPNE, UM5, UM16, UM28 and UM59), Elizabeth Lawlor (A673), Nouri Neamati (HAP1, mia-PaCa), Theodore H Welling III (HEPB3, HEPG2 and U87), Mark Day (HME and UMUC9), Sami Barmada (iPSC), Mircea Ivan (MCF7), Erika Newman (SHEP1), Gary Luker (T47D) and Yosef Shiloh (U2OS). We also thank personnel at University of Michigan Sequencing Core for professional technical assistance and Fan Meng and Manhong Dai for administration and maintenance of the University of Michigan Molecular and Behavioral Neuroscience Institute (MBNI) computing cluster. We also acknowledge former and current members of the Ljungman lab for their input and contributions to this work.

FUNDING

National Human Genome Research Institute [1R01HG006786, 1UM1HG009382].

Conflict of interest statement. None declared.

REFERENCES

- Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
- Ebert,M.S. and Sharp,P.A. (2012) Roles for microRNAs in conferring robustness to biological processes. *Cell*, **149**, 515–524.
- Wang,X., Tian,G., Li,Z. and Zheng,L. (2015) The crosstalk between miRNA and mammalian circadian clock. *Curr. Med. Chem.*, **22**, 1582–1588.
- Bartel,D.P. (2018) Metazoan MicroRNAs. *Cell*, **173**, 20–51.
- Lujambio,A. and Lowe,S.W. (2012) The microcosmos of cancer. *Nature*, **482**, 347–355.
- Rosenfeld,N., Aharonov,R., Meiri,E., Rosenwald,S., Spector,Y., Zepeniuk,M., Benjamin,H., Shabes,N., Tabak,S., Levy,A. *et al.* (2008) MicroRNAs accurately identify cancer tissue origin. *Nat. Biotechnol.*, **26**, 462–469.
- Takahashi,R.U., Miyazaki,H. and Ochiya,T. (2014) The role of microRNAs in the regulation of cancer stem cells. *Front. Genet.*, **4**, 295.
- Takahashi,R.U., Miyazaki,H. and Ochiya,T. (2015) The roles of MicroRNAs in breast cancer. *Cancers*, **7**, 598–616.
- Taft,R.J., Pang,K.C., Mercer,T.R., Dinger,M. and Mattick,J.S. (2010) Non-coding RNAs: regulators of disease. *J. Pathol.*, **220**, 126–139.
- Ha,M. and Kim,V.N. (2014) Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.*, **15**, 509–524.
- Wilczynska,A. and Bushell,M. (2015) The complexity of miRNA-mediated repression. *Cell Death Differ.*, **22**, 22–33.
- Chen,C.Y. and Shyu,A.B. (2011) Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA*, **2**, 167–183.
- Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
- Lee,Y., Kim,M., Han,J., Yeom,K.H., Lee,S., Baek,S.H. and Kim,V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.*, **23**, 4051–4060.
- Cai,X., Hagedorn,C.H. and Cullen,B.R. (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, **10**, 1957–1966.
- Denli,A.M., Tops,B.B., Plasterk,R.H., Ketting,R.F. and Hannon,G.J. (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature*, **432**, 231–235.
- 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.
- Han,J., Lee,Y., Yeom,K.H., Kim,Y.K., Jin,H. and Kim,V.N. (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.*, **18**, 3016–3027.

19. Winter, J., Jung, S., Keller, S., Gregory, R.I. and Diederichs, S. (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.*, **11**, 228–234.
20. Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J. and Plasterk, R.H. (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.*, **15**, 2654–2659.
21. Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363–366.
22. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T. and Zamore, P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, **293**, 834–838.
23. Knight, S.W. and Bass, B.L. (2001) A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science*, **293**, 2269–2271.
24. Houseley, J., LaCava, J. and Tollervy, D. (2006) RNA-quality control by the exosome. *Nat. Rev. Mol. Cell Biol.*, **7**, 529–539.
25. Meister, G. (2013) Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.*, **14**, 447–459.
26. Schanen, B.C. and Li, X. (2011) Transcriptional regulation of mammalian miRNA genes. *Genomics*, **97**, 1–6.
27. Ruegger, S. and Grosshans, H. (2012) MicroRNA turnover: when, how, and why. *Trends Biochem. Sci.*, **37**, 436–446.
28. Ozsolak, F., Poling, L.L., Wang, Z., Liu, H., Liu, X.S., Roeder, R.G., Zhang, X., Song, J.S. and Fisher, D.E. (2008) Chromatin structure analyses identify miRNA promoters. *Genes Dev.*, **22**, 3172–3183.
29. Marsico, A., Huska, M.R., Lasserre, J., Hu, H., Vucicevic, D., Musahl, A., Orom, U. and Vingron, M. (2013) PROMiRNA: a new miRNA promoter recognition method uncovers the complex regulation of intronic miRNAs. *Genome Biol.*, **14**, R84.
30. Nepal, C., Coolen, M., Hadzhiev, Y., Cussigh, D., Mydel, P., Steen, V.M., Carninci, P., Andersen, J.B., Bally-Cuif, L., Muller, F. *et al.* (2016) Transcriptional, post-transcriptional and chromatin-associated regulation of pri-miRNAs, pre-miRNAs and moRNAs. *Nucleic Acids Res.*, **44**, 3070–3081.
31. Chang, T.C., Perte, M., Lee, S., Salzberg, S.L. and Mendell, J.T. (2015) Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. *Genome Res.*, **25**, 1401–1409.
32. Liu, Q., Wang, J., Zhao, Y., Li, C.I., Stengel, K.R., Acharya, P., Johnston, G., Hiebert, S.W. and Shyr, Y. (2017) Identification of active miRNA promoters from nuclear run-on RNA sequencing. *Nucleic Acids Res.*, **45**, e121.
33. Tank, E.M., Figueroa-Romero, C., Hinder, L.M., Bedi, K., Archbold, H.C., Li, X., Weskamp, K., Safren, N., Paez-Colasante, X., Pacut, C. *et al.* (2018) Abnormal RNA stability in amyotrophic lateral sclerosis. *Nat. Commun.*, **9**, 2845.
34. Paulsen, M.T., Veloso, A., Prasad, J., Bedi, K., Ljungman, E.A., Tsan, Y.C., Chang, C.W., Tarrier, B., Washburn, J.G., Lyons, R. *et al.* (2013) Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 2240–2245.
35. Paulsen, M.T., Veloso, A., Prasad, J., Bedi, K., Ljungman, E.A., Magnuson, B., Wilson, T.E. and Ljungman, M. (2014) Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods*, **67**, 45–54.
36. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357–359.
37. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
38. Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J. *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, **47**, D766–D773.
39. Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
40. Fromm, B., Domanska, D., Høye, E., Ovchinnikov, V., Kang, W., Aparicio-Puerta, E., Johansen, M., Flatmark, K., Mathelier, A., Hovig, E. *et al.* (2019) MirGeneDB 2.0: the metazoan microRNA complement. *Nucleic Acids Res.*, doi:10.1093/nar/gkz885.
41. Alexa, A. and R.J. (2019) 2.36.0 ed, pp. R Package, <https://www.rdocumentation.org/packages/topGO/versions/2.24.0>.
42. Lizio, M., Harshbarger, J., Shimoji, H., Severin, J., Kasukawa, T., Sahin, S., Abugessaisa, I., Fukuda, S., Hori, F., Ishikawa-Kato, S. *et al.* (2015) Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.*, **16**, 22.
43. Lizio, M., Harshbarger, J., Abugessaisa, I., Noguchi, S., Kondo, A., Severin, J., Mungall, C., Arenillas, D., Mathelier, A., Medvedeva, Y.A. *et al.* (2017) Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals. *Nucleic Acids Res.*, **45**, D737–D743.
44. de Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Astrom, G., Babina, M., Bertin, N., Burroughs, A.M. *et al.* (2017) An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.*, **35**, 872–878.
45. Lizio, M., Abugessaisa, I., Noguchi, S., Kondo, A., Hasegawa, A., Hon, C.C., de Hoon, M., Severin, J., Oki, S., Hayashizaki, Y. *et al.* (2019) Update of the FANTOM web resource: expansion to provide additional transcriptome atlases. *Nucleic Acids Res.*, **47**, D752–D758.
46. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A., Iovino, N., Aravin, A., Pfeffer, S., Rice, A., Kamphorst, A.O., Landthaler, M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
47. Magnuson, B., Veloso, A., Kirkconnell, K.S., de Andrade Lima, L.C., Paulsen, M.T., Ljungman, E.A., Bedi, K., Prasad, J., Wilson, T.E. and Ljungman, M. (2015) Identifying transcription start sites and active enhancer elements using BruUV-seq. *Sci. Rep.*, **5**, 17978.
48. Morlando, M., Ballarino, M., Gromak, N., Pagano, F., Bozzoni, I. and Proudfoot, N.J. (2008) Primary microRNA transcripts are processed co-transcriptionally. *Nat. Struct. Mol. Biol.*, **15**, 902–909.
49. Fuchs, G., Voickek, Y., Benjamin, S., Gilad, S., Amit, I. and Oren, M. (2014) 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol.*, **15**, R69.
50. Jonkers, I., Kwak, H. and Lis, J.T. (2014) Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife*, **3**, e02407.
51. Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E. and Ljungman, M. (2014) Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. *Genome Res.*, **24**, 896–905.
52. Kirkconnell, K.S., Paulsen, M.T., Magnuson, B., Bedi, K. and Ljungman, M. (2016) Capturing the dynamic nascent transcriptome during acute cellular responses: The serum response. *Biol. Open*, **5**, 837–847.
53. Jonkers, I. and Lis, J.T. (2015) Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.*, **16**, 167–177.
54. Richard, P. and Manley, J.L. (2009) Transcription termination by nuclear RNA polymerases. *Genes Dev.*, **23**, 1247–1269.
55. Lykke-Andersen, S. and Jensen, T.H. (2007) Overlapping pathways dictate termination of RNA polymerase II transcription. *Biochimie*, **89**, 1177–1182.
56. Rondon, A.G., Mischo, H.E. and Proudfoot, N.J. (2008) Terminating transcription in yeast: whether to be a ‘nerd’ or a ‘rat’. *Nat. Struct. Mol. Biol.*, **15**, 775–776.
57. Core, L.J., Waterfall, J.J. and Lis, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**, 1845–1848.
58. Lloret-Llinares, M., Mapendano, C.K., Martlev, L.H., Lykke-Andersen, S. and Jensen, T.H. (2016) Relationships between PROMPT and gene expression. *RNA Biol.*, **13**, 6–14.
59. Lacadie, S.A., Ibrahim, M.M., Gokhale, S.A. and Ohler, U. (2016) Divergent transcription and epigenetic directionality of human promoters. *FEBS J.*, **283**, 4214–4222.
60. Kim, Y.K., Kim, B. and Kim, V.N. (2016) Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E1881–E1889.
61. Bommer, G.T., Gerin, I., Feng, Y., Kaczorowski, A.J., Kuick, R., Love, R.E., Zhai, Y., Giordano, T.J., Qin, Z.S., Moore, B.B. *et al.* (2007) p53-Mediated Activation of miRNA34 Candidate Tumor-Suppressor Genes. *Curr. Biol.*, **17**, 1298–1307.

62. Card,D.A., Hebbar,P.B., Li,L., Trotter,K.W., Komatsu,Y., Mishina,Y. and Archer,T.K. (2008) Oct4/Sox2-regulated miR-302 targets cyclin D1 in human embryonic stem cells. *Mol Cell Biol.*, **28**, 6426–6438.
63. Kirkconnell,K.S., Magnuson,B., Paulsen,M.T., Lu,B., Bedi,K. and Ljungman,M. (2017) Gene length as a biological timer to establish temporal transcriptional regulation. *Cell Cycle*, **16**, 259–270.
64. Monteys,A.M., Spengler,R.M., Wan,J., Tecedor,L., Lennox,K.A., Xing,Y. and Davidson,B.L. (2010) Structure and activity of putative intronic miRNA promoters. *RNA*, **16**, 495–505.
65. Marzi,M.J., Ghini,F., Cerruti,B., de Pretis,S., Bonetti,P., Giacomelli,C., Gorski,M.M., Kress,T., Pelizzola,M., Muller,H. *et al.* (2016) Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *Genome Res.*, **26**, 554–565.
66. Baccarini,A., Chauhan,H., Gardner,T.J., Jayaprakash,A.D., Sachidanandam,R. and Brown,B.D. (2011) Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells. *Curr. Biol.*, **21**, 369–376.
67. de la Mata,M., Gaidatzis,D., Vitanescu,M., Stadler,M.B., Wentzel,C., Scheffele,P., Filipowicz,W. and Grosshans,H. (2015) Potent degradation of neuronal miRNAs induced by highly complementary targets. *EMBO Rep.*, **16**, 500–511.