

# Figures and figure supplements

Nascent-Seq reveals novel features of mouse circadian transcriptional regulation

Jerome S Menet, et al.



**Figure 1**. Genome-wide assay of transcription in the mouse liver using Nascent-Seq. (**A**): Distribution of high-throughput sequencing signal within introns (green), exons (blue) and intergenic regions (grey) for Nascent-Seq and RNA-Seq datasets. (**B**): Visualization of Nascent-Seq and RNA-Seq signal at chr4: 40,730,000–41,002,500. Genes above the scale bar are transcribed from left to right and those below the scale bar are transcribed from right to left. Nascent-Seq signal exhibits increased intron signal and a 5' to 3' gradient signal (arrow). Moreover, differences between Nascent-Seq signal and RNA-Seq signal are observed for many genes (e.g., *Bag1* and *B4galt1*). (**C**): Nascent-Seq signal (brown), but not RNA-Seq signal (red), extends past the annotated 3'end of the genes *B4galt1* and *Nfx1*. (**D**): Gene ontology of genes with high Nascent-Seq and low RNA-Seq signals (and inversely) is indicative of RNA with short or long half-lives, respectively (see 'Materials and methods' for details). (**E**): Distribution of the Nascent-Seq signal *Figure 1*. Continued on next page



### Figure 1. Continued

ratio for the classes of genes enriched in (**D**). (**F**): Nascent-Seq/RNA-Seq signal ratio significantly correlates with mRNA half-lives (values from *Sharova et al., 2009*), and genes with high ratio display shorter half-lives and inversely. (**G**) and (**H**): Strategy used to determine the gene signal cut-off threshold used in our analysis. Variation of gene signal coming from the sequencing of a Nascent-Seq library (G; ZT8, replicate 1) sequenced in two Illumina flow-cell lanes was assessed by calculating the z-score (**H**). Less than 5% of the genes with a read per base pair superior to three exhibit a 1.3-fold gene signal variation. See 'Materials and methods' for more details.



**Figure 2**. Genome-wide analysis of rhythmic transcription in the mouse liver. (**A**): Visualization of *Npas2* Nascent-Seq signal at six time points of the light:dark cycle (first replicate). *Npas2* Nascent-Seq signal is rhythmic and peaks at ZT20-ZT0, contrary to the signal within the adjacent gene *Rpl31*. (**B**): Quantification of the number of genes that are rhythmically transcribed in the mouse liver. Genes with more than three reads per base pair for at least one time point were included for the analysis. Genes are considered to be rhythmically transcribed if signal amplitude (Amp) is greater than 1.5, if signals for the 12 time points follow a sinusoid curve ( $F_{24} > 0.45$ ) and if the  $F_{24}$  value is in the top 5% of all  $F_{24}$  values calculated after time points were permutated 10,000 times (p<0.05). A rhythm was considered to be strong (dark red) if  $F_{24} > 0.6$  and Ampl > 1.75. (**C**): Heatmap representation of Nascent-Seq signal for the 963 genes that are rhythmically transcribed in the mouse liver. High expression is displayed in yellow (z-score > 1), low expression in blue (z-score < 1). (**D**): Expression phase of rhythmically expressed nascent RNA (n = 936) was separated by bins of 2 hr. Analysis of their distribution reveals that fewer genes are transcribed at ZT16-20. (**E**) and (**F**): Rhythmic Nascent-Seq signal was detected for many precursors of non-coding RNAs such as pri-miRNA (d, *pri-miR122a*) and long non-coding RNA(e, lin-ncRNAs *BC019819*, *AK157581*, *BC049268*, *BC056646*). DOI: 10.7554/eLife.00011.004



**Figure 2—figure supplement 1**. Rhythmic transcription of IncRNA ENSMUSG0000098984 in the mouse liver. Visualization of Nascent-Seq signal (brown; six time points of replicate 1) for the long non-coding RNA (IncRNA) precursor ENSMUSG0000098984. Genes above the scale bar are transcribed from left to right and those below the scale bar are transcribed from right to left. DOI: 10.7554/eLife.00011.006



**Figure 2—figure supplement 2**. Rhythmic transcription of IncRNA ENSMUSG0000086813 in the mouse liver. Visualization of Nascent-Seq signal (brown; six time points of replicate 1) for the long non-coding RNA (IncRNA) precursor ENSMUSG0000086813. Genes above the scale bar are transcribed from left to right and those below the scale bar are transcribed from right to left. DOI: 10.7554/eLife.00011.007



Figure 2—figure supplement 3. Rhythmic transcription of IncRNA ENSMUSG0000086771 in the mouse liver. Visualization of Nascent-Seq signal (brown; six time points of replicate 1) for the long non-coding RNA (IncRNA) precursor ENSMUSG0000086771. Genes above the scale bar are transcribed from left to right and those below the scale bar are transcribed from right to left.

DOI: 10.7554/eLife.00011.008



Figure 2—figure supplement 4. Rhythmic transcription of pri-miRNA pri-Mir17hg in the mouse liver. Visualization of Nascent-Seq signal (brown; six time points of replicate 1) for the pri-miRNA pri-Mir17hg. Enlargement of pri-miRNA signal reveals that pri-miRNA transcription units are not well annotated, precluding a rigorous quantification of the signal Genes above the scale bar are transcribed from left to right and those below the scale bar are transcribed from right to left.



Figure 2—figure supplement 5. Rhythmic transcription of pri-miRNA ENSMUSG0000077856 in the mouse liver. Visualization of Nascent-Seq signal (brown; six time points of replicate 1) for the pri-miRNA ENSMUSG0000077856. Enlargement of pri-miRNA signal reveals that pri-miRNA transcription units are not well annotated, precluding a rigorous quantification of the signal. Genes above the scale bar are transcribed from left to right and those below the scale bar are transcribed from right to left.

DOI: 10.7554/eLife.00011.010



Figure 2—figure supplement 6. Rhythmic transcription of pri-miRNA ENSMUSG0000093077 in the mouse liver. Visualization of Nascent-Seq signal (brown; six time points of replicate 1) for the pri-miRNA ENSMUSG0000093077. Enlargement of pri-miRNA signal reveals that pri-miRNA transcription units are not well annotated, precluding a rigorous quantification of the signal. Genes above the scale bar are transcribed from left to right and those below the scale bar are transcribed from right to left. DOI: 10.7554/eLife.00011.011



**Figure 3**. Post-transcriptional events account for a significant fraction of rhythmic gene expression in the mouse liver. (**A**): Rhythmic gene expression was assessed as in *Figure 2B* for genes sufficiently expressed in both Nascent-Seq and RNA-Seq datasets. Four categories of rhythmically expressed genes were determined by comparing the Nascent-Seq and RNA-Seq datasets: rhythmic nascent RNA and mRNA (R-R), rhythmic nascent RNA only (R-AR), rhythmic nascent RNA only (AR-R) and arrhythmic nascent RNA and mRNA (AR-AR). (**B**): Heatmap representation of genes with rhythmic nascent RNA and mRNA expression (n = 342). Classification is based on the phase of nascent RNA oscillations, and each lane corresponds to one gene. (**C**): Double-plotted phase distribution of rhythmic nascent RNA expression (brown) and rhythmic mRNA expression (red) for genes of the R-R gene set. Both phases are highly correlated (r = 0.92). (**D**): Distribution of the difference between the phase of mRNA expression rhythm and the phase of nascent RNA expression rhythm for the 342 R-R genes. (**E**): Amplitude of mRNA expression rhythms are correlated with nascent RNA expression rhythms (r = 0.76). (**F**) and (**G**): Similar representation to (**B**) for rhythmically transcribed genes with no mRNA expression rhythms (**C**, n = 480), and genes that exhibit mRNA oscillations but no rhythms of transcription (**D**, n = 862). For all three heatmaps, high expression is displayed in yellow (z-score > 1), low expression in blue (z-score < 1).



**Figure 4**. Clock genes nascent RNA and mRNA expression in the mouse liver. Clock genes nascent RNA levels (brown; time points every 4 hr starting at ZT0) and mRNA levels (red; time points every 4 hr starting at ZT2) from the Nascent-Seq and RNA-Seq datasets. Relative levels between nascent RNA and mRNA expression profiles are identical for all genes to allow direct comparison. DOI: 10.7554/eLife.00011.015



**Figure 5**. Analysis of the different classes of rhythmically expressed genes in the mouse liver. (**A**): Nascent-Seq/RNA-Seq signal ratio (used as inferred half-life) is similar for the four categories of rhythmically expressed genes: rhythmic nascent RNA and mRNA (R-R), rhythmic nascent RNA only (R-AR), rhythmic nascent RNA only (AR-R) and arrhythmic nascent RNA and mRNA (AR-AR). (**B**): Similar as (**A**), using the RNA half-life values from *Sharova et al., 2009*. (**C**): Nascent-Seq rhythms of 25 of the 480 R-AR genes can be attributed to the rhythmic transcription of an adjacent gene. This applies to *Sphk2* Nascent-Seq rhythm, which likely results from rhythmic *Dbp* nascent RNA signal that extend the 3'end of *Dbp* gene and read through *Sphk2*. Genes above the scale bar are transcribed from left to right and those below the scale bar are transcribed from right to left. (**D**): Gene ontology of three categories of rhythmically expressed genes: rhythmic nascent RNA and mRNA (R-R), rhythmic nascent RNA only (R-AR), rhythmic mRNA only (AR-R). DOI: 10.7554/eLife.00011.016



**Figure 6**. Transcriptional variability of AR-R genes contributes to rhythmic mRNA expression. (**A**) and (**B**): Nascent RNA levels (brown; time points every 4 hr starting at ZT0) and mRNA levels (red; time points every 4 hr starting at ZT2) from the Nascent-Seq and RNA-Seq datasets for six genes of the AR-R gene set. While the majority of the AR-R genes exhibit variable nascent RNA expression (**A**), some of them exhibit a relatively constant transcription when compared to mRNA expression (**B**). (**C**): Standard deviation (SD; calculated using the 12 time points and normalized to the mean) of nascent RNA *Figure 6*. *Continued on next page* 



### Figure 6. Continued

expression is higher than the SD normalized to the mean of mRNA levels for most AR-R genes. (**D**) and (**E**): Higher transcriptional variability (SD) of arrhythmically transcribed genes is associated with higher occurrence of rhythmic mRNA expression (**D**), but not to nascent RNA expression levels (**E**). (**F**): Higher variability of transcription for the genes of the AR-R group is associated with increase amplitude of rhythms at both Nascent RNA (brown) and mRNA (red) level. Genes of the AR-R group (n = 862) were binned into five quintiles of equal size (q1-q5). (**G**): Heatmap representation of 86 AR-R genes that exhibit high level of transcription at only one time point, and with rhythmic mRNA expression. High expression is displayed in yellow (z-score > 1), low expression in blue (z-score < 1). (**H**): Nascent RNA levels (brown) and mRNA levels (red) for four AR-AR genes with variable nascent RNA expression that is not associated to rhythmic mRNA expression. (**I**): Number of predicted miRNA target sites of AR-R genes with high transcriptional variability (q1, top 20% of the 826 AR-R genes) and low transcriptional variability (q5, bottom 20%). (**J**): Gene ontology of AR-R genes with high transcriptional variability (top 25%) when compared to all AR-R genes. Significant enrichment (top) and depletion (bottom) of biological functions for these genes are displayed. Values correspond to the number of genes within this top 25% of genes, when compared to all AR-R genes. DOI: 10.7554/eLife.00011.017



Figure 7. Continued on next page



### Figure 7. Continued

**Figure 7**. Characterization of CLK and BMAL1 target genes in the mouse liver. (**A**) and (**B**): Visualization (**A**) and quantification (**B**) of BMAL1 ChIP-Seq, CLK ChIP-Seq and Input signals at BMAL1 and CLK significant peaks (analysis using MACS algorithm). BMAL1 ChIP-Seq, CLK ChIP-Seq and Input signals were retrieved based on the location of the BMAL1 peaks (center  $\pm$  1kb, for CLK:BMAL1 peaks and BMAL1 only peaks) or the CLK peaks (center  $\pm$  1kb, for CLK only peaks). Normalization was performed on the entire datasets by calculating the z-score ((x – mean)/SD). Heatmap displays high expression in red and low expression in blue. Quantification (**B**) was performed by averaging the z-score by bins of 25 bp for all CLK:BMAL1 peaks (n = 211), BMAL1 only peaks (n = 1368) and CLK only peaks (n = 548). (**C**): Enrichment of e-boxes (perfect CACGTG in red, degenerated e-boxes [one nucleotide mismatch, in orange]) within  $\pm$ 500 bp of CLK:BMAL1, BMAL1 only and CLK only peak centers. (**D**): Motifs enriched within CLK:BMAL1 peaks, BMAL1 only peaks and CLK only peaks, as revealed by MEME analysis. (**E**)–(**H**): Visualization of BMAL1 ChIP-Seq (blue), CLK ChIP-Seq (green) and Nascent-Seq (brown; six time points of replicate 1) signals for *Rev-Erba* (**E**), *Per1* (**F**), *Cry1* (**G**) and a cluster of 4 lncRNA (*AK079377*, *AK007907*, *AK036974*, *AK087624*) (**H**) targeted by CLK:BMAL1. Genes above the scale bar are transcribed from left to right and those below the scale bar are transcribed from right to left. DOI: 10.7554/eLife.00011.019



**Figure 8**. Disconnect between rhythmic BMAL1 DNA binding and its transcriptional output. (**A**): Heatmaps representing BMAL1 ChIP-Seq signal (from *Rey et al., 2011*), Nascent-Seq and RNA-Seq signal for CLK:BMAL1 target genes (six time points in duplicate). Genes were classified in four categories: rhythmic nascent RNA and mRNA (R-R), rhythmic nascent RNA only (R-AR), rhythmic mRNA only (AR-R) and arrhythmic nascent RNA and mRNA (AR-AR). High expression is displayed in yellow, low expression in blue. (**B**): Peak phase distribution of rhythmic BMAL1 DNA binding (blue, from *Rey et al., 2011*), of nascent RNA (black) and of mRNA (red) for the direct target genes that are rhythmically expressed at both the nascent RNA and mRNA levels. (**C**): Distribution of CLK:BMAL1 target genes within the 4 different classes of rhythmically expressed genes and its comparison to the genome-wide distribution. Rhythmic nascent RNA and mRNA: R-R; rhythmic nascent RNA only: R-AR; rhythmic mRNA only: AR-R; arrhythmic nascent RNA and mRNA: AR-AR. (**D**): qPCR quantification of *Rev-Erba*, *Per1*, *Per2* and *Cry1* pre-mRNA every 4 hr throughout the day in wild-type (black, n = 4 per time points) and *Bmal1-/-* mice (blue, n = 3 per time points). Error bar: s.e.m. (**E**): Visualization of BMAL1 ChIP-Seq (blue), CLK ChIP-Seq (green), Nascent-Seq (brown; six time points of replicate 1), Pol II ChIP-Seq signal (purple) at ZT10 and ZT22 (from *Feng et al., 2011*) and strand-specific Nascent-Seq signal for *Per2* (plus strand, top; minus strand, bottom). *Per2* is rhythmically transcribed (minus strand) with a peak at ZT16. An antisense transcript is rhythmically transcribed to *Per2* RNA (plus strand), peaking at ZT4.



**Figure 9**. Post-transcriptional events contribute to rhythmic mRNA expression in the mouse liver. Although rhythmic transcription plays a major role for approximately 30% of the genes that exhibit rhythmic mRNA expression, post-transcriptional events significantly contribute to the generation of mRNA rhythms for the majority of genes (~70%). Many post-transcriptional cyclers exhibit highly variable transcription that is buffered to generate robust rhythmic mRNA expression. Few genes exhibit a relatively constant transcription when compared to mRNA expression. These post-transcriptional events may include roles for RNA binding proteins and miRNAs to regulate RNA stability, 3' end formation and nuclei export.