/

Studying the regulation and development of circadian clock by systems biology approaches

Dissertation

for the award of the degree

"Doctor rerum naturalium" (Dr.rer.nat.)

of the Georg-August-Universität Göttingen

within the doctoral program "Biology"

of the Georg-August University School of Science (GAUSS)

submitted by

Haifang Wang

from Hebei, China

Göttingen, 2020

Thesis Committee

Prof. Dr. Gregor Eichele, Genes and Behavior, Max-Planck-Institute for Biophysical Chemistry, Göttingen (Referent)

Prof. Dr. Burkhard Morgenstern, Department of Bioinformatics (IMG), University of Göttingen, Göttingen (Co-Referent)

Prof. Dr. Christian Griesinger, NMR-based Structural Biology, Max-Planck-Institute for Biophysical Chemistry, Göttingen

Members of the Extended Examination Board

Dr. Ufuk Günesdogan, Department of Developmental Biology, University of Göttingen, Göttingen

Dr. Gerd Vorbrüggen, Department of Molecular Developmental Biology, Max-Planck-Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Tim Beißbarth, Department of Medical Bioinformatics, University Medical Center Göttingen, Göttingen

Date of oral examination: 18.09.2020

Acknowledgements

Contents

List of Figures

Abbreviations

CAGE	cap-analysis gene expression
ChIP	chromatin immunoprecipitation
CLIP-seq	crosslinking immunoprecipitation sequencing
DBTSS	database of transcriptional start site
DD	constant dark
DEGs	differentially expressed genes
dpf	days post-fertilization
ESCs	embryonic stem cells
FACs	flow-activated cell sorting
FPKM	fragments per kilobase of exon model per million mapped reads
GRO-seq	global run-on sequencing
GSEA	gene set enrichment analysis
GTF	gene transfer format
GWAS	genome-wide association studies
HTS	high-throughput screening
IGB	integrated genome browser
IGV	integrative genomics viewer
ipRGCs	intrinsically photosensitive retinal ganglion cells
iPSCs	induced pluripotent stem cells
IRES	ribosomal entry site
LD	light-dark
miRNAs	microRNAs
Nascent-seq	genome-wide sequencing of nascent RNA
NGS	next-generation sequencing
PCA	principal component analysis
piRNAs	PIWI-interacting RNAs
Pol II	RNA polymerase II
PWM	positional weight matrix
RBPs	mRNA-binding proteins
RHT	retinohypothalamic tract
RORE	retinoic acid-related orphan receptorresponse element
RPKM	reads per kilobase of exon model per million reads
SBS	sequencing by synthesis
SCN	suprachiasmatic nuclei
scRNA-seq	single-cell RNA-sequencing
siRNAs	short-interfering RNAs
SMART-seq	switching mechanism at the end of the 5´-end of the RNATranscript sequencing
SMRT	single-molecule real-time
sRNAs	small RNAs
TF	transcription factor

TFBS	transcription factor binding sites
TMM	trimmed mean of the M-values
TPM	transcripts per million
t-SNE	t-distributed stochastic neighbor embedding
UMAP	uniform manifold approximation and projection
UMIs	unique molecular identifiers
VNP	Venus-NLS-PEST
WGBS	whole-genome bisulfite sequencing

1 Summary

1.1 Summary in English

System biology approaches, coupled with emerging omics tools, have been widely used and greatly accelerated biological and medical research and discovery. In this thesis, I will focus on applying those tools and technologies to the post-transcriptional regulation of the circadian clock by miRNAs in mice and the development of a single-cell clock model in zebrafish, respectively.

The circadian clock is an endogenous oscillator that regulates a wide range of cellular, metabolic, physiological, and behavioral rhythms in animals, plants and micro-organisms. Gene regulatory circuits driving circadian rhythms in higher vertebrates and insects are mainly generated through transcriptional/translational feedback loops of core clock genes (*Bmal1*, *Clock*, *Per* gene family, *Cry* gene family, and *Rev-erb α/β* (*Nr1d1/2*)). Recent advance in clock regulation suggests that post-transcriptional mechanisms play essential roles in the regulation of circadian gene expression. By integrative analysis of various types of published data, we identified 57 circadian miRNA primary transcripts and in mouse liver 56 of them are regulated by one or more core circadian transcription regulator. We found that the mature miRNAs arising from these circadian primary transcripts were either not oscillating at all or oscillating but with low amplitudes. The over-expression of one of them, miR-378, in mouse liver by adenovirus injection downregulated a significant proportion of the circadian oscillating target genes. In particular, we observed that miR-378 participates in the circadian control of both, cell cycle and metabolism. It does so by forming either coherent or incoherent feed-forward loops with different circadian regulators.

Although the molecular mechanism of circadian clock is well studied, how and when these cellular rhythms arise during development is largely unknown. We sought to study the development of circadian clock using zebrafish (*Danio rerio*), a well-established model organism for the study of animal development and circadian rhythms. Due to the transparency of the fish larval stages, zebrafish is also a suitable system for *in vivo* live imaging. We constructed a transgenic zebrafish line, Venus-NLS-PEST (VNP), driven by the promoter of *nr1d1*, which encodes the core circadian regulator Rev-Erbα. This system allows us to monitor single-cell circadian rhythm *in vivo* for the first time. We carried out single-cell *in vivo* imaging of *nr1d1*:VNP fish over embryonic development, and found that circadian expression of pineal *nr1d1*:VNP positive cells was superimposed on their cell-type-specific developmental expression trends. Furthermore, the fish line helped us to address a long-standing issue in circadian clock ontogeny: is the clock development by synchronizing existing oscillators or by initializing coherent single-cell clocks? We found that early exposure to the light-dark (LD) cycle during development initializes the oscillators of individual cells rather than synchronizing existing oscillators of individual cells as researchers in the field have previously believed. This shows the necessity of the single-cell circadian reporter.

1.2 Zusammenfassung auf Deutsch

Systembiologische Ansätze, gepaart mit neuen -omics-Tools, sind weit verbreitet und haben die biologische und medizinische Forschung und Entdeckung erheblich beschleunigt. In meiner Doktorarbeit konzentrierte ich mich auf die Anwendung dieser Werkzeuge und Technologien auf (1) die posttranskriptionelle Regulation der circadianen Uhr durch miRNAs in Mäusen und (2) die Entwicklung der circadianen Uhr in individuellen Zellen des Zebrafisches.

Die circadiane Uhr ist ein endogener Oszillator, der eine Vielzahl von zellulären, metabolischen, physiologischen und Verhaltensrhythmen bei Tieren, Pflanzen und Mikroorganismen reguliert. Gen-Netzwerke, die den circadianen Rhythmus bei Wirbeltieren und Insekten steuern, bestehen aus Transkriptions- bzw. Translations-Rückkopplungsschleifen. Diese Schleifen bestehen aus Uhren-Proteinen (Bmal1, Clock, Per-Proteine, Cry-Proteine und Rev-erb α/β (kodiert von *Nr1d1/2*)). Jüngste Fortschritte bei der Erforschung der circadianen Uhr legen nahe, dass posttranskriptionelle Mechanismen eine wesentliche Rolle bei der Regulation der circadianen Genexpression spielen. Meine integrative Analyse verschiedener Arten von bereits veröffentlichter Daten, identifizierte 57 circadiane miRNA-Primärtranskripte, von denen 56 in der Mausleber durch circadiane Uhren-Proteine reguliert werden. Wir fanden heraus, dass die reifen miRNAs, die sich von diesen circadianen Primärtranskripten ableiten, entweder überhaupt nicht oder mit geringen Amplituden oszillieren. Die Überexpression eines von ihnen (miR-378) in der Mausleber mittels Adenovirus-Injektion, reduzierte bei einem signifikanten Anteil das circadiane Expressionsmuster der miRNA Zielgene. Insbesondere beobachteten wir, dass miR-378 an der Steuerung der circadianen Oszillation von Genen, die am Zellzyklus und am Stoffwechsel beteiligt sind. Diese Kontrolle basiert sowohl auf kohärenten oder inkohärente Feed-Forward-Schleifen, die jeweils aus den verschiedenen Uhren-Proteinen bestehen.

Obwohl der molekulare Mechanismus der circadianen Uhr gut untersucht ist, bleibt weitgehend unbekannt, wie und wann diese zellulären Rhythmen während der Entwicklung initialisiert werden. Ich wollte daher die Entwicklung der circadianen Uhr im Zebrafischen (Danio rerio) erforschen, einem etablierten Modellorganismus zu dem es umfangreiche Untersuchung sowohl der Entwicklung als auch der circadianen Rhythmik gibt. Weil Fischlarvenstadien durchsichtig sind. ist dieser Organismus gut geeignet für die in vivo live-Bildgebung. Wir konstruierten eine transgene Zebrafischlinie (Venus-NLS-PEST; VNP), die vom Promotor von circadianen Gens *nr1d1* gesteuert wird. Dieses Gen kodiert das Uhren-Protein Rev-Erbα. Mit diesem System konnten wir zum ersten Mal den circadianen Rhythmus einzelner Zellen im lebenden Organismus sichtbar machen. Wir führten eine *in vivo* Einzelzell Bildgebung an nr1d1: VNP-Fischen durch, die sich über die Embryonalentwicklung erstreckte. Ich fand heraus, dass in Zirbeldrüsenzellen die circadiane Rhythmik der nr1d1:VNP Expression und die entwicklungsbedingte nr1d1:VNP Expression sich überlagerten. Darüber hinaus hat uns die Fischlinie ermöglicht, eine seit langem bestehende Frage der Ontogenese der circadianen Uhr anzugehen. Entwickelt sich die Uhr durch Synchronisierung bereits vorhandener Oszillatoren oder durch Initialisierung kohärenter Einzelzellenuhren? Wir fanden heraus, dass frühes Exponieren von Fischembryonen an einen Hell-Dunkel-Zyklus, die Oszillatoren einzelner Zellen im Verband andreht. Es findet also keine Synchronisation bereits vorhandener unkoordiniertr Einzelzell-

Oszillatoren statt, wie Forscher auf diesem Gebiet ursprünglich angenommen hatten. Diese Erkenntnis illustriert den Nutzen nr1d1:VNP Reporterlinie.

2 Introduction

With the rapid development of omics technologies and bioinformatic tools, especially the next-generation sequencing (NGS) technologies, the study of the biological process at system level become feasible (see 2.1). This work aims to take advantage of those technologies and tools to study the regulation and development of the circadian clock (see 2.2). Circadian clock is an endogenous timing system that regulates almost all the physiological process in the body. By integrating all the related NGS data, including RNA-seq, miRNA-seq, and ChIP-seq data in mouse liver, we identified cohort of miRNAs derived from circadian miRNA primary transcripts that can have circadian functions. Afterwards, over-expressing of one of them, miR-378, coupled with transcriptome analysis by RNA-seq, we found that miR-378 participates in the circadian control of both cell cycle and metabolism (see 3.1). In the second project, through a systematic screening, we established a clock reporter zebrafish line that allows us to monitor circadian clock function at single-cell level in larva zebrafish *in vivo*. Using a state-of-art technology of single-cell sequencing (scRNA-seq), we systematically characterized the cell types where the clock reporter expressed. Then by two-photon imaging and 3D imaging analysis, we studied the ontogeny of the circadian clock and we found that cellular clocks require light exposure in order to start ticking, which is different than what was previously thought (see 3.2).

2.1 High-throughput sequencing and data analysis

Traditional Sanger sequencing (Sanger and Coulson, 1975; Sanger et al., 1977) is called first generation sequencing, and can only sequences a single DNA fragment at a time. NGS, also called high-throughput screening (HTS), achieved by massively parallel sequencing, refers to a number of modern sequencing technologies. This includes Illumina (Solexa) sequencing (Illumina, 2008), Roche/454's pyrosequencing (Margulies et al., 2005), ABI's SOLiD sequencing (Valouev et al., 2008), Ion Torrent's semiconductor sequencing (Thermo Fisher) (Rothberg et al., 2011), Single-molecule real-time (SMRT) sequencing (Pacific Biosciences) (Eid et al., 2009), and so on. NGS all follows a similar general paradigm: fragmentation, clonal amplification, and sequencing of clonal amplicons in a flow cell (Figure 2.1). At the clonal amplification step, 454 and SOLiD sequencing are based on emulsion PCR on the surface of a gel bead, while Illumina is based on a unique bridged PCR that takes place on the surface of a flow cell. For the sequencing, all three technologies are based on sequencing by synthesis (SBS) technology. 454 sequencing is called pyrosequencing, which utilizes the enzymes ATP sulfurylase and luciferase. Addition of the nucleotides by DNA polymerase induces a light-producing reaction, and the light signal strength is proportional to the number of incorporated nucleotides. The light signal is recorded with a CCD (charge-optic device) camera. On the Illumina platform, a mixture of modified nucleotides is added during synthesis. Each of them carries a base-specific removable fluorescent label that allows a larger output at lower cost. In addition, each modified nucleotide has the 3' OH group blocked which ensures to add one nucleotide at a time. Then the emitted fluorescent light spectrum is recorded with a microscope that shows which base is incorporated. In contrast to 454 and Illumina sequencing, on SOLiD platform, sequencing by synthesis is performed by utilizing DNA ligase, rather than DNA polymerase. Each cycle of sequencing involves the incorporation of a fluorescently labeled dinucleotide primer (Kulski, 2016; Mardis, 2008; Shendure and Ji, 2008).
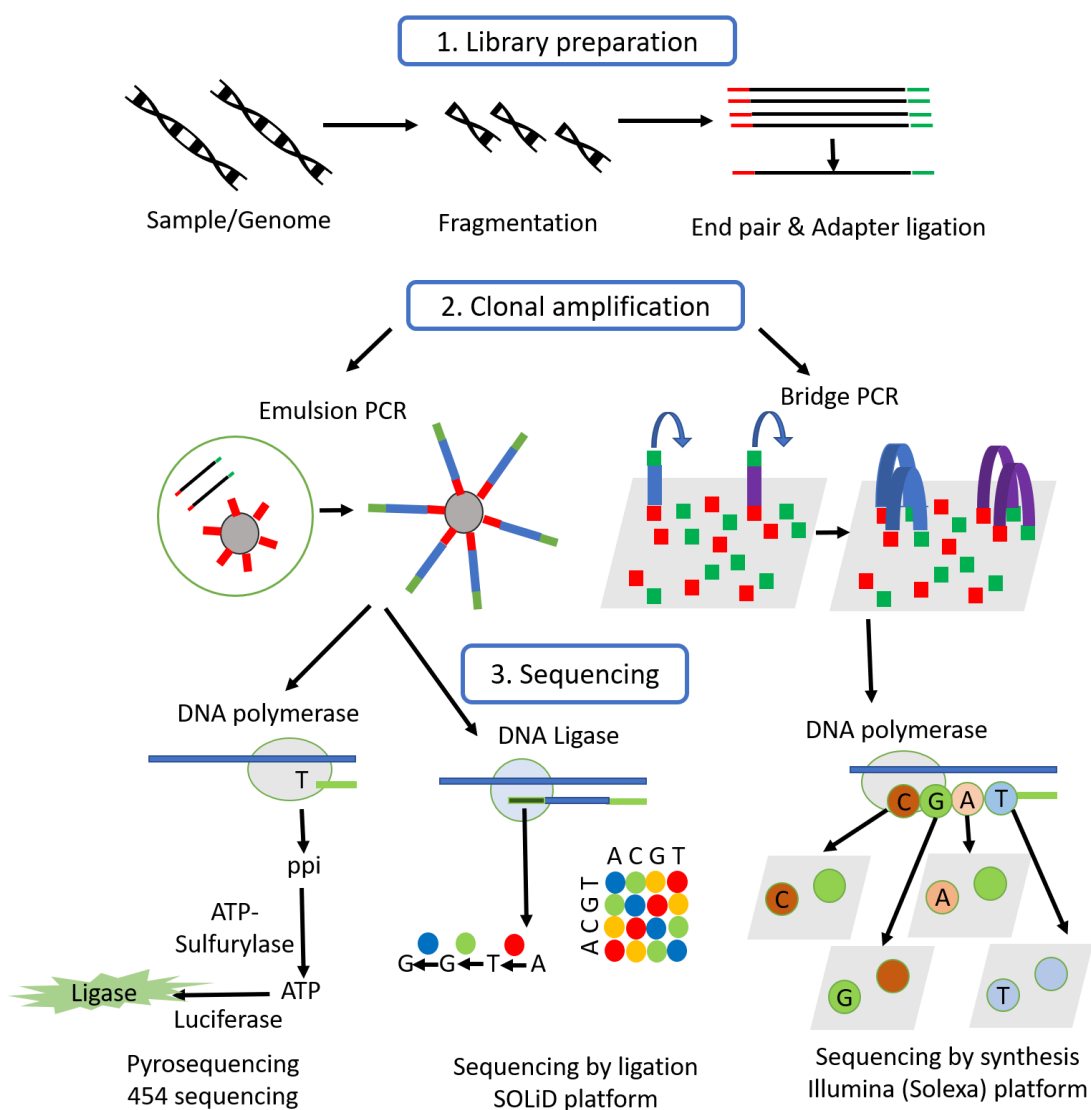
Figure 2.1 Overview of the basic steps involved in NGS for three widely used platforms. Generally, NGS workflows on all the platforms include three steps: fragmentation, clonal amplification, and sequencing.

Since the release in 2006, Illumina (Solexa) Genome Analyzer has largely improved both in boosting the output and in reducing the cost and become market-dominant platform among the NGSs. All the NGS data used in this thesis were performed on the Illumina (Solexa) platform. Therefore, in this section the introduction of NGS technology and its application are all based on the Illumina sequence system.

2.1.1 The basics of Illumina next-generation sequencer

Illumina NGS utilizes SBS technology - tracking the incorporation of labelled nucleotide as the DNA chain is copied - in a massively parallel fashion (Illumina, 2008). The Illumina NGS workflow can be separated into four steps.

a. library preparation: First, the DNA and cDNA samples are randomly cut to 200-500bp short fragments by hydrodynamic shearing. Then the 5' and 3' adapters are ligated at both ends of the DNA fragment, followed by purification and size-selection using agarose gel electrophoresis. Next the selected fragments are PCR amplified using the primers complementary to the adapters.

b. Cluster generation: The amplified DNA library is fixed on a proprietary glass slide flow cell surface, which contains millions of adapter-complimentary oligos lining the bottom. Then each fragment is amplified through bridge amplification.

c. Sequencing: Primers bound to the forward strands and the DNA polymerase add a fluorescently labeled-dNTPs containing a reversible polymerase enzyme terminator (3′-O-azidomethyl-dNTPs), which ensure that only one base can be added per reaction. Following are washing off the unincorporated bases, image scanning of the fluorescent signal with CCD camera and terminator removal allowing a next cycle of dNTP addition.

d. Data Analysis: The experimentally identified sequence reads are first aligned to a reference genome or *de novo* assembled to novel genomes or transcriptomes. Following that step, many types of analyses are possible depending on how the DNA or RNA samples were obtained. For Chromatin immunoprecipitation (ChIP) followed by HTS sequencing (ChIP-seq), binding peak and motif are identified. For RNA-sequencing (RNA-seq) data, differentially expressed genes (DEGs) across different conditions can be identified. This is typically followed by pathway analysis and functional enrichment analysis.

2.1.2 The application of next-generation sequencing

Illumina NGS platforms provide a wide variety of methods that have been comprehensively applied to many research areas in biology and medicine, including whole genome sequencing, target DNA/RNA sequencing, transcriptome profiling, ChIP-seq, DNA methylation sequencing, and so on.

2.1.2.1 Genome sequencing

Genome sequencing includes *de novo* sequencing and re-sequencing of existing genomes. *De novo* sequencing means to sequence a genome without prior knowledge of the sequence. NGS allows for faster, cheaper, and more precisely sequencing any species and has been applied to both *de novo* sequencing the large genomes (>5mb), including animals (Li et al., 2010) and plants (Nakasugi et al., 2013; Rastogi et al., 2014), as well as small genome such as bacteria, viruses, and other microbes. Re-sequencing existing genome can provide valuable information for disease and population-level studies. One widely used application is genome-wide association studies

(GWAS). GWAS is frequently used to identify genetic variations between two populations of organisms-one with a specific disease and one without, which allows researchers to map genes that affect specific outcomes or traits (Cantor et al., 2010). Comparing with the traditional microarray-based GWAS studies, which ultimately examined ~4 million markers for each sample, NGS approach allows examining all of the 3.2 billion bases of the human genome.

2.1.2.2 Transcriptome studies

Not so long ago, microarrays that are solid surfaces with a collection of surface-attached probes to which one hybridizes a cDNA, were the most dominant technology for investigating transcriptome changes. Yet, array technology has several obvious disadvantages, such as only known genes can be analyzed, and background hybridization adversely affects the accuracy of the measurements. RNA-Seq, on the other hand, is the direct sequencing of transcripts by NGS technology. It can quantify and profile any active gene or transcript both coding and non-coding without prior knowledge of the transcript sequence. As the costs of NGS are coming down significantly, RNA-seq now becomes a predominant platform for transcriptome studies (Zhao et al., 2014).

Illumina developed a variety of RNA-seq methods for mRNA sequencing, total RNA sequencing, exome sequencing, microRNA sequencing, ribosome profiling, and single-cell RNA-seq. mRNA sequencing, targeting all polyadenylated RNAs, is most widely used. It is applied to measure gene and transcript abundance, identify alternative splicing, detect gene fusions and other features. Total RNA-seq, on the other hand, detects all the transcripts both mRNA and multiple forms of noncoding RNA, including lncRNA, small RNA, circular RNA, and so on. Exome sequencing, which is an alternative for whole-genome sequencing in WGAS studies, captures only the variants in the coding regions of DNA, so it allows sequencing more samples with low sequencing depth. Ribosome profiling is a method for purification and sequencing the mRNA fragments that are being translated, which can be used for detecting cellular translation processes and quantification of protein expression changes (McGlincy and Ingolia, 2017). All the RNA-seq methods above are applied to analyze bulk tissue samples, while single-cell sequencing is highly sensitive RNA-seq methods designed for gene expression analysis of very low-input samples and even single cells, which enable the exploration of the distinct biology of different cells within an organ or a tumor and understanding subpopulation responses to various environmental cues.

2.1.2.3 DNA Methylation Sequencing

DNA methylation can significantly modify gene expression and chromatin state (Tost, 2010). Whole-genome bisulfite sequencing (WGBS) is a well-established technology to detect cytosine methylation, which provides a comprehensive view of methylation patterns at single-base resolution across the genome (Stadler et al., 2011).

2.1.2.4 ChIP-sequencing

ChIP-Seq is a powerful method for genome-wide profiling of transcription factors binding sites and histone modification sites, which is essential for a full understanding of transcriptional regulation. ChIP-seq have been widely applied to study gene regulation events, which are involved in the regulation of various diseases and biological pathways, such as development, aging and cancer progression (Mundade et al., 2014).

2.1.3 Single-cell RNA-sequencing (scRNA-seq)

Traditional bulk RNA-seq can provide the averaged transcriptome of a tissue or a group of cells. However it is well known that cell-to-cell differences are common to all organisms, which is contributing to particular function or to improve survival rate (Altschuler and Wu, 2010), so the explosive development of recently scRNA-seq technology has made a big progress in the understanding of cellular heterogeneity and its implications. Typically, single-cells are first isolated by limiting dilution, micromanipulation, flow-activated cell sorting (FACs), and most popular recently, by microfluidics (Hwang et al., 2018). For the library construction, SMART-Seq (Switching mechanism at the end of the 5´-end of the RNA transcript sequencing) protocol was developed to generate full-length transcripts with high level of mappable reads at single-cell level (Ramsköld et al., 2013), so it is suitable for the discovery of both alternative-splicing events and allele-specific expression. With the extensive optimization for sensitivity, accuracy, and cost (Picelli et al., 2014), SMART-Seq protocol has become widely used (Tasic et al., 2018; Tirosh et al., 2016). As profiling of a large number of cells is under heavy demand, many protocols have sacrificed full-length coverage in favor of sequencing part of the transcripts. Droplet microfluidics technology enables the high-throughput generation of water-in-oil droplets (Brouzes et al., 2009), which enabled the sequencing of hundred cells at once with a dramatic reduction in cost of scRNA-seq. InDrop (Klein et al., 2015), Drop-seq (Macosko et al., 2015) and 10X Genomics Chromium (10X) (Zheng et al., 2017) are the three mainly used protocol for droplet-based scRNA-seq. Those systems adopt similar principles, where the primers are embedded in the beads contain cell barcodes ensuring that each read can be mapped back to their original cell, and unique molecular identifiers (UMIs) helping to reduce sequencing errors introduced by PCR amplification. However, they differ in the material of the bead, barcode length, and method to amplify the cDNA, therefore affect the characteristics of the technology (Zhang et al., 2019). Drop-seq use small hard brittle resin beads, which results in only quite small proportion of beads being encapsulated with cells. InDrop and 10X use deformable hydrogel beads, which in turn enable beads being encapsulated to reach 80%. For cDNA amplification, inDrop uses the protocol as CEL-seq (Hashimshony et al., 2012), while10X and Drop-seq adopts the template-switching protocol like Smart-seq (Ramsköld et al., 2013).

2.1.4 Analysis of the high-throughput sequencing data

Here we mainly focus on software or methods to analyze the RNA-seq data, including mRNA-seq, miRNA-seq, and scRNA-seq. The analysis includes quality control, read alignment, quantification

of gene and transcript levels, visualization of the results, differential gene expression, functional analysis, promoter analysis, as illustrated in Figure 2.2.
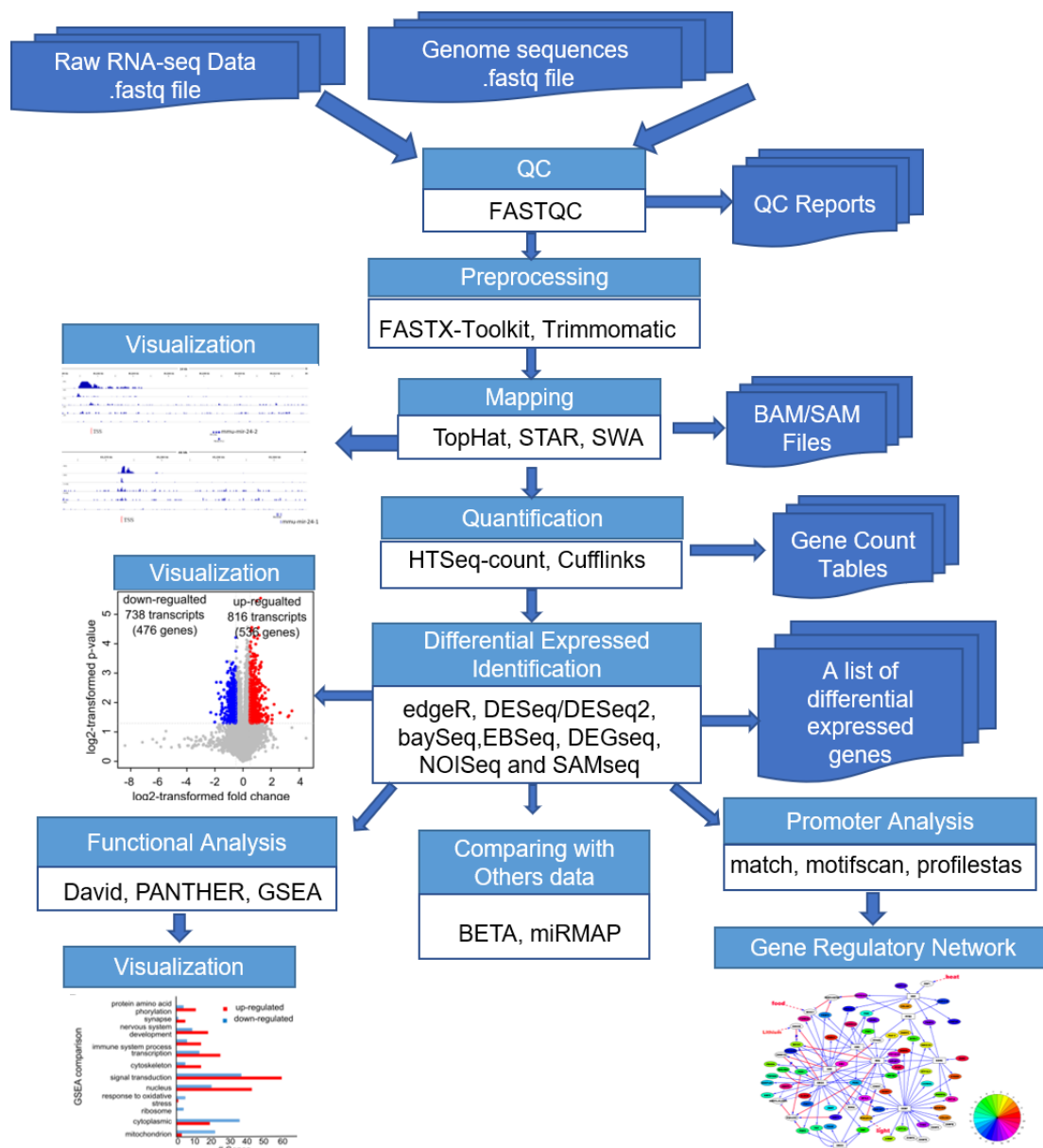
Figure 2.2 Workflow of the RNA-seq data analysis. The software used for each step is indicated in the white boxes.

2.1.4.1 Quality control and reads mapping

FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) is the most popular tool to perform quality control of Illumina reads, including the analysis of sequence quality, GC content, the presence of adaptors, overrepresented sequence, overrepresented k-mers, and duplicated reads. Software tools such as the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and

Trimmomatic (Bolger et al., 2014) are then applied to pre-process the data, such as removing the adapts, the low quality reads or nucleotides.

After pre-processing, reads are mapped to either reference genomes or transcriptomes. For the mapping of RNA-seq data, TopHat/TopHat2 (Kim et al., 2013; Trapnell et al., 2009) and STAR (Dobin et al., 2013) are most commonly applied. They can both map the reads to the genome with or without junction annotation. TopHat is based on the ultra-high-throughput short read aligner-Bowtie/Bowtie2 (Langmead and Salzberg, 2012; Langmead et al., 2009). First the RNA-seq reads are mapped to the genome using Bowtie/Bowtie2. Based on this initial mapping information, TopHat builds a database of putative splice junctions, then a second round of mapping is applied to confirm the junctions that were identified (Trapnell et al., 2009). While STAR is based on a novel RNA-seq alignment algorithm that uses seed search in uncompressed suffix arrays, that is to sequentially search for the maximum mappable prefix (MMP), followed by seed clustering and stitching. STAR is extremely fast comparing to other aligners (>50 times faster than TopHat), with the cost of high RAM usage (Dobin et al., 2013). The aligner will output all in a file in SAM/BAM format, which is a standard file format for saving the alignment information of the reads.

With the mapping result, the percentage of mapped reads is calculated, which could indicate the sequencing accuracy as well as the existing of sample contamination. For the model organisms, with well annotated genome, we expect more than 75% of RNA-seq reads map to the genome.

Genome browsers, such as UCSC browser (Raney et al., 2013), Integrated Genome Browser (IGB) (Freese et al., 2016) and Integrative Genomics Viewer (IGV) (Robinson et al., 2011), have been widely used to visualize the RNA-seq data. They all support the visualization of the SAM/BAM file, bed file, and also the processed coverage file such as BedGraph and BigWig files.

2.1.4.2 Quantification of the RNA expression

Quantification of the gene and transcript expression from RNA-seq data is based on the number of reads mapped to them. The most straight-forward approach is to sum up all the mapped reads using programs such as HTSeq-count, which produces a file containing the counts per gene. Given a SAM/BAM file and a gene transfer format (GTF) file, HTSeq-count counts the number of reads aligned to the exons for each gene, and discards multireads to avoid the possible errors that may be introduced in the differential expression identification. HTseq-count is implemented in STAR. There are also algorithms that calculate the counts on transcript level, which need to take into account the problem that transcripts from the same gene share most of their reads. Cufflinks (Roberts et al., 2011; Trapnell et al., 2010), for example, use a statistical model to estimate transcript abundance based on the assumption that the probability of observing a fragment is proportional to the abundances of the transcripts that the fragment belongs to (Trapnell et al., 2010). Besides quantification of the transcripts using the annotation from GTF file, Cufflinks can also identify novel transcripts from the mapping data alone (Trapnell et al., 2010).

Prior to comparing expression levels between samples, raw read counts need to be normalized to remove unwanted variation. Reads per kilobase of exon model per million reads (RPKM),

fragments per kilobase of exon model per million mapped reads (FPKM), and transcripts per million (TPM) are the most commonly used normalization methods. They all take the gene length and the library size into consideration (Mortazavi et al., 2008). RPKM is the read counts divided by the total number of reads and the length of a gene. FPKM is quite similar to RPKM. The only difference is that FPKM takes pair-end sequencing into account, and a pair of reads are mapped to one fragment. Instead of dividing by the total number of sequenced reads, TPM is calculated by dividing the reads counts by the total number of length-normalized reads, which result in the same sum of TPMs among samples, therefore makes it easier to compare across samples (Abbas-aghababazadeh et al., 2018). There are also methods that normalize the RNA-seq samples by comparing the distributions of the total or effective reads counts between samples. Those methods are implemented in the RNA-seq analysis toolkits, such as DESeq/DESeq2 (Anders and Huber, 2010; Love et al., 2014), Cufflinks (Trapnell et al., 2010), and edgeR (Robinson et al., 2010). DESeq uses the median ratio methods (Anders and Huber, 2010) to scale samples. CuffDiff implements both median ratio methods as DESseq and quartile normalization (Trapnell et al., 2010). EdgeR implements the trimmed mean of the M-values (TMM) approach (Robinson and Oshlack, 2010), that is the read counts are normalized by a weighted trimmed mean (Robinson et al., 2010).

2.1.4.3 Identify differentially expressed genes

In general, the methods for the identification of differentially expressed genes fall into two categories: parametric and non-parametric. Parametric methods assume that the expression value for each gene follows a particular distribution. The most widely used approach were initially *t* tests for comparison of two groups, and ANOVA (analysis of variance) for the comparison of multiple groups or factors. They both assume that the expression data follow a normal distribution. However, *t*-statistics has been criticized by researches because those genes that have a very low variance but no obvious expression changes will be falsely selected as differentially expressed (Jeanmougin et al., 2010). Therefore, numerous alternative methods were developed to model the expression variance. The two most popular RNA-seq data analysis software, edgeR (Robinson et al., 2010) and DESeq/DESeq2 (Anders and Huber, 2010; Love et al., 2014), adopt the negative binomial model as the main approach. baySeq (Hardcastle and Kelly, 2010) and EBSeq (Leng et al., 2013) apply an empirical Bayesian approaches, which are also based on the negative binomial model. DEGseq is based on the Poisson distribution, which is more suitable for studies that compare two samples with no or few replicates (Wang et al., 2009).

Although parametric tests are normally more powerful, they may lead to unreliable results if the assumed distribution does not hold. Non-parametric methods, on the other hand, do not require distributional assumptions. Mann-Whitney test and Wilcoxon rank-sum test are the two most widely used non-parametric methods, which compare two samples by using the rank order of the expression data rather than the expression values themselves. Non-parameteric methods have also been implemented in several RNA-seq data analysis toolkits. SAMseq (Li and Tibshirani, 2013) uses Wilcoxon rank-sum test. Other software tools, such as NOISeq implements an empirical approach to model the noise of the counting data (Tarazona et al., 2011, 2015).

2.1.4.4 Small RNA-seq analysis

Small RNAs (sRNAs) are usually 18–34 nucleotides in length, including microRNAs (miRNAs), short-interfering RNAs (siRNAs), PIWI-interacting RNAs (piRNAs), and other classes of regulatory molecules. Typically, sRNA-seq libraries were sequenced with relatively low sequence depth (around 2-10 million reads) due to the low complexity of sRNAs. Bioinformatics analysis of sRNA-seq are quite similar as standard RNA-seq, but with slight modifications. First, adaptor sequences for sRNA-seq are trimmed. Then length distribution of the trimmed reads is calculated, which is a good indicator of the sequence quality. For miRNAs in animals, they are usually peak at 22 and 23 nucleotides, whereas in plants the peaks are at 21- and 24-nucleotide. For piRNA the peak is around 20 to 40 nucleotides. After that, the pre-processed reads were aligned to a reference genome or transcriptome sequences using standard RNA-seq analysis tools, such as Bowtie2 (Langmead and Salzberg, 2012), STAR (Dobin et al., 2013), or BWA (Li and Durbin, 2009). The mapping of sRNA-seq is normally performed without mismatches due to the extremely short read length of sRNA-seq. There are specific tools for miRNA-seq analysis, such as miRDeep (Friedländer et al., 2011) and miRDeep-P (Yang and Li, 2011). In the case of animal and plant miRNAs, respectively, these tools have been largely used to quantify known and novel miRNA from sRNA-seq data.

2.1.4.5 Single cell RNA-seq data analysis

Once scRNA-seq reads are obtained from Illumina sequencer, quality control (QC) and pre-processing is performed. Low-quality nucleotides and adapter sequences will be removed at this step. For the droplet based scRNA-seq data, next step, called demultiplexing, is assigning reads to the corresponding cells according their cell barcodes and the corresponding mRNAs according the UMI. After demultiplexing, the reads are mapped to the genome using BWA (Li and Durbin, 2009) or STAR (Dobin et al., 2013). This results in the gene expression matrix where each row corresponds to a gene and each column corresponds to a cell. 10X Genomics provided Cell Ranger (Zheng et al., 2017) pipeline and Drop-seq team provided a Java tools to pre-process the data and generate the gene expression matrix (Macosko et al., 2015). Then the matrix is normally loaded into a R (R Core Team, 2019) toolkit for single cell genomics (Seurat) (Butler et al., 2018; Stuart et al., 2019), the most popular scRNA-seq analysis package in R, for clustering, marker gene identification, differential expression, and comparison with other datasets. The low-quality cells have to be removed before any downstream analysis, including the cells with small number of expressed genes and high expression of mitochondrial genes that are quite likely apoptotic cells, and the cells with extremely high number of reads and detected genes that may represent doublets. Genes expressed in quite small fraction of cells are also removed. The workflow of Seurat is as follows. First, the raw expression counts are normalized by scaling the number of reads across cells. Then Seurat calculates highly variable genes and focuses on these for downstream analysis. The data are scaled to remove unexpected sources of variation before clustering the data. For dimension reduction, a widely used unsupervised linear transformation technique, principal component analysis (PCA), is applied. Then T-distributed Stochastic Neighbor Embedding (t-SNE), or more recently, uniform manifold approximation and projection

(UMAP) (McInnes et al., 2018) are applied to visualize and summarize the data. Clusters are then obtained by grouping cells based on the expression similarity (distance) calculated from PCA analysis using K-Nearest Neighbour approach (KNN graph). A series of differential expression tests are implemented in the FindMarker function of Seurat, such as Wilcoxon rank sum test, likelihood-ratio test, standard AUC classifier, student's t-test, test based on the negative binomial distribution as DESeq2, and so on, for marker gene identification and differential expression analysis. Seurat also implemented many methods to integrate diverse types of single-cell data from different labs, technologies, platform, and even species (Stuart et al., 2019).

2.1.4.6 Function and Pathway analysis

Characterization of the biological functions or pathways that DEGs are involved in is critical for understanding and interpreting the transcriptome data. First, function and pathway analysis require functional and pathway annotation databases. Gene Ontology (GO) (Ashburner et al., 2000; The Gene Ontology Consortium, 2018) and KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2015), are the two most comprehensive database for function and pathway annotation, respectively. Bioconductor (Huber et al., 2015), DAVID (Huang et al., 2008; Toyama et al., 2009), PANTHER (Mi et al., 2009; Thomas et al., 2003), and gene set enrichment analysis (GSEA) (Mootha et al., 2003; Subramanian et al., 2005) are the most commonly used tools for function and pathway enrichment analysis. They all include the annotation from GO, KEGG, and other gene annotation databases such as protein domains, protein interactions, and so on. They are different with regard to the statistical methods used for the enrichment analysis. Bioconductor (Huber et al., 2015), PANTHER (Mi et al., 2009; Thomas et al., 2003) and DAVID (Huang et al., 2008; Toyama et al., 2009) compare the DEGs with the rest of the transcriptome for over representation in annotation terms, while GSEA (Mootha et al., 2003; Subramanian et al., 2005) is based on the ranking of the genes according to the significance of difference in expression.

2.1.4.7 Promoter analysis and gene regulatory network

It is well-known that transcriptome changes are achieved primarily by transcription regulation, which depends on the interaction between TFs and *cis*-acting sequence elements in promoter regions. Hence, it is of great importance to investigate the TFs that regulate transcriptome changes. Promoter analysis, which is referring to prediction of regulatory elements on the gene promoter, is normally applied for that purpose. First promoter region were extracted based on the transcription start sites (TSSs) information from databases such as DataBase of Transcriptional Start Site (DBTSS) (Suzuki et al., 2018; Yamashita et al., 2011), the Cap-Analysis Gene Expression (CAGE) database of the Fantom3 project (Takahashi et al., 2012), and the NCBI RefSeq database (O'Leary et al., 2016). Then there are many algorithms, such as Match (http://gene-regulation.com/pub/programs/match/help_match.html) and MotifScan (http://bioinfo.sibs.ac.cn/shaolab/motifscan/index.php), that can be applied to scan for hits of the positional weight matrix (PWM, a matrix of weights to represent the composition of a motif) to identify potential transcription factor binding sites (TFBS) on the extracted promoter

regions. The known vertebrate PWMs are well-annotated in TRANSFAC database (Matys et al., 2006). With the help of ChIP-seq technology, the predicted TFs can be verified by TF Chip-seq experiments. By combining the promoter analysis result and TF ChIP-seq data, one can construct a gene regulatory network for specific temporal, spatial, and condition, which can be further analyzed and visualized by Cytoscape (Shannon et al., 2003).

2.1.4.8 Integration of multi-level HTS data

Here, integration has two meanings: one is integrating the same type of HTS data from different studies, which may allow for the identification of the molecular signatures with increasing accuracy and robustness compared to those obtained from individual dataset. Integration of this type means comparison of RNA-seq data across species, tissues, conditions, patients, and studies whereby integration can provide new insight to biological and clinical studies. For example, by integrating the RNA-seq data across different species and tissues, Sudmant et al. confirmed that the gene expression difference between homologous organs is much smaller than that between different organs from one species (Sudmant et al., 2015). R packages such as metaRNASeq (Rau et al., 2014), metaSeq (Dale et al., 2014) have been developed for the meta-analysis of RNA-seq data.

On the other hand, integration also means combining different types of HTS data for the same samples. Direct target of TFs can be inferred by combining the transcriptome data and ChIP-seq data as described in BETA (Wang et al., 2013). Integration of RNA-seq and miRNA-seq data will help to identify the regulatory effects of miRNAs on post-transcriptional regulation and construct the regulatory network for miRNA-mRNA, which is implemented in module analysis program (miRMAP) (Vejnar and Zdobnov, 2012). By integrating proteomics and RNA-seq, researchers can identify novel isoforms, and investigate whether genes are under translational control and also explore post-translational regulation (Haider and Pal, 2013).

2.2 Circadian rhythms

Almost all organisms are synchronized to the light and dark cycle of the environment. These daily rhythms are referred as 'circadian' and are driven by endogenous clocks. Circadian clocks will adjust to changes in the light and dark cycle and allow organisms to anticipate and prepare for the changes in the environment (Johnson et al., 2003). The circadian clock controls the rhythmicity of nearly all the physiological and behavioral processes, such as sleep and physical activity, body temperature, blood pressure, circulating hormones, metabolism, brain wave activity, and so on (Buhr and Takahashi, 2013; Dibner et al., 2010). Circadian rhythm disruptions increase the incidence of various illnesses. Animal studies showed that circadian disruption increases tumor progression and impairs survival and immune function in young mice (Adams et al. 2003, Brager et al., 2013). Epidemiologic studies have found elevated rates of breast cancer in women and prostate cancer in men whose circadian rhythms were disrupted as a result of working in the rotating day and night shifts (Sigurdardottir et al., 2012; Wang et al., 2011).

2.2.1 Molecular basis of the circadian clock

At the organismal level, circadian physiology is created by a hierarchical network of central and peripheral clocks (Lowrey and Takahashi, 2004; Mohawk et al., 2012). In mammals, the central clock, which is also called the "master clock", resides in suprachiasmatic nuclei (SCN) of the hypothalamus. The SCN receive the light signal from retinal ganglion cells (ipRGCs) by means of the retinohypothalamic tract (RHT) (Doyle and Menaker, 2007), and synchronizes clocks in peripheral tissues through neural and humoral output signals (Lowrey and Takahashi, 2004; Mohawk et al., 2012). In non-mammalian vertebrates like zebrafish, the pineal complex is a photoreceptive organ that directly receives the light input. It contains an intrinsic circadian oscillator, which can drive the circadian expression of genes and the rhythmicity of the production of melatonin.

At the molecular level, the mechanism of the circadian clock depends on transcriptional/translational feedback loops that involve set of evolutionarily conserved core circadian regulators (Figure 2.3). This molecular clock not only characterizes the master clock but by and large is replicated in the so-called "peripheral" clocks present in all tissues and organs of the body (Yoo et al., 2004). In vertebrates, the core circadian regulators are BMAL1, CLOCK, PERs, CRYs, REV-ERBα (NR1D1), REV-ERBβ (NR1D2), and DBP (Reppert and Weaver, 2001; Takahashi et al., 2008). TFs BMAL1 and CLOCK form a complex that positively regulates the transcription of *Per* and *Cry* genes through activating the cis-regulatory element E-box in *Per* and *Cry* promoters. PER and CRY in turn form a complex inhibiting BMAL1/CLOCK transcriptional activity (Reppert and Weaver, 2001; Takahashi et al., 2008). Degradation of PER and CRY is required to stop this negative feedback and allows the restart of a new cycle of transcription the next day. Casein kinase 1ε/δ (CK1ε/δ) was found to phosphorylate the PER protein, then the phosphorylated PER is ubiquitylated by βTrCP and degraded by the 26S complex (Eide et al., 2005; Lowrey et al., 2000; Shirogane et al., 2005; Vanselow et al., 2006). It is thought that in the PER-CRY complex, CRYs are the major repressors (Ye et al., 2014). CRYs are ubiquitylated by SCF (Skp1-Cul1-FBXL3) complex and degraded by proteasome pathway (Xing et al., 2013). This core molecular clockwork is fine-tuned by intertwined negative/positive feedback loops. CLOCK/BMAL1 dimers binds to the E-box of the orphan nuclear-receptor genes *Rev-Erbα/ß* (*Nr1d1/2*) and *RORα/ß*. The TF proteins REV-ERB and ROR then compete for retinoic acid-related orphan receptor response element (RORE) within the promoter of *Bmal1* where ROR proteins initiate *Bmal1* transcription and REV-ERB proteins inhibits it (Guillaumond et al., 2005; Preitner et al., 2002). DBP and NFIL3, two TFs which binds the D-box element, have also been shown to play an important role in the regulation of mammalian circadian rhythm (Yamaguchi et al., 2000).
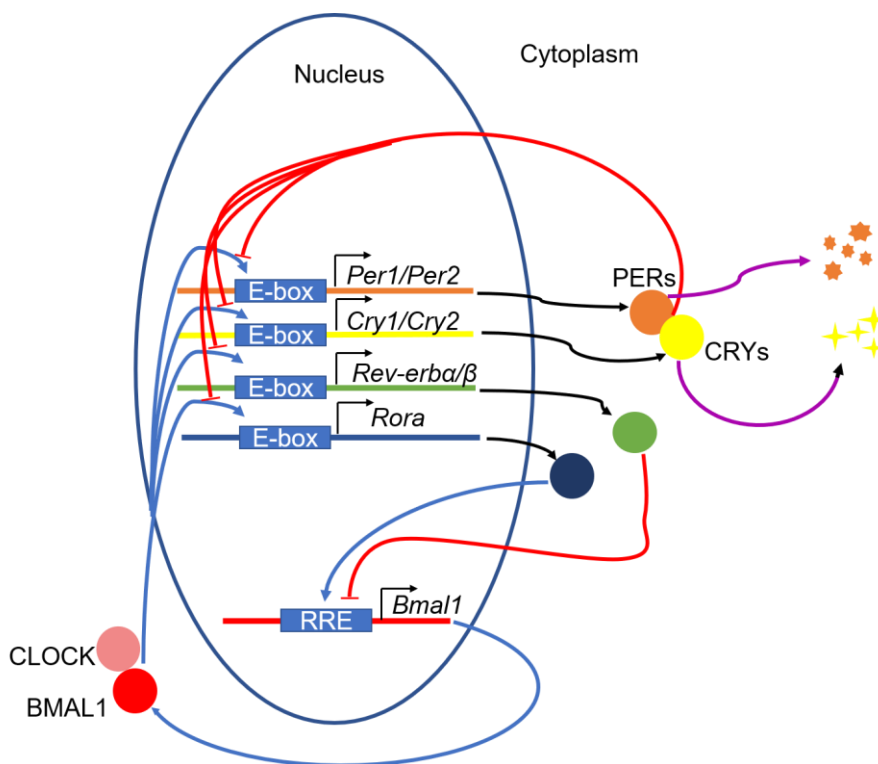
Figure 2.3. Molecular mechanisms of the circadian clock in mammals. BMAL1 and CLOCK form a hetrodimeric complex that binds DNA of clock targets at E-boxes and positively regulates target transcription. The resulting PER and CRY, in turn, form a complex inhibiting BMAL1/CLOCK transcriptional activity. Two other clock proteins, REV-ERBs and ROR, bind to the ROR/REV-ERB-response element (RORE) binding site on the *Bmal1* promoter and repress and activate its expression, respectively. Repression is terminated by the degradation of PER and CRY and a new cycle of transcription can start then.

2.2.2 Post-transcriptional regulation of the circadian clock

Post-transcriptional regulation plays an important role in mRNA metabolism, and includes splicing, RNA modification, maturation, subcellular localization, and stability regulation (Keene, 2007). By applying Nascent-seq in mouse liver, Menet et al. observed that only less than one third of the rhythmic mRNAs are rhythmically transcribed (Menet et al., 2012). Other studies have looked at circadian proteomics and found that the mRNAs for more than 50% of the circadian oscillating proteins are not rhythmic (Robles et al., 2014). Later on, Wang et al. estimated that about one-third of rhythmically accumulating mRNA are under the regulation of rhythmic mRNA degradation. These investigators applied a mathematical model to systematically investigate the contributions of transcriptional and posttranscriptional regulation (Wang et al., 2018). Emerging evidence suggests that many, if not all, post-transcriptional steps were under the regulation of circadian clocks (Green, 2018; Kojima et al., 2011; Lim and Allada, 2013; Preußner and Heyd, 2016) (Figure 2.4). It was hypothesized that post-transcriptional control of circadian clocks enhanced robustness and adaptability to different environments (Lim and Allada, 2013). As

illustrated in Figure 2.4, mRNA-binding proteins (RBPs) and microRNAs (miRNAs) are the two main trans-acting factors involved in the post-transcriptional regulation of the circadian clock.
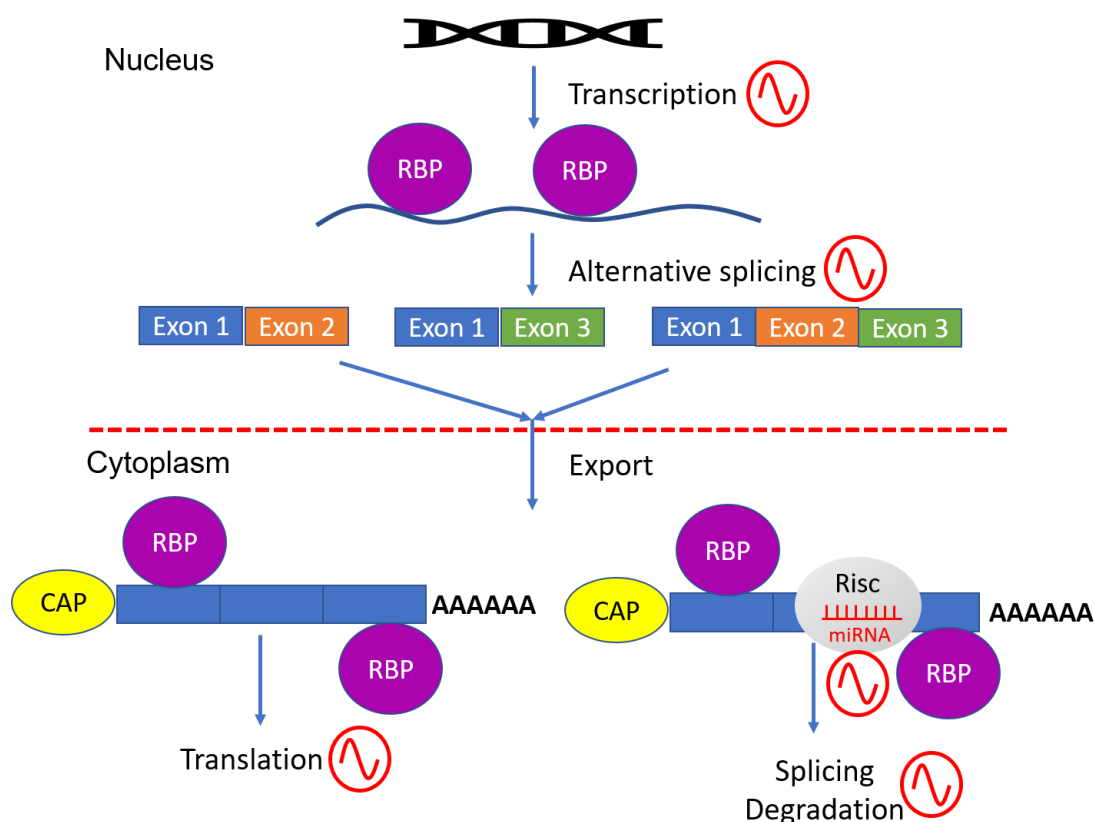
Figure 2.4 Post-transcriptional steps under the regulation of circadian clocks. In the nucleus, the pre-mRNAs are transcribed, followed by polyadenylation and splicing. Alternative splicing can give rise to different mature mRNAs from the same pre-mRNA. After export to the cytoplasm, mRNAs are translated or degraded. Red oscillators symbols indicate the steps that are regulated by circadian clock. How RBPs and miRNAs are involved in the circadian regulation of those steps is illustrated.

2.2.2.1 RBPs in post-transcriptional regulation of the circadian clock

RBP proteins bind to the double or single strand RNA and play a major role in post-transcriptional control of RNAs (Glisovic et al., 2008; Keene, 2007). Then, more and more studies pointed out the importance of RBPs in mediating posttranscriptional regulation in the mammalian circadian clock (Green, 2018; Kojima et al., 2011; Lim and Allada, 2013; Preußner and Heyd, 2016). RBM4 (also known as Lark), exhibiting circadian expression in the nervous systems of both flies and mice (Kojima et al., 2007; Zhang et al., 2000), can rhythmically promote the translation of *Per1* by binding to the 3′UTR of *Per1* in mice (Kojima et al., 2007) . HnRNP D, HnRNP Q and PTB, all belonging to the HnRNP family, are another type of RBPs that are implied in post-transcriptional regulation of circadian clock. HnRNP D promotes the degradation of *Cry1* by binding to a U-rich region of *Cry1* 3'UTR (Woo et al., 2010). hnRNP Q binding to both 5′ and 3′

UTR of *Per3* gene, can promote the translation of PER3 protein and acceleration *Per3* mRNA decay (Kim et al., 2011). PTB can bind the 3'UTR region of *Per2* and promote its degradation (Woo et al., 2009). PTB and HnRNPQ can both bind to the ribosomal entry site (IRES) region in the 5' UTR of *Rev-erbα* and promote its translation (Kim et al., 2010). Paraspeckles, built on long coding RNA *Neat1* with RBPs such as SFPQ, NONO, RBM14, and PSPC1, are thought to bind and prevent certain mRNAs from leaving the nucleus in a circadian manner (Torres et al., 2016). In addition, CIRBP and RBM3, two cold-induced RNA binding protein, were shown to modulate circadian gene expression post-transcriptionally (Liu et al., 2013; Morf et al., 2012). Neurodegeneration-associated RNA/DNA binding protein, FUS, was identified as a novel modulator of circadian gene expression, which established a link between circadian control and neurodegeneration-associated proteins (Jiang et al., 2018).

2.2.2.2 miRNAs in post-transcriptional regulation of circadian clock

miRNAs are a class of small non-coding RNA molecules, which can regulate the expression of target genes post-transcriptionally by mRNA degradation or by translational inhibition (Valinezhad et al., 2014). The biogenesis of miRNAs involved multiple steps. First miRNA genes are transcribed by RNA polymerase II (Pol II) into primary transcripts in the nucleus. Then the primary transcripts are cleaved into miRNA precursors by Drosha and Pasha. These precursors are in turn translocated into the cytosol and processed into mature miRNAs. The roles of miRNA in circadian regulation have been reported for many species, such as fruit fly, mouse, human, and chicken (Mehta and Cheng, 2013). In mammals, miR-219 and miR-132, two brain-specific miRNAs, regulated by the circadian oscillator in the SCN, affecting the circadian period and involving in the photic entrainment of SCN, respectively (Cheng et al., 2007). Using a forward genetic screen, Nagel et al. found that miR-192/194 cluster could target *Per* gene family and inhibit their expression (Nagel et al., 2009). Gatfield *et al.* claimed that miR-122, a liver specific miRNA, is regulated by REV-ERBα and associate with circadian control of hepatic metabolism, although the expression of mature miR-122 is constant (Gatfield *et al.*, 2009). Later, Kojima et al. showed that miR-122 can regulate the rhythmic expression of Nocturnin, which is a deadenylase under the control of circadian clock and plays an essential role in the posttranscriptional control of circadian rhythm (Kojima et al., 2010). In addition, a report by Tan et al. showed that miR-142-3p directly targeted *Bmal1* and its expression was in turn under the regulation of BMAL1/CLOCK (Tan et al., 2012). Chen et al. showed that miRNAs, including miR-24, miR-29a, and miR-30a, regulates *Per1/2* expression and are essential in generating a time delay that are critical for the circadian oscillator (Chen et al., 2013). Further *in vivo* experiments also revealed that miR-24, and slightly miR-30, suppressed PER2 protein translation (Yoo et al., 2017). However, they claimed that those miRNAs themselves are not rhythmic. Genome-wide profiling has also been applied to identify the circadian miRNAs. Xu et al. identified a retina-specific circadian miRNAs cluster including miR-96, miR-182 and miR-183 (Xu et al., 2007). Na et al. identified 85 liver circadian miRNAs in mouse using miRNA microarray (Na et al., 2009). Vollmers et al. identified 30 liver circadian miRNAs in mouse using miRNA-seq (Vollmers et al., 2012). However, the overlap of circadian miRNAs identified from those two mouse liver datasets is very small, which raised the questions about how to search for the miRNAs involved in regulating the circadian rhythm.

2.2.3 Development of the circadian clock

During animal development, circadian oscillators are gradually established at molecular, cellular and tissue levels. Here I shall review the advances in the understanding of the circadian clock development during ontogeny mainly on the cellular and molecular level in mammals and in zebrafish.

2.2.3.1 Development of circadian clock in mammals

In vitro studies in mammalian cell lines showed that no circadian oscillation can be detected in germ cells, stem cells, and induced pluripotent stem cells (iPSCs) (Alvarez et al., 2003; Morse et al., 2003; Yagita et al., 2010). The clock genes start to oscillate only if embryonic stem cells (ESCs) or iPSCs differentiate (Umemura et al., 2017, 2019; Yagita et al., 2010). These studies all imply that molecular circadian oscillations gradually appear during development. By *ex vivo* luciferase recordings of *Per2* expression in mouse fetal SCN, Carmona-Alcocer et al. found that only a few cells start oscillations at E14.5, while by E15.5 almost all of the recorded cells show synchronized circadian oscillation. By postnatal day P2, a dorsal-ventral phase wave of *Per2* expression was established (Carmona-Alcocer et al., 2017). However, they did not observe SCN cells become circadian spontaneously *in vitro*. *In vivo*, the daily rhythms of rat SCN were first detected between E19 and E21 (Sládek et al., 2004). In the mouse SCN, the daily rhythms of *Per1* transcript levels were detected at E17, while the rhythmic protein expression of PER1 and PER2 were detected by E18 (Ansari et al., 2009; Shimomura et al., 2001). Studies of SCN development showed that TFs, such as *Rora, Lhx1*, *Shh*, *Six3*, and *Six6,* participate in the embryonic specification, neurogenesis, and neuronal differentiation of the SCN (Bedont and Blackshaw, 2015; Landgraf et al., 2014). However, the study of circadian clock development has significantly lagged behind, though recent studies have begun to shed light on the topic. Using ESCs and iPSC, Umemura et al. showed that posttranscriptional control of CLOCK protein contributing to the starting of the circadian oscillating in both mouse and human cells (Umemura and Yagita, 2020; Umemura et al., 2017, 2019).

2.2.3.2 Development of circadian clock in zebrafish

Zebrafish (*Danio rerio*) is a well-established model organism for the study of animal development. Due to its transparent nature, it is also a suitable system for *in vivo* live imaging. Luciferase-based circadian reporter systems have been constructed and widely used to study the cellular clock both *in vivo* and *ex vivo* (Carr and Whitmore, 2005; Dekens and Whitmore, 2008; Kaneko and Cahill, 2005; Ono et al., 2013). But luciferase system has its limitation in gaining single-cell resolution *in vivo*.

In zebrafish, a mature circadian clock is only established after hatching (Vallone et al., 2007). It has been shown that early exposure to LD cycle is essential for the development of circadian rhythm in zebrafish (Dekens and Whitmore, 2008; Kaneko and Cahill, 2005; Ziv et al., 2005). However, it is yet unclear whether the onset of circadian clock *in vivo* is a synchronization of the

existing cellular oscillators or a *de novo* initiation of the cellular clock. *In vitro* study by Carr and Whitmore showed that zebrafish cell line maintained for two-month in constant darkness (DD) became desynchronized and that light can synchronize the clocks of these cells (Carr and Whitmore, 2005). However, *in vitro* studies may not replicate the conditions inside the organism. By whole mount fluorescent *in situ* hybridization of the *Per1* gene, Dekens and Whitmore observed that under DD condition, zebrafish siblings collected at both CT3 (the peak time of *Per1* expression under LD condition) and CT15, have similar number of *Per1* expressed cells (Dekens and Whitmore, 2008). In line with their observation, Ziv et al. found that under DD condition, the expression of *aanat2* in the zebrafish embryos is constant and the value is an average of the expression value under LD condition, so they proposed that under DD condition the expression of *aanat2* is oscillating in each cell, but not synchronized (Ziv et al., 2005). However, these studies did not trace the single cell from the same fish over time, so their observations could also be explained by the highly variation in baseline expression among different cells as we observed. Taking together, an *in vivo* single-cell study of the effect of light on the clock development is urgently needed.

3. Cumulative part of the Dissertation

3.1 Oscillating primary transcripts harbor miRNAs with circadian functions

Reference: **Wang, H.,** Fan, Z., Zhao, M., Li, J., Lu, M, Liu, W., Ying, H., Liu, M., Yan, J,. (2016) Oscillating primary transcripts harbor miRNAs with circadian functions. Sci. Rep. 6, 21598

3.1.1 Synopsis

Gene regulatory circuits representing and driving circadian rhythms have been well studied at the transcriptional level. However, recent advance suggests that post-transcriptional mechanisms play important roles in modulating circadian gene expression. Such modulation seem to be essential for circadian functions.

Several studies have indicated a significant role of miRNAs as post-transcriptional regulators of the circadian clock. However, the search for circadian miRNAs by miRNA-seq and microarray analysis in mouse liver has led to inconsistent results, which raised the question about how to reliably identify circadian miRNAs. Gatfield et al. found that miR-122, the most highly expressed miRNA in mouse liver, is regulated by NR1D1 (REV-ERB). miR-122 exhibits robust circadian oscillation at the primary transcript level but not as a mature miRNA. They showed that miR-122 can modulate circadian gene expression post-transcriptionally and participates in circadian regulation of metabolism in liver (Gatfield et al., 2009). Inspired by their observations, my colleagues and I sought to systematically search for circadian miRNA from both expression profiles of primary and of mature miRNAs.

To this end, I built a bioinformatics pipeline to analyze and integrate the published multi-omics data of mouse liver (Figure 3.1). First, by applying the Vespucci program (Allison et al., 2014), I constructed a comprehensive liver-specific primary miRNA transcript database using strand-specific circadian global run-on sequencing (GRO-seq) data of mouse liver (Fang et al., 2014). Then mouse liver-specific ChIP-seq of promoter marker (H3K4me3) and Pol II marker were used to validate and fine-tune the annotation on transcriptional start sites of those miRNA primary transcripts. The sequence reads of two liver-specific circadian nascent RNA transcriptome datasets, Nascent-seq (genome-wide sequencing of nascent RNA) (Menet et al., 2012) and GRO-seq (Fang et al., 2014), were then mapped to mouse genome by Bowtie2 and the expression value of each miRNA primary transcripts were quantified and normalized to RPKM by a customized R script. The circadian miRNAs transcripts were identified by fitting the data to cosine functions with 24h period and shifting phases. Using summed log2-tranformed cosine fitting p-value>9 in the two datasets as cutoff, I identified 57 circadian miRNA primary transcripts in mouse liver, which includes the primary transcripts of miR-122 and miR-24 that have been previously reported to be involved in the regulation of circadian clock. Six of them were selected for validating by my colleagues using qPCR. They showed that all the six primary transcripts were significant

oscillating and in the same phase as I calculated from the transcriptome data. Furthermore, by integrating all the published ChIP-seq data of core circadian regulators of mouse liver, I found that 56 out of 57 circadian miRNA primary transcripts are under the regulation of one or more core circadian TFs, such as BMAL1/CLOCK and NR1D1/NR1D2. I, with the help of my colleagues, proved that BMAL/CLOCK binds on the promoter of all the three selected circadian primary transcripts in mouse liver by ChIP-PCR. To further investigate if the physically binding inferred from TF's ChIP-seq data on those primary transcripts are really functional, my colleagues and I selected five of them to quantify their expression in liver-specific *Bmal1* cKO mice by PCR. I observed that four out of five primary transcripts were significantly under-expressed in *Bmal1* cKO mice compared to the control mice. This indicated that the binding of BMAL/CLOCK on those primary transcripts is functional. Taken together, we identified 57 circadian miRNA primary transcripts and most of them are under the control of one or more core circadian regulators. However, when I checked the miRNA expression at mature level by analysis of the published circadian miRNA-seq data, I observed that the mature miRNAs from the 53 circadian primary transcripts were not oscillating, only four mature miRNAs from those circadian primary transcripts were oscillating but with relatively low amplitudes comparing to their primary transcripts. This could explain why there is hardly any overlap among different circadian miRNA transcriptome studies. It is reported that miRNAs have much longer half-live comparing to mRNAs (Gantier et al., 2011), which may explain why most of the circadian miRNAs primary transcripts do not lead to the oscillating mature miRNAs.
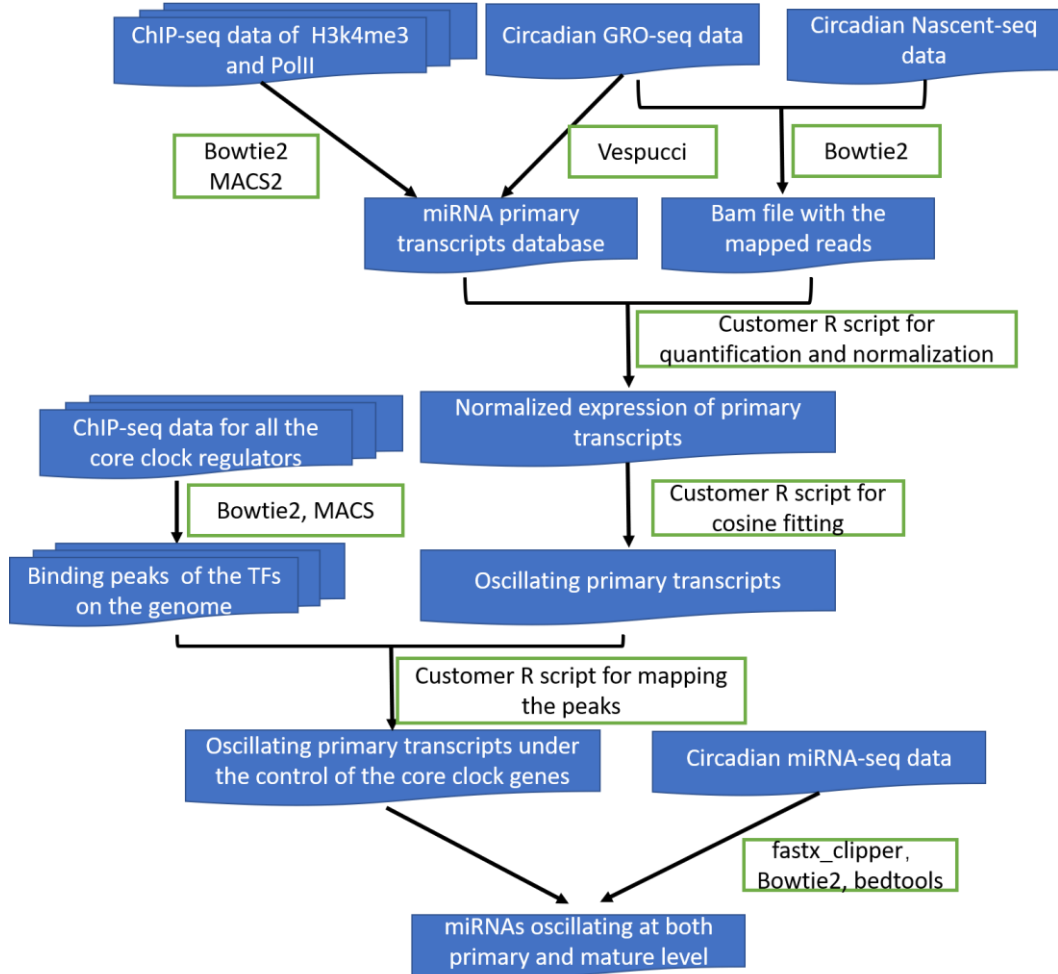
Figure 3.1 A bioinformatics pipeline to analyze and integrate the published multi-omics data of mouse liver to search for the circadian miRNAs. The software and programs used are named in the boxes with green border.

Among the four miRNAs that exhibited strong oscillating as primary transcripts and relatively lower but significant oscillation as mature transcripts, miR-122 has been reported to be involved in the circadian regulation of hepatic metabolism. Moreover, miR-24 was shown to regulate *Per* gene expression and shorten the circadian period. Those studies suggest that miRNAs rhythmic at primary transcript level could also participate in circadian regulation. I wondered if other miRNA showing significant oscillating as primary transcripts play a role in circadian regulation. To address this question, I selected miR-378, derived from pri-miR-378 transcript, for further study, since this miRNA had not yet been linked to circadian regulation. To study miR-378 function *in vivo*, my colleagues and I overexpressed miR-378 by injecting adenoviral miR-378/miR-378* as reported in Liu et al. (Liu et al., 2014). Briefly, DNA fragments of the miR-378 hairpin precursor were inserted into an adenovirus vector, which can overexpress both miR-378 (the dominant product of miR-378 precursor) and miR-378* (the minor product of miR-378 precursor) as much as 50-fold in mouse liver by tail vein injection. Then RNA-seq was applied to measure global gene expression profiles evoked by miR-378 over-expression. This analysis was carried out at CT10 and CT22 that correspond to the expression peak and trough of miR-378. It is well-known that the genes peaking at those two time points are mainly regulated by BMAL/CLOCK and

NR1D1/NR1D2, the two sets of core clock regulators, respectively (Fang et al., 2014). I use two-way ANOVA to examine the effect of two factors, that is the circadian sampling points (CT10 vs. CT22) and treatment (adenovirus vs. null virus). Using ANOVA p-value for treatment < 0.05 as cutoff, I found an eight-fold over-representation of the downregulated genes among those that are affected by ectopic miR-378, which is in accordance with the inhibiting role of miRNAs. By integrating with the miRNA targets databases, I observed a significant enrichment of the putative miR-378 targets in the cohort of downregulated genes. Functional enrichment analysis of the downregulated genes showed that cell cycles genes were mostly enriched, which agrees with the extensive studies on the effects of miR-378 in cell cycle and apoptosis (Wang et al., 2015). Taken together, adenoviral injection indeed over-expressed miR-378, which then suppressed the expression of its targets in liver.

Except for cell cycle related genes, circadian genes were also enriched in the under-expressed genes resulting from miR-378 overexpression. Using ANOVA p-value for circadian sampling points less than 0.05 as cutoff, I identified 2266 genes showing significantly differential expression between CT10 and CT22. Comparing with the liver circadian phase data from our mouse circadian gene database (Yan et al., 2008), we observed that the genes showing higher expression at CT10 than CT22 in this study were mainly peaking around CT10 in published circadian data, while the genes showing higher expression at CT22 than CT10 were mainly peak around CT22. Therefore, in the following analysis, the genes with ANOVA p-value < 0.05 for circadian sampling points were defined as the circadian oscillating genes. Furthermore, using ANOVA p-value for both treatment and circadian sampling time points less than 0.05 as cutoff, I identified 691 genes, showing significantly differential expression between CT10 and CT22, were downregulated by miR-378 over-expression. They were defined as miR-378 circadian targets, including *Cry2*, which is involved in the main feedback loop of circadian regulation. Functional enrichment analysis of those genes showed that cell cycle and apoptosis are the mostly enriched terms in the circadian targets peak at CT10 and CT22, respectively. Then five cell cycle genes from the list of miR-378 circadian targets were selected and validated by my colleagues using qPCR. Consistent with the RNA-seq data, all of them were significantly downregulated by miR-378 over-expression. Two of them, *CCne1* requiring for cell cycle G1/S transition and *Cdkn1a* inhibiting G1/S transition, with circadian peak at CT10 and CT22 respectively, were both downregulated upon miR-378 over-expression, so miR-378 may be bifunctional in the circadian control of cell cycle.

As illustrated by Gatfield et al., the constant expression of miRNAs can affect the circadian expression of the target genes by constantly promoting their mRNA decay, which in turn results in decreased baseline expression and increased circadian amplitudes of their circadian targets (Gatfield et al., 2009). Therefore, I further identified 89 miR-378 circadian targets with increased relative circadian amplitude, defined as the log2-transformed fold change between CT10 and CT22, upon miR-378 over-expression. Those genes were defined as the targets that are under circadian regulation of miR-378. Among them, cell cycle related genes are significantly enriched as before, especially, *Cdkn1a* exhibited more than four-fold increasing in relative circadian amplitude upon miR-378 over-expression. So miR-378 can regulate the degradation rate of cell cycle genes, and in turn affect their relative circadian amplitudes. We noticed that around 80% of the miR-378 circadian targets do not show significant amplitude changes, which may due to the

fact that they already are strongly regulated by miR-378 so that over-expression of miR-378 can hardly enhance their mRNA decay further.

Furthermore, by integrating with the published TFs's binding sites data from ChIP-seq and functional data from TF knockdown/knockout RNA-seq or microarray, I investigated how miR-378 crosstalk with the core circadian TFs to regulate circadian gene expression. As the circadian genes peaking at CT10 and CT22 are more likely to be regulated by BMAL/CLOCK and NR1D1/2, respectively (Fang et al., 2014), I focused on the crosstalk between mi-378,BMAL/CLOCK and NR1D1/2. Two types of regulatory motifs were identified by module analysis. miR-378 cooperated with BMAL1/CLOCK forming an incoherent feed-forward loop, which regulates circadian genes including *Gadd45a* and *Por* peaking around CT10. While miR-378 cooperated with NR1D1/2 forming a coherent feed-forward loop, which regulates the circadian genes including *Cdkn1a* and *Bcl2* peaking around CT22. Our analysis implied that miR-378 mediates the circadian control of cell cycle (*Cdkn1a*) and metabolism (*Por*, involved in the lipid metabolism) by forming either coherent or incoherent feed-forward loops with different circadian transcription factors.
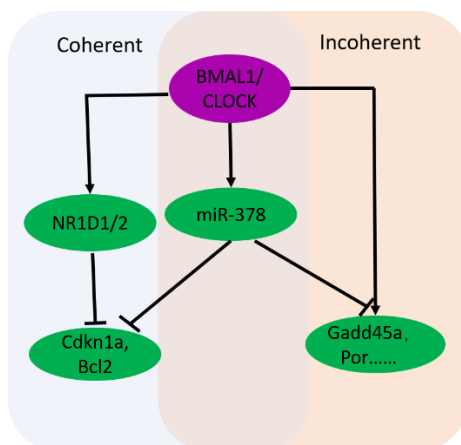
Figure 3.2 Two types of feed-forward loops that regulate circadian gene expression formed by core circadian TFs and miR-378. The color of the nodes represents their circadian peaks. Genes peak around CT10 are green, while genes peak around CT22 are purple. TFs, transcriptional factors.

In conclusion, my colleagues and I demonstrated that miRNAs derived from circadian miRNA primary transcripts can have circadian functions, which is a new strategy to search the miRNAs that may have circadian function. By a combination of system biology and experimental approaches, we showed that miR-378 derived from a circadian miRNA primary transcript involved in the circadian controlling of cell cycle and metabolism. It should be noted that for the transcriptome study the circadian sample were collected only at two time points, which could not provide us enough information about the effect of miR-378 over-expression on circadian genes. The effect of miR-378 over-expression on circadian period and circadian phase cannot be inferred, and the effect on circadian amplitude is also under-estimated. So, to fully investigate the circadian function of miR-378, collecting samples at more circadian time points is necessary in the future. In this study the over-expression of miR-378 is ectopic, so it will be valuable to do the RNA-seq

on the miR-378 knockout mice, which have been reported (Liu et al., 2014), to explore the circadian role of endogenous miR-378.

3.1.2 Declaration of my contribution

Conceptualization: Jun Yan, Haifang Wang.

Wet-lab experiments: Zenghua Fan, Meng Zhao, Juan Li, Minghua Lu, Wei Liu, Haifang Wang.

Data curation: Haifang Wang, Zenghua Fan.

Formal analysis: Haifang Wang, Zenghua Fan.

Investigation: Haifang Wang, Zenghua Fan, Jun Yan

Methodology: Haifang Wang, Jun Yan

Visualization: Haifang Wang

Writing – original draft: Haifang Wang, Jun Yan

Writing – review & editing: Haifang Wang, Jun Yan

SCIENTIFIC REP⚙RTS

Oscillating primary transcripts harbor miRNAs with circadian functions

Haifang Wang[1,2], Zenghua Fan[1,5], Meng Zhao[1,2], Juan Li[1], Minghua Lu[3], Wei Liu[4], Hao Ying[4], Mofang Liu[3] & Jun Yan[1,2]

The roles of miRNAs as important post-transcriptional regulators in the circadian clock have been suggested in several studies. But the search for circadian miRNAs has led to disparate results. Here we demonstrated that at least 57 miRNA primary transcripts are rhythmically transcribed in mouse liver. Most of these transcripts are under the regulation of circadian transcription factors such as BMAL1/ CLOCK and REV-ERBα/β. However, the mature miRNAs derived from these transcripts are either not oscillating or oscillating at low amplitudes, which could explain the inconsistency of different circadian miRNA studies. In order to show that these circadian primary transcripts can give rise to miRNAs with circadian functions, we over-expressed one of them, miR-378, in mouse by adenovirus injection. We found a significant over-representation of circadian oscillating genes under-expressed by miR-378 over-expression in liver. In particular, we observed that miR-378 modulates the oscillation amplitudes of *Cdkn1a* in the control of cell cycle and *Por* in the regulation of oxidation reduction by forming partnership with different circadian transcription factors. Our study suggests that circadian transcription of miRNA at primary transcript level can be a good indicator for circadian miRNA functions.

The daily changes in physiology and behaviors can be observed in almost all organisms. In mammals, a master clock located in the suprachiasmatic nucleus (SCN) of the hypothalamus drives circadian rhythms, adjusts itself according to the light input from the eyes, and synchronizes clocks in peripheral tissues such as the liver and kidney. At the molecular level, it has been shown that the negative transcriptional–translational feedback loops formed by a set of key circadian regulators (BMAL1, CLOCK, PER1, PER2, PER3, CRY1, CRY2, REV-ERBα, REV-ERBβ) are responsible for giving rise to the circadian physiology[1,2]. It is now known that transcriptional regulation is not the only way of circadian control. Comparing the nascent and mature mRNA transcripts on the whole genome level in mouse liver, Menet *et al.* claimed that only 28.4% of the rhythmic mRNAs are accompanied by rhythmic transcription[3]. Recently published circadian proteomic data showed that as much as 50% of the circadian oscillating proteins are derived from non-rhythmic mRNAs[4,5]. These observations strongly suggest that post-transcriptional regulation plays a substantial role in modulating temporal gene expression for proper circadian function.

MicroRNAs (miRNAs) are small non-coding RNA molecules that can regulate the expression of target genes by mRNA degradation or translational repression[6,7]. The roles of miRNAs in the circadian clock have been suggested in several organisms including fruit fly, mouse, human, and chicken[8]. In mammals, two brain specific miRNAs, miR-219 and miR-132, regulated by BMAL1/CLOCK and CREB proteins respectively, affect the circadian period and are involved in light dependent resetting of the clock in SCN[9]. MiR-192/194 cluster could inhibit the expression of *Per* family genes in HeLa cells[10]. MiR-142-3p directly targets *Bmal1* and is in turn regulated by BMAL1/CLOCK in the blood stream of human[11].

During miRNA biogenesis, miRNAs are first transcribed by RNA polymerase II (Pol II) in the nucleus from DNA into primary transcripts, which are subsequently cleaved by Drosha and Pasha into miRNA precursors

[1]CAS-MPG Partner Institute for Computational Biology, Shanghai 200031, China. [2]Institute of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai 200031, China. [3]Institute for Biochemistry and Cell Biology, Shanghai 200031, China. [4]Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [5]University of Chinese Academy of Sciences, Shanghai 200031, China. Correspondence and requests for materials should be addressed to J.Y. (email: junyan@picb.ac.cn)

of around 80 nt. Then the miRNA precursors were translocated into the cytosol by RAN-GTP, Exportin-5 and processed into mature miRNAs[6]. To date, there have been several studies using genome-wide profiling experiments to identify the circadian miRNAs at the mature level. Xu *et al.* identified a sensory organ-specific miRNA cluster including miR-96, miR-182 and miR-183, oscillating in the mouse retina in a circadian manner[12]. By microarray-based expression profiling of both miRNA and mRNA in mouse liver, Na *et al.* identified 85 liver circadian miRNAs[13]. Vollmers *et al.* identified 30 mouse liver circadian miRNAs by miRNA-sequencing (miR-seq)[14]. However, the overlapped circadian miRNAs between the two high-throughput profiling studies in mouse liver are limited. These discrepancies raised the question about the approach to identify circadian miRNAs from their circadian expression. Gatfield *et al.* revealed that a liver specific miRNA, miR-122 regulated by REV-ERBα/β, shows strong circadian oscillation at the primary transcript level but not at the mature level. Nevertheless, they demonstrated that miR-122 can influence the circadian accumulation of their targets and plays an important role in circadian regulation of metabolism in liver[15]. Later, it was shown that miR-122 can also modulate the rhythmic expression of *Nocturnin*, which is a circadian clock-regulated deadenylase and is important for post-transcriptional control in circadian rhythm[16]. Using MEF cells derived from DICER knockout mice, Chen *et al.* showed that miRNAs including miR-24, miR-29a, and miR-30a are required for generating a time delay for the circadian oscillator although these miRNAs themselves do not show circadian expression[17]. Recently, Du *et al.* showed that the binding sites of miR-24 on the 3′ Untranslated Region (3′UTR) of *Per1 and Per2* are important for *Per1 and Per2*'s circadian expression in DICER liver conditional knockout (cKO) mice[18]. These studies suggest that the circadian oscillations of miRNAs at mature level may not be necessary for miRNA's circadian functions.

In this study, we embarked a different approach to systematically investigate the circadian expression of primary transcripts of miRNAs in mouse liver. By integrating recently published circadian Nascent-sequencing (Nascent-seq) and global run-on sequencing (GRO-seq) data in mouse liver, we first identified 57 circadian primary transcripts of miRNAs exhibiting consistent circadian oscillations across the two datasets, including miR-122, miR-24 and miR-29a with known circadian functions. Next by integrating ChIP-seq data of circadian regulators, we found that most of these primary transcripts are under the regulation of core circadian regulators. Comparing with the sequencing data of mature miRNAs, we only identified four miRNAs showed consistent circadian phases and periods between primary and mature miRNA transcripts. We observed that the amplitudes of circadian oscillations of mature miRNAs were much lower than those of the miRNA primary transcripts most likely due to the long half-lives of miRNAs. The absence of high amplitude circadian oscillations of miRNAs at mature level can explain the inconsistency between various circadian miRNA datasets. To explore the circadian functions of the miRNAs derived from these circadian primary transcripts, we focused on miR-378, a miRNA without known circadian functions, and examined the transcriptome changes in mouse livers upon miR-378 over-expression at two circadian time points, CT10 and CT22. First, we observed a significant over-representation of the circadian oscillating genes affected by miR-378 over-expression. Cell cycle genes including *p21* (*Cdkn1a*) were further enriched in the affected circadian oscillating genes, which implied that miR-378 could mediates the circadian control of cell cycle. Our examination of circadian regulatory network involving miR-378 suggests that miR-378 mediates the circadian control of cell cycle and metabolism by forming partnership with circadian transcription factors (TFs). Taken together, our study shows that circadian oscillations of miRNAs at primary transcript level rather than mature level can give rise to important circadian functions.

Results

57 miRNA primary transcripts are circadianly transcribed in mouse liver.

To systematically identify circadian primary transcripts of miRNAs, we integrated the recently published data of Nascent-seq[3] and GRO-seq[19] of mouse livers collected in circadian cycles. First, we used Vespucci program[20] to assemble the primary miRNA transcripts in mouse liver *de novo* from strand-specific mouse liver circadian GRO-seq data[19] (see materials and methods). Out of the 1186 mouse miRNAs with known precursor sequences (mirbase release 21), we identified 558 miRNA primary transcripts expressed in mouse liver corresponding to 611 miRNA precursor sequences (Supplementary Table S1). In comparison, Saini *et al.* only identified 39 mouse intergenic miRNA primary transcripts based on the transcription features including CpG island, 5′ CAGE, transcription start site (TSS) and poly-A signals as well as public cDNA and EST sequences in mouse[21]. Among their nine primary transcripts overlapped with our result, our approach has recovered longer primary transcripts with extended 5′ and/or 3′ ends (Supplementary Fig. S1 and Supplementary Table S2). We then mapped the sequencing reads from GRO-seq and Nascent-seq onto our assembled primary transcripts by Bowtie program[22]. We quantified the expression levels of miRNA primary transcripts and identified those showing significant circadian oscillations consistently from both GRO-seq and Nascent-seq by fitting the expression values to cosine functions with shifting phases respectively (materials and methods). 57 miRNA primary transcripts corresponding to 60 miRNA precursors were defined as circadian miRNA primary transcripts (Table 1 and Fig. 1a). Six of them were selected to be validated by qPCR using specific primers. In our independently collected circadian mouse liver samples, all six miRNA primary transcripts exhibited significant circadian oscillations and consistent circadian phases with the two high-throughput sequencing datasets (Fig. 1b). Also consistent with Gatfield *et al.*'s result[15], the primary transcript of miR-122 showed significant circadian oscillation with a peak around CT0 and a trough at CT12. Several other miRNAs that have been previously implicated in circadian regulation were also identified in our result. These include miR-24 and miR-29a that show no circadian expression at mature levels but were reported to be involved in the regulation of circadian period[17].

Circadian regulation of oscillating miRNA primary transcripts.

To investigate if the oscillating miRNA primary transcripts are under the control of circadian TFs, we conducted the promoter analysis for circadian miRNA primary transcripts. First, we confirmed that our *de novo* assembled primary miRNA transcripts indeed captured the 5′ ends of the transcripts by comparing their 5′ ends with the mouse liver specific promoter

| Primary miRNA transcript	Significance score	Circadian phase (CT)
mmu-mir-6935	11.33	0.33
mmu-mir-340	14.42	0.42
mmu-mir-190a	10.31	0.50
mmu-mir-7223	9.78	2.67
mmu-mir-5104	17.86	3.17
mmu-mir-6371	9.15	3.50
mmu-mir-1933	9.05	3.83
mmu-mir-3073a	9.40	4.00
mmu-mir-7036b	13.08	4.17
mmu-mir-1962	9.31	4.50
mmu-mir-1927	10.31	4.83
mmu-mir-2139	14.53	5.00
mmu-mir-7664	11.05	5.67
mmu-mir-1249	9.18	6.00
mmu-mir-5120	12.53	6.00
mmu-mir-1950	10.74	7.17
mmu-mir-7072	10.26	7.17
mmu-mir-101b	13.80	7.67
mmu-mir-1892	10.97	8.50
mmu-mir-7046	10.92	9.00
mmu-mir-378	9.25	9.17
mmu-mir-342	10.34	9.83
mmu-mir-29b-1~29a	16.18	10.00
mmu-mir-1190	9.64	10.33
mmu-mir-7048	13.09	10.33
mmu-mir-101a	11.64	10.50
mmu-mir-6896	15.58	11.33
mmu-mir-6908	12.16	11.83
mmu-mir-7661	13.57	12.00
mmu-mir-8114	10.05	12.00
mmu-mir-1947	13.31	12.67
mmu-mir-6953	9.63	12.83
mmu-mir-23b~27b~24-1	9.52	13.00
mmu-mir-1906-1	9.32	13.17
mmu-mir-6972	13.48	13.50
mmu-mir-6930	9.57	13.67
mmu-mir-455	10.44	14.00
mmu-mir-6971	9.52	14.50
mmu-mir-21a	10.18	16.33
mmu-mir-1191b	9.11	17.00
mmu-mir-484	10.18	17.00
mmu-mir-7687	9.46	17.17
mmu-mir-1936	12.34	19.67
mmu-mir-6353	10.62	19.67
mmu-mir-8107	10.06	20.00
mmu-mir-7063	10.25	20.33
mmu-mir-3087	13.90	20.50
mmu-mir-7022	11.37	20.50
mmu-mir-8093	9.64	20.50
mmu-mir-7052	10.12	20.67
mmu-mir-1955	11.44	20.83
mmu-mir-5131	13.49	21.17
mmu-mir-122	9.20	22.25
mmu-mir-130a	9.63	22.50
mmu-mir-6963	12.34	22.83
mmu-mir-3962	25.12	23.83
mmu-mir-6964	11.93	23.83

Table 1. List of circadian oscillating miRNA primary transcripts sorted according to their circadian phases in circadian time (CT). The significance score is defined as summed − (log2-transformed cosine fitting p-value).
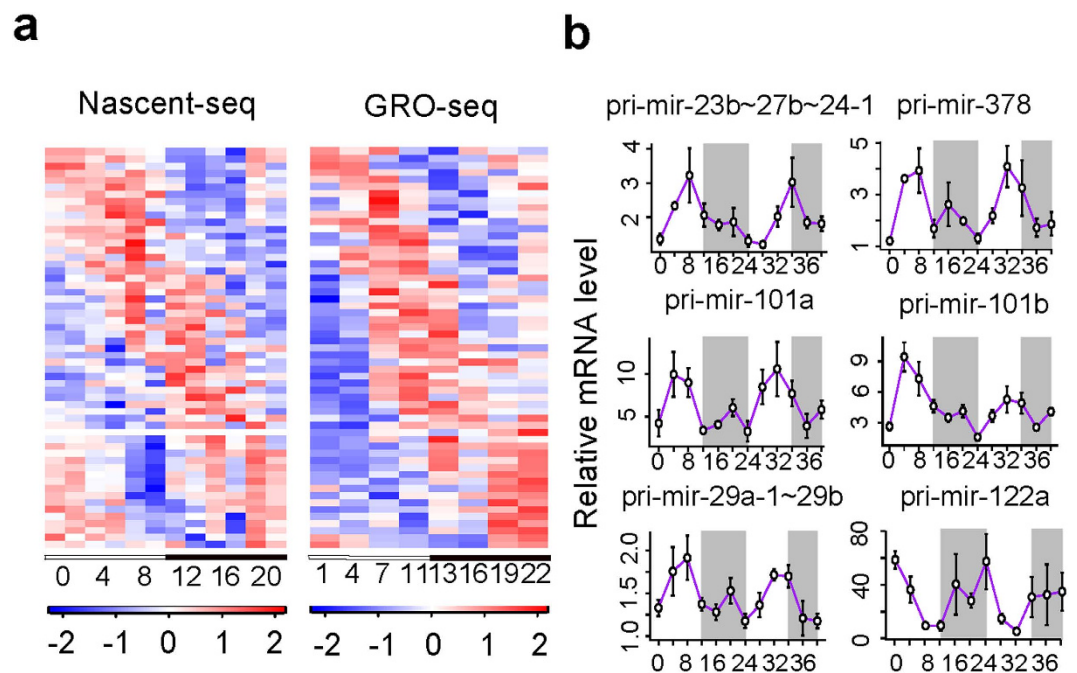
Figure 1. Circadian expression of miRNA primary transcripts. (**a**) Two heatmaps showed the expression patterns of circadian miRNA primary transcripts from Nascent-seq and GRO-seq respectively. High (red) and Low (blue) expression values as Z-scored normalized ratios are indicated in the color scale bar at the bottom. Black/white bars indicate the circadian day (white) and night (black) of the consecutive circadian cycles. (**b**) Expression patterns of circadian miRNA primary transcripts validated by qPCR. The x-axis represents the circadian time (CT), while the y-axis represents the relative expression values normalized by those of *Actb*. The white regions represent subjective day while the gray regions represent subjective night.

marker (H3K4me3) and Pol II marks. For 42 out of 57 circadian miRNA primary transcripts, we can find both H3K4me3 and Pol II marks on the 5′ ends of our assembled primary miRNA transcripts. For seven of them, we can only find H3K4me3 marks. This indicates that the 5′ ends of our assembled primary transcripts extended to the *bona fide* promoter regions of miRNAs. For the eight transcripts that do not overlap with H3K4me3 and Pol II marks, we corrected the 5′ ends of transcripts to the centers of the nearest H3K4me3 marks. Previously, Figueredo *et al.* have tried to identify the transcriptional regulators of the circadian miRNAs just by TF binding site prediction based on position weight matrix (PWM)[23]. However, the PWM-based TF binding site prediction tends to have high false positive and false negative rates. In the recent years, the ChIP-seq data of the core circadian regulators (BMAL1, CLOCK, PER1, PER2, CRY1, CRY2, REV-ERBα, REV-ERBβ, RORA, E4BP4) have been published in mouse liver[3,19,24–27]. By searching for circadian regulator binding sites on the promoters of circadian miRNA primary transcripts based on these ChIP-seq data, we identified potential circadian regulators of these miRNA primary transcripts (Supplementary Table S5). Comparing to all miRNA primary transcripts, we found that the promoters of circadian miRNA primary transcripts are more likely to be bound by most of the circadian regulators (Fig. 2a) and all promoters of the circadian miRNA primary transcripts, except for pri-mir-6971, contain at least one core circadian regulator binding site (Supplementary Table S5). Our BMAL1 ChIP-PCR experiment in mouse livers validated the physical binding of BMAL1 on the promoters of three selected circadian miRNA primary transcripts, pri-mir-23b~27b~24-1, pri-mir-101a, and pri-mir-378 (Fig. 2b). To further test if the regulatory relationships inferred from circadian TF ChIP-seq data are functional, we examined the expression changes of the circadian miRNA primary transcripts between liver-specific *Bmal1* cKO mice and wild-type mice by qPCR. We observed that four out of five tested BMAL1 regulated circadian miRNA primary transcripts containing BMAL1 binding sites are significantly under-expressed in liver-specific *Bmal1* cKO mice compared to wild-type mice at CT12 (Fig. 2c). It has been reported that miR-24 regulates *Per1* and *Per2* expression and is required for generating a time delay for the circadian oscillator[17]. Here we showed that miR-24 primary transcript is regulated by BMAL1/CLOCK. In summary, our results indicated that circadian miRNA primary transcripts are indeed regulated by circadian clock through core circadian TFs in mouse liver.

Dampening of circadian oscillation of miRNA expression at mature level. To date, two high-throughput studies have examined circadian expression of miRNAs at the mature level including miRNA
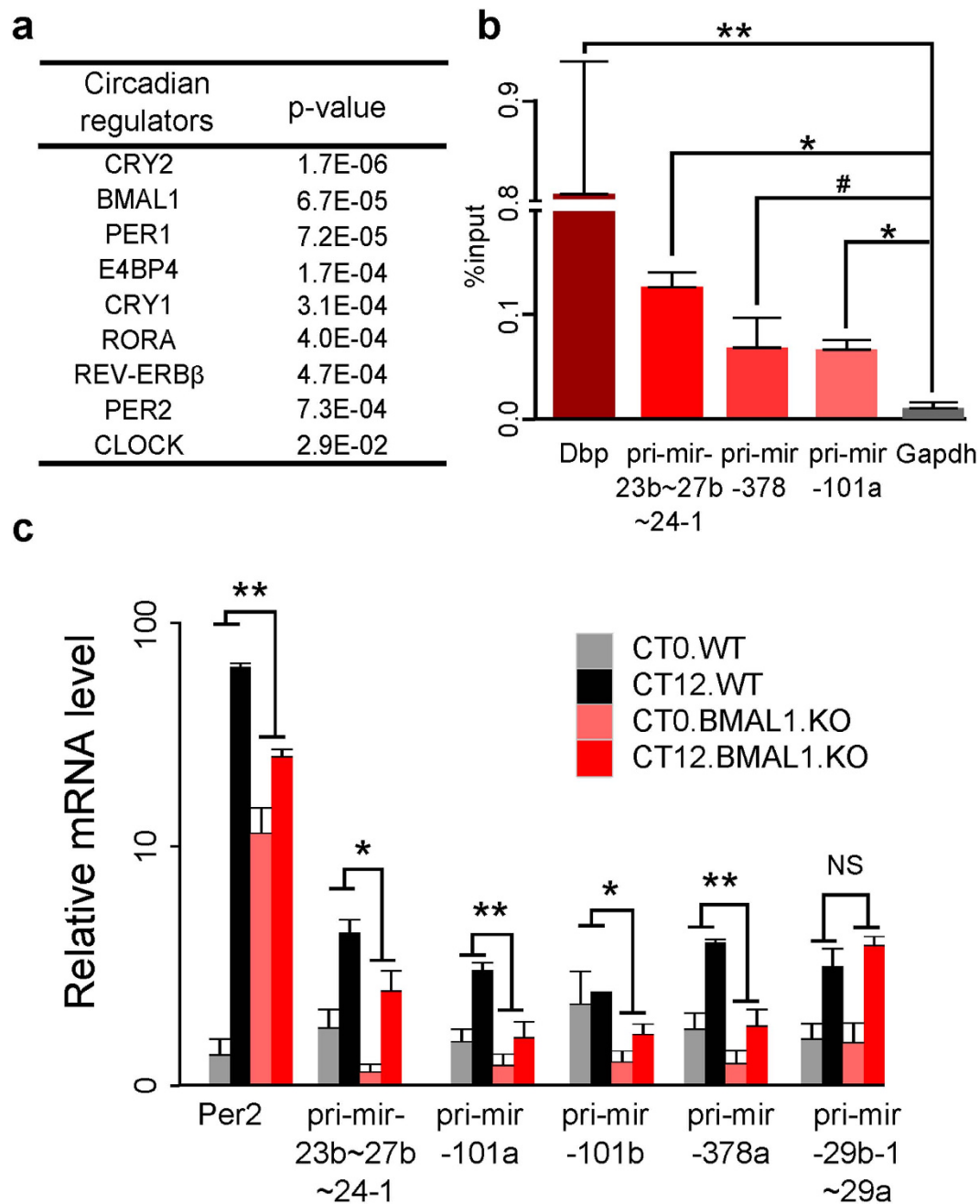
Figure 2. The regulation of circadian miRNA primary transcripts by circadian regulators. (**a**) circadian regulator binging sites are enriched on the promoters of the circadian oscillating miRNA primary transcripts comparing to all miRNA primary transcripts. Enrichment p-values were calculated by proportion test. (**b**) BMAL1 ChIP-PCR showed that the promoter regions of pri-mir-23b~27b~24-1, pri-mir-378, and pri-mir-101a are bound by BMAL1/CLOCK. *Gapdh* is used as the negative control, while Dbp is used as the positive control. **represents Student's t-test (unpaired, two-sided) p-value < 0.01, *represents Student's t-test (unpaired, two-sided) p-value < 0.05. #represents Student's t-test (unpaired, two-sided) p-value < 0.1 (marginally significant). (**c**) qPCR analysis of circadian miRNA primary transcripts of liver-specific *Bmal1* cKO mice and wild-type mice. *Per2* is used as the positive control. The y-axis has an exponential scale. Two-way ANOVA using circadian time (CT0 and CT12) and genotype (*Bmal1* cKO vs. wild-type) as two factors were applied to assess the statistical significance. **represents ANOVA p-value for genotype < 0.01, *represents ANOVA p-value for genotype < 0.05, NS stands for not significant p-value.

microarray data study by Na *et al.*[13] and miRNA-seq study by Vollmers *et al.*[14]. These two studies reported 85 and 30 circadian mature miRNAs respectively in mouse liver. Surprisingly, only two miRNAs (miR-26b and miR-150) appeared in common between both circadian miRNA lists. Although we were unable to obtain the raw data from
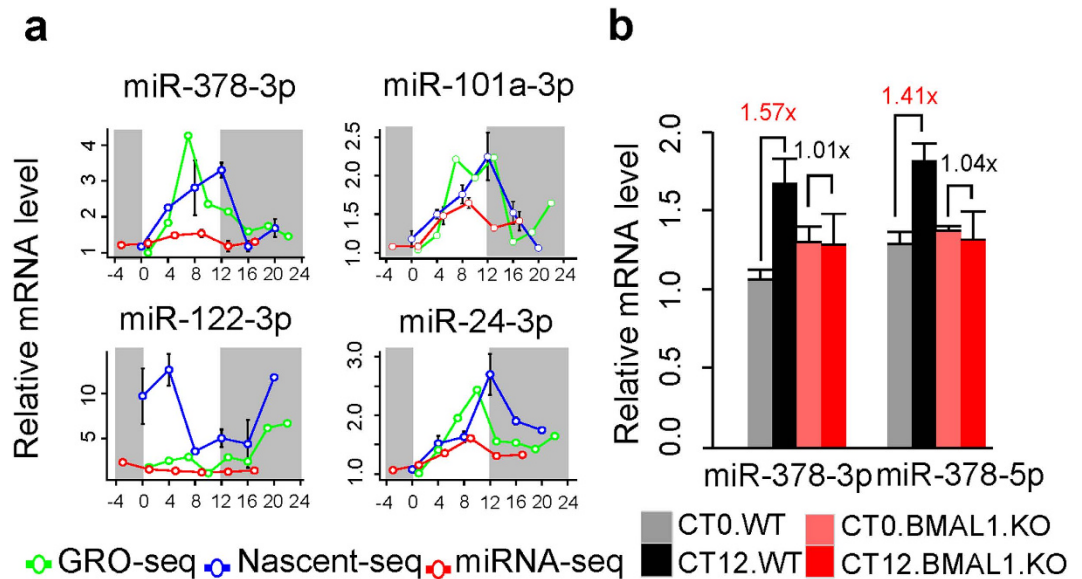
Figure 3. Circadian expression of mature miRNAs. (**a**) Comparison of the expression of circadian miRNA primary transcripts and their corresponding miRNAs at mature level. The green lines represent the expression values of primary transcripts from GRO-seq. The blue lines represent the expression values of primary transcripts from Nascent-seq. The red lines represent the expression values of mature transcripts from miRNA-seq. The white regions represent subjective day while the gray regions represent subjective night. (**b**) Expression patterns of miR-378-3p and miR-378-5p in liver-specific *Bmal1* cKO mice and wild-type mice at CT0 and CT12. Student's t-test (unpaired, two-sided) p-value < 0.05 are highlighted in red color for comparisons between liver-specific *Bmal1* cKO mice with wild-type mice.

Na *et al.*, a detailed examination of the original expression profiles of miR-26b and miR-150 in Vollmers *et al.*'s data suggested that miR-26b, i.e. miR-26b-5p, does not appear to be circadian oscillating (Supplementary Fig. S2a). Only miR-150-5p showed a moderate circadian oscillation in Panda's data with a peak time around CT8, which is close to that reported in Na *et al.*'s data (Supplementary Fig. S2b). This lack of consistency between the two circadian miRNA studies at mature level is in stark contrast to the overall consistency of circadian primary transcripts based on Nascent-seq and GRO-seq data. To investigate if circadian primary transcripts give rise to miRNAs with circadian expression at mature level at all, we re-analyzed the circadian miRNA-seq data from Vollmers *et al.*' using relaxed criteria (one-way ANOVA p-value < 0.05 and cosine fitting p-value < 0.05) (materials and methods). We found that four miRNAs (miR-24-3p, miR-101a-3p, miR-378-3p and miR-122-3p) showed significant circadian oscillations at mature levels with peak times close to those of their primary transcripts. However, the relative amplitudes of circadian oscillation at mature level are around 2-3 folds lower than those in primary transcripts (Fig. 3a). We next investigated whether mature miRNA expression are also regulated by BMAL1/CLOCK. We examined the expression changes of mature miR-378-3p and miR-378-5p in liver-specific *Bmal1* cKO mice and wild-type mice by miRNA specific qPCR (materials and methods). We found that the weak oscillations of miR-378-3p as well as miR-378-5p were abolished in liver-specific *Bmal1* cKO mice (Fig. 3b). These observations suggested that miR-378 is under the regulation of BMAL1/CLOCK at both primary and mature levels.

Genome-wide effects of miR-378 over-expression. In spite of weak or undetectable circadian expression of mature miRNAs, we found that at least 57 circadian miRNA primary transcripts including pri-mir-122 and pri-mir-24 do show strong circadian expression and are under circadian regulation. As miR-122 and miR-24 have been previously reported to be involved in circadian rhythm, we wondered if other circadian miRNA primary transcripts also harbor miRNAs with important circadian functions. To address this question, we selected miR-378 derived from pri-mir-378 transcript for further investigation. The expression of miR-378 shows circadian oscillation peaking at CT10 and is regulated by BMAL1/CLOCK. While miR-378 has been suggested to be related to cell cycle, cell proliferation and transformation, metabolism, adipocyte differentiation and stress responses[28–31], the functions of miR-378 in circadian rhythm have not yet been reported. To explore the circadian functions of miR-378, we injected adenovirus over-expressing both miR-378-3p and miR-378-5p into mice (termed Ad-378) while the control mice were injected with null virus (termed Ad-null, materials and methods). The efficiency of miR-378 over-expression in liver was about 50 folds as examined by miRNA specific qPCR (Supplementary Fig. S3). The liver samples of Ad-378 and Ad-null mice were collected at CT10 and CT22 corresponding to the expression peak and trough of miR-378 respectively. The global gene expression profiles were measured by RNA-sequencing (RNA-seq). We applied two-way ANOVA using circadian time (CT10 and CT22) and types of treatment (Ad-378 vs. Ad-null) as two factors to identify the genes that are affected by miR-378 over-expression.

Using ANOVA p-value for the factor of treatment, i.e. Ad-378 vs. Ad-null, less than 0.05 as the cutoff, we obtained 3,718 genes affected by miR-378 over-expression (Fig. 4a). Among these genes, the under-expressed genes in Ad-378 compared to Ad-null were over-represented by as much as eight-fold, consistent with the inhibitory role of miR-378 in gene regulation. Functional annotations of these genes showed that the most enriched function for the under-expressed genes is cell cycle (DAVID enrichment p-value $= 2.6 \times 10^{-44}$, using all mouse genes as the background). This is also consistent with the known function of miR-378 in cell cycle. Integrating with the miRNA-target interaction data based on both computational prediction and large-scale AGO CLIP-sequencing data from starBase (http://starbase.sysu.edu.cn/), we found a significant enrichment of the putative miR-378 targets in the under-expressed genes (Fisher's exact test p-value $= 1.43 \times 10^{-12}$, odd ratio $= 1.82$). Comparing with miR-378 target genes identified by another genome-wide study of miR-378-3p in NIH-3T3 cells[32], we again found a significant overlap between their result and our under-expressed genes by miR-378 over-expression in mouse liver (Fig. 4b).

Circadian oscillating genes are significantly affected by miR-378 over-expression.

We observed that the under-expressed genes were significantly enriched for liver circadian oscillating genes complied from six published circadian microarray datasets in mouse liver (mouse liver circadian database, materials and methods). Namely, a significantly elevated proportion (14.2%) of the under-expressed genes are circadian oscillating in mouse liver as compared to only 10% of the un-affected genes upon miR-378 over-expression (Fig. 4c). Using the p-value for the circadian time factor less than 0.05 as the cutoff, 2,266 genes showing significant differences in expression between CT10 and CT22 were identified. Comparing with the circadian peak time data from our mouse liver circadian database (materials and methods), the genes over-expressed at CT10 compared to CT22 (CT10 > CT22) in our study predominantly have circadian peak times around CT10 in published data while the genes over-expressed at CT22 compared to CT10 (CT10 < CT22) in our study have circadian peak times around CT20 (Fig. 4d). The core circadian genes (*Bmal1, Clock, Rev-ervα, Rev-ervβ, Per1, Per2, Per, Cry1, Cry2, and Cry3*) all showed significant differences in expression between CT10 and CT22 in our result, which is consistent with the literatures (Supplementary Fig. S4).

Furthermore, 833 genes showing significant difference in expression between CT10 and CT22 were affected by miR-378 over-expression (ANOVA p-value for treatment < 0.05 and ANOVA p-value for time < 0.05) (Fig. 4e). These include at least one core circadian gene, *Cry2* (Supplementary Fig. S4). These genes were clustered into four groups based on the hierarchical clustering of their gene expression. Group I genes peaked around CT10 and were under-expressed in Ad-378 (377 genes). Group II genes peaked around CT10 and were over-expressed in Ad-378 (96 genes). Group III genes peaked around CT22 and were under-expressed in Ad-378 (314 genes). Group IV genes peaked around CT22 and were over-expressed in Ad-378 (46 genes). Since the under-expressed genes are more likely to be the direct targets of miR-378, we then focused on the genes in Group I and Group III and defined the genes in these two groups as miR-378 directly targeted circadian oscillating genes, denoted as miR-378 circadian targets. We applied DAVID program[33] to identify the enriched biological functions for each group of genes using all mouse genes as the background. Interestingly, we found that the most significantly enriched function for Group I is cell cycle while the most significantly enriched function for Group III is positive regulation of programmed cell death and apoptosis (Supplementary Table S9). Five miR-378 circadian targets involved in cell cycle regulation (Ccne1, Runx3, Cdkn1a, Bbc3, and Bcl2) were selected to be validated by qPCR. Consistent with the RNA-seq data, all of them showed significant changes upon miR-378 over-expression. For all of them we can find the putative miR-378 binding sites on their 3′ UTRs using Miranda algorithm (Fig. S5). miR-378 has been mostly described as an oncogene-like miRNA in different cancer cell lines[34,35] and was reported to promote tumor growth, cell survival, and cell transformation[36,37]. We observed that both cyclin E1 (*Ccne1*) peaking around CT10, which is required for cell cycle G1/S transition and *Cdkn1a* peaking around CT22, which inhibits G1/S transition, are miR-378 circadian targets. It suggests that miR-378 may have dual roles in the circadian regulation of cell cycle progression.
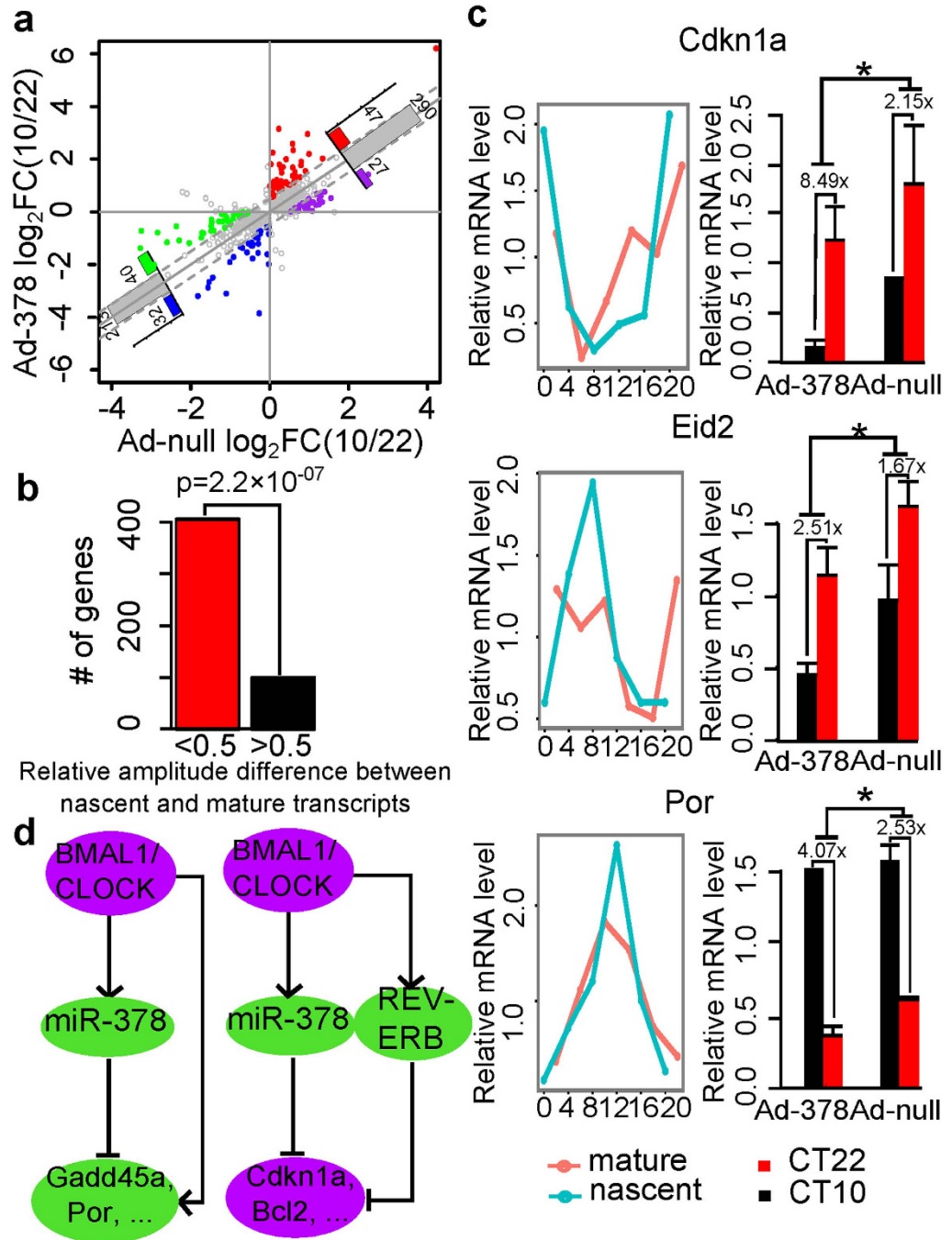
miR-378 is involved in circadian gene regulation.

For the mRNA transcripts under circadian transcriptional regulation, the inclusion of miRNA regulation will affect their mRNA degradation and influence their baseline levels as well as their amplitudes of circadian oscillations. Comparing the relative amplitude defined as the log2-transformed fold changes between CT10 and CT22, i.e. $\log_2 FC(CT10/CT22)$, of Ad-378 and Ad-null, we observed that 80% of miR-378 circadian targets showed little or no differences in relative amplitudes (Fig. 5a), i.e. relative amplitude differences less than 0.5. These genes also tend to show little or no differences in their relative amplitudes between their nascent transcripts measured in Nascent-seq and mature mRNA transcripts measured in RNA-seq (Fig. 5b). This suggests that these genes may be already under strong regulation by miR-378 such that the over-expression of miR-378 does not further increase their circadian amplitudes. However we still identified 47 circadian oscillation genes peaking around CT10 (red dots in Fig. 5a) and 32 genes peaking around CT22 (blue dots in Fig. 5a) showing increased relative amplitudes upon miR-378 over-expression. Cell cycle related genes are significantly enriched among those genes (DAVID enrichment p-value $= 3.3 \times 10^{-5}$, using all mouse genes as the background). In particular, *Cdkn1a* showed more than four-fold relative amplitude increase upon miR-378 over-expression. Therefore, miR-378 can influence the relative circadian amplitudes of cell cycle related genes by regulating their mRNA degradation rates. For these genes, we reasoned that the relative amplitudes of their premature transcripts may be also higher than those of their mature transcripts. Indeed, we identified that three genes (*Cdkn1a*, *Eid2*, and *Por*) showed significant increased relative amplitudes in their nascent transcripts comparing to the mature transcripts in a similar fashion as in Ad-378 comparing to Ad-null (Fig. 5c).

We next examined the potential cooperation between core circadian TFs and miR-378 in circadian gene regulation. It has been known that the circadian oscillating genes peaking around CT10 are more likely regulated by BMAL1/CLOCK while the genes peaking around CT22 are more likely regulated by REV-ERBα/β[19].
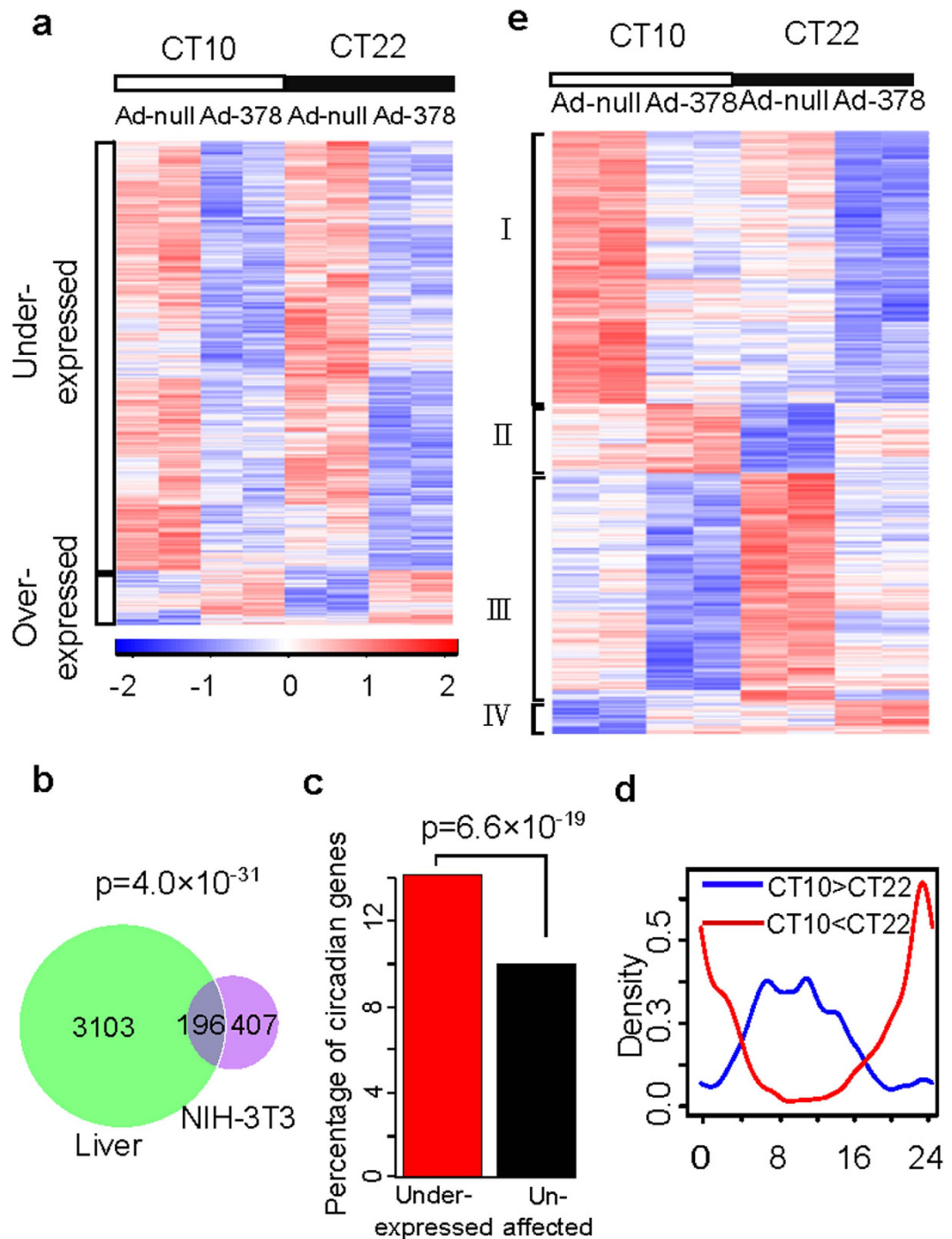
Figure 4. Transcriptome profile of miR-378 over-expression in mouse liver. (a) Heatmap of all genes affected by miR-378 over-expression. The genes are clustered into two groups based on their expression patterns. High (red) and low (yellow) expression values as Z-scored normalized ratios are indicated in the color scale bar at the bottom. Black/white bars indicate the circadian day (white) and night (black) of the consecutive circadian cycles. **(b)** The number of overlapping genes between under-expressed genes in Ad-378 comparing to Ad-null (green circle) and miR-378-3p targets from Ruckerl *et al.*'s data (purple circle). Enrichment p-value was determined by Fisher's exact test. **(c)** Enrichment of circadian transcripts in under-expressed genes upon miR-378 over-expression. Enrichment p-value was determined by proportion test. **(d)** Comparison of the circadian peak times in our data with the peak times from mouse liver circadian database. The red curve represents the circadian peak time distribution for genes over-expressed at CT22 in our data. The blue curve represents the circadian peak time distribution for genes over-expressed at CT10 in our data. **(e)** Heatmap of the circadian oscillating genes affected by miR-378 over-expression. The transcripts are clustered into four groups based on their expression patterns. High (red) and low (yellow) expression values as Z-scored normalized ratios are indicated in the color scale bar at the bottom. Black/white bars indicate the circadian day (white) and night (black) of the consecutive circadian cycles.

Figure 5. miR-378 in circadian gene regulation. (**a**) Comparison of the log2-transformed fold changes of CT10 and CT22 (log$_2$FC(CT10/CT22)) between Ad-null and Ad-378 of miR-378 circadian targets (Group I and Group III genes in Fig. 4e). The solid grey line represents y = x. The dashed grey lines represent y = x ± 0.5. The colored dots represent the genes with log$_2$FC(CT10/CT22) differences between Ad-378 and Ad-null larger than 0.5, while the grey nodes represent the genes with the difference less than 0.5. Number of genes in each group is indicated in the bar plot. (**b**) Comparison of the relative amplitudes (log2-transformed peak to trough ratio) between nascent and mature transcripts of the grey dots in 5a. Red bar represents the genes that do not show relative amplitude changes (difference of the relative amplitude less than 0.5). Black bar represents the genes with relative amplitude changes (difference of the relative amplitude larger than 0.5). Enrichment p-values were calculated by Fisher's exact test using the relative amplitude differences between nascent and mature transcripts of all mouse genes as the background. (**c**) The expression patterns of the three genes (*Cdkn1a*, *Eid2*, and *Por*) that showed increased relative amplitudes in nascent transcript comparing to mature transcript and in Ad-378 comparing to Ad-null. The left figures show the expression patterns of nascent and mature transcripts and the right figures show the expression patterns upon miR-378 over-expression. *represents ANOVA p-value for treatment < 0.05. (**d**) Two gene regulatory motifs formed by miR-378 and circadian TFs (BMAL1/CLOCK, REV-ERBα/β). The green nodes represent the genes peaking around CT10, while the purple nodes represent the genes peaking around CT22.

Therefore, we overlapped the Group I genes with the BMAL1/CLOCK regulated genes and Group III genes with the REV-ERBα/β regulated genes. Here we defined BMAL1/CLOCK and REV-ERBα/β regulated genes by both physical TF binding data from ChIP-seq and functional data from TF knock-out RNA-seq or microarray from ours and published data (materials and methods). As such, we constructed a gene network involving miR-378 and circadian TFs (Fig. S7). We observed two types of regulatory motifs (Fig. 5d). In the first motif, which is a feed-forward loop (FFL), miR-378 cooperated with BMAL1/CLOCK to regulate the genes including *Gadd45a* and *Por* with the circadian peak around CT10. In the second motif, miR-378 cooperated with REV-ERBα/β to regulate the genes including *Cdkn1a* and *Bcl2* with circadian peaks around CT22.

For the miR-378 circadian target genes that lack circadian transcriptional regulation, miR-378 alone can lead to their circadian expression by rhythmic post-transcriptional regulation. Lück *et al.* have previously identified 493 circadian oscillating genes under rhythmic post-transcriptional regulation[38]. We overlapped the miR-378 circadian targets in our study with Lück *et al.*'s result. We found that there was a significant overlap between these two gene sets (Fisher's exact test p-value $= 4.74 \times 10^{-6}$, odd ratio $= 2.65$). Interestingly, five genes involved in the regulation of transcription (*Rdbp, Elp3, Fus, Irf2, Pnrc2*) were among the overlapped genes. These genes either showed significant increases in oscillation amplitudes of mature transcripts comparing to nascent transcripts such as *Fus* or large phase shift between nascent and mature transcripts such as *Rdbp* and *Pnrc2*. Meanwhile, they were also significantly under-expressed in Ad-378 vs. Ad-null in our study (Supplementary Fig. S6). *Fus* is a circadian oscillating gene that shows consistent circadian expression across many different tissues in mouse[39]. Consistent with our result, Ma *et al.* showed that miR-378 promotes the migration and metastasis of liver cancer cells by down-regulating *Fus* expression[29]. Therefore, circadian post-transcriptional regulation of *Fus* by miR-378 can further contribute to the circadian regulation of cell cycle. The post-transcriptional regulation of these genes by miR-378 may play a predominant role in their circadian regulation.

Discussion

In mouse, liver is the most extensively studied peripheral tissue in circadian rhythm so far. Various high-throughput technologies including microarray, RNA-seq, and ChIP-seq have been applied to the study of circadian rhythm in mouse liver. Several papers have reported the functions of miRNAs in circadian rhythm in mouse liver. But the studies on specific miRNAs showed that the miRNAs of circadian functions are either arrhythmic or only rhythmic[17] at primary transcript level[15], while two systematic investigations of circadian miR-NAs in mouse liver at mature level resulted in quite different miRNA lists[13,14]. In this study, we first identified 57 miRNA primary transcripts showing significant and consistent circadian oscillations by integrating two independent next-generation sequencing data of nascent transcripts in mouse liver. We investigated their regulations by core circadian regulators by integrating the circadian ChIP-seq data. Notably, three of them have already been reported to be involved in circadian regulation, which suggests that the miRNAs rhythmic at primary level could harbor circadian functions. We then compared the primary miRNA expression with the mature miRNA expression and found only four miRNAs that showed consistent circadian phases with their primary transcripts while the circadian amplitudes of the mature miRNAs are only around one third of the primary transcripts. The average half-life of mature miRNAs in mammals was estimated to be around 119h[40]. Under such long half-lives, most circadian primary transcripts will not lead to miRNAs with detectable circadian oscillation at mature level, which may also explain the inconsistency of the high-throughput studies designed to identify the mouse liver circadian miRNAs. Therefore, we proposed that the circadian expression of miRNAs at primary level can be a better criterion to define circadian miRNAs.

In the light of our result, there are two scenarios of how miR-378 could exert its impact on circadian expression. First, Gatfield *et al.*[15] suggested that miRNAs such as miR-122 that are arrhythmic at mature level could contribute to the large amplitudes of oscillation of target mRNA transcripts by maintaining their high but constant degradation rates. Supporting this model, we identified 79 miR-378 circadian targets that showed increased relative amplitudes upon miR-378 over-expression. For the rest of miR-378 circadian targets that do not show significant amplitude changes, it is still possible that they are strongly regulated by miR-378 in wild type mice so that miR-378 over-expression can not increase their amplitudes further. This model can explain how the miR-NAs that are either weakly rhythmic or arrhythmic at mature level could contribute to the circadian regulation. Second, we observed that the genes under rhythmic post-transcriptional regulation significantly enriched in miR-378 circadian targets. As the mature miRNA-378-3p and miRNA-378-5p still showed low amplitudes of circadian oscillation, the circadian expression of these targets may be driven by rhythmic post-transcriptional regulation through miRNA-378. In this scenario, miRNAs rhythmically regulate the degradation of the target mRNAs and thus are crucial for their circadian expression.

Circadian regulation is an essential control mechanism of cell cycle[41]. Our functional annotation has pointed to the cross talk between circadian rhythm and cell cycle through miR-378. Among the cell cycle genes targeted by miR-378, *Ccne1* peaking around CT10 is required for cell cycle G1/S transition, while *Cdkn1a* , an inhibitor of G1/S transition, shows circadian peak around CT22 and has long been shown to be repressed by REV-ERBα[42]. In NIH-3T3 cells, *Cdkn1a* is also significantly under-expressed in miR-378 over-expression and up-regulated in miR-378 inhibition[32]. We suggested that miRNA-378 might regulate different groups of target genes by cooperating with different circadian TFs (Fig. 5d). It has been shown that one of our motifs constituting a FFL is particularly prevalent in the miRNA-mediated regulatory network[43]. Through both of our motifs, miR-378 can further fine-tune the circadian amplitudes of target genes by increasing their degradation rates. In addition, miR-378 has been reported to be involved in metabolic process. In human breast cancer cells, miR-378 induces metabolic shift by inhibiting the expression of two PGC1β partners, ERRγ and GABPA. This leads to the decrease of gene expression in oxidative phosphorylation and the increase in glycolysis[44]. In mouse liver, it has been shown that miR-378 targets p110α and controls glucose and lipid homeostasis by modulating hepatic insulin signaling[30]. In our study, two metabolic genes (*Pdk4* and *Por*) showing robust circadian expression in mouse liver are under-expressed

upon miR-378 over-expression. *Pdk4* is involved in glucose metabolism while *Por* is involved in oxidative reduction. Our findings may provide the explanation of the time-of-day dependent function of miR-378 in the regulation of cell cycle and metabolism.

In summary, our study suggests a new way of searching for miRNAs with circadian functions, i.e. among the miRNAs showing circadian expression in their primary transcripts. We have provided evidences for the role of miR-378 in circadian regulation. Future studies will be necessary to explore the circadian functions of the rest of circadian primary miRNA transcripts.

Material and Methods

Assembly of miRNA primary transcripts. Vespucci program[20] developed for the identification of *de novo* transcripts from GRO-sequencing data, was applied to assemble the mouse liver primary miRNA transcripts from strand-specific mouse liver circadian GRO-seq data[19] with two modified parameters, DENSITY_MULTIPLER = 100 and MAX_EDGE = 1000. The primary transcripts containing known miRNA precursors were extracted. The identified mouse liver primary transcripts were listed in Supplementary Table S1.

Identification of circadian miRNA primary transcript. Sequencing reads from GRO-seq[19] and Nascent-seq[3] were mapped to mouse genome (mm9) by Bowtie2 program (default parameter)[22]. The read number of each primary transcript was normalized to reads per kilobase per million tags (RPKM).We searched for the circadian oscillating miRNA primary transcripts that showed significant and consistent circadian oscillation on both GRO-seq and Nascent-seq data by fitting them to the common cosine functions with 24 hours' period and shifting phases as described previously[39]. The circadian primary transcripts were defined as the summed log2-transformed cosine fitting p-value > 9 and circadian phase difference < 4. False discovery rate (FDR) was calculated by randomly shuffling the gene 1000 times. The cutoffs that we used corresponded to an overall FDR = 29% as computed from random permutation.

Animal preparation and sample collection. All mice were males aged 8–12 weeks and were maintained on a C57BL/6J background. The mice were housed under a 12-h light/12-h dark regimen with food and water available ad libitum for one week before being switched to complete darkness. The mice were switched to complete darkness two days before execution. CT0 is defined as the time when the lights are turned on. The animal handing protocols have been approved by Institutional Animal Care and Use Committee of Institute for Nutritional Sciences. All experimental procedures performed followed the guidelines of the Institutional Animal Care and Use Committee of Institute for Nutritional Sciences. Ad-378 and Ad-null mice were generated as previously described[30]. Basically, DNA fragments encoding miR-378 were introduced into pENTR/U6 vector under the control of the human U6 promoter. 5×10^8 plaque-forming unit viruses were administered to four Ad-378 mice through tail vein injection for miR-378 over-expression. Control viruses were injected to four Ad-null mice. Two Ad-378 and two Ad-null mice were sacrificed at CT10 and CT22 respectively eight days after injection. To validate the circadian expression of miRNA primary transcripts, liver samples from wild-type mice were collected. Starting at CT0, three mice were killed every four hours for 48 hours. Livers of all mice were quickly dissected and snap-frozen in liquid nitrogen and stored at −80 °C.

RNA samples of livers from three liver-specific *Bmal1* cKO mice and three wild-type mice (Bmal1$^{flox/flox}$) sacrificed at CT0 and CT12 respectively were provided by Dr. Yi Liu (Institute for Nutritional Sciences). Basically, Bmal1$^{flox/flox}$ mice were obtained from Jackson's Lab (stock no. 007668). These mice have loxP sites flanking exon 8 of the *Bmal1* gene. Then the *Bmal1*$^{flox/flox}$ mice were bred to mice expressing CRE recombinase with liver specific promoter. The resulting offsprings have the exons encoding BMAL1 basic helix-loop-helix (bHLH) domain deleted in the liver.

qPCR of miRNA primary transcripts. Total RNA was isolated from mouse liver by Trizol (Invitrogen). Total RNA quantities were measured by a Nanodrop spectrometer. RNA quality was assessed by an Agilent Bioanalyzer. 500 ng of total RNA was reverse transcribed using random hexamer and the Superscript II reverse transcriptase (Invitrogen). 5 μl of RT product (1:10 diluted) and SYBR Green I Master Mix (Roche) were used in qPCR on a LightCycler 480 (Roche). The specificity of PCR was checked by melting curve analysis. In every qPCR assay, *Actb* was used as the control for any significant bias of starting materials across samples. Specific primers for miRNA primary transcripts were listed in Supplementary Table S3. The same qPCR procedure was applied to validate the expression pattern of cell cycle genes upon miR-378 over-expression, and the primers for the cell cycle genes were listed in Supplementary Table S10.

Promoter analysis of circadian miRNAs. Mouse liver H3k4me3 and PolII ChIP-seq data from Bing Ren's lab[45] were used to confirm that our *de novo* assembled primary miRNA transcripts indeed captured the 5′ end of the transcripts. The sequencing reads were mapped to mouse genome (mm9) using bowtie2 program (default parameter)[22]. MACS program[46] was applied to identify the binding peaks of ChIP-seq data (MACS 1.4.2, default parameter). For the six transcripts that do not overlap with H3K4me3 and Pol II marks, we corrected their 5′ ends of transcripts to the centers of the nearest H3K4me3 or Pol II marks.

The fastq files containing the raw sequence reads of the circadian ChIP-seq data were downloaded from GEO database (http://www.ncbi.nlm.nih.gov/geo) and listed in Supplementary Table S4. For the ChIP-seq experiments conducted across multiple circadian time points, we selected the time point corresponding to the binding peak of each regulator. The reads were mapped to mouse mm9 genome using bowtie2 (default parameters)[22] while bowtie1 were used for to Koike *et al.*'s color-space SOLID RNA-seq data. MACS program[46] was applied to identify the binding sites of each regulator (MACS 1.4.2, default parameter). The regulators having binding sites between upstream 5,000 bp of 5′ end and 3′ end of the transcripts were defined as the regulators of the miRNA primary transcripts. For the six BMAL1 ChIP-seq datasets, we defined the physical binding of BMAL1 on the transcripts

if the binging events appear in at least two datasets. For the other regulators, we defined their binding as long as their binding events appear in at least one dataset. The number of datasets supporting the binding of regulators on each miRNA primary transcript was listed in Supplementary Table S5.

BMAL1 ChIP-PCR analysis, analysis of mouse liver circadian miRNA-seq data, qPCR of mature miR-378-3p and miR-378-5p expression, circadian miRNA target identification, RNA-sequencing and data analysis, miR-378-3p targets in NIH-3T3 cell, mouse liver circadian database, comparison of the relative circadian amplitudes of nascent and mature transcripts, and regulation of circadian TFs are available in the Supplementary information.

References

1. Takahashi, J. S., Hong, H. K., Ko, C. H. & McDearmon, E. L. The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat Rev Genet* **9,** 764–75 (2008).
2. Reppert, S. M. & Weaver, D. R. Molecular analysis of mammalian circadian rhythms. *Annu Rev Physiol* **63,** 647–76 (2001).
3. Menet, J. S., Rodriguez, J., Abruzzi, K. C. & Rosbash, M. Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *Elife* **1,** e00011 (2012).
4. Robles, M. S., Cox, J. & Mann, M. *In-vivo* quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genet* **10,** e1004047 (2014).
5. Mauvoisin, D. *et al.* Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. *Proc Natl Acad Sci USA* **111,** 167–72 (2014).
6. Shukla, G. C., Singh, J. & Barik, S. MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. *Mol Cell Pharmacol* **3,** 83–92 (2011).
7. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136,** 215–33 (2009).
8. Mehta, N. & Cheng, H. Y. Micro-managing the circadian clock: The role of microRNAs in biological timekeeping. *J Mol Biol* **425,** 3609–24 (2013).
9. Cheng, H. Y. *et al.* microRNA modulation of circadian-clock period and entrainment. *Neuron* **54,** 813–29 (2007).
10. Nagel, R., Clijsters, L. & Agami, R. The miRNA-192/194 cluster regulates the Period gene family and the circadian clock. *FEBS J* **276,** 5447–55 (2009).
11. Zhou, L. *et al.* Clock-controlled mir-142-3p can target its activator, *Bmal1*. Vol. 13, 27–27 (2012).
12. Xu, S., Witmer, P. D., Lumayag, S., Kovacs, B. & Valle, D. MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster. *J Biol Chem* **282,** 25053–66 (2007).
13. Na, Y. J. *et al.* Comprehensive analysis of microRNA-mRNA co-expression in circadian rhythm. *Experimental & molecular medicine* **41,** 638–647 (2009).
14. Vollmers, C. *et al.* Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome. *Cell Metab* **16,** 833–45 (2012).
15. Gatfield, D. *et al.* Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes & Development* **23,** 1313–1326 (2009).
16. Kojima, S, Gatfield, D., Esau, C. C. & Green, C. B. MicroRNA-122 modulates the rhythmic expression profile of the circadian deadenylase Nocturnin in mouse liver. *PloS one* **5,** e11264–e11264 (2010).
17. Chen, R., D'Alessandro, M. & Lee, C. miRNAs Are Required for Generating a Time Delay Critical for the Circadian Oscillator. *Current Biology* **23,** 1959–1968 (2013).
18. Du, N. H., Arpat, A. B., De Matos, M. & Gatfield, D. MicroRNAs shape circadian hepatic gene expression on a transcriptome-wide scale. *Elife* **3,** e02510 (2014).
19. Fang, B. *et al.* Circadian enhancers coordinate multiple phases of rhythmic gene transcription *in vivo*. *Cell* **159,** 1140–52 (2014).
20. Allison, K. A., Kaikkonen, M. U., Gaasterland, T. & Glass, C. K. Vespucci: a system for building annotated databases of nascent transcripts. *Nucleic Acids Res* **42,** 2433–47 (2014).
21. Saini, H. K., Enright, A. J. & Griffiths-Jones, S. Annotation of mammalian primary microRNAs. *BMC Genomics* **9,** 564 (2008).
22. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–9 (2012).
23. Figueredo, D. D. S., Barbosa, M. R., Gitaí, D. L. G. & de Andrade, T. G. Predicted MicroRNAs for Mammalian Circadian Rhythms. *Journal of biological rhythms* **28,** 107–16 (2013).
24. Annayev, Y. *et al.* Gene model 129 (Gm129) encodes a novel transcriptional repressor that modulates circadian gene expression. *J Biol Chem* **289,** 5013–24 (2014).
25. Koike, N. *et al.* Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* **338,** 349–54 (2012).
26. Cho, H. *et al.* Regulation of circadian behaviour and metabolism by REV-ERB-alpha and REV-ERB-beta. *Nature* **485,** 123–7 (2012).
27. Rey, G. *et al.* Genome-wide and phase-specific DNA-binding rhythms of BMAL1 control circadian output functions in mouse liver. *PLoS Biol* **9,** e1000595 (2011).
28. Xing, Y. *et al.* microRNA-378 promotes mesenchymal stem cell survival and vascularization under hypoxic-ischemic conditions *in vitro*. *Stem Cell Res Ther* **5,** 130 (2014).
29. Ma, J. *et al.* MiR-378 promotes the migration of liver cancer cells by down-regulating Fus expression. *Cell Physiol Biochem* **34,** 2266–74 (2014).
30. Liu, W. *et al.* Hepatic miR-378 targets p110alpha and controls glucose and lipid homeostasis by modulating hepatic insulin signalling. *Nat Commun* **5,** 5684 (2014).
31. Knezevic, I. *et al.* A novel cardiomyocyte-enriched microRNA, miR-378, targets insulin-like growth factor 1 receptor: implications in postnatal cardiac remodeling and cell survival. *J Biol Chem* **287,** 12913–26 (2012).
32. Ruckerl, D. *et al.* Induction of IL-4Ralpha-dependent microRNAs identifies PI3K/Akt signaling as essential for IL-4-driven murine macrophage proliferation *in vivo*. *Blood* **120,** 2307–16 (2012).
33. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4,** 44–57 (2009).
34. Wang, X. *et al.* microRNAs are biomarkers of oncogenic human papillomavirus infections. *Proc Natl Acad Sci USA* **111,** 4262–7 (2014).
35. Redova, M. *et al.* Circulating miR-378 and miR-451 in serum are potential biomarkers for renal cell carcinoma. *J Transl Med* **10,** 55 (2012).
36. Feng, M. *et al.* Myc/miR-378/TOB2/cyclin D1 functional module regulates oncogenic transformation. *Oncogene* **30,** 2242–51 (2011).
37. Lee, D. Y., Deng, Z., Wang, C. H. & Yang, B. B. MicroRNA-378 promotes cell survival, tumor growth, and angiogenesis by targeting SuFu and Fus-1 expression. *Proc Natl Acad Sci USA* **104,** 20350–5 (2007).
38. Luck, S., Thurley, K., Thaben, P. F. & Westermark, P. O. Rhythmic degradation explains and unifies circadian transcriptome and proteome data. *Cell Rep* **9,** 741–51 (2014).
39. Yan, J., Wang, H., Liu, Y. & Shao, C. Analysis of gene regulatory networks in the mammalian circadian rhythm. *PLoS Comput Biol* **4,** e1000193 (2008).

40. Gantier, M. P. *et al.* Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. *Nucleic Acids Res* **39,** 5692–703 (2011).
41. Masri, S., Cervantes, M. & Sassone-Corsi, P. The circadian clock and cell cycle: interconnected biological circuits. *Curr Opin Cell Biol* **25,** 730–4 (2013).
42. Grechez-Cassiau, A., Rayet, B., Guillaumond, F., Teboul, M. & Delaunay, F. The circadian clock component BMAL1 is a critical regulator of p21WAF1/CIP1 expression and hepatocyte proliferation. *J Biol Chem* **283,** 4535–42 (2008).
43. Tsang, J., Zhu, J. & van Oudenaarden, A. MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Mol Cell* **26,** 753–67 (2007).
44. Eichner, L. J. *et al.* miR-378(*) mediates metabolic shift in breast cancer cells via the PGC-1beta/ERRgamma transcriptional pathway. *Cell Metab* **12,** 352–61 (2010).
45. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–80 (2012).
46. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9,** R137 (2008).

Acknowledgements

Author Contributions

H.W. and J.Y. conceived, designed and interpreted the experiments. M.Z. and J.L. performed experiments, including sample collection, qPCR. H.W. and Z.F. performed all bioinformatics analyses. W.L. and H.Y. generated the miR-378 over-expression mice. M.G.L. and M.L. provided the expertise and instrumentation required for the experiments. H.W. and J.Y. wrote the manuscript. All authors reviewed the manuscript.

Additional Information

Accession codes: RNA-seq data of miR-378 over-expression and liver-specific *Bmal1* cKO were submitted to GEO under accession numbers GSE73308 and GSE73271.

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Wang, H. *et al.* Oscillating primary transcripts harbor miRNAs with circadian functions. *Sci. Rep.* **6**, 21598; doi: 10.1038/srep21598 (2016).

3.2 Single-cell in vivo imaging reveals light-initiated circadian oscillators in zebrafish

Reference: **Wang, H.**, Yang, Z., Li, X., Huang, D., Yu, S., He, J., Li, Y., Yan, J. (2020) Single-cell in vivo imaging of cellular circadian oscillators in zebrafish. PLoS Biol. 18, e3000435.

3.2.1 Synopsis

Studying the circadian clock at single-cell level is of fundamental importance to circadian research. Genetically encoded calcium indicator (GCaMP) has been widely used for monitoring single-cell neuronal activity *in vivo*, however there is no circadian reporter system at the single cell level in a live animal.

In this study, through a systematic screening, my colleagues and I obtained a transgenic zebrafish line, *nr1d1*:VNP, driven by the promoter of *nr1d1*, which encodes the core circadian regulator Rev-Erbα. VNP stand for Venus-NLS-PEST, which is an engineered short-lived nuclear fluorescent protein. Venus is a yellow fluorescent protein with high fluorescence intensity, NLS is a nuclear localization sequence, and PEST is a signal peptide for rapid protein degradation. This system allows us to monitor single-cell circadian clocks in zebrafish larva *in vivo* for the first time. qPCR of both endogenous *nr1d1* and transgenic *nr1d1*:VNP were conducted by my colleagues and me, which showed that the endogenous and transgenic expression patterns exhibited consistent circadian rhythms during development. This indicates that *nr1d1*:VNP can be used to report the endogenous expression of *nr1d1*.

Next, my colleagues and I carried out single-cell *in vivo* imaging of the whole-brain of *nr1d1*:VNP transgenic fish during embryonic development using a two-photon microscope. I observed that *nr1d1*:VNP expression first starts from the photoreceptor cells in pineal gland, then spreads to cells in other brain regions, such as optic tecum and cerebellum. Interestingly, *nr1d1*:VNP signal is strongest in the pineal gland. This developmentally regulated rhythmic expression of *nr1d1*:VNP closely mimics that of endogenous *nr1d1* from whole mount ISH reported by Delaunay et al (Delaunay et al., 2000).

Using scRNA-seq technology, I characterized the expression of *nr1d1*:VNP in different cell types in the whole brain (Figure 3.3). In line with the imaging data, I found that the mRNA of *nr1d1*:VNP was most highly enriched in cell clusters of photoreceptors of the pineal gland. Therefore, we focused on the pineal gland in the subsequent analyses. Note, however, that the mRNA of *nr1d1*:VNP was also enriched in many other cell clusters, such as the granule cells in the cerebellum, the glutamatergic neurons in the forebrain and the hypothalamus, the dorsal habenula cells, as well as some non-neuronal cells.
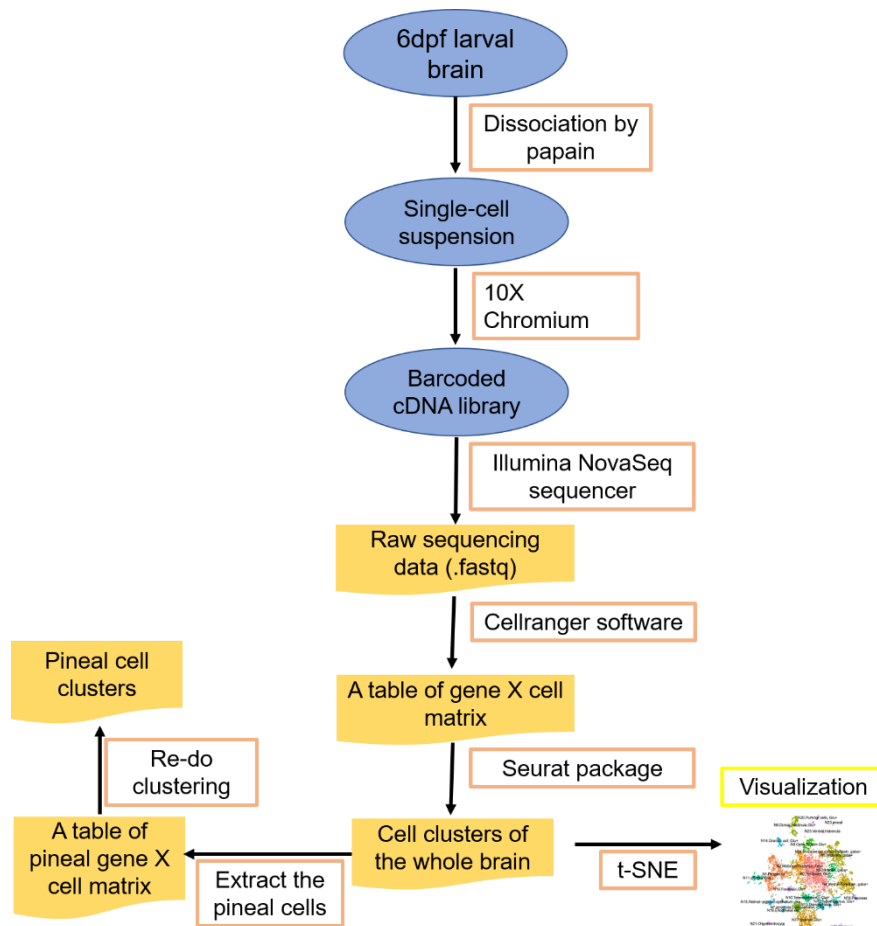
Figure 3.3 Workflow of scRNA-seq analysis of 6dpf larva zebrafish brain. dpf, days post-fertilization.

To study the development of single-cell clock *in vivo* in the pineal gland, we imaged the pineal starting from 3.5 days post-fertilization (dpf) to 6.5dpf every 12 hours using two-photon microscope (Figure 3.4). I applied CMTK image registration to align the 3D imaging of each time point from the same individual, therefore I could trace each single cell across the time points. I observed that VNP-positive cells within the pineal gland showed two components of the changes: a general increase in the base line fluorescence and in the circadian oscillation of *nr1d1*:VNP signals during development. Furthermore, clustering analysis of the *nr1d1*-VNP positive cells using t-SNE revealed that *nr1d1*:VNP-positive cells can be classified into two clusters. Cluster 1 cells showed stronger increase in the baseline level fluorescence during development and cluster 2 cells showed stronger circadian oscillations of fluorescence. To quantify the circadian and development components, a regression model that combines a stepwise function representing the circadian components and a linear function representing the development components were applied to fit the expression data of each single cell (Figure 3.4). I observed that the regression coefficients for both developmental and circadian components differ greatly across *nr1d1*:VNP-positive cells, and they are significantly negative correlated. Furthermore, consistent with the clustering result, cluster 1 cells showed significant larger regression coefficient of developmental

effect than oscillating cells, while cluster 2 cells showed significant larger regression coefficient of circadian effect than developmental cells.
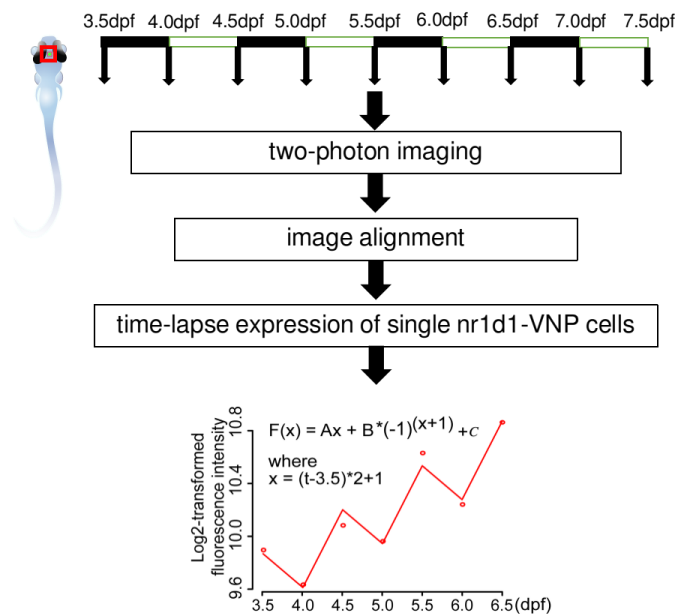

Figure 3.4 The workflow for imaging the pineal cells of *nr1d1*:VNP fish during development.

According to the scRNA-seq data, *nr1d1*:VNP were expressed in both rod-like and cone-like photoreceptor cells in the pineal gland. By crossing the *nr1d1*:VNP fish with Tg(*xops*:nfsB-mCherry) fish line labeling the rod cell and Tg(*lws2*:nfsB-mCherry) labelling the cone cells, I confirmed that *nr1d1*:VNP co-expressed with both cone cell marker (*lws2*) and rod cell marker (*xops*). Furthermore, we did cell-type-specific imaging with the fish labeled by both *nr1d1*:VNP and *xops*:nfsB-mCherry. I found that the baseline expression level of rod-like cells were significantly higher compared to the non-rod-like cells. In summary, during early development of zebrafish, circadian oscillation of *nr1d1*:VNP positive cells in the pineal was superimposed on developmental trends that exhibit significant cell-type specificity.

In zebrafish, it has been shown that exposure to several LD cycles is essential for the development of the circadian clock (Kaneko and Cahill, 2005). However, it is still unclear whether the onset of circadian clock is a synchronization of existing cellular oscillators or a *de novo* initiation of the individual cellular clock. To address this question, my colleagues and I imaged *nr1d1*:VNP fish raised under DD condition. I observed that the cellular circadian oscillations in the fish raised under DD condition were extremely dampened, while the developmental trend of clock formation was not affected. I did not observe pineal cells become circadian spontaneously under DD condition *in vivo* at single-cell level. Our data suggested that early exposure to the LD cycle during development gets the oscillators of individual cells started rather than synchronizes the existing oscillators of individual cells as people in the field have previously believed. This shows the usefulness of the single-cell circadian reporter.

In conclusion, we provided not only an important tool for studying circadian clock at the single-cell level *in vivo* but also got novel insights on the development of circadian clock in general.

3.2.2 Declaration of my contribution

Conceptualization: Haifang Wang, Jie He, Yuanhai Li, Jun Yan.

Data curation: Haifang Wang, Xingxing Li.

Formal analysis: Haifang Wang, Zeyong Yang.

Funding acquisition: Haifang Wang, Zeyong Yang, Jun Yan.

Investigation: Haifang Wang, Zeyong Yang, Xingxing Li, Dengfeng Huang, Shuguang Yu.

Methodology: Haifang Wang, Zeyong Yang, Xingxing Li, Shuguang Yu.

Validation: Haifang Wang, Xingxing Li.

Visualization: Haifang Wang.

Writing – original draft: Haifang Wang, Jun Yan.

Writing – review & editing: Haifang Wang, Jie He, Yuanhai Li, Jun Yan.

Single-cell in vivo imaging of cellular circadian oscillators in zebrafish

Haifang Wang[ID]1,2☯, Zeyong Yang3,4,5☯, Xingxing Li6☯, Dengfeng Huang1,2, Shuguang Yu1,2, Jie He[ID]1,2*, Yuanhai Li6*, Jun Yan[ID]1,2,7*

1 Institute of Neuroscience, State Key Laboratory of Neuroscience, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China, **2** Shanghai Center for Brain Science and Brain-Inspired Intelligence Technology, Shanghai, China, **3** Department of Anesthesiology, International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, **4** Shanghai Key Laboratory of Embryo Original Disease, Shanghai, China, **5** Shanghai Municipal Key Clinical Specialty, Shanghai, China, **6** Department of Anesthesiology, First Affiliated Hospital of AnHui Medical University, Hefei, Anhui, China, **7** School of Future Technology, University of Chinese Academy of Sciences, Beijing, China

☯ These authors contributed equally to this work.
* junyan@ion.ac.cn (JY); liyuanhai-1@163.com (YL); jiehe@ion.ac.cn (JH)

Abstract

The circadian clock is a cell-autonomous time-keeping mechanism established gradually during embryonic development. Here, we generated a transgenic zebrafish line carrying a destabilized fluorescent protein driven by the promoter of a core clock gene, *nr1d1*, to report in vivo circadian rhythm at the single-cell level. By time-lapse imaging of this fish line and 3D reconstruction, we observed the sequential initiation of the reporter expression starting at photoreceptors in the pineal gland, then spreading to the cells in other brain regions at the single-cell level. Even within the pineal gland, we found heterogeneous onset of *nr1d1* expression, in which each cell undergoes circadian oscillation superimposed over a cell type–specific developmental trajectory. Furthermore, we found that single-cell expression of *nr1d1* showed synchronous circadian oscillation under a light–dark (LD) cycle. Remarkably, single-cell oscillations were dramatically dampened rather than desynchronized in animals raised under constant darkness, while the developmental trend still persists. It suggests that light exposure in early zebrafish embryos has significant effect on cellular circadian oscillations.

Introduction

Circadian rhythm evolves to align animal behaviors to periodic daily environmental changes. At the molecular level, the vertebrate circadian clock is mainly generated through transcriptional/translational feedback loops of core clock genes [1]. Among them, two transcription factors (TFs), BMAL1 (also known as ARNTL or MOP3) and CLOCK form heterodimers to bind to E-boxes in the promoters and initiate the transcription of their target genes [2–4], including Per family genes (*Per1*, *Per2*, and *Per3*) and Cry family genes (*Cry1* and *Cry2*). The activation of these genes results in the formation of the PER/CRY complex and thereby inhibits CLOCK/

BMAL1 transcriptional activity, forming a negative feedback loop [5]. Nuclear receptor, REV-ERBα (also known as NR1D1), represses the transcription of *Bmal1* and itself is under the transcriptional regulation of BMAL1/CLOCK, giving rise to the second negative feedback loop of the circadian clock [6]. The genome-wide regulation by circadian TFs such as BMAL1/CLOCK and REV-ERBα typically leads to thousands of genes showing circadian expression in a given tissue. Although the basic network of core circadian genes is present in almost every cell, many of the circadian-controlled genes are tissue specific or cell type specific. Their circadian expression is a result of either tissue-specific binding of circadian TFs [7] or transcriptional cascade from tissue-specific TFs regulated by circadian TFs [8]. At the organismal level, the overt circadian rhythm is governed by an intricate network of circadian oscillators in which the master pacemaker, such as suprachiasmatic nuclei (SCN) in mammals or the pineal gland in zebrafish, is believed to play a pivotal role [1,9]. Traditional in vivo or ex vivo studies of the cellular circadian clock have relied on luciferase reporter systems driven by core clock gene promoters [10,11]. But luciferase imaging can only achieve single-cell resolution in organotypic slices in culture with a high-resolution charge-coupled device (CCD) camera. Transgenic zebrafish lines carrying luciferase reporters driven by core clock gene promoters have been developed in larval zebrafish. But in vivo luciferase imaging of these fish lines lacked single-cell resolution and can only report the population-level circadian rhythm [10,12]. Genetically encoded calcium indicator (GCaMP) has been widely used to monitor in vivo single-cell calcium activity. In comparison, there is still a lack of zebrafish line to report the circadian expression at the single-cell level in vivo.

Circadian rhythm has to be established at molecular, cellular, tissue, and behavioral levels during animal development. Day-night rhythms in the fetal rat SCN are first detected between embryonic day E19 and E21 [13]. Rhythmic expression of circadian clock genes is not detected in both in vitro and in vivo mouse embryonic stem cells (ESCs) but only appears when ESCs differentiate into neural stem cells [14,15]. By ex vivo luciferase imaging of mouse fetal SCN, Carmona-Alcocer and colleagues showed that a few cells in SCN start circadian oscillations on E14.5, widespread synchronized oscillations were formed on E15.5, and then a dorsal-ventral phase wave was established at postnatal day P2 [16]. Zebrafish (*Danio rerio*) embryos that are in vitro fertilized and transparent provide an accessible model organism for in vivo live imaging, and thereby are being widely used in the study of animal development. A functional circadian clock, characterized by free-running activity, rhythmic cell cycle, and circadian gene expression, is established after hatching in zebrafish [17]. However, how single-cell circadian clocks were established during embryonic development is still unclear. It is well documented that early exposure to a light–dark (LD) cycle is required for the development of the circadian clock in zebrafish larva [10]. But it is still under debate whether the effect of light on clock development is a synchronization of already existing oscillators or an initiation of single-cell clocks. Questions like these can only be addressed by time-lapse in vivo imaging of single-cell circadian clock reporter.

Here, we report a transgenic zebrafish line using destabilized fluorescent protein, Venus-NLS-PEST (VNP), driven by the promoter of a key circadian clock gene, *nr1d1*. This system allows us to monitor the development of single-cell circadian rhythm in live zebrafish larva in a cell type–specific manner. We observed that VNP reporter expression undergoes stepwise onset starting at the photoreceptor cells in the pineal gland, then spreads to cells in other brain regions. Using single-cell RNA-seq (scRNA-seq), we characterized the cell types expressing VNP in the whole brain. Within the pineal gland, we found that each cell undergoes circadian oscillation superimposed over cell type–specific developmental trajectories. Under LD cycle, cellular expression of VNP shows synchronous circadian oscillation. However, the circadian expression of *nr1d1*:VNP-positive cells was dramatically dampened rather than

desynchronized, while the developmental trend was still present at the single-cell level in fish raised under constant darkness. Our result suggests that the early exposure of LD cycle is crucial for the ontogeny of functional single-cell oscillators.

Results

Screening for in vivo circadian reporters in zebrafish

To monitor circadian rhythm at the single-cell level in live animals, we have screened for in vivo circadian reporters among various combinations of destabilized fluorescent proteins driven by core clock gene promoters in larval zebrafish. We first tested peTurboGFP-dest1 (TGFPD1) encoding a destabilized variant of green fluorescent protein TurboGFP. The plasmids of *bmal1a/bmal2/per2/nr1d1*:TGFPD1 were constructed by homologous recombination. We observed that none of these plasmids were expressed in F1 embryos. We then tested the plasmids containing core clock gene promoters driven VNP, another form of destabilized fluorescent protein. We found that the plasmids of *cry2b/bmal1a/per2*:VNP were not expressed in F1 zebrafish embryos, while *per1a*:*VNP* has too low expression to be used for in vivo imaging. For *nr1d1*:VNP, it has been reported that zebrafish *nr1d1* gene has two promoters, i.e., ZfP1 (distal) and ZfP2 (proximal). ZfP1 is conserved and functionally similar to mammalian *Nr1d1* promoter [18]. We found that the F1 fish containing only the proximal *nr1d1* promoter ZfP2 (1.5 kbp) failed to drive VNP expression. But *nr1d1*:VNP containing both ZfP1 and ZfP2 (6.2 kbp) showed robust expression in zebrafish embryos (Fig 1A and Table A in S2 Data). Therefore, we chose *nr1d1*:VNP containing ZfP1 and ZfP2 promoters as the circadian reporter and generated the transgenic fish line Tg(*nr1d1*:VNP) by the transposable element of *Oryzias latipes* (Tol) 2 system.

The 6.2-kbp ZfP1 and ZfP2 promoter region to drive VNP expression includes the entire set of known *cis*-regulatory elements: E-box, RRE, Crx-, and Otx5-binding sites (Fig 1A). We measured the mRNA levels of endogenous *nr1d1* and VNP expression in the Tg(*nr1d1*:VNP) transgenic fish from 3.5 days postfertilization (dpf) to 7.5 dpf under the 12-hour/12-hour LD cycles using real-time PCR. *nr1d1* and VNP showed highly correlated expression patterns (Pearson's r = 0.9) indicating that VNP expression faithfully reported the expression of *nr1d1* at the mRNA level. In addition, both genes showed a higher expression level at the dawn than dusk over days (Fig 1B), which is consistent with our previous result that *nr1d1* mRNA shows circadian expression peaking at ZT23 (Zeitgeber time) in larval zebrafish [8]. Previous studies have shown that the rhythmic *period* gene expression starts from the second day after birth [19,20]. We found that *nr1d1* mRNA also starts oscillating from the second day of the development, as early as the *per1b* gene (S1A Fig). Therefore, *nr1d1* reporter exhibits similar developmental profiles as *per1b* and can be used for studying when and how the cellular clock is developed.

We examined the spatial distribution of fluorescence-labeled cells in *nr1d1*:VNP at 7.5 dpf using an in vivo two-photon imaging system and found *nr1d1*:VNP-positive cells in many brain regions, including the pineal gland, the optic tectum, and the cerebellum (Fig 1C, S1 and S2 Movies). In particular, we observed that *nr1d1*:VNP-positive cells are most highly expressed in the pineal gland. We next monitored in vivo expression of *nr1d1*:VNP at the whole-brain scale from 3.5 dpf to 7.5 dpf at ZT0 and ZT12 (Fig 1D) by time-lapse imaging. We aligned the time series of 3D images using CMTK toolkit (see Methods) so that we can trace each single cell in the same fish over time. In this way, we observed that the expression of *nr1d1*:VNP was sequentially turned on and underwent gradual increase in distinct brain regions during development, as early as 3.5 dpf in the pineal gland, followed by the optic tectum at 5.5 dpf, and other brain regions such as the cerebellum at later time points (Fig 1E, Table B in S2 Data, S1
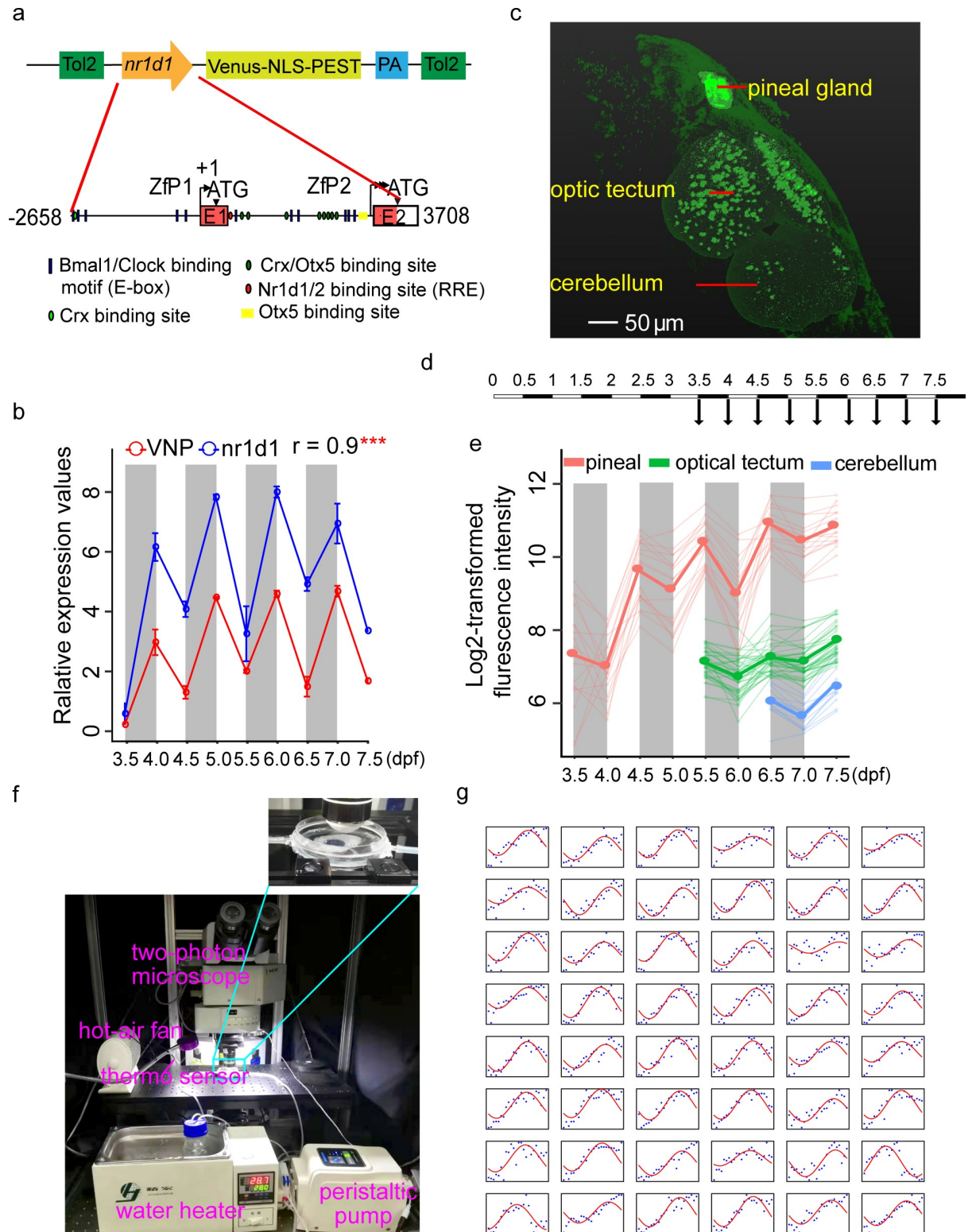
Fig 1. Construction of in vivo circadian reporter in zebrafish. (a) The upper graph shows the schematic of *nr1d1*:VNP construct design. The lower graph shows a magnified view of the *nr1d1* promoter sequence used for driving the circadian expression of VNP. The putative RRE (Nr1d1/2 binding site), E-box (Bmal1/Clock binding site), Crx, Otx5, and Crx/Otx5 binding sites were indicated by red oval, blue rectangle, light green oval, yellow rectangle, and dark green oval, respectively. (b) Plot of real-time PCR results of *nr1d1* and *Venus-pest* expression. Each time point has 2 replicates, and each replicate is the pool of 7–10 fish. The dots show the original value, while the solid lines and the bars

represent mean ± SEM. (c) Whole-brain two-photon image shows the spatial distribution of *nr1d1*:VNP-positive cells in different brain regions at 7.5 dpf. (d) Experimental design to examine the developmental dynamics of *nr1d1*:VNP expression in the whole zebrafish brain. (e) Single-cell tracing result of *nr1d1*:VNP-positive cells in the pineal gland (red), the optic tectum (green), and the cerebellum (blue) for one fish. The cells were tracked from 3.5 dpf in the pineal gland, 5.5 dpf in optic tectum, and 6.5 dpf in cerebellum, respectively. Each thin line represents one cell, while the thick line represents the mean value of cells in each brain region. (f) Illustration of the setup for time-lapse imaging. The fish were fixed in a chamber with circulating egg-water without anesthesia at temperature of 28 ± 0.5°C. (g) Single-cell tracing results of all 48 *nr1d1*:VNP-positive cells imaged at 1-hour resolution. The blue dots represent the original fluorescence signals, while the solid red line represents the smoothed curve fitted by the cosine functions. The numerical values for panels b, e, and g are in S1 Data. dpf, days postfertilization; PA, poly(A) site; Tol2, transposable element of *Oryzias latipes* 2; VNP, Venus-NLS-PEST.

and S2 Movies). In addition, single-cell tracing revealed that circadian oscillations of *nr1d1*:VNP expression appear to peak at ZT12 across brain regions (Fig 1E). The expression pattern of *nr1d1*:VNP closely resembled that of endogenous *nr1d1* reported by Delaunay and colleagues [19]. Thus, *nr1d1*:VNP fish can be used as a reporter for both developmental expression and circadian expression of endogenous *nr1d1* gene in zebrafish. To perform high temporal resolution imaging, we imaged the fish every hour for one day starting at 5 dpf (Fig 1F and 1G, S3 Movie, Table C in S2 Data) by a time-lapse imaging system, as illustrated in Fig 1F. The fishes were embedded in agarose gel and kept alive during the entire imaging process. We observed that 39 out of 48 traced cells in the pineal gland showed circadian oscillations using JTKcycle adjusted $P < 0.05$ and absolute oscillating amplitude >100 as the cutoff (Fig 1G and S1B and S1C Fig). This shows that *nr1d1*:VNP fish can report cellular circadian oscillations under continuous imaging.

Characterization of *nr1d1*:VNP-expressing cells

To identify the cell types expressing *nr1d1*:VNP in the whole brain, we conducted single-cell RNA-seq (scRNA-seq) of approximately 15,000 cells dissociated from the brain of Tg(*nr1d1*:VNP) larval fish at 6.5 dpf. Among them, 6,514 cells detected with more than 500 genes were used for the downstream analysis. In total, 26 cells clusters were classified from scRNA-seq. They were manually annotated by comparing the marker genes with those from whole-brain scRNA-seq data in adult zebrafish [21] (Fig 2A). Wilcoxon Rank Sum test implemented in Seurat package was applied to calculate the significance of enrichment of VNP signal in each brain region. Using adjusted $P < 0.01$ as the cutoff, we found that the mRNA of *nr1d1*:VNP was most highly enriched in cell clusters of photoreceptors in the pineal gland, which is consistent with the imaging data and leads us to focus on the pineal gland in the following analysis. The mRNA of *nr1d1*:VNP was also enriched in glutamatergic neurons in forebrain and hypothalamus, granule cells in cerebellum, dorsal habenula cells, as well as some non-neuron cells such as oligodendrocyte, retinal-pigment epithelium-like cells, and endothelial cells (Fig 2B).

We next examined the expression of *nr1d1*:VNP in the pineal gland at the single-cell resolution. Fig 2C showed an example of 3D reconstruction of *nr1d1*:VNP signals in the pineal gland. At 4.5 dpf, each pineal gland contains about 60 *nr1d1*:VNP-positive cells. To better understand the cell type of these cells, we crossed Tg(*nr1d1*:VNP) with Tg(*aanat2*:mRFP) [12]. Averaged 3D cell density distribution of *nr1d1*:VNP-positive cells across six fish (Methods section) revealed that *nr1d1*:VNP-positive cells were mostly distributed around the lumen of the pineal gland, while the cell density was relatively low in the central part of the pineal gland (Fig 2D). We observed that *nr1d1*:VNP signals were highly overlapped with the *aanat2*:mRFP signals (94%) indicating that most *nr1d1*:VNP-positive cells were melatonin-synthesizing photoreceptor cells in the pineal gland (Fig 2C, S4 Movie). According to our scRNA-seq data, VNP is expressed in both rod-like photoreceptors and cone-like photoreceptors (Fig 2E). Indeed, we observed co-labeling of VNP with rod cell marker (*xops*) (10%) and cone cell marker (*lws2*)
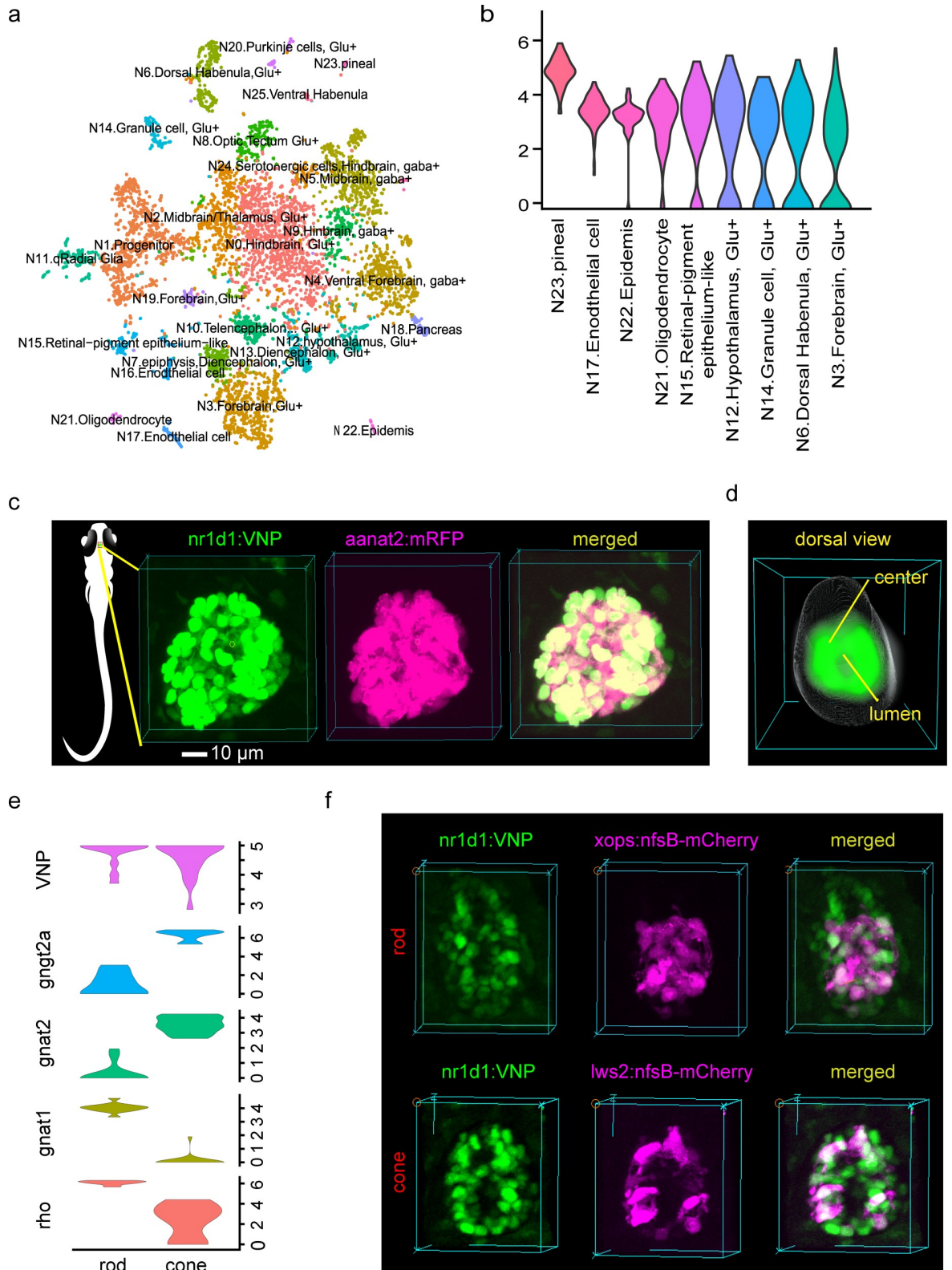
Fig 2. Characterization of *nr1d1*:VNP expressing cells. (a) t-SNE visualization of brain cell clusters. The clusters were annotated by comparing to the adult zebrafish scRNA-seq data. (b) Violin plot demonstrated the clusters with enriched expression of *nr1d1*:VNP. *Y* axis denotes the normalized expression value. Note the highest levels in the pineal gland. (c) Fluorescence images show the co-expression of

nr1d1:VNP (nuclear signal) and *aanat2*:mRFP (cytoplasmic signal) in the zebrafish pineal gland. The left graph illustrates the 3D location of the pineal gland in zebrafish larva. (d) Three-dimensional reconstruction of a common zebrafish pineal gland by aligning and averaging of six fish. The gray sphere represents the boundary of the pineal gland and the green color represents the density distribution of *nr1d1*:VNP-positive cells in the pineal gland. (e) Violin plot showed the expression of rod cell markers (*rho*, *gnat1*), cone markers (*gnat2*, *gngt2a*), and *nr1d1*:VNP in photoreceptor clusters. *Y* axis denotes the normalized expression value. (e) Fluorescence images showed the co-expression of *nr1d1*:VNP (nuclear signal) with the Tg(*xops*:nfsB-mCherry) fish line (cytoplasmic signal) and Tg(*lws2*:nfsB-mCherry) fish line (cytoplasmic signal) in the zebrafish pineal gland. The numerical values for panel b and e are in S1 Data. scRNA-seq, single-cell RNA-seq; t-SNE, *t*-distributed stochastic neighbor embedding; VNP, Venus-NLS-PEST.

https://doi.org/10.1371/journal.pbio.3000435.g002

(21%) in the pineal gland by crossing the Tg(*nr1d1*:VNP) fish line with Tg(*xops*:nfsB-mCherry) and Tg(*lws2*:nfsB-mCherry) fish lines, respectively (Fig 2F). Our scRNA-seq data also suggested that *nr1d1*:VNP is expressed in proliferative cells. Indeed, when we crossed our VNP fish line with Tg(*her4*:DsRed) zebrafish marking proliferative cells [22], we found co-labeling of VNP and red fluorescent protein from *Discosoma* (DsRed) in the pineal at 5.5 dpf (26%) (S1D Fig). This result indicated that a small number of photoreceptors could be still proliferating at 5 dpf. The co-labeled cells may include a small number of glial cells, which also express high levels of *her4*, as scRNA-seq data suggested that *nr1d1* is also expressed in non-photoreceptor cells such as glial cells.

Developmental dynamics of *nr1d1*:VNP expression within the pineal gland

To examine the development of circadian oscillations more closely in the pineal gland, we imaged *nr1d1*:VNP signals in the pineal gland from 3.5 dpf to 6.5 dpf using two-photon imaging (Fig 3A and 3B, S5–S10 Movies, Table D in S2 Data). The time series of 3D images were first aligned using CMTK toolkit. Then, the fluorescent intensities of each single cell in the aligned images were traced by Trackmate (a Fiji plug-in) [23] (details in Methods). The majority of cells already showed higher *nr1d1*:VNP signals in dusk than dawn when the signals first become detectable in the pineal gland (S2 Fig). This pattern does not change significantly during development (ANOVA *P* = 0.78) (S2 Fig). However, we found that VNP-positive cells within the pineal gland display heterogeneous temporal patterns of *nr1d1*:VNP expression over development (Fig 3C). Some cells showed rapid increase in the baseline level, while other cells showed more robust circadian oscillations. We then quantified the circadian and developmental components in each cell by fitting the time-series data with a regression model that combines a stepwise function of time, $(B^*(-1)^{(x+1)})$, denoting circadian oscillation, with a linear function of time, $(Ax + C)$, denoting the developmental increase (Fig 3D). After fitting the model, we found that the regression coefficients of developmental effect (A) and circadian oscillation (B) both vary widely across cells, but there is a negative correlation between them (Fig 3E).

To reveal if the heterogeneity of single-cell temporal profiles within the pineal gland is due to the differences in cell type, we next conducted cell type–specific imaging by crossing *nr1d1*:VNP with the fish line, labeling rod-like cells in red fluorescent protein Tg(*xops*:nfsB-mCherry). After imaging *nr1d1*:VNP signals in the rod-like and non-rod-like cells simultaneously in the same fish during development from 3.5 dpf to 6.5 dpf (Table E in S2 Data, S11–S16 Movies), we found that rod-like cells have higher level of baseline expression than non-rod-like cells (Fig 3F and 3G). Taken together, pineal photoreceptor cells of zebrafish larvae undergo circadian oscillations superimposed over cell type–specific developmental trajectories.

Single-cell circadian oscillations in the pineal gland

We next monitored the *nr1d1*:VNP signals in the pineal glands of live zebrafish larvae with temporal resolution every 2 hours at 5 dpf using two-photon imaging (Fig 4A, S17 and S18

a

b

3.5dpf 4.0dpf 4.5dpf 5.0dpf 5.5dpf 6.0dpf 6.5dpf

20μm

c

d

$F(x) = Ax + B*(-1)^{(x+1)} + C$

where
$x = (t-3.5)*2+1$

f

rod   non-rod

e

$r = -0.35$ ***

g

***
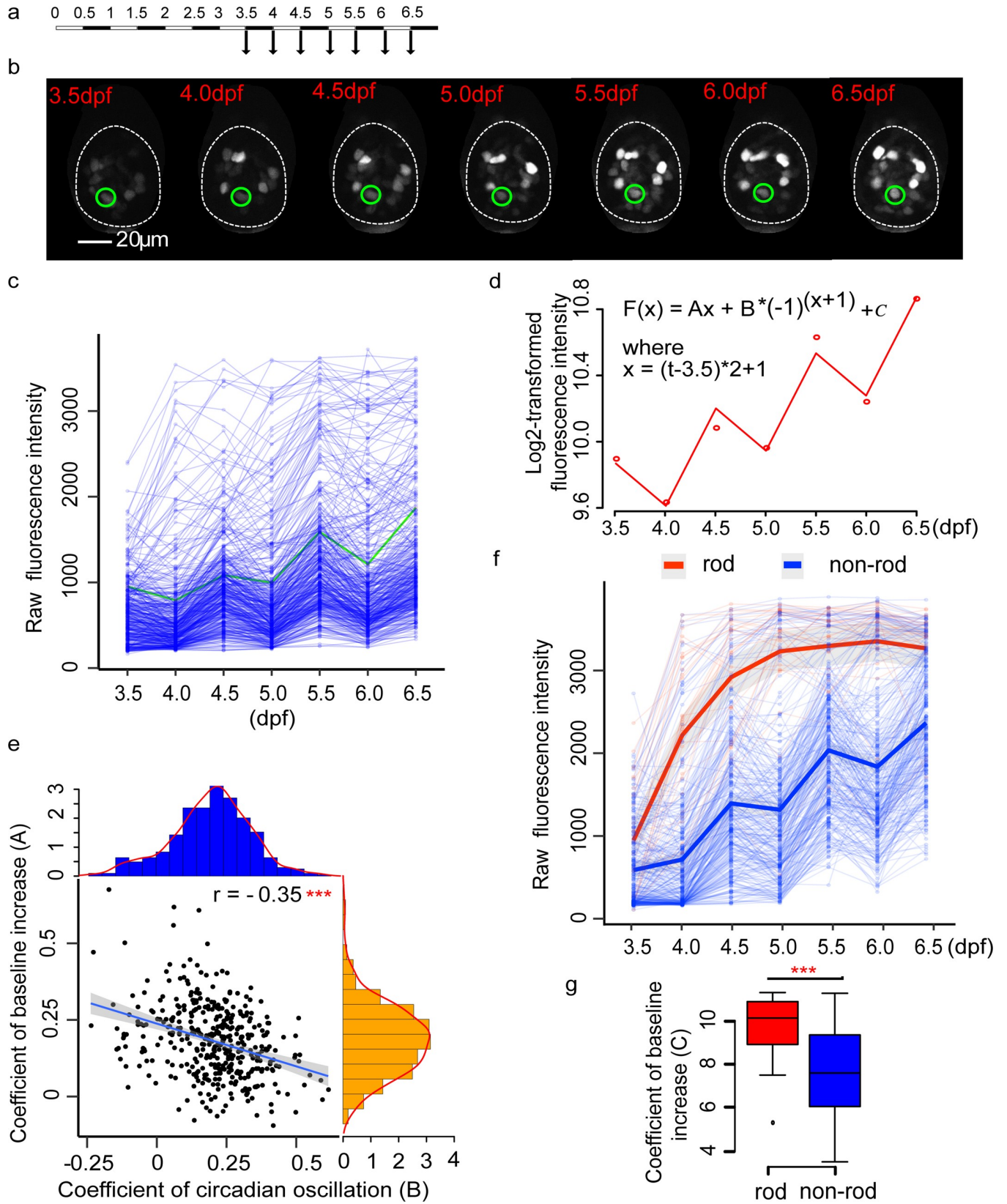
Fig 3. Developmental dynamics of *nr1d1*:VNP expression. (a) Experimental design to examine the developmental dynamics of *nr1d1*:VNP expression. (b) Fluorescence images illustrate the tracing of one *nr1d1*:VNP cell during development from 3.5 dpf to 6.5 dpf. This cell was highlighted in green in Fig 3C. (c) Raw fluorescence intensity of the traced cells (six fish). (d) Regression model used to fit the fluorescence intensity (the cell is illustrated in Fig 3B and highlighted in Fig 3C). The dots represent the fluorescence intensity, while the line represents the fitting curve. (e) Scatterplot demonstrated the relationship between coefficients of developmental effect (A) and circadian oscillation (B). The histograms in orange and blue showed the distribution of coefficients of developmental effect (A) and the distribution of coefficients of circadian oscillations (B), respectively. ***Pearson's correlation *P* < 0.001. (f) Raw fluorescence intensity of rod-like cells and non-rod-like cells (six fish). Each thin line represents one cell and each dot represents raw fluorescence intensity. The thick lines represent the loess-smoothed curves for all rod-like cells in red and non-rod-like cells in blue, respectively. The shaded areas show the 95% confidence level of the smoothed curve. (g) A comparison of baseline expression (C) between rod-like and non-rod-like cells. Two-tailed Student *t* test was applied to calculate the levels of significance between the two types of cells. **P* < 0.01. The numerical values for panels c–g were in S1 Data. dpf, days postfertilization.

https://doi.org/10.1371/journal.pbio.3000435.g003

Movies). We traced every *nr1d1*:VNP-positive cell at a different time of the day (Fig 4B, Table F in S2 Data). Fig 4C showed the trace plots of individual cells of two zebrafish throughout the day. JTKcycle was applied to calculate the circadian phase and amplitude of each cell. Using JTKcycle adjusted *P* < 0.05 and absolute oscillating amplitude >100 as the cutoff, we identified 99 cells showed circadian oscillations in a total of 117 *nr1d1*:VNP-positive cells in the pineal gland. Their circadian phases were distributed around ZT12 within a narrow range (Fig 4D), with Kuramoto order parameter R = 0.86 indicating the phase coherence of single-cell circadian oscillators [24]. Clustering analysis of single-cell VNP traces identified two distinct clusters of cells that can be distinguished by their baseline fluorescence intensities (Fig 4E and 4G). Such difference in baseline level of expression at 5 dpf can be explained by the cell type–specific developmental trajectories that we observed above. We also observed that the cluster with higher baseline level (cluster 2) also showed lower relative amplitude of oscillation (Fig 4H). Similar negative correlation between mean transcriptional activity and relative amplitude of circadian oscillations has been reported in mouse [25]. However, there is no significant difference in circadian phase between the two clusters (S3 Fig). In short, circadian clocks in pineal photoreceptor cells are oscillating synchronously under LD in spite of the large difference in baseline level.

LD cycle is essential for cellular clocks

It is known that LD cycle is required for the development of the circadian clock in zebrafish [10]. However, it is unclear whether the role of LD cycle in clock ontogeny is a synchronization of individual oscillators or an initiation of oscillation. To address this question, we imaged *nr1d1*:VNP signals in zebrafish larvae raised under constant dark (DD) from 3.5 dpf to 6.5 dpf using two-photon imaging (Fig 5A, S19–S22 Movies, Table G in S2 Data). Examination of the expression patterns of all *nr1d1*:VNP-positive cells from four DD fish showed that, although the expression of *nr1d1*:VNP still increased during development, no obvious oscillations were observed in DD cells (Fig 5B). Indeed, regression analysis showed that the oscillation coefficients defined as the stepwise function of the majority of DD cells are not significantly different from zero (Fig 5C). However, DD cells have even higher developmental coefficients than LD cells (Fig 5D). In comparison, when we imaged the fish transferred into DD condition after six LD cycles (LD_DD cells) (Fig 5E–5G, Tables H–I in S2 Data, S23–S28 Movies), we observed that LD_DD cells were still oscillating into two days of darkness (Fig 5F) and the amplitude of circadian oscillation showed no significant difference compared to the fish kept under LD condition (LD_LD cells) (Fig 5G). To further examine the circadian oscillation in dark-raised fish, we imaged *nr1d1*:VNP signals in the pineal gland every 2 hours at 5 dpf (Fig 5H, Table J in S2 Data, S29–S31 Movies). JTKcycle (from MetaCycle package in R) was applied to calculate the circadian phase and amplitude of each cell. We found that DD cells showed much lower oscillating amplitudes than those in LD cells (Fig 4I and 4J). Using JTKcycle adjusted *P* < 0.05 and absolute amplitude >100 as the cutoff, only 3 out of 142 DD cells showed significant

a

b

c

d Distribution of circadian phase

e cluster 1 cluster 2

f

g

h

i

j

Fig 4. Circadian dynamics of the pineal gland at higher temporal resolution (two fish). (a) Experimental design to examine the developmental dynamics of *nr1d1*:VNP expression at higher temporal resolution. (b) Fluorescence images illustrate the tracing of one *nr1d1*:VNP example cell every 2 hours. (c) Single-cell tracing results of all 117 *nr1d1*:VNP-positive cells in two zebrafish pineal glands (two fish). The blue dots represent the original fluorescence signals, while the solid red line represents the smoothed curve fitted by the cosine functions. The example cell in (b) is highlighted by a green square. (d) Circadian phase distributions of the 117 *nr1d1*:VNP-positive cells. (e) t-SNE visualization of the clustering result of the 117 *nr1d1*:VNP-positive cells. (f) Raw fluorescence intensity traces of the two types of *nr1d1*:VNP-positive cells in (e). (g and h) The comparison of the baseline expression (g) and relative circadian amplitude (h) between the two types of cells. The colors of the boxes correspond to (e). Two-tailed Student *t* test was applied to calculate the levels of significance between the two types of cells. ***$P < 0.001$. (i and j) Scatterplot demonstrates the relationship between baseline intensity and relative oscillation amplitude for fish1 (i) and fish2 (j), respectively. ***Pearson's correlation $P < 0.001$. The numerical values for panels c–j are in S1 Data. dpf, days postfertilization; t-SNE, *t*-distributed stochastic neighbor embedding; ZT, Zeitgeber time.

oscillations, compared to 99 out of 117 LD cells showing significant oscillations (Fig 5K). Cosine fitting with 24-hour period and shifting phases as described in our previously paper [26] was also applied to confirm this result, using cosine fitting $P < 0.05$ and relative oscillating amplitude >0.05 as the cutoff. Only 1 out of 142 DD cells showed significant oscillation, consistent with the JTKcycle result (S4A–S4C Fig). Time-lapse imaging in every hour at 5 dpf of DD cells further confirmed this result (S32 Movie, Table K in S2 Data). In fact, no oscillating cells were identified using JTKcycle adjusted $P < 0.05$ and absolute amplitude >100 as the cutoff (S4D–S4G Fig). Therefore, our result suggests that the cellular circadian oscillations in DD fish were severely dampened rather than desynchronized during the development.

Discussion

The mouse version of *nr1d1*:VNP was originally developed by Nagoshi and colleagues in NIH 3T3 cells. VNP was chosen for its high fluorescence intensity and high folding efficiency [27]. With this reporter, they were able to show that each NIH 3T3 fibroblast cell has a self-sustained circadian oscillator that can be entrained by serum shock. Also using this reporter, two independent studies have demonstrated the tight coupling between circadian rhythm and cell cycle in proliferating NIH 3T3 cells [28,29]. But the application of *nr1d1*:VNP in circadian research has so far been limited in cell lines. In our study, we constructed live zebrafish carrying *nr1d1*:VNP fluorescence reporter, and for the first time, we can monitor the circadian gene expression at the single-cell resolution in zebrafish larvae. Using this reporter fish, we revealed the interplays among circadian clock, cell type–specific development, and LD cycle. It is noted that caveats of using a transgenic technology rather than a site-specific knock-in strategy may exist, such as the position effects of the reporter insertion. As fluorescent, ubiquitination-based cell cycle indicator (FUCCI) lines to label different stages of cell cycle are available in zebrafish [30], one can cross *nr1d1*:VNP fish with cell-cycle reporter lines to examine the relationship between circadian clock and cell cycle in vivo. By crossing fish lines containing cell type–specific fluorescent markers with our *nr1d1*:VNP fish, one can conduct cell type–specific imaging of circadian rhythm, as we have demonstrated for rod-like cells in the pineal gland. Larval zebrafish has also been used in drug screening for chemical compounds affecting circadian rhythm and sleep [31]. Our *nr1d1*:VNP fish could be used to investigate the effect of selected drug targets on the synchronization of single-cell oscillators in vivo. With advanced microfluidic systems to capture and recapture larval zebrafish, one can conduct more intense imaging on the same fish to achieve higher temporal resolution while minimizing the perturbations to the fish [32]. Using a newly developed movable platform to track zebrafish larvae and a novel volume imaging system, live imaging of cellular circadian rhythm on free-moving zebrafish will become possible [33]. In summary, our single-cell circadian reporter zebrafish line will have broad applications in circadian research.
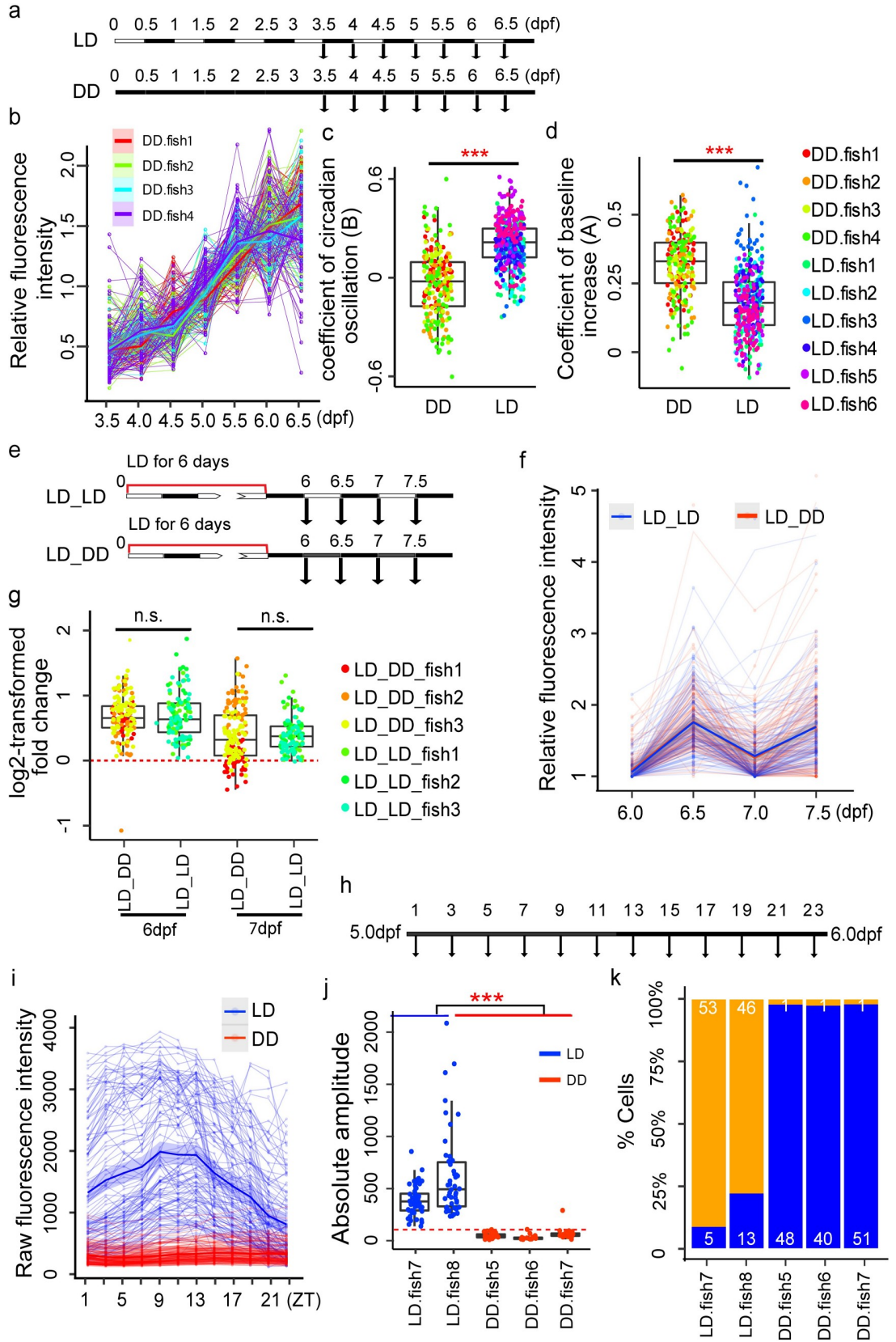
Fig 5. LD cycle is essential for *nr1d1* oscillation. (a) Experimental design to investigate the effect of light on the onset of circadian clock development. (b) Expression patterns of all cells under DD condition during development (four fish). Each thin line represents one cell. Each thick line represents the loess-smoothed curve of all cells in each individual fish. The shaded area shows the 95% confidence level of the smoothed curve. (c) Comparison of oscillating coefficient (B) between LD (six fish) and DD (four fish) cells. Two-tailed Student *t* test was applied to calculate the levels of significance between the two types of cells. ***$P < 0.001$. (d) Comparison of developmental coefficient (A) between LD (six fish) and DD (four fish) cells. Two-tailed Student *t* test was applied to calculate the levels of significance between the two types of cells. ***$P < 0.001$. (e) Experimental design to examine the single-cell circadian clocks under DD after transferred from LD (LD_DD). (f) Expression patterns of LD_LD (three fish) and LD_DD cells (three fish) from 6.0 to 7.5 dpf. Each thin line represents one cell and each dot represents raw fluorescence intensity. Thick lines represent the loess-smoothed curves for all the LD_DD cells in red and LD_LD cells in blue, respectively. The shaded areas show the 95% confidence level of the smooth curve. (g) Log2-transformed dusk–dawn ratios of VNP fluorescence intensities of all the cells in (f). Each data point represents one cell and the colors correspond to different fish. The red dashed line represents $y = 0$. n.s. represents no significant difference in dusk–dawn ratios between different time points (two-tailed ANOVA test). (h) Experimental design to examine the expression pattern of DD cells at higher temporal resolution. (i) Expression patterns of LD (two fish) and DD cells (three fish) across one day at 2-hour resolution. Each thin line represents one cell and each dot represents raw fluorescence intensity. Thick lines represent the loess-smoothed curves for all the LD in blue and DD cells in red, respectively. The shaded areas show the 95% confidence level of the smooth curve. (j) Comparison of the absolute amplitudes of LD and DD cells in (i) from JTKcycle. The red dashed line represents $y = 100$. (k) Percentages of oscillating cells (JTKcycle adjusted $P < 0.05$ and absolute amplitude >100) in each LD and DD fish. The orange bars represent the percentage of oscillating cells, while the blue bars represent the percentage of non-oscillating cells. The numerical values for panels b, c, d, f, g, i, j, and k were in S1 Data. DD, constant dark; dpf, days postfertilization; LD, light–dark; VNP, Venus-NLS-PEST; ZT, Zeitgeber time.

https://doi.org/10.1371/journal.pbio.3000435.g005

It has been shown that physiological and behavioral rhythms in zebrafish appear during the larval stages of zebrafish [34–36]. In our study, the expression of *nr1d1* appeared gradually in development and stepwise in different brain regions. As the major component of the second negative feedback loop in circadian clock, *nr1d1* is directly regulated by BMAL1 and represses the expression of *bmal1* [1]. The onset of *nr1d1* expression may represent the maturation of a functional circadian clock during development. Our single-cell study revealed the developmental trajectory of *nr1d1* expression varies greatly among individual cells, and the onset of circadian rhythm is dependent on early light exposure. Ziv and colleagues have shown that *per2* expression can be induced by light and is necessary for light-dependent onset of circadian oscillation of *aanat2* in the pineal gland [37]. In addition, *otx5*, a TF constantly expressed in circadian cycle, has been shown to regulate the expression of *nr1d1* in the pineal gland [38]. It is likely that *otx5/crx* are responsible for cell type–specific developmental components of *nr1d1* expression. Nagoshi and colleagues have suggested that the robust expression of *Nr1d1*: VNP in mouse NIH 3T3 cells is due to high transcriptional activity of *Nr1d1* promoter in these cells, as *Nr1d1* is more highly expressed in tissues such as liver compared to other core clock genes such as *Bmal1* [27]. In our case, the presence of *otx5* and *crx* binding sites in *nr1d1*:VNP promoter may have given rise to its robust expression in pineal photoreceptors and cells in other brain regions. Therefore, one can replace *otx5* and *crx* binding sites by those of other cell type–specific TFs in *nr1d1*:VNP to report the circadian rhythm in cell types of interest. This will enable single-cell imaging of circadian clock in a broad range of cell types in live zebrafish.

Several studies have shown that, in the absence of light, the overall circadian oscillation of gene expression was absent in developing zebrafish embryos [10,20,37]. But it is unclear if this is due to the lack of initiation or synchronization of cellular clocks. By imaging *per1-luc* transgenic fish, Dekens and colleagues suggested that cells in the embryos kept under constant environmental conditions may undergo asynchronous oscillations [20]. Ziv and colleagues also proposed that the expression of *aanat2* is asynchronous in individual pineal cells under constant darkness [37]. Carr and Whitmore showed that light-sensitive cells of the zebrafish DAP20 cell line kept under long-term darkness became desynchronized, and light can synchronize the clocks in these cells [11]. To our surprise, we did not observe asynchronous oscillations of *nr1d1*:VNP among individual cells in our reporter fish raised under constant

darkness. In Dekens and colleagues' study, separate zebrafish siblings fixed at circadian time (CT) 3 and CT15, respectively, under DD conditions have an intermediate number of cells expressing *per1* at the RNA level [20]. Similarly, Ziv and colleagues observed an average level of *aanat2* in the zebrafish embryos kept in constant darkness compared with the fish kept in LD cycle. However, these observations are not the direct evidence for asynchronous oscillations, as the gene expression was not traced over time in the same fish. Instead, they can be explained by the highly variable baseline gene expression between single cells within the same fish observed in our study. The difference between our study and Carr and Whitmore's study may reflect the difference between in vivo versus in vitro systems. During our time of imaging, zebrafish embryos are undergoing rapid changes of cellular states through differentiation and proliferation, which may preclude the existence of a self-sustained cellular clock. The zebrafish cell line used by Carr and Whitmore is a stable transfected cell line and can harbor free-running autonomous clocks as demonstrated in some stable mammalian cell lines [39]. However, it has also been shown that mammalian induced pluripotent stem cells (iPSCs) and ESCs are non-rhythmic [14,40], suggesting that the difference of cellular states plays an important role in the state of cellular clocks. Our study thus suggests that early light exposure in zebrafish larvae has a dramatic effect on the development of self-sustained cellular clocks, demonstrating the utility of the single-cell circadian zebrafish reporter. As *nr1d1* is mainly involved in the second circadian feedback loop, one cannot rule out the possibility that the development of cellular circadian oscillations that we observed in this study is specific to this second loop. Therefore, it will be important to further investigate the single-cell expression of other circadian clock genes such as *Per* family genes during development in future studies.

Methods

Ethics statement

All experiments with zebrafish were done in accordance with protocols approved by the Animal Use Committee of Institute of Neuroscience, Chinese Academy of Sciences (permission number NA-045-2019).

Construction of the *nr1d1*:VNP (Venus-NLS-PEST) expression vector

The zebrafish *nr1d1*:VNP reporter plasmid was constructed as follows. The mouse *Nr1d1*:VNP reporter plasmid was generously provided by Emi Nagoshi [27] (Department of Genetics and Evolution, Sciences III, University of Geneva). The plasmid carries an intact VNP sequence. VNP is a nuclear fluorescent protein with a high folding efficiency and short half-life [28]. The zebrafish *nr1d1*:VNP plasmid was obtained by homologous recombination with four cassettes. The targeting cassette was generated by four-step PCR amplification. The first PCR step was performed to amplify the tol2-polyA-ampicillin gene with an additional 3′ sequence that is homologous to the 5′ end region of zebrafish *nr1d1* promoter (template: 394-tcf-lef-TuboGFPdest1, forward primer: 5′-GCTTCTGCTAGGATCAATGTGTAG-3′, reverse primer: 5′-CCTATAGTGAGTCGTATTACCAACTT-3′). The second and third PCR steps were performed to amplify the promoter of zebrafish *nr1d1* gene (6.2 kbp) (template sequences from zebrafish genome: second PCR forward primer: 5′- TACGACTCACTATAGGTTTTCCACCCGCTGGGCTGCCTCTCACGTG-3′, reverse primer: 5′-TATTCCTTGCTTCCGTCTGTAGTGCCCAAT-3′; third PCR forward primer: 5′-GGCAGACATCACGGGTTAAGACACAGTGTT-3′; reverse primer: 5′-CAGCAGGCTGAAGTTAGTAGCTCCGCTTCCTCCCGGAGGGGATGGTGGGAATGATGATCC-3′). The fourth PCR step was performed to amplify the cassette of VNP (template sequences from mouse *Nr1d1*:VNP reporter plasmid forward primer: 5′-ACTAACTTCAGCCTGCTGATGGTGAGCAAGGGCGAGGA-3′, reverse primer: 5′-CTACACAT

TGATCCTAGCAGGAGCACAG-3′). The 3′ arm of the first PCR was homologous to the 5′ arm of the second PCR, the 3′ arm of the second PCR was homologous to the 5′ arm of the third PCR, the 3′ arm of the third PCR was homologous to the 5′ arm of the fourth PCR, and the 3′ arm of the fourth PCR was homologous to the 5′ arm of the first PCR. The vector construct was verified by sequencing. Fig 1A shows the schematic of the *nr1d1*:VNP construct design. The structure of the 6.2-kbp *nr1d1* promoter containing multiple E-box, RRE, and pineal-specific elements was also described.

Generation of transgenic Tg(*nr1d1*:VNP) zebrafish reporter line

Generation of transgenic fish using the Tol2 system was carried out according to the published protocol [41]. Constructs were injected together with capped RNA encoding transposase (10 ng/μL each of DNA and RNA) into fertilized eggs at the one-cell stage. The injected fish (F0 generation) were raised and screened for integration of the transgene into the germline. Isolation of transgene-positive progeny (F1) was carried out using EGFP imaging using fluorescence stereomicroscpe (Olymous, SZX16). For in vivo imaging, *Tg(nr1d1*:VNP*)* zebrafish was crossed with a pigmentation mutant strain (*casper* mutant), which is transparent in the whole brain except the eyes.

Generation of transgenic Tg(lws2:nfsB-mCherry) and Tg(xops:nfsB-mCherry) zebrafish reporter line

The 1.77-kbp lws2 and 1.38-kbp xops (Xenopus rhodopsin) promoter were amplified by PCR. Specific primers for the lws2 promoter: 5′-GGCCAGATGGGCCCTGTTGTGCACCAGAT CTGAGT-3′ and 5′-TGGTCCAGCCTGCTTTTTGGAAACCCTGAAGATCA-3′. The xops promoter (1,370 bp) was amplified from pFIN-XOPS-tdTOMP (Semple-Rowland SL and colleagues, 2010, addgene #44359) using forward (5′-TATAGGGCGAATTGGGGCCGCAGAT CTTTATACATTGC-3′) and reverse (5′-CCGGTGGATCCCAAACCCTCGAGATCCCTA GAAGCCTGTGAT-3′) primers containing homogenous recombination sites, respectively. The products were subcloned into pTol-uas:nfsB-mCherry plasmid to substitute the uas promoter by using homologous recombination kit (ClonExpress MultiS One Step Cloning Kit, C113; Vazyme, China). The pTol-lws2:nfsB-mCherry plasmid and pTol-xops:nfsB-mCherry were co-injected into AB embryos with Tol2 transposase mRNA at the one-cell stage. The mCherry signal was screened to identify F1.

Zebrafish husbandry

Adult zebrafish (NCBI taxonomy ID: 7955) were raised and maintained in fully automated zebrafish housing systems (Aquazone, Tzofit, Israel; temperature 28˚C, pH 7.0, conductivity 300 mS) under 14-hour/10-hour LD cycles, and fed with paramecium twice a day. Larvae were fed twice a day starting from 5 dpf. For the experiments under normal LD condition, embryos were produced by natural spawning in the morning and raised in egg-water containing methylene blue (0.3 ppm) in a light-controlled incubator under 12-hour/12-hour LD cycles at 28˚C. ZT0 is defined as the time when the lights are turned on (9 AM). For the experiments under DD condition, embryos were collected within 30 minutes after birth and put into a black box in a dark incubator at 28˚C. All the experimental protocols were approved by the Animal Use Committee of Institute of Neuroscience, Chinese Academy of Sciences.

We use Tg(*nr1d1*:VNP), Tg(*aanat2*:mRFP), Tg(*lws2*:nfsB-mCherry), Tg(*xops*:nfsB-mCherry), and Tg(*her4*:DsRed) zebrafish lines, which were maintained on an AB background. Sex includes male and female. We used the larva fish aged from 3.5 dpf to 7.5 dpf.

In vivo imaging

Fish were randomly selected to be imaged under LD or DD. We were not blinded to LD and DD fish group allocation. For time-lapse imaging every two hours lasting for one day starting at ZT0 (Figs 1G, 4 and 5I–5K), 5.0 dpf zebrafish larvae were anesthetized with MS-222 (Sigma-Aldrich) (0.01%–0.02%) and mounted in low melting point agarose (1.3%–1.5%). The fish were maintained in a chamber with circulating egg-water without anesthesia at temperature of 28 ± 0.5˚C. For imaging from 3.5 dpf to 6.5 dpf every 12 hours at ZT0 and ZT12 (Figs 1E, 3 and 5A–5G) and single time point imaging (Fig 2C and 2F), the zebrafish larvae were anesthetized at each time point with MS-222 (sigma-Aldrich) (0.01%–0.02%) and embedded in low melting point agarose (0.8%–1.0%) before imaging. Three to four fish were imaged in each imaging session that lasts for one half to one hour. Fish were released immediately after the end of each imaging session and in a free-moving state between image sessions. All fish were imaged from dorsal view using two-photon microscope (Olympus) as described previously [42]. Imaging was performed using a 25× (numerical aperture [NA] = 1.05) water-immersion objective (Olympus). Excitation was provided by a Ti:Sapphire femtosecond pulsed laser system (Coherent) tuned to 900 nm, which allowed efficient simultaneous excitation of Venus fluorescent protein. Laser power was set to 33–35 mW for pineal imaging and 44.4 mW for whole-brain imaging. We used two-photon excitation microscopy for two reasons, as suggested by Carvalho and Heisenberg [43]. First, the near-infrared wavelength has undetectable effects on fish physiology and behavior. Second, as the two-photon excitation is only achieved near the focal plane, it minimizes photo bleaching and phototoxicity. The sample size estimate is based on our previous studies. Two fish under LD condition were imaged for the whole-brain imaging (Fig 1E). One fish under LD condition was imaged for one day every hour (Fig 1G). Two fish under LD condition and three fish under DD condition were imaged for one day every 2 hours (Figs 4 and 5I–5K). Six fish under LD condition and four fish under DD condition were imaged from 3.5 dpf to 6.5 dpf every 12 hours (Figs 1E, 3A–3E and 5A–5G). Three fish under LD_LD condition and three fish under LD_DD condition were imaged from 6.0 dpf to 7.5 dpf every 12 hours (Fig 5F). Six Tg(*xops*:nfsB-mCherry) fish (Fig 3F–3H) were imaged from 3.5 dpf to 6.5 dpf every 12 hours.

scRNA-seq and data analysis

Larval heads (6-dpf) were dissected on dissection medium (DMEF/F12 with 2% 100X penicillin-streptomycin) and pineal regions were enriched by pipetting the pineal with the whole brain into a tube placed on ice. Then the dissociated cells were obtained according to the protocol from Miguel A. Lopez-Ramirez and colleagues [44]. First, add 300 uL papain solution to the dissected tissue at 37˚C in water heater for 15 minutes, with gentle pipetting during digestion. Then, stop the digestion by adding 1.2 mL of washing solution (650 μL gluose [45%], 500 μl HEPES [1 M], and 5 mL FBP were added into 93.85 mL DEBS 1X) and wash the cells twice using washing solution. In the end, sterilize the cells using a 40-um pore size filter. To count the living cells, we stained cells using trypan blue and counted the cells using a hemocytometer. The resulting single cell suspension was promptly loaded on the 10X Chromium system using Chromium Single Cell 3' Reagents v2. The barcoded cDNA library was then sequenced on the Illumina NovaSeq 6000 system.

For data analysis, raw sequencing data were converted to matrices of expression counts using the cellranger software provided by 10X Chromium; zebrafish reference transcriptome (ENSEMBL Zv11, release 95) was used as reference genome. The gene expression matrix was then loaded into Seurat package [45] in R for the following analysis. Cells with less than 500 genes or a percentage of mitochondiral genes > 0.02 were excluded from the following

analysis. The graph-based method from Seurat was used to cluster the left cells (6,514 cells). The PCA was selected as the dimensional reduction technique in the construction of the shared nearest-neighbor (SNN) graph. The first 50 PCs were then used in an SSN clustering, with the resolution equal to 1. A *t*-distributed stochastic neighbor embedding (t-SNE) clustering was performed on the scaled matrix (with highly variable genes only) to obtain a 2D representation of the cell states. Finally, 26 clusters were identified. They were manually annotated by comparing the marker genes with those from whole-brain scRNA-seq data in adult zebrafish [21].

Image analysis

The image data were first converted to 'nrrd' format using a customer Fiji [46] macro code for downstream analysis. Time series of 3D images in 'nrrd' format for each fish were then aligned by CMTK toolkit (parameters: -awr 01 -l fa -g -T 8 -X 10 -C 1 -G 20 -R 2 -E 1 -A '—accuracy 5—dofs 12' -W '—accuracy 5' for the alignment of the pineal gland; -ar 01 -l a -A '—exploration 50—accuracy 5—dofs 12' for the alignment of whole brain) to facilitate the single cell tracing. TrackMate [23], an open source Fiji [46] plug-in, was applied to perform the single-cell nucleus identification and tracing on 3D images over time. Briefly, Laplacian of Gaussian (LoG) filter with a sigma suited to the estimated blob diameter (5.5 μm) was applied to detect the single-cell nuclei in each 3D image. Then HyperStack Displayer was applied to visualize the identified spots, which allows manual editing afterwards. A nearest-neighbor algorithm (parameter: maximal linking distance 5.5 μm) was applied to trace each single cell. After the automatic processing, the identification and tracing of each cell were manually curated using the manual tracking tools of TrackMate [23]. Combining the automated and manual tracking approaches can significantly increase the detection accuracy of single cells. In the end, the cell position and mean image intensities were exported for the following statistical analysis. For the whole-brain data, the pineal gland, the optic tectum, and the cerebellum were first manually extracted, then single cells were traced using manual tracking tools of TrackMate [23], and mean intensities of each brain region were used for the following analysis. The cell position and mean intensities for all the fish analyzed were listed in Tables B–J in S2 Data.

Three-dimensional reconstruction of zebrafish pineal gland

To systematically analyze the 3D distribution of the *nr1d1*:VNP-positive cells, a common 3D pineal gland structure was reconstructed by averaging six zebrafish pineal glands at 5.5 dpf. First, we manually reconstructed the 3D pineal structure of six individual 5.5 dpf zebrafish pineal glands, taking advantage of the clear pineal gland boundary in two-photon imaging. Secondly, all 3D pineal gland structures from different fish were aligned using rigid transformation using CMTK (parameters: -ar 01 -l a -A '—accuracy 5—dofs 12'). Thirdly, the signals from aligned 3D pineal glands were added and smoothed by Gaussian blur (Sigma = 30) using Fiji package [46]. In the end, a common 3D pineal gland structure was obtained by implementing Li's minimum cross entropy thresholding method using Fiji package [46].

The cells of each pineal gland were then registered to the common 3D pineal gland using the same transform matrix for the alignment of the 3D pineal gland structure. The 'ks' package in R (https://www.R-project.org/) was applied to calculate the cell density distribution. The common zebrafish pineal structure and *nr1d1*:VNP cell density distribution were visualized by NeuTube software [47] (Fig 2D).

Statistical analysis

All the statistical analysis was performed using the computing environment R (https://www.R-project.org/). The fluorescence intensities were log2-tranformed before fitting. Relative

amplitude was defined as absolute amplitudes divided by means. For the time-lapse imaging data from 5.0 dpf to 6.0 dpf every half hour, both JTKcycle (from MetaCycle package in R) and cosine fitting were applied to analysze the data. JTKcycle adjusted $P < 0.05$ and absolute amplitude >100 were used as the cutoff to define rhythmic cells. The circadian phase and amplitude of each cell were also calculated by fitting the mean intensity of each cell to cosine functions with a 24-hour period and shifting phases as described previously [26]. R package 'circular' was applied to calculate the average circadian peak (mean.circular) and Kuramoto's order parameter R (Rho.circular). For the imaging data of developmental time course from 3.5 dpf to 6.5 dpf, the fluorescence intensities were first normalized by dividing the mean intensity. A regression model combining a stepwise function denoting the circadian effect and a linear function denoting the developmental trend were used to fit the data. Namely, $F(x) = Ax + B^*(-1)^{(x+1)} + C$, A denotes the slope of developmental trend, B denotes the amplitude of circadian oscillation, and C denotes the baseline level of fluorescence intensity. Nonlinear least squares (nls) were applied to determine these three parameters and the level of significance of each coefficient. R package 'Seurat' was applied to perform t-SNE analysis and visualization of clusters. To classify LD cells imaged every 2 hours, all the 12 principle components were used and resolution was set to 0.4.

Real-time qPCR

For each time point, A total of 15 to 20 fish were separated into 2 tubes equally. Here, the sample size estimate is based on our previous studies. Trizol (Invitrogen) was immediately added to each tube. Then, total RNA was isolated from zebrafish larvae by Trizol (Invitrogen) according to the manufacturer's protocol. Total RNA quantities were measured by a Nanodrop spectrometer (Nanodrop 2000). Five hundred nanograms of total RNA was reverse transcribed using HiScript II Q RT SuperMix for qPCR (+gDNA wiper) (vazyme) according to the manufacturer's protocol. Two microliters of RT product (1:10 diluted) and 10 μL of SYBR Green Master Mix (vazyme) were used in qPCR on an ABI StepOne Plus System (Thermo Fisher) according to the manufacturer's protocol. The specificity of PCR was checked by melting curve analysis. In every qPCR assay, *eef1a1* was used as the control gene for any significant bias of starting materials across samples. Primers used in this study were listed in Table L in S2 Data.

Supporting information

S1 Fig. (a) Plot of real-time PCR results of *nr1d1* and *per1b* during early development (from 26 hpf to 68 hpf). Each time point is the pool of 10–20 fish. (b) Raw fluorescence values of all the traced *nr1d1*:VNP cells at 1-hour intervals for one day. (c) Circadian phase distributions of the 48 *nr1d1*:VNP-positive cells in b. (d) Fluorescence images show the co-expression of *nr1d1*:VNP (nuclear signal) and *her4*:DsRed (cytoplasmic signal) in the zebrafish pineal gland. hpf, hours post fertilization.
(EPS)

S2 Fig. Log2-transformed dusk–dawn ratios of VNP fluorescence intensities at different time points during development (i.e., 4.5 dpf/4.0 dpf, 5.5 dpf/5.0 dpf, 6.5 dpf/6.0 dpf). (a) Each data point represents one cell and the colors correspond to different fish. The red dashed line represents $y = 0$. n.s. represents no significant difference in dusk–dawn ratios between different time points (two-tailed ANOVA test). (b) Barplot shows the number of cells with dusk $>$ dawn (blue color, i.e., the number of cells above the red dash line in b) or dusk $<$ dawn (red color, i.e., the number of cells below the red dash line in b). dpf, days postfertilization; VNP, Venus-NLS-PEST.
(EPS)

S3 Fig. The comparison of the circadian phase between the two types of cells in Fig 4E: (a) fish1, (b) fish2. The color of the boxes corresponds to Fig 4E. Two-tailed Student *t* test was applied to calculate the levels of significance between the two types of cells. n.s. represents $P > 0.05$.
(EPS)

S4 Fig. LD cycle is essential for *nr1d1* oscillation. (a) Comparison of the relative amplitudes of LD and DD cells in Fig 5I from cosine fitting. The red dashed line represents $y = 0.05$. (b) Percentages of oscillating cells (cosine fitting $P < 0.05$ and relative amplitude > 0.05) in each LD and DD fish. The orange bars represent the percentage of oscillating cells, while the blue bars represent the percentage of non-oscillating cells. (c) Single-cell tracing results of all 142 *nr1d1*:VNP-positive cells in three zebrafish pineal glands raised under DD condition. The blue dots represent the original fluorescence signals, while the solid red line represents the smoothed curve fitted by the cosine functions. (d) Expression patterns of LD (1 fish) and DD cells (1 fish) across one day at 1-hour resolution. Each thin line represents one cell and each dot represents raw fluorescence intensity. Thick lines represent the loess-smoothed curves for all the LD in blue and DD cells in red, respectively. The shaded areas show the 95% confidence level of the smooth curve. (e) Comparison of the absolute amplitudes of LD and DD cells from JTKcycle (1-hour resolution). The red dashed line represents $y = 100$. (f) Percentages of oscillating cells (JTKcycle adjusted $P < 0.05$ and absolute amplitude > 100) in each LD and DD fish (1-hour resolution). The orange bars represent the percentage of oscillating cells, while the blue bars represent the percentage of non-oscillating cells. (g) Single-cell tracing results of all 24 *nr1d1*:VNP-positive DD cells imaged at 1-hour resolution. The blue dots represent the original fluorescence signals, while the solid red line represents the smoothed curve fitted by the cosine functions. DD, constant dark; LD, light–dark.
(EPS)

S1 Data.   Excel spreadsheet containing data values plotted in all main and supporting figures.
(XLSX)

S2 Data. Tables A–L. (A) Promoter sequences (in Fastq format) used in this study to screen the circadian reporters in zebrafish. **(B)** Single-cell tracing result of *nr1d1*:VNP cells in the pineal gland, the optical tectum, and the cerebellum during development based on the whole-brain two-photon imaging under LD condition. The cells were tracked starting at 3.5 dpf in pineal gland, 5.5 dpf in optical tectum, and 6.5 dpf in cerebellum, respectively. Fish were imaged every 12 hours until 7.5 dpf. **(C)** Single-cell tracing result of *nr1d1*:VNP cells in the pineal gland imaged every hour under LD condition based on the two-photon imaging of the pineal gland. The fish were imaged starting at 5 dpf for one day. **(D)** Single-cell tracing result of *nr1d1*:VNP cells in the pineal gland during development under LD condition based on the two-photon imaging of the pineal gland. The fish were imaged every 12 hours from 3.5 dpf to 6.5 dpf. **(E)** Single-cell tracing result of *nr1d1*:VNP cells in the pineal gland of Tg(nr1d1:VNP) × Tg(xops:nfsB-mCherry) fish from 3.5 dpf to 6.5 dpf under LD condition. The fish were imaged every 12 hours from 3.5 dpf to 6.5 dpf. **(F)** Single-cell tracing result of *nr1d1*:VNP cells in the pineal gland imaged every 2 hours under LD condition based on the two-photon imaging of the pineal gland. The fish were imaged starting at 5 dpf for one day. **(G)** Single-cell tracing result of *nr1d1*:VNP cells in the pineal gland during development under DD condition based on the two-photon imaging of the pineal gland. The fish were imaged every 12 hours from 3.5 dpf to 6.5 dpf. **(H)** Single-cell tracing result of nr1d1:VNP cells in the pineal gland from 6.0 dpf to 7.5 dpf under LD condition based on the two-photon imaging of the pineal

gland. The fish were imaged every 12 hours from 6.0 dpf to 7.5 dpf. Those are the control cells for Table I. **(I)** Single-cell tracing result of nr1d1:VNP cells in the pineal gland from 6.0 dpf to 7.5 dpf under LD_DD condition (fish were transferred to DD at 5.5 dpf) based on the two-photon imaging of the pineal gland. The fish were raised under LD from 0 dpf to 5.5 dpf and transferred to DD at 5.5dpf. The fish were imaged every 12 hours from 6.0 dpf to 7.5 dpf. **(J)** Single-cell tracing result of *nr1d1*:VNP cells in the pineal gland imaged every 2 hours under DD condition based on the two-photon imaging in the pineal gland. The fish were imaged starting at 5 dpf for one day. **(K)** Single-cell tracing result of *nr1d1*:VNP cells in the pineal gland imaged every hour under DD condition based on the two-photon imaging in the pineal gland. The fish were imaged starting at 5 dpf for one day. **(L)** Primers sequences used in RT-PCR experiment. In Tables A–K, TrackIDs are the identifiers unique for each individual fish. PositionX/Y/Z.* and Intensity.* are the *x/y/z* position and mean fluorescence intensity of the cell in each time point. DD, constant dark; dpf, days postfertilization; LD, light–dark; RT-PCR, real-time PCR.
(XLSX)

S1 Movie. The combined image stacks of the whole brain using two-photon imaging (whole brain fish 1). The fish was imaged from 3.5 dpf to 7.5 dpf every 12 hours (9 stacks). The fish was raised under LD condition. From left to right, top to bottom, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, 6.5 dpf, 7.0 dpf, and 7.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S2 Movie. The combined image stacks of the whole brain using two-photon imaging (whole brain fish 2). The fish was imaged from 3.5 dpf to 7.5 dpf every 12 hours (9 stacks). The fish was raised under LD condition. From left to right, top to bottom, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, 6.5 dpf, 7.0 dpf, and 7.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S3 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD fish 8). The fish was imaged at 5.0 dpf in every hour (24 stacks). The fish was raised under LD condition. dpf, days postfertilization; LD, light–dark.
(AVI)

S4 Movie. Confocal 3D reconstructions of zebrafish pineal gland. Zebrafish larvae were co-labeled with *nr1d1*:VNP (blue) and *aanat2*:mRFP (red).
(AVI)

S5 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD fish 1). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S6 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD fish 2). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S7 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD fish 3). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S8 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD fish 4). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S9 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD fish 5). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S10 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD fish 6). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S11 Movie. The combined image stacks of the pineal gland using two-photon imaging (xops:nfsB-mCherry fish 1). This is Tg(nr1d1:VNP) × Tg(xops:nfsB-mCherry) fish. The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S12 Movie. The combined image stacks of the pineal gland using two-photon imaging (xops:nfsB-mCherry fish 2). This is Tg(nr1d1:VNP) × Tg(xops:nfsB-mCherry) fish. The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S13 Movie. The combined image stacks of the pineal gland using two-photon imaging (xops:nfsB-mCherry fish 3). This is Tg(nr1d1:VNP) × Tg(xops:nfsB-mCherry) fish. The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S14 Movie. The combined image stacks of the pineal gland using two-photon imaging (xops:nfsB-mCherry fish 4). This is Tg(nr1d1:VNP) × Tg(xops:nfsB-mCherry) fish. The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S15 Movie. The combined image stacks of the pineal gland using two-photon imaging (xops:nfsB-mCherry fish 5). This is Tg(nr1d1:VNP) × Tg(xops:nfsB-mCherry) fish. The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S16 Movie. The combined image stacks of the pineal gland using two-photon imaging (xops:nfsB-mCherry fish 6). This is Tg(nr1d1:VNP) × Tg(xops:nfsB-mCherry) fish. The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under LD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S17 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD fish 7). The fish was imaged at 5.0 dpf in every two hours (12 stacks). The fish was raised under LD condition. dpf, days postfertilization; LD, light–dark.
(AVI)

S18 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD fish 8). The fish was imaged at 5.0 dpf in every two hours (12 stacks). The fish was raised under LD condition. dpf, days postfertilization; LD, light–dark.
(AVI)

S19 Movie. The combined image stacks of the pineal gland using two-photon imaging (DD fish 1). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under DD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. DD, constant dark; dpf, days postfertilization.
(AVI)

S20 Movie. The combined image stacks of the pineal gland using two-photon imaging (DD fish 2). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under DD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. DD, constant dark; dpf, days postfertilization.
(AVI)

S21 Movie. The combined image stacks of the pineal gland using two-photon imaging (DD fish 3). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under DD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5 dpf. DD, constant dark; dpf, days postfertilization.
(AVI)

S22 Movie. The combined image stacks of the pineal gland using two-photon imaging (DD fish 4). The fish was imaged from 3.5 dpf to 6.5 dpf every 12 hours (7 stacks). The fish was raised under DD condition. From left to right, 3.5 dpf, 4.0 dpf, 4.5 dpf, 5.0 dpf, 5.5 dpf, 6.0 dpf, and 6.5dpf. DD, constant dark; dpf, days postfertilization.
(AVI)

S23 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD_LD fish 1). The fish was imaged from 6.0 dpf to 7.5 dpf every 12 hours (4 stacks). The fish was raised under LD condition. From left to right, 6.0 dpf, 6.5 dpf, 7.0 dpf, and 7.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S24 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD_LD fish 2). The fish was imaged from 6.0 dpf to 7.5 dpf every 12 hours (4 stacks). The fish was raised under LD condition. From left to right, 6.0 dpf, 6.5 dpf, 7.0 dpf, and 7.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S25 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD_LD fish 3). The fish was imaged from 6.0 dpf to 7.5 dpf every 12 hours (4 stacks). The fish was raised under LD condition. From left to right, 6.0 dpf, 6.5 dpf, 7.0 dpf, and 7.5 dpf. dpf, days postfertilization; LD, light–dark.
(AVI)

S26 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD_DD fish 1). The fish was imaged from 6.0 dpf to 7.5 dpf every 12 hours (4 stacks). The fish was raised under LD condition from 0 dpf to 5.5 dpf and was transferred to DD condition from 5.5 dpf. From left to right, 6.0 dpf, 6.5 dpf, 7.0 dpf, and 7.5 dpf. DD, constant dark; dpf, days postfertilization; LD, light–dark.
(AVI)

S27 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD_DD fish 2). The fish was imaged from 6.0 dpf to 7.5 dpf every 12 hours (4 stacks). The fish was raised under LD condition from 0 dpf to 5.5 dpf and was transferred to DD condition from 5.5 dpf. From left to right, 6.0 dpf, 6.5 dpf, 7.0 dpf, and 7.5 dpf. DD, constant dark; dpf, days postfertilization; LD, light–dark.
(AVI)

S28 Movie. The combined image stacks of the pineal gland using two-photon imaging (LD_DD fish 3). The fish was imaged from 6.0 dpf to 7.5 dpf every 12 hours (4 stacks). The fish was raised under LD condition from 0 dpf to 5.5 dpf and was transferred to DD condition from 5.5 dpf. From left to right, 6.0 dpf, 6.5 dpf, 7.0 dpf, and 7.5 dpf. DD, constant dark; dpf, days postfertilization; LD, light–dark.
(AVI)

S29 Movie. The combined image stacks of the pineal gland using two-photon imaging (DD fish 5). The fish was imaged at 5.0 dpf in every two hours (12 stacks). The fish was raised under DD condition. DD, constant dark; dpf, days postfertilization.
(AVI)

S30 Movie. The combined image stacks of the pineal gland using two-photon imaging (DD fish 6). The fish was imaged at 5.0 dpf in every two hours (12 stacks). The fish was raised under DD condition. DD, constant dark; dpf, days postfertilization.
(AVI)

S31 Movie. The combined image stacks of the pineal gland using two-photon imaging (DD fish 7). The fish was imaged at 5.0 dpf in every two hours (12 stacks). The fish was raised under DD condition. DD, constant dark; dpf, days postfertilization.
(AVI)

S32 Movie. The combined image stacks of the pineal gland using two-photon imaging. The fish was imaged at 5.0 dpf every hour (22 stacks, CT3–CT24). The fish was raised under DD condition. CT, circadian time DD, constant dark; dpf, days postfertilization.
(AVI)

Acknowledgments

Author Contributions

Conceptualization: Haifang Wang, Jie He, Yuanhai Li, Jun Yan.

Data curation: Haifang Wang, Xingxing Li.

Formal analysis: Haifang Wang, Zeyong Yang.

Funding acquisition: Haifang Wang, Zeyong Yang, Jun Yan.

Investigation: Haifang Wang, Zeyong Yang, Xingxing Li, Dengfeng Huang, Shuguang Yu.

Methodology: Haifang Wang, Zeyong Yang, Xingxing Li, Shuguang Yu.

Supervision: Jie He, Yuanhai Li, Jun Yan.

Validation: Haifang Wang, Xingxing Li.

Visualization: Haifang Wang.

Writing – original draft: Haifang Wang, Jun Yan.

Writing – review & editing: Haifang Wang, Jie He, Yuanhai Li, Jun Yan.

References

1. Takahashi JS, Hong HK, Ko CH, Mcdearmon EL. The genetics of mammalian circadian order and disorder: implications for physiology and disease. Nat Rev Genet. 2008; 9: 764–775. https://doi.org/10.1038/nrg2430 PMID: 18802415

2. Gekakis N, Staknis D, Nguyen HB, Davis FC, Wilsbacner LD, King DP, et al. Role of the CLOCK protein in the mammalian circadian mechanism. Science. 1998; 280:1564–1569 https://doi.org/10.1126/science.280.5369.1564 PMID: 9616112

3. Yoo S-H, Ko CH, Lowrey PL, Buhr ED, Song E -j., Chang S, et al. A noncanonical E-box enhancer drives mouse Period2 circadian oscillations in vivo. Proc Natl Acad Sci USA. 2005; Feb 15; 102(7): 2608–2613. https://doi.org/10.1073/pnas.0409763102 PMID: 15699353

4. Ohno T, Onishi Y, Ishida N. A novel E4BP4 element drives circadian expression of mPeriod2. Nucleic Acids Res. 2007; 35(2): 648–655. https://doi.org/10.1093/nar/gkl868 PMID: 17182630

5. Buhr ED, Takahashi JS. Molecular components of the mammalian circadian clock. Handb Exp Pharmacol. 2013; 217:3–27

6. Preitner N, Damiola F, Zakany J, Duboule D, Albrecht U, Schibler U. The Orphan Nuclear Receptor REV-ERB Controls Circadian Transcription within the Positive Limb of the Mammalian Circadian Oscillator. Cell. 2002; 110: 251–260. https://doi.org/10.1016/s0092-8674(02)00825-5 PMID: 12150932

7. Perelis M, Marcheva B, Ramsey KM, Schipma MJ, Hutchison AL, Taguchi A, et al. Pancreatic β cell enhancers regulate rhythmic transcription of genes controlling insulin secretion. Science. 2015; 350: aac4250. https://doi.org/10.1126/science.aac4250 PMID: 26542580

8. Li Y, Li G, Wang H, Du J, Yan J. Analysis of a Gene Regulatory Cascade Mediating Circadian Rhythm in Zebrafish. PLoS Comput Biol. 2013; 9(2): e1002940. https://doi.org/10.1371/journal.pcbi.1002940 PMID: 23468616

9. Ben-moshe Z, Foulkes NS, Gothilf Y. Functional Development of the Circadian Clock in the Zebrafish Pineal Gland. Biomed Res Int. 2014; 2014: 1–8.

10. Kaneko M, Cahill GM. Light-Dependent Development of Circadian Gene Expression in Transgenic Zebrafish. PLoS Biol. 2005; 3(2): e34. https://doi.org/10.1371/journal.pbio.0030034 PMID: 15685291

11. Carr A-JF, Whitmore D. Imaging of single light-responsive clock cells reveals fluctuating free-running periods. Nat Cell Biol. 2005; 7: 319–321. https://doi.org/10.1038/ncb1232 PMID: 15738976

12. Vatine G, Vallone D, Appelbaum L, Mracek P, Ben-Moshe Z, Lahiri K, et al. Light directs zebrafish period2 expression via conserved D and E boxes. PLoS Biol. 2009; 7(10): e1000223. https://doi.org/10.1371/journal.pbio.1000223 PMID: 19859524

13. Sladek M, Sumova A, Kovacikova Z, Bendova Z, Kristyna L, Illnerova H. Insight into molecular core clock mechanism of embryonic and early postnatal rat suprachiasmatic nucleus. Proc Natl Acad Sci USA. 2004; 101: 6231–6236. https://doi.org/10.1073/pnas.0401149101 PMID: 15069203

14. Yagita K, Horie K, Koinuma S, Nakamura W, Yamanaka I, Urasaki A. Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. Proc Natl Acad Sci USA. 2010; 107: 3846–3851. https://doi.org/10.1073/pnas.0913256107 PMID: 20133594

15. Umemura Y, Koike N, Matsumoto T, Yoo S-H, Chen Z, Yasuhara N, et al. Transcriptional program of Kpna2/Importin-α2 regulates cellular differentiation-coupled circadian clock development in mammalian cells. Proc Natl Acad Sci USA. 2014; 111: E5039–5048. https://doi.org/10.1073/pnas.1419272111 PMID: 25389311

16. Carmona-Alcocer V, Abel JH, Sun TC, Petzold LR, Doyle FJ, Simms CL, et al. Ontogeny of Circadian Rhythms and Synchrony in the Suprachiasmatic Nucleus. J Neurosci. 2018; 38(6):1326–1334. https://doi.org/10.1523/JNEUROSCI.2006-17.2017 PMID: 29054877

17. Vallone D, Lahiri K, Dickmeis T, Foulkes NS. Start the Clock! Circadian Rhythms and Development. Dev Dyn. 2007; 236: 142–155. https://doi.org/10.1002/dvdy.20998 PMID: 17075872

18. Kakizawa T, Nishio SI, Triqueneaux G, Bertrand S, Rambaud J, Laudet V. Two differentially active alternative promoters control the expression of the zebrafish orphan nuclear receptor gene Rev-erba. J Mol Endocrinol. 2007; 38: 555–568. https://doi.org/10.1677/JME-06-0063 PMID: 17496157

19. Delaunay F, Thisse C, Marchand O, Laudet V, Thisse B. An inherited functional circadian clock in zebrafish embryos. Science. 2000; 289: 297–300. https://doi.org/10.1126/science.289.5477.297 PMID: 10894777

20. Dekens MPS, Whitmore D. Autonomous onset of the circadian clock in the zebrafish embryo. EMBO J. 2008; 27: 2757–2765. https://doi.org/10.1038/emboj.2008.183 PMID: 18800057

21. Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat Biotechnol. 2018; 36:442–450 https://doi.org/10.1038/nbt.4103 PMID: 29608178

22. Yeo SY, Kim MJ, Kim HS, Huh TL, Chitnis AB. Fluorescent protein expression driven by her4 regulatory elements reveals the spatiotemporal pattern of Notch signaling in the nervous system of zebrafish embryos. Dev Biol. 2007; 301(2): 555–567 https://doi.org/10.1016/j.ydbio.2006.10.020 PMID: 17134690

23. Tinevez JY, Perry N, Schindelin J, Hoopes GM, Reynolds GD, Laplantine E, et al. TrackMate: An open and extensible platform for single-particle tracking. Methods. 2017; 115: 80–90. https://doi.org/10.1016/j.ymeth.2016.09.016 PMID: 27713081

24. Kuramoto Y. Self-entrainment of a population of coupled non-linear oscillators. In: Araki H, editor. Mathematical Problems in Theoretical Physics. 1975. pp. 420–422.

25. Westermark PO. Linking Core Promoter Classes to Circadian Transcription. PLoS Genet. 2016; 12: 1–32.

26. Yan J, Wang H, Liu Y, Shao C. Analysis of gene regulatory networks in the mammalian circadian rhythm. PLoS Comput Biol. 2008; 4: e1000193. https://doi.org/10.1371/journal.pcbi.1000193 PMID: 18846204

27. Nagoshi E, Saini C, Bauer C, Laroche T, Naef F, Schibler U. Circadian gene expression in individual fibroblasts: Cell-autonomous and self-sustained oscillators pass time to daughter cells. Cell. 2004; 119: 693–705. https://doi.org/10.1016/j.cell.2004.11.015 PMID: 15550250

28. Bieler J, Cannavo R, Gustafson K, Gobet C, Gatfield D, Naef F. Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. Mol Syst Biol. 2014; 10: 739–739. https://doi.org/10.15252/msb.20145218 PMID: 25028488

29. Feillet C, Krusche P, Tamanini F, Janssens RC, Downey MJ, Martin P, et al. Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle. Proc Natl Acad Sci USA. 2014; 111: 9828–9833. https://doi.org/10.1073/pnas.1320474111 PMID: 24958884

30. Sugiyama M, Sakaue-Sawano A, Iimura T, Fukami K, Kitaguchi T, Kawakami K, et al. Illuminating cell-cycle progression in the developing zebrafish embryo. Proc Natl Acad Sci USA. 2009; 106(49): 20812–20817. https://doi.org/10.1073/pnas.0906464106 PMID: 19923430

31. Rihel J, Prober DA, Arvanites A, Lam K, Zimmerman S, Jang S, et al. Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. Science. 2010; 327(5963): 348–351. https://doi.org/10.1126/science.1183090 PMID: 20075256

32. Yang F, Gao C, Wang P, Zhang GJ, Chen Z. Fish-on-a-chip: Microfluidics for zebrafish research. Lab on a Chip. 2016; 16:1106–1125 https://doi.org/10.1039/c6lc00044d PMID: 26923141

33. Cong L, Wang Z, Chai Y, Hang W, Shang C, Yang W, et al. Rapid whole brain imaging of neural activity in freely behaving larval zebrafish (Danio rerio). Elife. 2017; 6:e28158 https://doi.org/10.7554/eLife.28158 PMID: 28930070

34. Prober DA, Rihel J, Onah AA, Sung R, Schier AF. Hypocretin / Orexin Overexpression Induces An Insomnia-Like Phenotype in Zebrafish. J Neurosci. 2006; 26: 13400–13410. https://doi.org/10.1523/JNEUROSCI.4332-06.2006 PMID: 17182791

35. Cahill GM. Automated Video Image Analysis of Larval Zebrafish Locomotor Rhythms. Circadian Rhythms: Methods and Protocols. Totowa, NJ: Humana Press; 2007. pp. 83–94.

36. Zhdanova I V, Wang SY, Leclair OU, Danilova NP. Melatonin promotes sleep-like state in zebrafish. Brain Res. 2001; 903: 263–268. https://doi.org/10.1016/s0006-8993(01)02444-1 PMID: 11382414

37. Ziv L, Levkovitz S, Toyama R, Falcon J, Gothilf Y. Functional development of the zebrafish pineal gland: Light-induced expression of period2 is required for onset of the circadian clock. J Neuroendocrinol. 2005; 17: 314–320. https://doi.org/10.1111/j.1365-2826.2005.01315.x PMID: 15869567

38. Gamse JT, Shen YC, Thisse C, Thisse B, Raymond PA, Halpern ME, et al. Otx5 regulates genes that show circadian expression in the zebrafish pineal complex. Nat Genet; 2002; 30: 117–121. https://doi.org/10.1038/ng793 PMID: 11753388

39. Nagoshi E, Brown SA, Dibner C, Kornmann B, Schibler U. Circadian Gene Expression in Cultured Cells. Methods in Enzymol; 2005. pp. 543–557.

40. Inada Y, Uchida H, Umemura Y, Nakamura W, Sakai T, Koike N, et al. Cell and tissue-autonomous development of the circadian clock in mouse embryos. FEBS Lett. 2014; 588: 459–465. https://doi.org/10.1016/j.febslet.2013.12.007 PMID: 24374345

41. Kawakami K, Takeda H, Kawakami N, Kobayashi M, Matsuda N, Mishina M. A transposon-mediated gene trap approach identifies developmentally regulated genes in zebrafish. Dev Cell. 2004; 7: 133–144. https://doi.org/10.1016/j.devcel.2004.06.005 PMID: 15239961

42. Appelbaum L, Wang G, Yokogawa T, Skariah GM, Smith J, Mourrain P, et al. Circadian and Homeostatic Regulation of Structural Synaptic Plasticity in Hypocretin Neurons Lior. Neuron. 2011; 68: 87–98.

43. Carvalho L, Heisenberg C. Imaging Zebrafish Embryos by Two-Photon Excitation Time-Lapse Microscopy. Zebrafish, Methods in Molecular Biology. 2009. pp. 273–287.

44. Lopez-Ramirez MA, Calvo C-F, Ristori E, Thomas J-L, Nicoli S. Isolation and Culture of Adult Zebrafish Brain-derived Neurospheres. J Vis Exp. 2016; https://doi.org/10.3791/53617 PMID: 26967835

45. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018; 36: 411–420 https://doi.org/10.1038/nbt.4096 PMID: 29608179

46. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: An open source platform for biological image analysis. Nat Methods. 2012; 9: 676–682. https://doi.org/10.1038/nmeth.2019 PMID: 22743772

47. Feng L, Zhao T, Kim J. neuTube 1.0: A New Design for Efficient Neuron Reconstruction Software Based on the SWC Format. eNeuro. 2007; 27: 9247–51.

3.3 Conclusion

Twenty years ago, through extensive molecular and genetic studies, the core clock genes (BMAL1, CLOCK, PERs, CRYs, REV-ERBα (NR1D1), REV-ERBβ (NR1D2)) and the transcription-translation feedback loops formed by them that drive the rhythmic expression of thousands of genes were identified. After that, with the development of the genome-wide transcriptome studies, more and more genes were discovered involving in circadian regulation. By integrating all the published circadian transcriptome data in mice, we have identified 41 circadian genes that showed consistent circadian oscillation in more than seven mouse tissues and constructed a comprehensive gene regulatory network for mouse circadian clock (Yan et al., 2008). Then by integrating the spatial gene expression atlas information and transcriptome data, we contracted a circadian gene regulatory network for zebrafish circadian rhythm, which illustrated how circadian clock genes can control diverse physiological functions (Li et al., 2013). Meanwhile, more and more evidences showed that regulation at the level of chromatin, posttranscriptional, translational and posttranslational affects nearly all aspects of circadian rhythms. We systematically searched for the circadian miRNAs (Wang et al. 2016), LncRNAs (Fan et al. 2017), and RBPs (Jiang et al., 2018), and explored their role in circadian regulation. We found that they all involved in the modulation of circadian clock, which may enhance the robustness and precision of circadian regulation. Although the molecular mechanism of circadian clock is largely known, how these cellular rhythms arise during development is not well understood. So we systematically screened for the clock reporter among the promoter of all the core clock genes, and got one clock reporter fish line (*nr1d1*:VNP) eventually. Using this fish line, we found that light can initialize the single-cell clock during early development in zebrafish (Wang et al., 2020). Individual molecular oscillators in the SCN need to cooperate with each other to function as a coherent pacemaker, while the cells in the SCN are heterogeneous and they fall into different cell types. To investigate how SCN cells are synchronized as a coherent pacemaker, we first applied stat-of-art technologies, scRNA-seq and smFISH, to character the cell identities and 3D organization of the SCN (Wen et al., 2020).

3.4 Outlook

We constructed live zebrafish carrying *nr1d1*:VNP fluorescence reporter, which allows for monitoring the circadian gene expression at a single-cell resolution in zebrafish larvae for the first time. This fish line could be used for many aspects of circadian research. As fluorescence ubiquitin cell cycle indicator (FUCCI) fish lines have been constructed to visualize the cell cycle stages at single cell level, we are planning to cross *nr1d1*:VNP fish with the FUCCI fish lines to investigate the cross-talk between circadian clock and cell cycle *in vivo*. Also, we want to cross *nr1d1*:VNP fish with fish lines labelled by cell-type specific markers, so we can study cell-type specific expression of circadian rhythms as we have demonstrated for rod-like photoreceptor cells. Furthermore, this fish line can also be applied to screen for small molecules that affect circadian rhythm at single-cell level *in vivo*.

By tracing the development of *nr1d1*, we found that light is essential for the establishment of single-cell oscillating. However, the circadian clock system of zebrafish consists of at least two

negative feedback loops, and *nr1d1* is only specific to the second loop. It is possible that the rapid dampening in DD condition during early development is specific to the "second" loop, whereas the classic BMAL/CLOCK – CRY/PER loop might behave differently. Therefore, it will be of great importance for the future studies to investigate single-cell expression of other circadian clock genes such as *period* genes during development. Furthermore, how the development of circadian clock is regulated at cellular and molecular level is still largely unknown. Recently, *in vitro* studies by Umemura et al. showed that posttranscriptional control of CLOCK protein contributing to the staring of the circadian clock in both mouse and human cells. However, an *in vivo* single-cell study on the regulation of circadian clock development is still lacking. By systematic searching, we have found several candidate regulators (TFs, RBPs, miRNAs, and hormones) that might be involved in the regulation of circadian clock development. Next, by crossing the clock reporter fish with the fish carrying mutations of those regulators, we can screen for the regulators involved in the regulation of circadian clock development, and then investigate the molecular mechanisms behind it. Last, but not least, *in vivo* study on the development of the master clock in the SCN at single-cell level have never been touched. I am planning to apply modern scRNA-seq technology to investigate the development of SCN at single-cell level.

4. References

Abbas-aghababazadeh, F., Id, Q.L., and Id, B.L.F. (2018). Comparison of normalization approaches for gene expression studies completed with high- throughput sequencing. PLoS one 13, e0206312.

Altschuler, S.J., and Wu, L.F. (2010). Essay Cellular Heterogeneity : Do Differences Make a Difference ? 141, 559–563.

Alvarez, J.D., Chen, D., Storer, E., and Sehgal, A. (2003). Non-cyclic and Developmental Stage-Specific Expression of Circadian Clock Proteins During Murine Spermatogenesis1. Biol. Reprod. 69, 81–91.

Allison, K.A., Kaikkonen, M.U., Gaasterland, T., and Glass, C.K. (2014). Vespucci: A system for building annotated databases of nascent transcripts. Nucleic Acids Res. 42, 2433-2447

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. 11, R106.

Ansari, N., Agathagelidis, M., Lee, C., Korf, H.-W., and Von Gall, C. (2009). Differential maturation of circadian rhythms in clock gene proteins in the suprachiasmatic nucleus and the pars tuberalis during mouse ontogeny. Eur. J. Neurosci. 29, 477–489.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.

Bedont, J.L., and Blackshaw, S. (2015). Constructing the suprachiasmatic nucleus: a watchmaker's perspective on the central clockworks. Front. Syst. Neurosci. 9, 74.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Brouzes, E., Medkova, M., Savenelli, N., Marran, D., Twardowski, M., Hutchison, J.B., Rothberg, J.M., Link, D.R., Perrimon, N., and Samuels, M.L. (2009). Droplet microfluidic technology for single-cell high-throughput screening. Proc. Natl. Acad. Sci. U.S.A. 106,14195-14200.

Buhr, E.D., and Takahashi, J.S. (2013). Molecular components of the mammalian circadian clock. Handb. Exp. Pharmacol. (217), 3–27.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420.

Cantor, R.M., Lange, K., and Sinsheimer, J.S. (2010). Prioritizing GWAS Results: A Review of Statistical Methods and Recommendations for Their Application. Am. J. Hum. Genet. 86, 6–22

Carmona-Alcocer, V., Abel, J.H., Sun, T.C., Petzold, L.R., Doyle, F.J., Simms, C.L., and Herzog, E.D. (2017). Ontogeny of circadian rhythms and synchrony in the suprachiasmatic nucleus. J. Neurosci. 38, 2006–2017.

Carr, A.J., and Whitmore, D. (2005). Imaging of single light-responsive clock cells reveals fluctuating free-running periods. Nat. Cell Biol. 7, 319–321.

Chen, R., D'Alessandro, M., and Lee, C. (2013). miRNAs are required for generating a time delay critical for the circadian oscillator. Curr. Biol. 23, 1959–1968.

Cheng, H.Y., Papp, J.W., Varlamova, O., Dziema, H., Russell, B., Curfman, J.P., Nakazawa, T., Shimizu, K., Okamura, H., Impey, S., et al. (2007). microRNA modulation of circadian-clock

period and entrainment. Neuron 54, 813–829.

Dale, R.K., Matzat, L.H., and Lei, E.P. (2014). metaseq: a Python package for integrative genome-wide analysis reveals relationships between chromatin insulators and associated nuclear mRNA. Nucleic Acids Res. 42, 9158–9170.

Dekens, M.P.S., and Whitmore, D. (2008). Autonomous onset of the circadian clock in the zebrafish embryo. EMBO J. 27, 2757–2765.

Delaunay, F., Thisse, C., Marchand, O., Laudet, V., and Thisse, B. (2000). An inherited functional circadian clock in zebrafish embryos. Science 289, 297–300.

Dibner, C., Schibler, U., and Albrecht, U. (2010). The Mammalian Circadian Timing System: Organization and Coordination of Central and Peripheral Clocks. Annu. Rev. Physiol. 72, 517–549.

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.

Doyle, S., and Menaker, M. (2007). Circadian Photoreception in Vertebrates. Cold Spring Harb. Symp. Quant. Biol. 72, 499–508.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., et al. (2009). Real-time DNA sequencing from single polymerase molecules. Science 323, 133-138 Eide, E.J., Woolf, M.F., Kang, H., Woolf, P., Hurst, W., Camacho, F., Vielhaber, E.L., Giovanni, A., and Virshup, D.M. (2005). Control of Mammalian Circadian Rhythm by CKIε-Regulated Proteasome-Mediated PER2 Degradation. Mol. Cell. Biol. 25, 2795–2807.

Fan, Z., Zhao, M., Joshi, P.D., Li, P., Zhang, Y., Guo, W., Xu, Y., Wang, H., Zhao, Z., and Yan, J. (2017). A class of circadian long non-coding RNAs mark enhancers modulating long-range circadian gene regulation. Nucleic Acids Res. 45, 5720–5738.

Fang, B., Everett, L.J., Jager, J., Briggs, E., Armour, S.M., Feng, D., Roy, A., Gerhart-Hines, Z., Sun, Z., and Lazar, M.A. (2014). Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. Cell. 159, 1140-1152

Freese, N.H., Norris, D.C., and Loraine, A.E. (2016). Integrated genome browser: visual analytics platform for genomics. Bioinformatics 32, 2089–2095.

Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W., and Rajewsky, N. (2011). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 40, 37–52.

Gantier, M.P., McCoy, C.E., Rusinova, I., Saulep, D., Wang, D., Xu, D., Irving, A.T., Behlke, M.A., Hertzog, P.J., MacKay, F., et al. (2011). Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. Nucleic Acids Res. 39, 5692-5703

Gatfield, D., Le Martelot, G., Vejnar, C.E., Gerlach, D., Schaad, O., Fleury-Olela, F., Ruskeepää, A.L., Oresic, M., Esau, C.C., Zdobnov, E.M., et al. (2009). Integration of microRNA miR-122 in hepatic circadian gene expression. Genes Dev. 23, 1313–1326.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS Lett. 582, 1977–1986.

Green, C.B. (2018). Circadian posttranscriptional regulatory mechanisms in mammals. Cold Spring Harb. Perspect. Biol. 10, 1–12.

Guillaumond, F., Dardente, H., Giguère, V., and Cermakian, N. (2005). Differential Control of Bmal1 Circadian Transcription by REV-ERB and ROR Nuclear Receptors. J. Biol. Rhythms 20, 391–403.

Haider, S., and Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. Curr. Genomics 14, 91–110.

Hardcastle, T.J., and Kelly, K.A. (2010). baySeq: Empirical Bayesian methods for identifying differential expression in sequence count data. BMC Bioinformatics 11, 422.

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. Cell Rep. 2, 666–673.

Johnson, C. H., Elliott, J.A., and Foster, R. (2003). Entrainment of Circadian Programs. Chronobiol. Int. 20, 741–774.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37, 1–13.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods 12, 115–121.

Hwang, B., Lee, J.H., and Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp. Mol. Med. 50, 96

Illumina, I. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53-39

Jeanmougin, M., de Reynies, A., Marisa, L., Paccard, C., Nuel, G., and Guedj, M. (2010). Should we abandon the t-test in the analysis of gene expression microarray data: a comparison of variance modeling strategies. PLoS One 5, e12336–e12336.

Jiang, X., Zhang, T., Wang, H., Wang, T., Qin, M., Bao, P., Wang, R., Liu, Y., Chang, H.C., Yan, J., et al. (2018). Neurodegeneration-associated FUS is a novel regulator of circadian gene expression. Transl. Neurodegener. 7, 1–11.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457–D462.

Kaneko, M., and Cahill, G.M. (2005). Light-dependent development of circadian gene expression in transgenic zebrafish. PLoS Biol. 3, e34–e34.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. 8, 533–543.

Kim, D.Y., Woo, K.C., Lee, K.H., Kim, T.-D., and Kim, K.T. (2010). hnRNP Q and PTB modulate the circadian oscillation of mouse Rev-erb α via IRES-mediated translation. Nucleic Acids Res. 38, 7068–7078.

Kim, D.Y., Kwak, E., Kim, S.H., Lee, K.H., Woo, K.C., and Kim, K.T. (2011). hnRNP Q mediates a phase-dependent translation-coupled mRNA decay of mouse Period3. Nucleic Acids Res. 39, 8901–8914.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2 : accurate alignment of transcriptomes in the presence of insertions , deletions and gene fusions. Genome Biol. 14, R36

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 161, 1187-1201

Kojima, S., Matsumoto, K., Hirose, M., Shimada, M., Nagano, M., Shigeyoshi, Y., Hoshino, S.,

Ui-Tei, K., Saigo, K., Green, C.B., et al. (2007). LARK activates posttranscriptional expression of an essential mammalian clock protein, PERIOD1. Proc. Natl. Acad. Sci. U. S. A, 104, 1859 LP – 1864.

Kojima, S., Gatfield, D., Esau, C.C., and Green, C.B. (2010). MicroRNA-122 modulates the rhythmic expression profile of the circadian deadenylase Nocturnin in mouse liver. PLoS One 5, e11264–e11264.

Kojima, S., Shingle, D.L., and Green, C.B. (2011). Post-transcriptional control of circadian rhythms. J. Cell Sci. 124, 311–320.

Kulski, J.K. (2016). Next-Generation Sequencing - An Overview of the History, Tools, and "Omic" Applications. Next Generation Sequencing - Advances, Applications and Challenges, 3-60.

Jiang, X., Zhang, T., Wang, H., Wang, T., Qin, M., Bao, P., Wang, R., Liu, Y., Chang, H.C., Yan, J., et al. (2018). Neurodegeneration-associated FUS is a novel regulator of circadian gene expression. Transl. Neurodegener. 7, 1–11

Landgraf, D., Koch, C.E., and Oster, H. (2014). Embryonic development of circadian clocks in the mammalian suprachiasmatic nuclei. Front. Neuroanat. 8, 143.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–360.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10, R25

Leng, N., Dawson, J.A., Thomson, J.A., Ruotti, V., Rissman, A.I., Smits, B.M.G., Haag, J.D., Gould, M.N., Stewart, R.M., and Kendziorski, C. (2013). EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. Bioinformatics 29, 1035–1043.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, J., and Tibshirani, R. (2013). Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Stat. Methods Med. Res. 22, 519–536.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., et al. (2010). The sequence and de novo assembly of the giant panda genome. Nature 463, 311–317.

Li, Y., Li, G., Wang, H., Du, J., and Yan, J. (2013). Analysis of a Gene Regulatory Cascade Mediating Circadian Rhythm in Zebrafish. PLoS Comput. Biol. 9, e1002940.

Lim, C., and Allada, R. (2013). Emerging roles for post-transcriptional regulation in circadian clocks. Nat. Neurosci. 16, 1544–1550.

Liu, W., Cao, H., Ye, C., Chang, C., Lu, M., Jing, Y., Zhang, D., Yao, X., Duan, Z., Xia, H., et al. (2014). Hepatic miR-378 targets p110α and controls glucose and lipid homeostasis by modulating hepatic insulin signalling. Nat. Commun. 5, 5684

Liu, Y., Hu, W., Murakawa, Y., Yin, J., Wang, G., Landthaler, M., and Yan, J. (2013). Cold-induced RNA-binding proteins regulate circadian gene expression by controlling alternative polyadenylation. Sci. Rep. 3, 2054.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550.

Lowrey, P.L., and Takahashi, J.S. (2004). Mammalian circadian biology: elucidating genome-wide levels of temporal organization. Annu. Rev. Genomics Hum. Genet. 5, 407–441.

Lowrey, P.L., Shimomura, K., Antoch, M.P., Yamazaki, S., Zemenides, P.D., Ralph, M.R.,

Menaker, M., and Takahashi, J.S. (2000). Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. Science 288, 483–492.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 161, 1202-1214

Mardis, E.R. (2008). Next-Generation DNA Sequencing Methods. Annu. Rev. Genomics Hum. Genet. 9, 387–402.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376-380.

Matys, V., Kel-Margoulis, O. V, Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., et al. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res. 34, D108–D110.

McGlincy, N.J., and Ingolia, N.T. (2017). Transcriptome-wide measurement of translation by ribosome profiling. Methods. 126, 112-129

McInnes, L., Healy, J., and Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv E-Prints arXiv:1802.03426.

Mehta, N., and Cheng, H.-Y.M. (2013). Micro-Managing the Circadian Clock: The Role of microRNAs in Biological Timekeeping. J. Mol. Biol. 425, 3609–3624.

Menet, J.S., Rodriguez, J., Abruzzi, K.C., and Rosbash, M. (2012). Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. Elife 1, e00011–e00011.

Mi, H., Dong, Q., Muruganujan, A., Gaudet, P., Lewis, S., and Thomas, P.D. (2009). PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res. 38, D204–D210.

Mohawk, J.A., Green, C.B., and Takahashi, J.S. (2012). Central and Peripheral Circadian Clocks in Mammals. Annu. Rev. Neurosci. 35, 445–462.

Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. 34, 267–273.

Morf, J., Rey, G., Schneider, K., Stratmann, M., Fujita, J., Naef, F., and Schibler, U. (2012). Cold-Inducible RNA-Binding Protein Modulates Circadian Gene Expression Posttranscriptionally. Science. 338, 379–383.

Morse, D., Cermakian, N., Brancorsini, S., Parvinen, M., and Sassone-Corsi, P. (2003). No Circadian Rhythms in Testis: Period1 Expression Is Clock Independent and Developmentally Regulated in the Mouse. Mol. Endocrinol. 17, 141–151.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628.

Mundade, R., Ozer, H.G., Wei, H., Prabhu, L., and Lu, T. (2014). Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond. Cell Cycle 13, 2847–2852.

Na, Y.J., Sung, J.H., Lee, S.C., Lee, Y.J., Choi, Y.J., Park, W.Y., Shin, H.S., and Kim, J.H. (2009). Comprehensive analysis of microRNA-mRNA co-expression in circadian rhythm. Exp.

Mol. Med. 41, 638–647.

Nagel, R., Clijsters, L., and Agami, R. (2009). The miRNA-192/194 cluster regulates the Period gene family and the circadian clock. FEBS J. 276, 5447–5455.

Nakasugi, K., Crowhurst, R.N., Bally, J., Wood, C.C., Hellens, R.P., and Waterhouse, P.M. (2013). De Novo Transcriptome Sequence Assembly and Analysis of RNA Silencing Genes of Nicotiana benthamiana. PLoS One 8, e59534.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44, D733–D745.

Ono, D., Honma, S., and Honma, K. (2013). Cryptochromes are critical for the development of coherent circadian rhythms in the mouse suprachiasmatic nucleus. Nat. Commun. 4, 1666.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research, 13, 2498–2504.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181.

Preitner, N., Damiola, F., Lopez-Molina, L., Zakany, J., Duboule, D., Albrecht, U., and Schibler, U. (2002). The Orphan Nuclear Receptor REV-ERBα Controls Circadian Transcription within the Positive Limb of the Mammalian Circadian Oscillator. Cell 110, 251–260.

Preußner, M., and Heyd, F. (2016). Post-transcriptional control of the mammalian circadian clock: implications for health and disease. Pflugers Arch. 468, 983–991.

R Core Team (2019) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at https://www.R-project.org/.

Ramsköld, D., Luo, S., Wang, Y., Li, R., Deng, Q., Omid, R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2013). Full-Length mRNA-Seq from single cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 30, 777–782.

Raney, B.J., Dreszer, T.R., Barber, G.P., Clawson, H., Fujita, P.A., Wang, T., Nguyen, N., Paten, B., Zweig, A.S., Karolchik, D., et al. (2013). Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics 30, 1003–1005.

Rastogi, S., Meena, S., Bhattacharya, A., Ghosh, S., Shukla, R.K., Sangwan, N.S., Lal, R.K., Gupta, M.M., Lavania, U.C., Gupta, V., et al. (2014). De novo sequencing and comparative analysis of holy and sweet basil transcriptomes. BMC Genomics 15, 558

Rau, A., Marot, G., and Jaffrézic, F. (2014). Differential meta-analysis of RNA-seq data from multiple studies. BMC Bioinformatics 15, 91.

Reppert, S.M., and Weaver, D.R. (2001). Molecular Analysis of Mammalian Circadian Rhythms. Annu. Rev. Physiol. 63, 647–676.

Roberts, A., Trapnell, C., Donaghey, J., Rinn, J.L., and Pachter, L. (2011). Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 12, R22.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for

differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Robles, M.S., Cox, J., and Mann, M. (2014). In-Vivo Quantitative Proteomics Reveals a Key Contribution of Post-Transcriptional Mechanisms to the Circadian Regulation of Liver Metabolism. PLOS Genet. 10, e1004047.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature 475, 348-352.

Sanger, F., and Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 94, 441-448.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors.   Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. Nat. Biotechnol. 26, 1135–1145.

Shimomura, H., Moriya, T., Sudo, M., Wakamatsu, H., Akiyama, M., Miyake, Y., and Shibata, S. (2001). Differential daily expression of Per1 and Per2 mRNA in the suprachiasmatic nucleus of fetal and early postnatal mice. Eur. J. Neurosci. 13, 687–693.

Shirogane, T., Jin, J., Ang, X.L., and Harper, J.W. (2005). SCFβ-TRCP Controls Clock-dependent Transcription via Casein Kinase 1-dependent Degradation of the Mammalian Period-1 (Per1) Proteinm. J. Biol. Chem. 280, 26863–26872.

Sigurdardottir, L.G., Valdimarsdottir, U.A., Fall, K., Rider, J.R., Lockley, S.W., Schernhammer, E., and Mucci, L.A. (2012). Circadian disruption, sleep loss, and prostate cancer risk: a systematic review of epidemiologic studies. Cancer Epidemiol. Biomarkers Prev. 21, 1002–1011.

Sládek, M., Sumová, A., Kováčiková, Z., Bendová, Z., Laurinová, K., and Illnerová, H. (2004). Insight into molecular core clock mechanism of embryonic and early postnatal rat suprachiasmatic nucleus. Proc. Natl. Acad. Sci. U. S. A. 101, 6231–6236.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., Nimwegen, E. van, Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480, 490–495.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550.

Sudmant, P.H., Alexis, M.S., and Burge, C.B. (2015). Meta-analysis of RNA-seq expression data across species, tissues and studies. Genome Biol. 16, 287.

Suzuki, A., Kawano, S., Mitsuyama, T., Suyama, M., Kanai, Y., Shirahige, K., Sasaki, H., Tokunaga, K., Tsuchihara, K., Sugano, S., et al. (2018). DBTSS/DBKERO for integrated analysis of transcriptional regulation. Nucleic Acids Res. 46, D229–D238.

Takahashi, H., Kato, S., Murata, M., and Carninci, P. (2012). CAGE (cap analysis of gene expression): a protocol for the detection of promoter and transcriptional networks. Methods Mol. Biol. 786, 181–200.

Takahashi, J.S., Hong, H.K., Ko, C.H., and McDearmon, E.L. (2008). The genetics of mammalian

circadian order and disorder: implications for physiology and disease. Nat. Rev. Genet. 9, 764–775.

Tan, X., Zhang, P., Zhou, L., Yin, B., Pan, H., and Peng, X. (2012). Clock-controlled mir-142-3p can target its activator, Bmal1. BMC Mol. Biol. 13, 27.

Tarazona, S., García, F., Ferrer, A., Dopazo, J., and Conesa, A. NOIseq: a RNA-seq differential expression method robust for sequencing depth biases. EMBnet.Journal 17, 18–19.

Tarazona, S., García-Alcalde, F., Dopazo, J., Ferrer, A., and Conesa, A. (2011). Differential expression in RNA-seq: a matter of depth. Genome Res. 21, 2213–2223.

Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A. Di, Nueda, M.J., Ferrer, A., and Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res. 43, e140–e140.

Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature 563, 72–78.

The Gene Ontology Consortium (2018). The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47, D330–D338.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: A library of protein families and subfamilies indexed by function. Genome Res. 13, 2129–2141.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth II, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352, 189–196.

Torres, M., Becquet, D., Blanchard, M.-P., Guillen, S., Boyer, B., Moreno, M., Franc, J.-L., and François-Bellan, A.-M. (2016). Circadian RNA expression elicited by 3'-UTR IRAlu-paraspeckle associated elements. Elife 5, e14837.

Tost, J. (2010). DNA Methylation: An Introduction to the Biology and the Disease-Associated Changes of a Promising Biomarker. Mol. Biotechnol. 44, 71–81.

Toyama, R., Chen, X., Jhawar, N., Aamar, E., Epstein, J., Reany, N., Alon, S., Gothilf, Y., Klein, D.C., and Dawid, I.B. (2009). Transcriptome analysis of the zebrafish pineal gland. Dev. Dyn. 238, 1813–1826.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515.

Umemura, Y., and Yagita, K. (2020). Development of the Circadian Core Machinery in Mammals. J. Mol. Biol. In Press.

Umemura, Y., Koike, N., Ohashi, M., Tsuchiya, Y., Meng, Q.J., Minami, Y., Hara, M., Hisatomi, M., and Yagita, K. (2017). Involvement of posttranscriptional regulation of Clock in the emergence of circadian clock oscillation during mouse development. Proc. Natl. Acad. Sci. 114, E7479 LP-E7488.

Umemura, Y., Maki, I., Tsuchiya, Y., Koike, N., and Yagita, K. (2019). Human Circadian Molecular Oscillation Development Using Induced Pluripotent Stem Cells. J. Biol. Rhythms 34,

525–532.

Valinezhad Orang, A., Safaralizadeh, R., and Kazemzadeh-Bavili, M. (2014). Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation. Int. J. Genomics 2014, 970607.

Vallone, D., Lahiri, K., Dickmeis, T., and Foulkes, N.S. (2007). Start the Clock ! Circadian Rhythms and Development. Dev. Dyn. 236, 142–155.

Valouev, A., Ichikawa, J., Tonthat, T., Stuart, J., Ranade, S., Peckham, H., Zeng, K., Malek, J.A., Costa, G., McKernan, K., et al. (2008). A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res. 18, 1051–1063.

Vanselow, K., Vanselow, J.T., Westermark, P.O., Reischl, S., Maier, B., Korte, T., Herrmann, A., Herzel, H., Schlosser, A., and Kramer, A. (2006). Differential effects of PER2 phosphorylation: molecular basis for the human familial advanced sleep phase syndrome (FASPS). Genes Dev. 20, 2660–2672.

Vejnar, C.E., and Zdobnov, E.M. (2012). MiRmap: comprehensive prediction of microRNA target repression strength. Nucleic Acids Res. 40, 11673–11683.

Vollmers, C., Schmitz, R.J., Nathanson, J., Yeo, G., Ecker, J.R., and Panda, S. (2012). Circadian oscillations of protein-coding and regulatory RNAs in a highly dynamic mammalian liver epigenome. Cell Metab. 16, 833–845.

Wang, J., Symul, L., Yeung, J., Gobet, C., Sobel, J., Lück, S., and Westermark, P.O., Molina N., Naef F. (2018). posttranscriptional regulation underlies temporal mRNA accumulation in mouse liver. Proc. Natl. Acad. Sci. U. S. A. 115, E1916-E1925

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2009). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics 26, 136–138.

Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C.A., Zhang, Y., and Liu, X.S. (2013). Target analysis by integration of transcriptome and ChIP-seq data with BETA. Nat. Protoc. 8, 2502–2515.

Wang, X.S., Armstrong, M.E.G., Cairns, B.J., Key, T.J., and Travis, R.C. (2011). Shift work and chronic disease: the epidemiological evidence. Occup. Med. (Lond). 61, 78–89.

Wang, Z., Ma, B., Ji, X., Deng, Y., Zhang, T., Zhang, X., et al. (2015). MicroRNA-378-5p suppresses cell proliferation and induces apoptosis in colorectal cancer cells by targeting BRAF. Cancer Cell Int. 15, 40.

Wen, S., Ma, D., Zhao, M., Xie, L., Wu, Q., Gou, L., Zhu, C., Fan, Y., Wang, H., and Yan, J. (2020). Spatiotemporal single-cell analysis of gene expression in the mouse suprachiasmatic nucleus. Nat. Neurosci. 23, pages456–467

Woo, K.C., Kim, T.D., Lee, K.H., Kim, D.Y., Kim, W., Lee, K.Y., and Kim, K.T. (2009). Mouse period 2 mRNA circadian oscillation is modulated by PTB-mediated rhythmic mRNA degradation. Nucleic Acids Res. 37, 26–37.

Woo, K.C., Ha, D.C., Lee, K.H., Kim, D. Y., Kim, T.-D., and Kim, K.T. (2010). Circadian Amplitude of Cryptochrome 1 Is Modulated by mRNA Stability Regulation via Cytoplasmic hnRNP D Oscillation. Mol. Cell. Biol. 30, 197–205.

Xing, W., Busino, L., Hinds, T.R., Marionni, S.T., Saifee, N.H., Bush, M.F., Pagano, M., and Zheng, N. (2013). SCF(FBXL3) ubiquitin ligase targets cryptochromes at their cofactor pocket. Nature 496, 64–68.

Xu, S., Witmer, P.D., Lumayag, S., Kovacs, B., and Valle, D. (2007). MicroRNA (miRNA)

Transcriptome of Mouse Retina and Identification of a Sensory Organ-specific miRNA Cluster. J. Biol. Chem. 282, 25053–25066.

Yagita, K., Horie, K., Koinuma, S., Nakamura, W., Yamanaka, I., Urasaki, A., Shigeyoshi, Y., Kawakami, K., Shimada, S., Takeda, J., et al. (2010). Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. Proc. Natl. Acad. Sci. U. S. A. 107, 3846–3851.

Yamaguchi, S., Mitsui, S., Yan, L., Yagita, K., Miyake, S., and Okamura, H. (2000). Role of DBP in the Circadian Oscillatory Mechanism. Mol. Cell. Biol. 20, 4773–4781.

Yamashita, R., Sathira, N.P., Kanai, A., Tanimoto, K., Arauchi, T., Tanaka, Y., Hashimoto, S.-I., Sugano, S., Nakai, K., and Suzuki, Y. (2011). Genome-wide characterization of transcriptional start sites in humans by integrative transcriptome analysis. Genome Res. 21, 775–789.

Yang, X., and Li, L. (2011). miRDeep-P: a computational tool for analyzing the microRNA transcriptome in plants. Bioinformatics 27, 2614–2615.

Yan, J., Wang, H., Liu, Y., and Shao, C. (2008). Analysis of gene regulatory networks in the mammalian circadian rhythm. PLoS Comput. Biol. 4, e1000193.

Ye, R., Selby, C.P., Chiou, Y.-Y., Ozkan-Dagliyan, I., Gaddameedhi, S., and Sancar, A. (2014). Dual modes of CLOCK:BMAL1 inhibition mediated by Cryptochrome and Period proteins in the mammalian circadian clock. Genes Dev. 28, 1989–1998.

Yoo, S.H., Yamazaki, S., Lowrey, P.L., Shimomura, K., Ko, C.H., Buhr, E.D., Siepka, S.M., Hong, H.-K., Oh, W.J., Yoo, O.J., et al. (2004). PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc. Natl. Acad. Sci. U. S. A. 101, 5339–5346.

Yoo, S.-H., Kojima, S., Shimomura, K., Koike, N., Buhr, E.D., Furukawa, T., Ko, C.H., Gloston, G., Ayoub, C., Nohara, K., et al. (2017). Period2 3'-UTR and microRNA-24 regulate circadian rhythms by repressing PERIOD2 protein accumulation. Proc. Natl. Acad. Sci. U. S. A. 114, E8855–E8864.

Zhang, X., McNeil, G.P., Hilderbrand-Chae, M.J., Franklin, T.M., Schroeder, A.J., and Jackson, F.R. (2000). Circadian regulation of the lark RNA-binding protein within identifiable neurosecretory cells. J. Neurobiol. 45, 14–29.

Zhang, X., Li, T., Liu, F., Systems, U.S.R., Zhang, X., Li, T., Liu, F., Chen, Y., Yao, J., Li, Z., et al. (2019). Comparative Analysis of Droplet-Based Ultra-High- Article Comparative Analysis of Droplet-Based. Mol. Cell 73, 130-142.

Zhao, S., Fung-Leung, W.P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 9, e78644.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049

Ziv, L., Levkovitz, S., Toyama, R., Falcon, J., and Gothilf, Y. (2005). Functional development of the zebrafish pineal gland: Light-induced expression of period2 is required for onset of the circadian clock. J. Neuroendocrinol. 17, 314–320.