

Computational analysis of circadian splicing events
in human cancer cell lines and mammalian tissues

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

vorgelegt von

Rukeia El-Athman

Berlin, Dezember 2019

Gutachter/innen:

Erstgutachterin:

Frau PD Dr. Angela Relógio

Institut für Theoretische Biologie und Molekulares Krebsforschungszentrum

Charité – Universitätsmedizin Berlin

Weitere Gutachter/innen:

Frau Prof. Dr. Heike Siebert

Fachbereich Mathematik und Informatik

Freie Universität Berlin

Herr Prof. Dr. Hanspeter Herzel

Institut für Theoretische Biologie

Charité – Universitätsmedizin Berlin

Tag der Disputation: 29. Juli 2020

Abstract

The circadian clock regulates physiology and behavior of various organisms in synchrony with daily environmental rhythms. At the cellular level, circadian rhythmicity is driven by the interplay of clock genes and proteins that interact via negative feedback loops, thereby causing oscillations with a period of 24 h in the expression of numerous target genes. The resulting rhythms in the abundance of proteins and other biomolecules are responsible for the temporal organization of diverse biological processes. Accumulating evidence suggests that alternative splicing might be one of these clock-controlled processes. Alternative splicing describes a versatile mechanism of gene regulation that generates several distinct protein isoforms from a single gene via the differential inclusion or exclusion of alternate RNA regions. Both disruptions of the circadian clock and aberrant splicing are associated with carcinogenesis and tumor progression.

This dissertation seeks to answer the question whether mammalian alternative splicing is regulated by the circadian clock, and whether the hypothesized regulation differs between cancer cells in different tumor stages. In particular, it tries to elucidate whether changes in circadian regulated splicing events could be responsible for the production of protein isoforms that contribute to the malignant development of cancer cells. The study is based on data from two human colon cancer cell lines, SW480 and SW620, that have been derived from a primary tumor and a metastasis of the same patient and thus serve as an *in vitro* model of colorectal tumor progression. A computational analysis was conducted to identify 24-h rhythmic genes and alternative splicing events on transcriptome-level based on the time-series data of both cell lines. As a reference, previously published time-series data of numerous healthy tissues from mouse and baboon organs were analyzed.

The analysis revealed differences in the circadian phenotype of the two cell lines, with the metastasis-derived cell line SW620 exhibiting a stronger dysregulation of circadian rhythmicity. Furthermore, this work shows that splicing-related genes and putative splicing events display 24-h rhythms that differ between primary tumor- and metastasis-derived cells. Both in healthy tissues and cancer cells, rhythmic splicing was found to affect many genes that are themselves involved in splicing, suggesting a partial autoregulation of the process. Several of the spliced candidate genes encode for protein isoforms that are involved in processes promoting tumor progression, such as migration and angiogenesis. Taken together, the results presented in this dissertation point to a circadian regulation of alternative splicing that plays a role in cancer development.

Acknowledgements

The work presented in this thesis was developed between April 2016 and November 2019 in the group of Dr. Angela Relógio (Systems Biology of Cancer) at the Institute for Theoretical Biology and the Molecular Cancer Research Center, Charité – Universitätsmedizin Berlin, and Humboldt-Universität zu Berlin. It was financed by the Bundesministerium für Bildung und Forschung (eBio-CIRSPLICE - FKZ031A316) and by the Dr. Rolf M. Schwiete Stiftung.

I would like to thank my supervisor, Dr. Angela Relógio, for giving me the opportunity to work on this exciting project, for her constant support and trust in my work, as well as for being part of my thesis advisory committee, and for reviewing this thesis. I am grateful to my second supervisor Prof. Heike Siebert for being part of my thesis advisory committee and would like to thank her for many helpful discussions and encouragement, and for reviewing this thesis. I thank Prof. Hanspeter Herzel for his insightful seminars on all things oscillating and for reviewing this thesis. Further thanks go to Prof. Thomas F. Meyer for being part of my thesis advisory committee.

I thank the IRI Graduate School Life Sciences for many interesting seminars and for financial support for travels. I am grateful to the Joachim Herz Stiftung whose Add-On Fellowship for Interdisciplinary Life Sciences enabled me, amongst other things, to participate in several courses and conferences in the last three years. My heartfelt thanks go to Prof. Suliana Manley who agreed to be my mentor in the framework of the DREAM mentoring program for female doctoral students of the Dahlem Research School of the Freie Universität Berlin.

To my former and current fellow members of the Systems Biology of Cancer group, I owe the deepest gratitude for their limitless support and for sharing many hours of laughter and science: Alireza Basti, Luise Fuhr, Mónica Abreu, Müge Yalçın, Nikolai Genov, Rosario Astaburuaga, and Yin Li, you rock around the clock! Special thanks go to Luise for generating the experimental data on which this thesis is based, and to Dora Knezevic for her support of the computational analysis during her internship in the group. I also thank the ITB lunch group, both from Invalidenstraße and House 4: Marjan Faizi, Philipp Burt, Janek Grzegorzewski, Rike Schuppner, Abhishek Upadhyay, and Bharath Ananthasubramaniam, you made going to the Mensa every day a true treat that I will greatly miss.

This work would not have been possible without the manifold support and encouragement from my friends and family. My special thanks go to my sisters Fatima, Amina, and Sumeia and to my friend Eileen who, I am sure, all learned a lot about the importance of the circadian clock and alternative splicing during the last years.

Contents

1	Introduction	1
1.1	The Circadian Clock.....	1
1.1.1	A Brief History of Chronobiology	2
1.1.2	Architecture of the Mammalian Circadian System.....	3
1.1.3	The Role of the Circadian Clock in Cancer	5
1.1.4	Ultradian Rhythms and Circadian Harmonics.....	8
1.1.5	Analysis of Circadian Rhythmicity on the Genome-Scale.....	10
1.2	Pre-mRNA Splicing	15
1.2.1	Mechanism and Regulation of Splicing.....	15
1.2.2	Alternative Splicing	18
1.2.3	The Role of Aberrant Splicing in Cancer	19
1.2.4	Analysis of Alternative Splicing on the Genome-Scale	21
1.3	Interplay between the Circadian Clock and mRNA Splicing.....	26
1.3.1	Regulation of the Circadian Clock via Alternative Splicing	26
1.3.2	Regulation of Alternative Splicing via the Circadian Clock	28
1.3.3	A Role for Clock-Regulated Splicing Events in Cancer?.....	30
1.4	Aim and Structure of the Thesis.....	31
2	Materials and Methods.....	33
2.1	Materials.....	33
2.1.1	Deposited Data.....	33
2.1.2	Software and Algorithms.....	34
2.2	Molecular and Cell Biology Methods	35
2.2.1	Sample Preparation.....	35
2.3	Bioinformatics and Statistical Methods	35
2.3.1	Processing of Microarray Transcriptome Data	36
2.3.2	Processing of RNA-seq Transcriptome Data	36
2.3.3	Rhythmicity Analysis	37
2.3.4	Correlation Analysis.....	38
2.3.5	Clustering Analysis	39
2.3.6	Detection of Alternatively Spliced Exons in Microarray Data.....	39
2.3.7	Detection of Differentially Rhythmic Splice Variants in RNA-seq Data	40
2.3.8	Prediction of Transcription Factor Binding Sites.....	40
2.3.9	Functional Enrichment Analysis.....	41
2.3.10	Compilation of Lists of Splicing-related Genes	42
2.3.11	Visualization.....	42

3	Results	43
3.1	Analysis of the Circadian Transcriptome of Human CRC Cell Lines.....	43
3.1.1	Genome-wide Analysis of Differential Rhythmicity between SW480 and SW620 Cells.....	43
3.1.2	Functional Annotation of Circadian Genes in CRC Cells Reveals Phase-shifted Pathways.....	48
3.1.3	Detection of Differential Rhythmicity of Splicing-related Genes.....	51
3.1.4	Identification of Differential and Rhythmic Splicing Events in SW480 and SW620 Cells.....	53
3.2	Comparison of Circadian Microarray and RNA-Seq Data of CRC Cell Lines.....	58
3.2.1	Correlation of Gene Expression between Microarray and RNA-seq Data.....	58
3.2.2	Normalization and Cross-Platform Concatenation of Circadian Datasets.....	60
3.2.3	Detection of Differentially Rhythmic Phase-shifted Splice Variants in CRC Cell Lines.....	65
3.3	Analysis of Rhythmic Splicing Events in Mammalian Tissues.....	70
3.3.1	Identification of Putative Ultradian and Circadian AS Events in Murine Organs.....	71
3.3.2	Comparison of Transcript-level and Gene-level Expression in Baboon Tissues.....	74
3.3.3	Detection of Differentially Rhythmic Phase-shifted Splice Variants in Baboon Tissues.....	77
3.3.4	Conserved Rhythmicity of Splicing-related Genes across Mammalian Tissues.....	83
4	Discussion	87
4.1	A CRC Cell Line Model Reveals Disruption of Clock Genes during Tumor Progression.....	87
4.2	Circadian Rhythmicity Differs between Primary Tumor- and Metastasis-derived Cells.....	88
4.3	Rhythmic AS Events in CRC are likely Regulated by Circadian Splicing Genes.....	90
4.4	Rhythmic Splicing Genes and AS Events are Widespread in Mammalian Tissues.....	94
4.5	Limitations of the Study.....	96
5	Conclusion and Outlook	99
6	Bibliography	103
7	Appendix	121
7.1	Supplementary Material.....	121
7.1.1	Supplementary Figures.....	121
7.1.2	Supplementary Tables.....	138
7.1.3	External Data Files.....	168
7.2	Zusammenfassung.....	169
7.3	Eidesstattliche Erklärung.....	171

Index of Figures and Tables

Index of Figures

Figure 1-1: Mammalian core circadian feedback loops.....	4
Figure 1-2: Molecular connections between the circadian clock and the cell cycle in mammals.	7
Figure 1-3: Schematic representation of the harmonic regression model.....	11
Figure 1-4: Geometric interpretation of the least squares estimation.	12
Figure 1-5: Schematic representation of the RAIN algorithm for rhythm detection.	14
Figure 1-6: Transesterification reactions during pre-mRNA splicing.	16
Figure 1-7: Stepwise spliceosome assembly.....	17
Figure 1-8: Types of AS events.	19
Figure 1-9: Microarray design.....	22
Figure 1-10: Types of RNA-seq reads.....	24
Figure 1-11: Schematic representation of the mapping strategies of STAR and Salmon.....	25
Figure 1-12: Interplay between the circadian clock and alternative splicing.	27
Figure 1-13: Structure of the thesis.....	31
Figure 2-1: Sampling of a cellular model of human CRC progression.....	35
Figure 2-2: Schematic representation of the FIRMA and FIRMAGene analysis.....	39
Figure 2-3: Schematic representation of the analysis of differentially rhythmic SVPs in RNA-seq data.....	40
Figure 3-1: Core clock genes and circadian transcription factors in SW480 and SW620 cells.	44
Figure 3-2: Transcriptome-level analysis of 24-h rhythmic genes in SW480 and SW620 cells.	45
Figure 3-3: Differential rhythmicity analysis between rhythmic genes in SW480 and SW620 cells.	46
Figure 3-4: Phase-clustered circadian pathways in SW480 and SW620 cells.	50
Figure 3-5: Spliceosome-associated genes are phase-shifted between SW480 and SW620 cells.....	51
Figure 3-6: Splicing-related genes display 24-h rhythms in SW480 and SW620 cells.....	52
Figure 3-7: Candidate differential AS events between SW480 and SW620 cells.	55
Figure 3-8: Comparison of selected candidate circadian AS events between SW480 and SW620 cells.	56
Figure 3-9: Correlation of gene expression between circadian microarray and RNA-seq data.	60
Figure 3-10: RNA-seq expression data of CRC cell lines before and after TDM-normalization.....	61
Figure 3-11: Cross-platform comparison of 24-h rhythmic gene sets of SW480 and SW620 cells.....	63
Figure 3-12: Cross-platform gene-wise comparison of circadian parameters of 24-h rhythmic genes.....	64
Figure 3-13: Expression of core clock genes in the concatenated microarray and RNA-seq time-series.	65
Figure 3-14: Differentially 24-h rhythmic phase-shifted splice variants of the same gene in CRC cell lines.....	66
Figure 3-15: Phase-shifted splice variants in SW480 and SW620 cells are linked with the splicing process.....	68
Figure 3-16: Candidate genes with phase-shifted splice variants are linked with apoptosis and migration.....	70
Figure 3-17: Re-analysis of the circadian transcriptome of twelve murine tissues.....	71
Figure 3-18: Candidate genes with 24-h and 12-h rhythmic AS events in murine tissues.	72
Figure 3-19: Enriched processes for candidate genes with rhythmic AS events in murine tissues.....	74
Figure 3-20: Re-analysis of the circadian transcriptome of 64 olive baboon tissues.....	75
Figure 3-21: Discrepancies between time-series expression identified on gene- and transcript-level.	76
Figure 3-22: Analysis of differentially rhythmic SVPs of the same gene in the same baboon tissue.	78
Figure 3-23: Enriched biological processes for genes with rhythmic phase-shifted splice variants in baboon. ...	79

Figure 3-24: Phase distributions of top candidate phase-shifted splice variants in baboon tissues.	80
Figure 3-25: Conserved 24-h rhythmicity of splicing-related genes in mouse and baboon.....	84
Figure 3-26: Phase distribution of robustly oscillating splicing-related genes across baboon tissues.....	86
Figure S 1: Time-series expression of core clock genes in SW480 and SW620 cells (microarray data).....	121
Figure S 2: Comparison of the circadian transcriptome between SW480 and SW620 cells.....	121
Figure S 3: Functional annotation for differentially rhythmic genes between SW480 and SW620 cells.....	122
Figure S 4: Functional annotation of clusters of 24-h rhythmic genes in SW480 cells.....	123
Figure S 5: Functional annotation of clusters of 24-h rhythmic genes in SW620 cells.....	124
Figure S 6: Time-series expression of 24-h rhythmic splicing-related genes in CRC cell lines.....	125
Figure S 7: Differential AS events in NCAM1 and the EMT-associated genes CD44 and FGFR2.....	126
Figure S 8: Candidate circadian AS events in SW480 and SW620 cells.....	127
Figure S 9: RNA-seq mapping rates and number of expressed genes in microarray and RNA-seq data.	128
Figure S 10: Rhythmic gene sets identified in SW480 and SW620 cells.	128
Figure S 11: Time-series expression of core clock genes in SW480 and SW620 cells (RNA-seq data).	129
Figure S 12: Expression heatmaps of 24-h rhythmic genes in SW480 and SW620 cells.....	129
Figure S 13: Rhythmic gene and transcript sets identified in SW480 and SW620 cells.	130
Figure S 14: Expression of 24-h rhythmic splicing-related genes in CRC cell lines.....	131
Figure S 15: Period distribution of rhythmic genes in mouse and baboon tissues.	132
Figure S 16: Number of 12-h and 24-h rhythmic genes in mouse and baboon tissues.	133
Figure S 17: Murine candidate genes with putative rhythmic AS events.	134
Figure S 18: Intersections between baboon genes identified as rhythmic on gene- and transcript-level.	135
Figure S 19: Comparison of rhythmic parameters between gene- and transcript-level in baboon.....	136
Figure S 20: Time-series expression of 24-h rhythmic splicing-related genes in mouse and baboon.....	137

Index of Tables

Table 1: Overview of candidate genes with circadian AS events in SW480 and SW620 cells.	57
Table 2: Topmost consistently 24-h rhythmic splicing-related genes in mouse and baboon tissues.....	85
Table S 1: Lists of human splicing-related genes.....	138
Table S 2: Candidate exons with differential AS events between SW480 and SW620 cells.....	149
Table S 3: GO terms enriched for genes with differential AS events between SW480 and SW620 cells.....	156
Table S 4: GO terms enriched for genes with 24-h rhythmic phase-shifted SVPs in SW480 and SW620 cells....	157
Table S 5: Tissues in the murine and baboon multi-organ circadian transcriptome datasets.	159
Table S 6: GO terms enriched for genes with rhythmic FIRMAGene scores in murine tissues.	161
Table S 7: GO terms enriched for genes with rhythmic phase-shifted SVPs in baboon tissues.	163
Table S 8: Overview of 24-h rhythmic phase-shifted SVPs detected in at least six baboon tissues.....	166

Index of Abbreviations

AME	Analysis of Motif Enrichment
AS	alternative splicing
BH	Benjamini-Hochberg
BPS	branch point sequence
CDF	chip definition file
CPM	counts per million
CRC	colorectal cancer
CRPC	castration-resistant prostate cancer
DAVID	Database for Annotation, Visualization, and Integrated Discovery
DODR	Detection of Differential Rhythmicity
EMT	epithelial-to-mesenchymal transition
ER	endoplasmic reticulum
FC	fold change
FDR	false discovery rate
FIRMA	Finding Isoforms using Robust Multichip Analysis
FPKM	fragments per kilobase million
GEO	Gene Expression Omnibus
GO	Gene Ontology
hnRNP	heterogeneous nuclear ribonucleoprotein
HTA	Human Transcriptome Array
IQR	interquartile range
JT	Jonckheere-Terpstra
JTK	Jonckheere-Terpstra-Kendall
MEM	maximal exact match
MESOR	midline estimating statistic of rhythm
MMP	maximal mappable prefix
mRNA	messenger RNA
NGS	next-generation sequencing
NMD	nonsense-mediated decay
PAR	proline and acidic amino acid-rich
pre-mRNA	precursor messenger RNA
PSEA	Phase Set Enrichment Analysis
RAIN	Rhythmicity Analysis Incorporating Nonparametric methods
RBM	RNA binding motif
RBP	RNA-binding protein
RHT	retinohypothalamic tract
RMA	Robust Multichip Average
RNA-seq	RNA-sequencing

RNP	ribonucleoprotein
RORE	retinoic acid-related orphan receptor-binding element
RPKM	reads per kilobase million
RSS	residual sum of squares
RXR	retinoid X receptor
SCN	suprachiasmatic nucleus
SF	splicing factor
SMEM	super maximal exact match
snRNA	small nuclear RNA
snRNP	small ribonucleoprotein particle
STAR	Spliced Transcripts Alignment to a Reference
SVP	splice variant pair
TMM	trimmed mean of M-values
TPM	transcripts per million
TSS	transcription start site
TTFL	transcriptional translational feedback loop

1 Introduction

” [...] there is apparently no organ and no function in the body which does not exhibit a similar daily rhythmicity.

– Jürgen Aschoff

German Physician and Chronobiologist

1.1 The Circadian Clock

Life on Earth is subject to diurnal changes of light and darkness caused by the planet's counter-clockwise rotation around its own axis. Organisms have adapted to these recurring and predictable changes in the environment by evolving a highly accurate internal time-keeping system known as the circadian clock, derived from the Latin words *circa* and *diem*, meaning “about a day”. Circadian clocks are responsible for the generation of ~24-h rhythms in physiology and behavior that can be found in virtually all light-sensitive organisms across the three domains of life, including cyanobacteria (Cohen and Golden, 2015), fungi (Liu and Bell-Pedersen, 2006), plants (Greenham and McClung, 2015), insects (Tomioka and Matsumoto, 2010), fish (Reebs, 2002), and mammals (Mohawk et al., 2012). Though the genetic components of the individual clocks differ, ranging from three genes forming a cell-autonomous oscillator in cyanobacteria to multiple interlocked oscillators in mammals (Bell-Pedersen et al., 2005), they all share the same defining properties: They generate innate, endogenous, and self-sustained rhythms with a period of ~24 h that persist even in the absence of cyclic environmental signals (free-running rhythms) (Pittendrigh, 1960; Roenneberg and Merrow, 2005). The circadian period is temperature-compensated, meaning that it is kept constant for temperatures within a physiological range, and undergoes only slight changes under free-running conditions (Aschoff, 1981; Pittendrigh, 1954). The phase, i.e., peak time of the rhythm, can be entrained to the environment by external timing cues called “Zeitgebers” (German for “time givers”) of which light is the most dominant one but which also include non-photoc stimuli such as temperature, activity, and feeding time (Refinetti, 2010). The entrainment of the internal clock to the external time facilitates the anticipation of and adaption to daily cycles in the environment and thus likely confers an evolutionary advantage to the organism (Pittendrigh, 1993).

1.1.1 A Brief History of Chronobiology

The earliest records of circadian processes can be traced back to the fourth century BC when the Greek admiral Androstenes, who served under Alexander the Great, noted the daily opening and nightly closing of the leaves of the tamarind tree *Tamarindus indicus* (Lee Phillips, 2005). Over 2000 years later, in 1729, the French astronomer Jean-Jacques d'Ortous de Mairan conducted the first recorded circadian experiment by placing a *Mimosa pudica* in constant darkness (de Mairan, 1729). De Mairan observed that the plant continued the daily folding and unfolding of its leaflets despite the lack of light and concluded that the plant "sensed the sun without seeing it". The endogenous character of the circadian clock in plants was first described in 1832 by the Swiss botanist Augustin Pyramus de Candolle who reported a free-running rhythm of 22 h for a mimosa placed under constant light (de Candolle, 1832).

In the 20th century, circadian research expanded from plants to encompass the exploration of daily rhythmicity in animals and other organisms. Important hallmarks include the discovery of the heritability of the circadian clock in plants by Erwin Bünning (Bünning, 1932), the temperature-compensation of its period in *Drosophila melanogaster* by Colin Pittendrigh (Pittendrigh, 1954), and the human circadian clock in temporal isolation by Jürgen Aschoff (Aschoff and Wever, 1962). The term "circadian" was first coined by Franz Halberg in the late 1950s to describe endogenous biological rhythms with a diurnal periodicity (Halberg et al., 1959). A further milestone in the research of mammalian circadian clocks was the identification of the suprachiasmatic nucleus (SCN) of the hypothalamus as the central circadian pacemaker in mammals by ablation experiments in rats (Moore and Eichler, 1972; Stephan and Zucker, 1972). While the existence of circadian clocks in various eukaryotic species from single-celled organisms to plants to animals was widely accepted by the 1960s (Roenneberg and Merrow, 2005), prokaryotes were long considered incapable of expressing circadian rhythms due to being "too simple" – a dogma that was finally overthrown by the discovery of a circadian rhythm of nitrogen fixation in photosynthesizing cyanobacteria (Mitsui et al., 1986). Research on the molecular mechanisms of the clock was fueled by the discovery of the first clock mutants in a *Drosophila* mutant screen (Konopka and Benzer, 1971) with the responsible gene aptly named *period* (*per*) and later successfully cloned by Michael Young and colleagues (Bargiello et al., 1984). The discovery of the self-suppressing role of the Period protein (PER) led Paul Hardin, Jeffrey Hall, and Michael Rosbash to propose a transcriptional translational feedback loop (TTFL) as the molecular mechanism responsible for the generation of circadian rhythmicity (Hardin et al., 1990). These and other findings raised the awareness of the importance of the circadian clock for human physiology which culminated in Hall, Rosbash and Young receiving the Nobel Prize in Physiology or Medicine 2017 "for their discoveries of molecular mechanisms controlling the circadian rhythm".

1.1.2 Architecture of the Mammalian Circadian System

The mammalian circadian clock can be conceptualized as a hierarchical system consisting of a main oscillator or “pacemaker” in a brain region known as the SCN and peripheral oscillators residing in organs and tissues throughout the body (Welsh et al., 2010). Located in the anteroventral hypothalamus of the brain, the SCN is composed of two clusters of tightly interconnected neurons (~10,000 each in mouse) that are situated immediately dorsal to the optic chiasm and receive direct photic input from the intrinsically photosensitive retinal ganglion cells of the inner retina via the retinohypothalamic tract (RHT) (Patton and Hastings, 2018). In short, a glutamate release at synaptic contacts between the RHT and SCN neurons leads to a depolarization of the neuronal membrane (Meijer and Schwartz, 2003). The resulting calcium influx likely activates cAMP response element-binding protein-mediated transcription of the core clock genes *Per1* and *Per2* that contain cAMP response elements in their promoters (Meijer and Schwartz, 2003). Since *Per* genes are only light-inducible in the SCN during the subjective night of an animal, this enables a reset of the circadian phase in accordance with external light conditions and thus allows for the entrainment of the SCN to solar time (Dibner et al., 2010).

Though cell-autonomously generated and self-sustained, circadian rhythms in the SCN additionally rely on electrophysiological coupling mechanisms between neurons that allow them to maintain their intercellular synchrony even in constant darkness (Patton and Hastings, 2018; Welsh et al., 2010). The SCN fulfills its pacemaker function by generating output signals that coordinate the synchronization of local clocks in other brain regions and peripheral tissues through a variety of direct and indirect pathways. Direct pathways include the circadian variation of neuronal and humoral signals, while indirect pathways include circadian changes in body temperature, as well as rest-activity and feeding-fasting cycles (Dibner et al., 2010). In addition, cell-autonomous oscillators reside in most peripheral tissues of the organism where they regulate local circadian physiology relevant to the respective cell type (Mohawk et al., 2012). The clocks of both peripheral tissues and SCN neurons share the same molecular makeup: They consist of a regulatory network of clock genes and proteins which interact via positive and negative TTFLs (Dibner et al., 2010) (**Figure 1-1**). At the core of the mammalian clock network, the activators CLOCK (and its paralogue NPAS2) and BMAL1 (also known as ARNTL) form a heterodimeric basic helix-loop-helix PER-ARNT-SIM transcription factor complex that rhythmically activates the expression of target genes from the *Per* (*Per1*, *Per2*, *Per3*) and *Cryptochrome* (*Cry1*, *Cry2*) families by binding to consensus E-box DNA motifs in their promoters (Takahashi, 2017). The translated PER and CRY proteins form a complex that translocates from the cytoplasm to the nucleus where it inhibits the CLOCK:BMAL1-mediated transcription of *Per* and *Cry* genes, thus forming a negative feedback loop (Shearman et al., 2000). Once the repressor proteins are degraded through ubiquitin-dependent pathways, transcription of the genes can resume again and the cycle starts anew (Partch et al., 2014). The rate at which the PER:CRY complex enters the nucleus and gets degraded is further controlled by the casein kinases CKI δ and CKI ϵ and by E3 ligase complexes which target PER and CRY proteins for ubiquitylation (Gallego and Virshup, 2007).

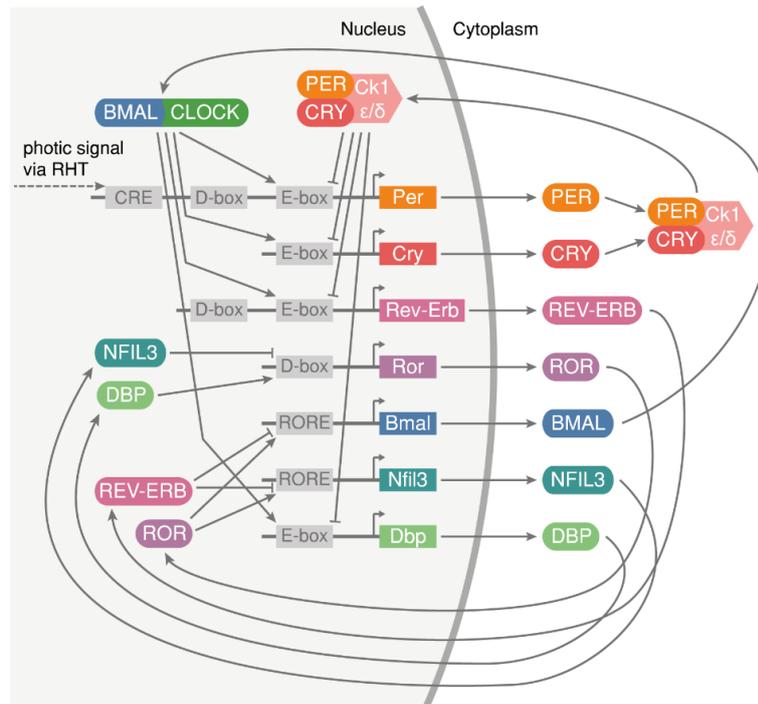


Figure 1-1: Mammalian core circadian feedback loops. Adapted from Takahashi (2017).

In a second negative feedback loop, the CLOCK:BMAL1 complex activates the transcription of genes from the *Rev-Erb* family (*Nr1d1*, *Nr1d2*) also via E-box binding (Triqueneaux et al., 2004). The translated orphan nuclear receptors REV-ERB α and REV-ERB β repress the transcription of *Bmal1* by binding to RevDR2 and retinoic acid-related orphan receptor-binding elements (ROREs) in its promoter region (Preitner et al., 2002). The resulting rhythmic expression of *Bmal1* mRNA is further modulated by orphan nuclear receptors from the ROR family (ROR α , ROR β , ROR γ) which compete with the REV-ERB proteins for ROREs binding but, in contrast to them, activate *Bmal1* transcription (Sato et al., 2004). A third loop involves the nuclear factor NFIL3 (also known as E4BP4) and three members of the proline and acidic amino acid-rich (PAR) basic leucine zipper transcription factor family: DBP, TEF, and HLF (Takahashi, 2017). Together, these transcription factors regulate circadian gene expression by interacting on D-box containing promoters and enhancers of target genes, with NFIL3 repressing and the clock-controlled PAR proteins activating their transcription (Mitsui et al., 2001). Analogous to *Bmal1*, *Nfil3* expression itself is repressed by REV-ERB α (Duez et al., 2008), whereas *Dbp* transcription is activated via CLOCK:BMAL1 E-box binding (Yamaguchi et al., 2000). Together, these interlocking feedback loops generate robust endogenous rhythms in the expression of so-called clock-controlled genes with a periodicity of approximately 24 h and a variety of phases. In addition to the direct regulation via core-clock protein binding to circadian *cis*-regulatory elements of clock-controlled genes (E/E'-boxes, D-boxes, ROREs), the circadian transcriptome is controlled by a multitude of downstream signaling circuits. Recent genome-wide multi-organ studies of the circadian transcriptome in nocturnal and diurnal mammals have shown that in any given tissue, ~5–20% of expressed genes undergo circadian oscillations in mRNA levels and that ~40–65% of protein-coding genes and 97% of ubiquitously expressed genes oscillate in at least one organ (Mure et al., 2018; Zhang et al., 2014).

Circadian transcriptional programs are highly diverse as well as tissue-specific, thus enabling the coordination of local physiology and organ function. Well-studied clock-controlled cellular processes include metabolism (Reinke and Asher, 2019), cell cycle (Gaucher et al., 2018), stem cell division and differentiation (Brown, 2014), as well as multiple post-transcriptional mechanisms (Torres et al., 2018). The clock further controls diverse physiological and behavioral processes such as variations in body temperature (Aschoff, 1983), blood pressure (Giles, 2006), hormone production (Hastings et al., 2007), immune activity (Scheiermann et al., 2013), sleep-wake and rest-activity cycles (Beersma and Gordijn, 2007), and memory consolidation (Eckel-Mahan and Storm, 2009). Indeed, Jürgen Aschoff proved to be quite far-sighted when he noted that “there is apparently no organ and no function in the body which does not exhibit a similar daily rhythmicity” (Aschoff, 1965).

1.1.3 The Role of the Circadian Clock in Cancer

Given the scope of biological processes under circadian control, it is not surprising that the clock has a profound impact on human health and that dysregulations of the circadian system are associated with several pathologies (Roenneberg and Mellow, 2016). These include sleep disorders (Toh et al., 2001), mood disorders (Grandin et al., 2006), metabolic syndrome and obesity (Bishehsari et al., 2016), cardiovascular disorders (Crnko et al., 2019) as well as susceptibility to cancer (Fu and Kettner, 2013; Fu and Lee, 2003). Dysregulations of the circadian clock can be caused both by environmental perturbations of the organism’s internal circadian rhythm and by genetic mutations within clock genes (Preußner and Heyd, 2016).

The first reports of a connection between disrupted circadian rhythms and tumor development date back to the 1960s and describe the formation of spontaneous mammary tumors in rodents induced via constant light exposure and ablation of the pineal gland (Hamilton, 1969; Jöchle, 1964). In a series of epidemiological studies, a higher incidence of breast, colorectal, and prostate cancer in addition to metabolic and gastrointestinal diseases were reported for long-term shift workers such as radio operators and night-time nurses (Hansen, 2001; Knutsson, 2003; Schernhammer et al., 2001; Schernhammer et al., 2003; Tynes et al., 1996). Apparently, the continuous misalignment of the internal circadian clock with the environment increases the risk of carcinogenesis which led the International Agency for Research on Cancer of the World Health Organization to classify shift work involving circadian disruption as “probably carcinogenic to humans” (exposure circumstances, class 2A carcinogen) (IARC, 2007). Disrupted circadian rhythms have also been shown to constitute a prognostic biomarker for survival and tumor response of cancer patients (Mormont et al., 2000; Sephton et al., 2000). Studies in animal models have shown that the disruption of circadian rhythms through ablation of the SCN or exposure to chronic jet-lag enhances the growth of tumors in mice (Filipski et al., 2002; Filipski et al., 2009). The above described findings suggest that a functioning clock may act as a tumor suppressor and that a disrupted clock might promote tumorigenesis (Fu and Kettner, 2013). However, the underlying mech-

anisms by which the dysregulation of circadian rhythmicity may affect tumor growth are still the subject of ongoing research and it is not yet clear whether a malfunctioning clock is a cause or a consequence of carcinogenesis and/or tumor progression.

At the molecular level, the mammalian circadian clock is coupled with a second biological oscillator that plays a fundamental role in controlling cell growth and proliferation: the cell division cycle (Gaucher et al., 2018). In analogy to findings in unicellular organisms where the circadian clock controls the timing of cell division (Yang et al., 2010), it has been suggested that the circadian clock may also act as a gating mechanism for cell cycle progression in mammalian cells, allowing cell division only at specific times of the day (Nagoshi et al., 2004). However, more recent studies have rejected the gating hypothesis in favor of either a unidirectional (Bieler et al., 2014) or a bidirectional coupling (Feillet et al., 2014) between the circadian clock and the cell cycle. A number of molecular connections between core circadian and cell cycle components have been identified in recent years (Feillet et al., 2015; Fu and Kettner, 2013; Shostak, 2017) (**Figure 1-2**). One of these connections is the WEE1 kinase which inhibits the cell cycle regulator CDK1 through phosphorylation and thus acts as a G2/M checkpoint by preventing the cell's entry into mitosis. In mice, *Wee1* transcriptional expression has been found to be directly activated by the Clock: Bmal1 heterodimer and to be suppressed by Per/Cry proteins (Matsuo et al., 2003). Though not reported as rhythmic itself, the multifunctional nuclear protein NONO is part of the mammalian circadian circuitry through binding to PER proteins and antagonizing their repressive activity (Brown et al., 2005). In mice, *Nono* was found to bind to the G1/S checkpoint and tumor suppressor gene *p16-Ink4a* and to activate its circadian transcription in a Per-dependent manner (Kowalska et al., 2013). A recent study further provided evidence for *p16-Ink4a* acting as a mediator for changes in the circadian period upon Ras-mediated oncogenic dysregulation of the clock (El-Athman et al., 2017). Additional links from the clock to cell cycle regulators include CRY2-dependent protein turnover of the oncogene and cell cycle regulator c-MYC (Huber et al., 2016), the rhythmic regulation of *p21* transcription via binding of elements from the ROR and REV-ERB protein families to ROREs motifs in its promoter (Gréchez-Cassiau et al., 2008), and Per2 preventing the Mdm2-mediated ubiquitination of the tumor suppressor p53 (Gotoh et al., 2016). Notably, p53 influences the circadian clock and the cell cycle in a bidirectional fashion which can be considered as further evidence for a coupling between the two oscillators: p53 represses *Per2* transcription by blocking the binding of Clock: Bmal1 (Miki et al., 2013), whereas *p53* transcription is in turn controlled by Bmal1 (Jiang et al., 2016). Overall, the precise molecular interplay between the clock and the cell cycle in healthy and tumor tissues is hard to uncover due to the plasticity of both cell cycle control and the circadian circuitry (Partch et al., 2014; Shostak, 2017).

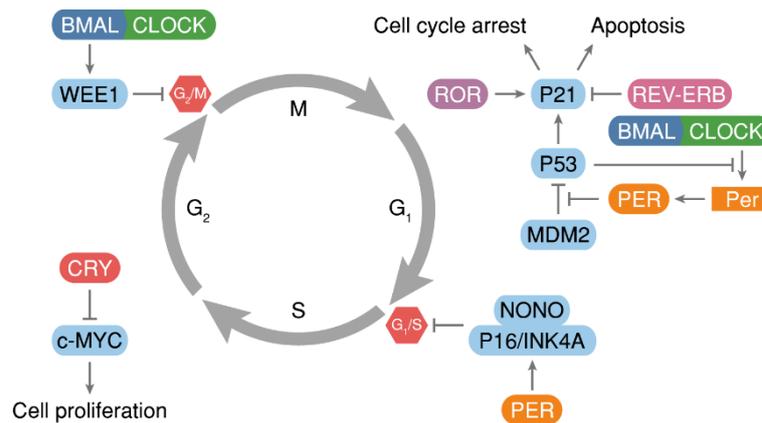


Figure 1-2: Molecular connections between the circadian clock and the cell cycle in mammals.

Mutations and alterations in the expression of core clock and clock-regulated genes are common throughout many types of cancers and have been found to be associated with patient survival, tumor stage, and clinical subtype (Benna et al., 2017; Ye et al., 2018). In a recent study, Ye et al. (2018) examined the expression levels of core clock genes across multiple human cancers and found them to be altered dependent on their role in the circadian TTFL circuitry, with transcriptional repressors tending to be downregulated, and transcriptional activators tending to be upregulated in cancer. For instance, dysregulations in the expression levels of members of the *PER* circadian repressor family have been linked to cancer in various clinical (Chen et al., 2005; Zhao et al., 2014a) and *in vitro* studies (Gery et al., 2006; Hua et al., 2006), predominantly in a tumor suppressor context. Overexpression of *Per2* has been found to induce apoptotic cell death in murine and human cancer cells via downregulation of the anti-apoptotic oncogenes *c-Myc*, *Bcl-XL*, and *Bcl-2* and the upregulation of the pro-apoptotic tumor suppressor gene *p53* and its target gene *Bax* (Hua et al., 2006; Oda et al., 2009). Despite the widely accepted tumor-suppressive role of the circadian clock (Davis et al., 2019), some studies also report tumor-promoting properties of clock genes that are partly in conflict with other findings. For example, on the one hand, the disruption of the clock via a simultaneous deletion of both *Cry* genes has been found to reduce the risk of carcinogenesis in a mice strain that is cancer-prone due to a *p53* mutation (Ozturk et al., 2009). On the other hand, the same deletion has been reported to enhance the formation of carcinomas in mice exposed to irradiation and carcinogenic compounds (Lee et al., 2010; Mteyrek et al., 2017), indicating that the role of the circadian clock in cancer is context-dependent (Shostak, 2017).

In light of the above described links between disrupted internal timing and cancer and the apparent tumor-suppressive role of the circadian clock, it seems only logical to apply this knowledge to the benefit of patients. Daily rhythms in human physiology modulate pharmacokinetics and pharmacodynamics and thus influence the effect of xenobiotics on the organism according to the time-of-day (Dallmann et al., 2016). Humans display large interindividual differences regarding their internal time and their sleep and wake preferences – ranging from extreme “larks” to extreme “owls” (Roenneberg et al., 2003). The concept of chronotherapy or chronopharmacology aims at adjusting the dosing-time to the internal time of patients in order to enhance treatment efficacy and minimize toxic side effects (Cederroth et al., 2019; Ozturk et al., 2017). Accordingly, new diagnostic tools for the determination of

internal circadian time are paving the way for personalized clinical approaches that take the chronotype of patients into account (Braun et al., 2018; Laing et al., 2017; Wittenbrink et al., 2018). Promising findings regarding chronotherapy have been made in studies with colorectal cancer (CRC) patients. CRC is the third most commonly diagnosed cancer in men and the second in women and remains a major cause of mortality, especially in Western societies (Bray et al., 2018). Most patients die from metastases to the liver and the lungs and the five-year survival rate for CRC found at the distant stage is only ~14% (Noone et al., 2018). Many processes of the gastrointestinal tract and the digestive system display circadian rhythms (Scheving, 2000) and disruption of the circadian clock has been implicated in colorectal carcinogenesis and CRC progression (Mazzoccoli et al., 2014). Alterations in clock gene expression have been found in tumor tissues of patients with CRC compared to healthy surrounding tissues (Wang et al., 2012; Wang et al., 2011) and are further associated with the formation of metastases and poorer survival rates (Mazzoccoli et al., 2011; Oshima et al., 2011). Chronomodulated delivery of chemotherapy has shown favorable results for male patients with metastatic CRC in a Phase III study, resulting in an increased survival time for men but not for women, indicating a sex dependency for optimal treatment schedules (Giacchetti et al., 2006).

In summary, dysregulated circadian clocks are associated with mammalian carcinogenesis and cancer progression. Through its coupling to the cell cycle, the circadian clock is responsible for the temporal regulation of key processes of the cell whose disruption can result in carcinogenesis, including cell proliferation and apoptosis, and other hallmarks of cancer (Sulli et al., 2019). Mutations and alterations in the expression of core clock and clock-regulated genes are commonly observed in cancer patients and cancer cell lines and studies in tumor models suggest a tumor suppressor role of the circadian system. However, the exact mechanisms of how a disrupted clock might favor malignant transformation are still a topic of ongoing research and might furthermore be context dependent. Translational clinical research tries to leverage the knowledge of links between the clock and cancer in chronotherapeutical treatment approaches that take the internal time of patients into account.

1.1.4 Ultradian Rhythms and Circadian Harmonics

In addition to circadian rhythms, mammals also exhibit so-called sub-circadian or ultradian rhythms with a period < 24 h in their physiology and behavior (Prendergast and Zucker, 2016). On the molecular level, ultradian rhythms in gene expression at the second harmonic (12 h period) and third harmonic (8 h period) of circadian rhythmicity have first been discovered *in vivo* in murine liver (Hughes et al., 2009). The authors identified hundreds of 12-h and dozens of 8-h harmonic genes in liver samples and validated 12-h rhythmic genes in several other tissues but found none in *ex vivo* cultured murine and human cells. Several of the 12-h rhythmic genes lost their oscillations in a restricted feeding experiment, leading the authors to hypothesize that 12-h rhythms might be regulated via feeding behavior and food metabolism. Based on these findings, they suggested a partly extrinsic origin of rhythms at the second harmonic which is based on interactions between autonomous circadian clocks and systemic cues that cannot be sufficiently replicated *in vitro* (Hughes et al., 2009). The existence of

12-h rhythms in mammalian transcription has been further validated in subsequent studies that were predominantly conducted in murine liver (Hughes et al., 2012b; Vollmers et al., 2009; Zhu et al., 2017). Interestingly, it was found that loss-of-function of the core clock gene *Clock* diminishes 12-h hepatic rhythmicity and that brain-specific *Clock* rescue converts 12-h rhythms into 24-h rhythms without loss of amplitude, suggesting that the central circadian clock is at least partly involved in their generation (Hughes et al., 2012b).

Ultradian rhythms have also been observed at the proteomic and metabolomic level. The *Ire1α* gene shows 12-h oscillations in murine liver, leading to 12-h rhythms in abundance of the endoplasmic reticulum (ER) membrane protein IRE1α and a rhythmic activation of the IRE1α-XBP1 pathway in mouse liver (Cretenet et al., 2010). The splicing efficiency of the transcription factor gene *Xbp1* and the expression of the spliced protein isoform XBP1s show robust 12-h rhythms both *in vivo* and *in vitro* (Cretenet et al., 2010; Zhu et al., 2017). Furthermore, it was found that siRNA-mediated knockdown of *Xbp1* in mouse embryonic fibroblasts severely impairs 12-h rhythms of several genes, implying a role for XBP1s as a transcriptional regulator of a purportedly independent mammalian 12-h clock (Zhu et al., 2017). Recently, third harmonics of circadian rhythms have been observed in metabolite accumulation in U2OS cells, including amino acids, methylation products, and vitamin B₁ (Krishnaiah et al., 2017).

The detection of ultradian rhythms requires sampling at a high resolution, e.g., every 1 h or 2 h (Hughes et al., 2009), and is further compounded by the fact that circadian harmonics are commonly masked by co-expressed circadian rhythms (van der Veen and Gerkema, 2016; Zhu et al., 2018). *Post hoc* analyses of previously published datasets using different methods for the detection of ultradian rhythmicity than Hughes et al. (2009) suggest that the prevalence of ultradian transcriptomic rhythms might have been underestimated and offer alternative interpretations of the data. In contrast to the previous findings by Hughes et al. (2009), a more recent study using a non-spectral analysis method for the separation of ultradian from circadian components reported more than 900 genes with ultradian expression in NIH 3T3 cells *in vitro*, supporting the hypothesis of intrinsically driven ultradian rhythms (van der Veen and Gerkema, 2016). Another *post hoc* analysis of the murine liver transcriptome data revealed ~4% of all hepatic genes to have dominant 12-h transcriptional rhythms and ~8% to have dominant 8-h rhythms (Zhu et al., 2017). A period was defined as dominant if the amplitude of the rhythm was the greatest among all identified oscillations (Zhu et al., 2017).

Despite these findings, only little is known about the molecular mechanisms that give rise to ultradian oscillations and that might consolidate the conflicting reports of clock-dependent and clock-independent ultradian rhythmicity. A theoretical analysis by Westermarck and Herzel uncovered a possible mechanism for the generation of 12-h rhythms that is based on an interplay between components of the circadian clock (Westermarck and Herzel, 2013). The authors propose a model of a circadian AND funnel where pairs of circadian transcription factors bind to the promoters of target genes in a non-competitive way. They showed that 12-h rhythms in the expression of a target gene can be produced, provided that both transcription factors are either activators or repressors of transcription and out of

phase by ~12 h or that one transcription factor is an activator and the other a repressor and both have approximately the same phase. Interestingly, they also showed that 8-h rhythms can theoretically be generated by a circadian AND funnel if one transcription factor is circadian and the other has a 12-h period (Westermarck and Herzel, 2013).

The function of ultradian rhythms is still largely unknown. Pathway analysis carried out based on the ultradian gene sets in murine liver identified by Hughes et al. (2009) revealed 12-h rhythmic genes to be involved in diverse processes including ER homeostasis, the regulation of cell division, and protein processing. 8-h rhythmic genes were found to be associated with NF- κ B signaling and lipid metabolism (Hughes et al., 2009). Later studies have suggested a metabolic relevance of ultradian rhythms, as well as a possible ultradian gating of the cell cycle (van der Veen and Gerkema, 2016; Ventre et al., 2015). It has further been hypothesized that 12-h rhythms in gene expression and protein and metabolite abundance might be in tune with a daily biphasic “metabolic stress cycle” that results from energy overdraft and excess at transition periods between feeding-fasting and sleep-wake cycles (Zhu et al., 2018; Zhu et al., 2017). According to another hypothesis, mammalian 12-h rhythms may have evolved from an ancestral circatidal clock which regulates the behavior of coastal and estuarine animals in sync with the ~12.4-h periodic ebb and flow of the tides (Wilcockson and Zhang, 2008; Zhu et al., 2018). There is still an ongoing debate about whether circatidal rhythms are generated via two antiphasic 24.8-h periodic circadian clocks (Palmer, 1995) or whether they are the product of a dedicated circatidal clock with a period of 12.4 h that is independent of the circadian clock (Zhang et al., 2013).

Overall, there is accumulating evidence that ultradian rhythms at the second and third harmonic of the circadian period are prevalent in mammalian tissues, though their physiological function and their molecular underpinnings still remain to be elucidated.

1.1.5 Analysis of Circadian Rhythmicity on the Genome-Scale

Detecting rhythms in a time-series, assessing their significance and estimating rhythmic parameters is a fundamental aspect of circadian studies. A typical experimental layout of a genome-scale circadian experiment consists of ~20,000 observations (e.g., genes) and 6–48 sparsely sampled, equidistant time points, usually taken every 2–4 h for 1–2 circadian cycles. For the detection of circadian features in genome-scale data, statistical algorithms are used that can generally be divided in parametric and non-parametric methods. A simple and widely used example of a parametric algorithm for rhythm detection is the cosinor or harmonic regression method (**Figure 1-3**). The single cosinor is a harmonic regression model for a single component that fits the first harmonics in a Fourier expansion to estimate amplitudes and phases of a time-series of oscillating data points by employing a least squares method (Cornelissen, 2014; Lück et al., 2014). It can be applied for the detection of rhythms in a time-series under the condition that the period is known beforehand, as can be assumed for most circadian studies (Cornelissen, 2014).

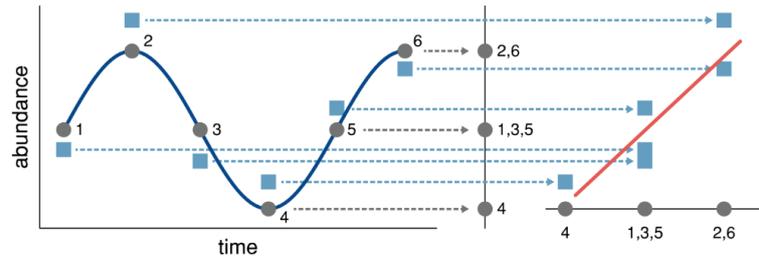


Figure 1-3: Schematic representation of the harmonic regression model. Time points (gray circles) are grouped according to their position in a sinusoidal curve with a pre-defined period (dark blue curve). Measurements (light blue squares) are binned according to the time point groups. Amplitude and significance of a rhythm can be inferred from the slope of the linear least squares approximation (red line). Adapted with permission from a presentation by Pål Westermark.

A sum of sinusoidal functions

$$y = m + a \cos\left(\frac{2\pi t}{\tau}\right) + b \sin\left(\frac{2\pi t}{\tau}\right) + \varepsilon \quad (1)$$

is fitted to the time-series where y represents the abundance of a molecule at time t , m is the rhythm-adjusted mean, also known as the midline estimating statistic of rhythm (MESOR), τ is the period (in radians), and ε is the error term (Cornelissen, 2014). Using the trigonometric angle sum identity $\cos(\alpha \pm \beta) = \cos(\alpha)\cos(\beta) \mp \sin(\alpha)\sin(\beta)$ and the coefficients $a = A \cos(\phi)$ and $b = A \sin(\phi)$, (1) can be rewritten as a single sinusoid function in the form

$$y = m + A \cos\left(\frac{2\pi t}{\tau} - \phi\right) + \varepsilon \quad (2)$$

With a relative amplitude $A = \pm\sqrt{a^2 + b^2}$ and phase (or acrophase) $\phi = \arctan\left(\frac{b}{a}\right)$ (Cornelissen, 2014; Lück et al., 2014).

The objective of the linear least squares method is to minimize the residual sum of squares (RSS), defined as the sum of the differences between the measurements y_i obtained at time t_i ($i=1, 2, \dots, N$) and the values \hat{y}_i estimated from the fitted model (Cornelissen, 2014)

$$RSS = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3)$$

Following an analytical approach, the least squares solution can be found by solving the normal equations for linear regression. In order to do this, the linear system can be represented by matrix notation and the normal equations are formulated using matrix calculus to compute a scalar-by-vector derivative.

A general regression model with n observations and k explanators can be written as

$$y = X\beta + e, \quad (4)$$

where y and e are $n \times 1$ vectors, β is a $k \times 1$ vector and X is an $n \times k$ matrix.

In the specialized case of a harmonic regression model for mean-centered data with $m = 0$, one has

$$y_i \approx a \cos\left(\frac{2\pi t_i}{\tau}\right) + b \sin\left(\frac{2\pi t_i}{\tau}\right) = A \cos\left(\frac{2\pi t_i}{\tau} - \phi\right). \quad (4)$$

Then let matrix X be

$$X = \begin{pmatrix} \cos\left(\frac{2\pi t_1}{\tau}\right) & \sin\left(\frac{2\pi t_1}{\tau}\right) \\ \cos\left(\frac{2\pi t_2}{\tau}\right) & \sin\left(\frac{2\pi t_2}{\tau}\right) \\ \dots & \dots \end{pmatrix}. \tag{5}$$

Estimates for a and b that minimize the RSS can be obtained by solving the normal equation system

$$X^T X \hat{\beta} = X^T y \tag{6}$$

$$\hat{\beta} = (a, b) = (X^T X)^{-1} X^T y, \tag{7}$$

with $X\hat{\beta}$ representing the orthogonal projection of y to the column space of X , and $\hat{\beta}$ representing the vector of the least squares prediction (Davison, 2003) (**Figure 1-4**).

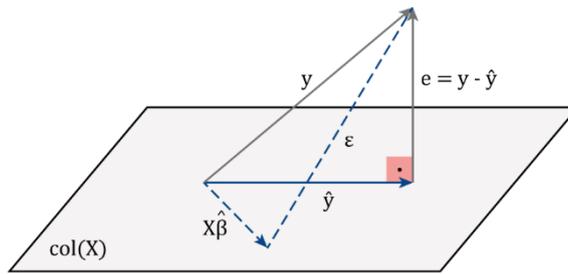


Figure 1-4: Geometric interpretation of the least squares estimation. The linear combination $X\hat{\beta}$ represents the orthogonal projection of the observation y to the column space of X . The fitted value $\hat{y} = X(X^T X)^{-1} X^T y$ is the point that is closest to y in the column space of X . The vector of residuals $e = y - \hat{y}$ is orthogonal to \hat{y} . The estimation by least squares amounts to minimizing the squared distance $(y - X\hat{\beta})^T (y - X\hat{\beta})$. Adapted from Davison (2003).

Statistical significance of the model is determined by an F-test for the null hypothesis that there is no rhythm and that a and b are randomly distributed according to a Gaussian probability distribution (Cornelissen, 2014; Halberg et al., 1967). The null hypothesis is rejected if the p -value for $A > 0$ (or more precisely either a or $b > 0$) is below the chosen probability level (Lück et al., 2014).

The covariance matrix of $\hat{\beta}$ can be determined by

$$\text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} = \sigma^2 \begin{pmatrix} 2/N & 0 \\ 0 & 2/N \end{pmatrix} \tag{8}$$

where σ^2 is the variance of the noise. Using the Jacobian matrices

$$J_A = (a/A \quad b/A) \quad \text{and} \tag{9}$$

$$J_\phi = (-b/A^2 \quad a/A^2), \tag{10}$$

the variances of amplitude and phase can be obtained by error propagation, resulting in

$$\text{Var}(A) = J_A \text{Var}(\hat{\beta}) J_A^T = \sigma^2 \frac{2}{N} \quad \text{and} \tag{11}$$

$$\text{Var}(\phi) = J_\phi \text{Var}(\hat{\beta}) J_\phi^T = \sigma^2 \frac{2}{A^2 N}. \tag{12}$$

The same framework can also be used to test for differences in rhythmicity, e.g., in phase or amplitude of features measured in different conditions, as implemented in the Detection of Differential Rhythmicity (DODR) software package (Thaben and Westermark, 2016).

In case of outliers and non-Gaussian variability of the measured data, non-parametric algorithms using rank sum testing can be applied instead. A prominent example of a non-parametric rank-based method for biological rhythm detection is the JTK_CYCLE method that applies the Jonckheere-Terpstra-Kendall (JTK) algorithm (Hughes et al., 2010). The Jonckheere-Terpstra (JT) test for ordered alternatives or Jonckheere trend test is a non-parametric statistical test that can be used to determine whether there is a significant monotonic trend between an ordinal or a continuous dependent variable (e.g., mRNA expression levels) and an ordinal independent variable (e.g., time) (Jonckheere, 1954; Terpstra, 1952).

Given a set of n different samples $(X_{11}, \dots, X_{1m_1}), \dots, (X_{n1}, \dots, X_{1m_n})$ of size m_1, \dots, m_n that is taken from independent populations $F_1(x), \dots, F_n(x)$, the null and alternative hypotheses of the JT test are given by (13) and (14), respectively:

$$H_0: F_1(x) = F_2(x) = \dots = F_n(x) \quad \text{and} \quad (13)$$

$$H_1: F_1(x) < F_2(x) < \dots < F_n(x), \quad (14)$$

with H_0 stating that the samples are drawn from populations with the same distribution function and H_1 assuming a strictly monotonic ordering.

The JT test statistic is expressed as

$$s = \sum_{i>j}^n U_{ij} = \sum_{i=1}^{n-1} \sum_{j=i+1}^n U_{ij}, \quad (15)$$

using the Mann-Whitney-U-statistic for comparison of two samples:

$$U_{ij} = \sum_{k=1}^{m_i} \sum_{l=1}^{m_j} q_{i_k, j_l} \quad \text{with} \quad (16)$$

$$q_{i_k, j_l} = \begin{cases} 1 & \text{if } X_{ik} < X_{jl}, \\ 0 & \text{else} \end{cases}. \quad (17)$$

The Kendall rank correlation coefficient, also referred to as Kendall's tau, is a measure of the ordinal association between two measured quantities and describes the similarity of orderings of the data when ranked (Kendall, 1938). JTK_CYCLE applies the JT test to compare the ranks of the measured values of a time-series to those of a pre-defined reference waveform, by default a cosine (Hughes et al., 2010). In a first step, all points in a time-series are compared in a pairwise manner to determine their ranks. The resulting increasing/decreasing pattern is compared to the increasing/decreasing pattern of a reference curve. Based on a user-defined range of period lengths, the algorithm then determines an optimal combination of period and phase, such that the p -value of Kendall's tau correlation between the values from the measured time-series and each tested cyclical ordering is minimized. Afterwards, all minimal p -values are Bonferroni-adjusted to correct for the multiple testing of possible periods and phases (Hughes et al., 2010).

Several other non-parametric algorithms for the genome-scale detection of rhythms have been built based on the idea of JTK_CYCLE, including empirical JTK (eJTK) that provides empirically-calculated p -values for arbitrary waveforms (Hutchison et al., 2015), its successor BooteJTK that additionally uses parametrically bootstrapped re-samplings of a time-series (Hutchison et al., 2018), and the Rhythmicity Analysis Incorporating Nonparametric methods (RAIN) algorithm (Thaben and Westermark, 2014). In contrast to JTK_CYCLE and similar to eJTK, RAIN also allows for non-sinusoidal, non-symmetric waveforms that may have differing shapes and lengths of the rising and the falling parts of the curve (Thaben and Westermark, 2014). In order to do this, RAIN applies a generalized version of the JT test, the so-called rank test for umbrella alternatives that allows for the detection of patterns with a rising and falling shape, termed a variable umbrella peak (Mack and Wolfe, 1981). Under the assumption that a largest population $F_e(x)$ and a smallest population $F_1(x)$ exist, the alternative hypothesis is described by

$$H_1: F_1(x) < F_2(x) < \dots < F_e(x) > \dots > F_n(x) > F_1(x) \quad (18)$$

and the test statistic is expressed as

$$s = \sum_{i=1}^{e-1} \sum_{j=i+1}^e U_{ij} + \sum_{i=e}^{n-1} \sum_{j=i+1}^n U_{ji} + \sum_{j=e+1}^n U_{j1}. \quad (19)$$

The data is first grouped by measurement time and subsequently binned such that repeated measurements or time point replicates (e.g., ZT0 and ZT24 for a 24-h period) are in the same bin (**Figure 1-5**) (Thaben and Westermark, 2014). The measured data is then tested against an oscillation model that consists of a rising and a falling slope, with the slopes being tested independently by a summation of Mann-Whitney-U-tests (Thaben and Westermark, 2014). Due to different phases, amplitudes, and peak shapes being considered, the resulting p -values are internally corrected for partially-dependent multiple testing using the Benjamini-Hochberg (BH) method (Thaben and Westermark, 2014).

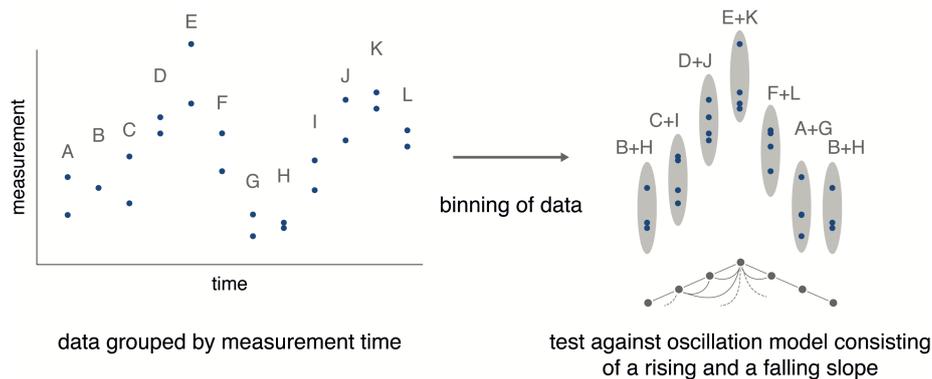


Figure 1-5: Schematic representation of the RAIN algorithm for rhythm detection. Adapted from Thaben and Westermark (2014).

For genome-scale studies, a further correction for multiple testing of the computed p -values should be considered based on the number of tested features (e.g., expressed genes) (Storey and Tibshirani, 2003). The p -value estimates the probability of finding the observed or more extreme results when the null hypothesis is true (e.g., for a truly non-rhythmic gene). The adjusted p -value or q -value obtained

by multiple testing correction estimates the expected false discovery rate (FDR) when rejecting the null hypothesis for results with an equal or smaller q -value. To filter for biologically significant rhythms, it is further recommended to exclude rhythmic features with very small amplitudes by additionally employing an amplitude cutoff as determined by harmonic regression, e.g., $A \geq 0.1$. To increase the detection power in genome-scale experiments, the amplitude cutoff can also be employed to pre-filter RAIN p -values before the correction for multiple testing, since the estimated p -value is independent from the amplitude (Bourgon et al., 2010).

In this thesis, both the parametric harmonic regression method and the non-parametric RAIN algorithm are applied for the genome-scale detection of circadian rhythmicity in transcription. Harmonic regression is further used for the estimation of rhythmic parameters, i.e., phase and amplitude of oscillations, and the DODR method is used to detect differences in rhythmicity of a gene measured in two conditions.

1.2 Pre-mRNA Splicing

During transcription in the nucleus, eukaryotic genes are synthesized as precursor messenger RNAs (pre-mRNAs) that require various processing steps such as 5' capping, splicing, and polyadenylation of the 3' end before being transported to the cytoplasm where they are subsequently translated into proteins. Splicing describes the process of excising the non-coding intervening sequences (introns) and concatenating the coding expressed sequences (exons) of a pre-mRNA transcript to produce a mature messenger RNA (mRNA) transcript. Most mammalian genes contain multiple introns, making splicing an essential step in pre-mRNA editing and gene expression.

1.2.1 Mechanism and Regulation of Splicing

The biochemical mechanism of pre-mRNA splicing has been extensively studied, e.g., in yeast and mammalian model systems, and has been found to be largely conserved between lower and higher eukaryotes (Wahl et al., 2009; Will and Lührmann, 2011). It consists of two consecutive transesterification reactions (i.e., phosphate transfer) occurring between RNA nucleotides in order to first cleave the intronic region and subsequently ligate the two flanking exons (Shi, 2017) (**Figure 1-6**). In the first reaction, the 2'-hydroxyl group of a conserved adenosine residue at the intronic branch point sequence (BPS) upstream of the 3' splice site of the pre-mRNA initiates a nucleophilic attack on the 5' terminal phosphate of the 5' splice site and forms a 2'-5' phosphodiester bond, thereby cleaving the upstream exon from the intron. The result is a released upstream exon with a 3'-hydroxyl group and a lariat intermediate consisting of the intron and the downstream exon. In the second step, the 3'-hydroxyl group of the released exon performs a nucleophilic attack on the phosphodiester bond at the last intronic nucleotide of the 3' splice site, yielding two ligated exons and an excised intron lariat. Since splice site sequences are usually short and often degenerate (Smith and Valcárcel, 2000), the accurate recognition of the exact sites for the transesterification reactions requires multiple proofreading mechanisms and

a complex network of RNA-protein, RNA-RNA, and protein-protein interactions to ensure the precision and specificity needed for a correct splicing process (Bonnal et al., 2012). In addition to splice site sequences, *cis*-acting splicing regulatory elements also include intronic and exonic splicing enhancers or silencers located in the vicinity of 3' and 5' splice sites that modulate splicing by binding *trans*-acting regulatory proteins that either stimulate or inhibit the splicing process (Will and Lührmann, 2011).

In most cases, pre-mRNA splicing is catalyzed by the major spliceosome, a large and highly dynamic ribonucleoprotein (RNP) complex comprised of the five small ribonucleoproteins (snRNPs) U1, U2, U5, U4/U6, and more than 200 additional non-snRNP proteins (Wahl et al., 2009) (**Figure 1-7**). Each snRNP consists of one or, in the case of U4/U6, two small nuclear RNAs (snRNAs), a group of highly abundant RNA molecules with a length of 100–300 nucleotides, and specific associated proteins (Wahl et al., 2009). While the snRNPs U1, U2, U4, and U5 have a binding site for the seven core Sm proteins B/B', D₁, D₂, D₃, E, F, and G, the U6 snRNP associate with a set of Sm-like LSm proteins (Will and Lührmann, 2011). Some metazoan species and plants have an additional minor spliceosome whose assembly is compositionally distinct with different but functionally analogous components (Will and Lührmann, 2011).

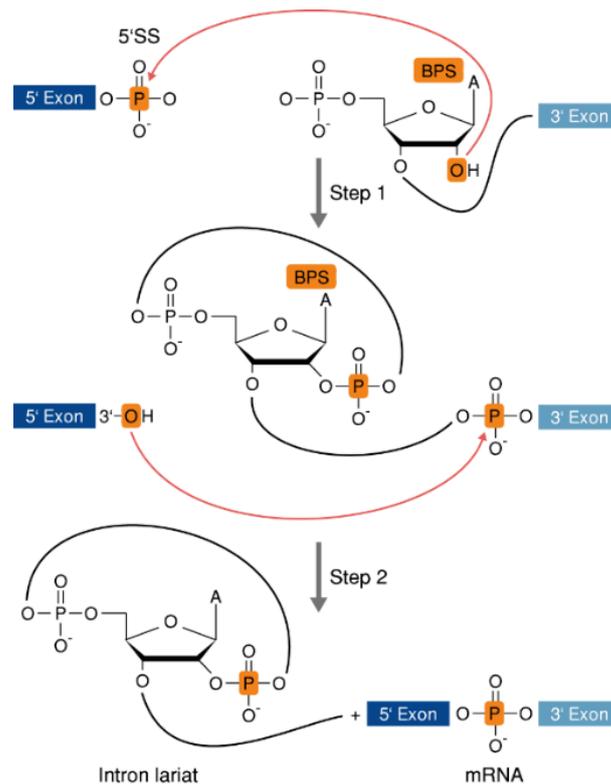


Figure 1-6: Transesterification reactions during pre-mRNA splicing. Adapted from Shi (2017).

In addition to the core spliceosomal components, a large number of *trans*-acting auxiliary splicing regulatory factors or splicing factors (SFs) is required for the assembly of the spliceosome and the splicing process. SFs include members of the SR protein family, a conserved group of proteins encoded by nine genes in humans that is characterized by an RS domain containing long repeats of serine and arginine

amino acid residues (Shepard and Hertel, 2009). SR proteins regulate exon inclusion by binding to exonic splicing enhancers, thereby facilitating the recruitment of other spliceosomal components to the regulated splice site (Long and Caceres, 2009). The splicing-promoting activities of SR proteins compete in an antagonistic way with another group of non-snRNPs involved in splicing known as heterogeneous nuclear RNPs (hnRNPs) (Zhu et al., 2001). HnRNPs are a structurally diverse and highly abundant group of primarily nuclear RNA-binding proteins (RBPs) that form complexes with transcripts produced by RNA polymerase II (Han et al., 2010). More than half of the major hnRNPs have been reported or proposed to be involved in splicing (Martinez-Contreras et al., 2008). Most hnRNPs repress splicing by either antagonizing the recognition of splice sites or by interfering with the binding of proteins bound to splicing enhancers (Martinez-Contreras et al., 2008). But there are also reports of hnRNPs playing a positive role in splicing control depending on the location of their binding sites relative to the regulated splice sites (Martinez-Contreras et al., 2008). Another essential SF is the heterodimer complex U2AF that consists of a 35-kDa and a 65-kDa subunit (Mollet et al., 2006). In conjunction with the SF SF1, U2AF65 mediates the recognition of the pre-mRNA BPS in order to define the 3' end of the intron during early spliceosome assembly (Selenko et al., 2003).

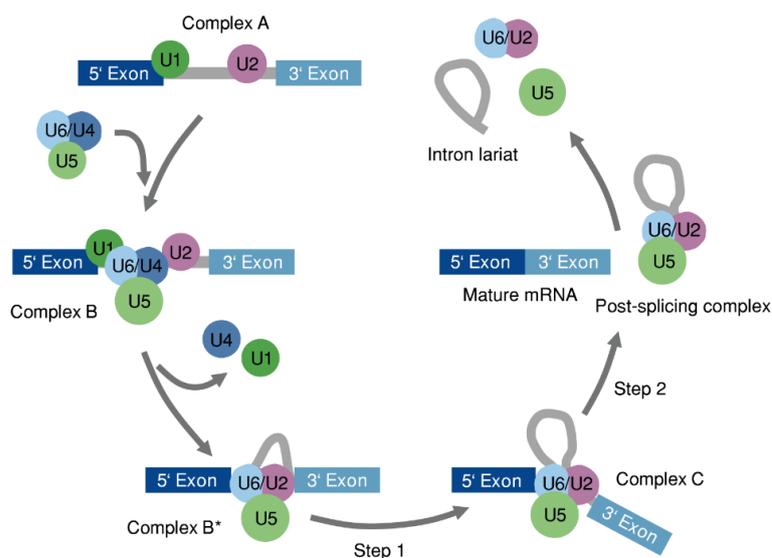


Figure 1-7: Stepwise spliceosome assembly. Adapted from Nguyen et al. (2016).

The core spliceosomal components and SFs assemble in a sequential manner on the pre-mRNA to catalyze the two transesterification steps (Wahl et al., 2009) (**Figure 1-7**). In short, the spliceosome assembly is initiated by the ATP-independent binding of the U1 snRNP to the 5' splice site and the binding of the 65-kDa and 35-kDa subunits of U2AF to the polypyrimidine tract and the conserved AG dinucleotide of the 3' splice site, respectively. The U1 snRNP binding is stabilized by members of the SR protein family while the U2AF65 binding is supported by SF1. Together, they form the spliceosomal complex E. U2AF in turn supports the binding of U2 snRNP and U2-related proteins to the BPS in an ATP-dependent manner which leads to the disassociation of SF1/mBBP, resulting in the formation of complex A. Subsequently, the preassembled U4/5/6 tri-snRNP bind to form the inactivated complex B. To become catalytically active, the spliceosome undergoes major conformational and compositional changes

including the release of U1 and U4, leading to the activated complex B*. The first transesterification reaction gives rise to the complex C which again undergoes rearrangements before the second transesterification reaction takes place and the spliceosome disassociates. During assembly and activation, the protein composition of the spliceosome undergoes drastic changes, including an extensive remodeling of spliceosomal subunits (Wahl et al., 2009). While some proteins are present throughout the whole splicing cycle, such as U2-associated proteins and members of the SR and hnRNP protein families, others seem to be only loosely associated or required in specific situations, e.g., different RBPs (Wahl et al., 2009). Mass spectrometric analyses of human spliceosomal complexes at different stages of assembly have revealed that over 170 proteins associate with the spliceosome at some point during the splicing process (Will and Lührmann, 2011). Overall, splicing is a highly complex and tightly regulated process that involves a multitude of components at specific points of the assembly and for specific target genes.

1.2.2 Alternative Splicing

In most multi-exon genes, there exist several alternative patterns of exon inclusion and intron removal for a single primary transcript which allows for the differential processing of pre-mRNAs, known as alternative splicing (AS). While constitutively spliced exons are present in every mRNA that is produced from a given pre-mRNA, alternatively spliced exons can either be excised or included fully or partly in the mature mRNA transcript (Blencowe, 2006). The production of multiple mRNAs with differing exonic compositions enables cells to create several protein isoforms with related, distinct or even antagonistic properties from a single gene locus and is thus assumed to increase not only the complexity of the transcriptome but also to engender proteome diversity (Blencowe, 2017; Gallego-Paez et al., 2017; Liu et al., 2017a). AS can also act as an on-off switch of gene expression by introducing premature stop codons into the mRNA transcripts, thus affecting the quantitative control of gene expression (Lopez, 1998). While AS events occur ubiquitously in eukaryotes, the rate of AS is higher in organisms with a large cell type diversity (e.g., birds and mammals) than in species with fewer distinct cell types, suggesting a link between AS and organism complexity (Chen et al., 2014b; Kim et al., 2007).

There are several types of AS events whose prevalence varies among species (Keren et al., 2010) (**Figure 1-8**). Exon skipping or cassette alternative exons denote alternatively spliced exonic sequences that can either be included or skipped in the mature mRNA, independent of the excision or retention of other exons. Exon skipping/cassette-type alternative exons are the most common type of AS in higher eukaryotes, accounting for nearly 40% of known AS events (Keren et al., 2010). The second and third most frequent events are the selections of alternative 3' and 5' splice sites that sometimes introduce changes in the coding sequence that differ only in a single codon (Blencowe, 2006). Together, they account for at least one quarter of AS events in mammals (Sugnet et al., 2004). Intron retention describes the inclusion of an otherwise intronic sequence in the mature mRNA transcript which contributes to less than 5% of known AS events in metazoans (Keren et al., 2010). Other less frequent and more complicated events have an impact on adjacent or distal AS events in the same transcript such as

the selection of a set of two or more mutually exclusive alternative exons (Blencowe, 2006). In addition to the above described AS events, other mechanisms of pre-mRNA processing exist, including alternative promoter usage which can lead to the inclusion of alternative first exons (Davuluri et al., 2008) and the selection of alternative polyadenylation sites which can lead to differing terminal exons (Lutz, 2008).

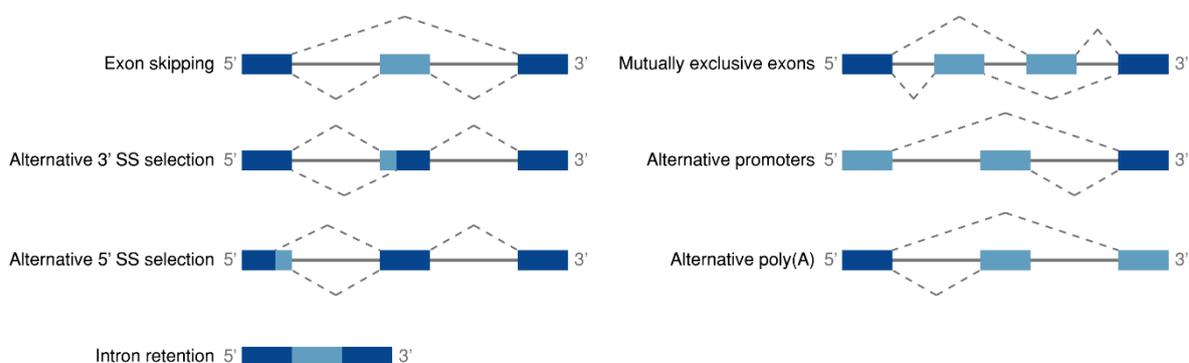


Figure 1-8: Types of AS events. Adapted from Keren et al. (2010).

According to genome-wide studies, nearly 95% of human multi-exon genes undergo AS, resulting in more than 100,000 distinct AS events, of which the majority has been reported to display tissue-dependent variations (Pan et al., 2008; Wang et al., 2008). It has been estimated that the majority of human genes generates almost eight different mRNAs (Chacko and Ranganathan, 2009) and the most recent release of the Ensembl genome database reports ~20,500 known human protein-coding genes in comparison to ~227,000 produced transcripts (Cunningham et al., 2018). However, not all reported AS events are necessarily functional and/or result in protein-coding transcripts but might also be due to errors in the splicing process (Melamud and Moul, 2009; Pickrell et al., 2010). Moreover, not all protein isoforms resulting from AS events are functionally different or important (Bhuiyan et al., 2018; Tress et al., 2017a, b). Nonetheless, AS plays an important role in controlling enzymatic properties, intracellular localization, and ligand interactions of proteins which have profound functional effects on diverse biological processes (Kelemen et al., 2013). While the effect of an individual splicing isoform is often small, so-called splicing programs seem to regulate global changes in AS to generate coordinated patterns of splicing events with a collective impact on physiology (Kelemen et al., 2013). For instance, AS enables differential gene expression between cell types during differentiation and development and thereby contributes to the tissue-specificity of multicellular organisms (Baralle and Giudice, 2017). However, the functional impact of many alternatively spliced isoforms is still unknown, necessitating further research.

1.2.3 The Role of Aberrant Splicing in Cancer

Aberrant pre-mRNA splicing is frequently implicated in human disease, notably in carcinogenesis and tumor progression (David and Manley, 2010; Ghigna et al., 2008; Kaida et al., 2012). Dysregulations in AS patterns are associated with various aspects of tumor biology, such as the control of metabolism,

proliferation, apoptosis, angiogenesis, and metastasis (Bonnal et al., 2012), and have even been proposed to constitute an additional hallmark of cancer (Ladomery, 2013). It has further been suggested that the high flexibility of alternatively spliced gene products might provide tumor cells with the ability to produce protein isoforms that are normally expressed in specific developmental stages but down-regulated in adult cells (David and Manley, 2010). In cancer, these isoforms could promote unchecked growth and survival, promoting the spread of tumor cells. Isoform switches in different tumor types have been linked to losses in functional protein domain families that are frequently mutated in cancer, possibly conferring a selective advantage to tumor cells similar to that caused by somatic mutations in cancer drivers (Climente-Gonzalez et al., 2017). However, despite large-scale alterations in AS having been observed in cancer, so far, only few AS events have been shown to be functionally relevant in cancer-related processes (David and Manley, 2010).

A well-studied example of a gene with cancer-relevant AS is the apoptosis regulator *Bcl-X* that can be alternatively spliced at two competing 5' splice sites, leading either to the production of the anti-apoptotic isoform *Bcl-XL* that is usually overexpressed, or the pro-apoptotic isoform *Bcl-XS* that is down-regulated in cancer cells (Kaida et al., 2012; Oltean and Bates, 2014). Another prominent example for aberrant AS in cancer is the SF2/ASF-regulated splicing of the vascular endothelial growth factor VEGF that has two main groups of isoforms: VEGF_{xxx} and VEGF_{xxx}b (Nowak et al., 2010). While the VEGF_{xxx} isoforms promote angiogenesis and are overexpressed in cancer, the VEGF_{xxx}b isoforms are anti-angiogenic and downregulated in cancer (Kaida et al., 2012). Aberrant splicing also affects oncogenes and tumor suppressors. Isoforms of the oncoprotein MDM2 arise through intricate patterns of AS and have been detected to be frequently overexpressed in human tumors (Jeyaraj et al., 2008). While MDM2 normally binds to and degrades p53 (**Figure 1-2**), the isoform MDM2-B lacks the p53-binding domain and interacts with full-length MDM2 instead, thereby inhibiting MDM2-mediated degradation of mutated p53 and promoting its accumulation and gain-of-function in tumorigenesis (Zheng et al., 2013).

Other examples for altered AS patterns with consequences in cancer include the genes encoding for the transmembrane protein CD44 and the fibroblast growth factor receptor FGFR2 which are both involved in epithelial-to-mesenchymal transition (EMT). EMT describes the process of polarized epithelial cells undergoing a transition to assume a mobile mesenchymal cell phenotype which plays a critical role in embryonic development (Kalluri and Weinberg, 2009). Epithelial cancer cells can hijack this process to gain enhanced migratory and invasive capacity as well as increased resistance to apoptosis (Pastushenko and Blanpain, 2018). The genomic structure of human *CD44* consists of 20 exons, ten of which are included or skipped in a variable fashion, defining the attachment properties of the molecule and its subsequent impact on cancer metastasis and EMT (Cooper and Dougherty, 1995; Warzecha and Carstens, 2012). *FGFR2* encodes for two isoforms due to mutually exclusive splicing of alternative cassette exons IIIb which is characteristic to epithelial cells and isoform IIIc which is characteristic to mesenchymal cells (Warzecha et al., 2009). A switch from the IIIb to the IIIc isoform correlates with aggressive tumor growth and invasive capability of cancers (Oltean et al., 2006; Yan et al., 1993).

Changes in AS patterns can be due to mutations in splicing regulatory sequences which usually result in an impaired recognition of the splice site (Brooks et al., 2014; Dorman et al., 2014; Kim et al., 2015b). However, many cancer-associated AS events have been found to occur in the absence of mutations in the affected genes (Grosso et al., 2008b). Instead, changes in the splicing machinery and splicing regulatory networks seem to be responsible for many aberrant splicing events. Various SFs have been reported to be mutated in cancer, including U2AF1, SF3B1, SRSF2, and RBM10 (Imielinski et al., 2012; Yoshida et al., 2011). SFs have also been observed to be upregulated or to a slighter extent downregulated in various human tumors, suggesting that they can act as proto-oncogenes and tumor suppressors, respectively (Grosso et al., 2008b; Sebestyén et al., 2016). For instance, the SR protein SF2/ASF that is encoded by *SRSF1* and regulates both constitutive and AS is upregulated in various human cancers and its overexpression can trigger malignant transformation, suggesting a proto-oncogenic role (Karni et al., 2007). It has been shown that alterations of AS events and the expression of particular SFs can be indicative of tumor grade and response to therapy and can thus be used as biomarkers for the classification of tumors (David and Manley, 2010; Grosso et al., 2008b; Relógio et al., 2005). Splicing-based therapeutic approaches also include the use of cancer-specific splicing isoforms as potential drug targets (Le et al., 2015). The splicing machinery itself has further been proposed as a target of anti-tumor drugs in order to inhibit and partly correct cancer-associated splicing abnormalities (Bonnal et al., 2012; Grosso et al., 2008b).

1.2.4 Analysis of Alternative Splicing on the Genome-Scale

The genome-wide screening of AS patterns requires the use of high-throughput technologies such as splicing-sensitive DNA microarrays or RNA-sequencing (RNA-seq). DNA microarrays are two-dimensional microchips that contain anchored arrays of short single-stranded DNA molecules, known as probes. For measuring gene expression, nucleic acid samples, e.g., DNA or RNA, are isolated, converted into complementary DNA and labelled with fluorescent dyes, before being applied to the array where they bind to complementary probes in a process known as hybridization. The strength of the fluorescence signal emitted from specific spots on the hybridized array upon laser excitation corresponds to the amount of bound nucleotides and can be used to infer expression values of genomic features. Depending on the design of the array, this enables the quantitative analysis of expression on gene-level (gene arrays) or exon-level (exon arrays), as well as the detection of AS events (exon and transcriptome arrays) (**Figure 1-9**).

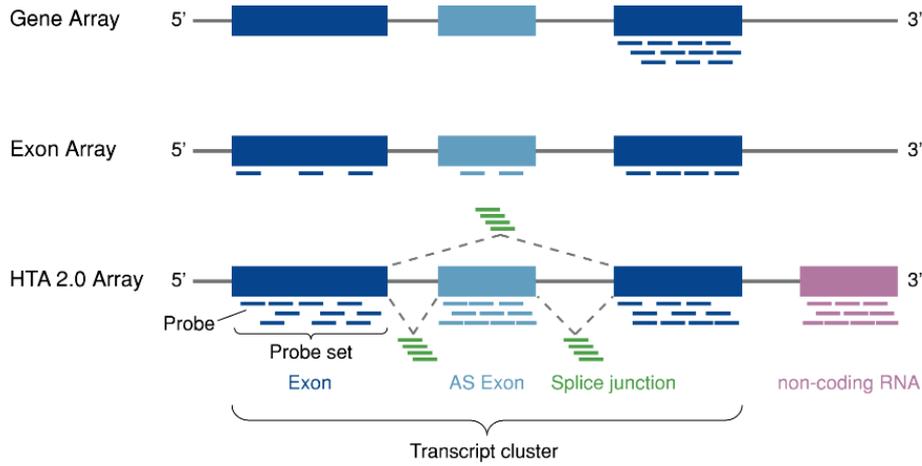


Figure 1-9: Microarray design. The Affymetrix Human Transcriptome Array (HTA) 2.0 interrogates over six million distinct probes for both coding and non-coding genes. On average, exons are measured with ten probes and splice junctions with four probes. Probes are arranged into probe sets that usually coincide with an exon. Each probe set is assigned to a transcript cluster that corresponds to a gene. In comparison, Affymetrix exon arrays measure up to four probes per exon, and Affymetrix gene arrays measure approximately 25 probes per transcript cluster. Adapted from Xu et al. (2011).

The Finding Isoforms using Robust Multichip Analysis (FIRMA) method can be used to detect AS of internal cassette exons from exon array data of single samples without replicates (Purdom et al., 2008). FIRMA detects exon-specific changes in expression levels and scores them depending on whether they deviate from the expected gene expression level by fitting of the robust multichip average (RMA) model. RMA is a widely accepted standard for preprocessing data from multiple microarrays to gain comparable gene expression values (Irizarry et al., 2003a). Starting with raw Affymetrix CEL files containing intensity values, the method employs three steps to produce so-called RMA expression values: In the first step, the background is adjusted to remove local artifacts and noise, followed by a quantile-normalization to remove the effects of individual arrays and make the measurements from different arrays comparable. In the final step, probe intensities are summarized, resulting in \log_2 -transformed measurements that represent gene expression levels. The expression measure is obtained by fitting the following linear additive model for each probe set:

$$Y_{ik} = c_i + p_k + \varepsilon_{ik}, \quad (20)$$

where Y_{ik} is the background-adjusted, normalized, and \log_2 -transformed intensity for probe j in sample i , c_i is the chip effect, p_k is the log-scale probe effect for probe k and ε_{ik} represents the error term that is assumed to be independent and identically distributed with mean 0 (Irizarry et al., 2003b). To make all probe effects relative and identifiable, $\sum_k p_{jk} = 0$ is assumed for all probe sets. Model parameters are estimated using robust procedures such as median polish or iteratively reweighted least squares. The expression measure of a gene in sample i is defined as the estimate of the chip effect c_i .

To include the possibility of different expression levels per exon of a gene, FIRMA considers the more general additive model

$$Y_{ijk(j)} = c_i + e_j + d_{ij} + p_{k(j)} + \varepsilon_{ijk(j)}, \quad (21)$$

where c_i is again the chip effect for sample i , e_j is the relative change in expression for exon j , d_{ij} is the relative change for sample i in exon j , and $p_{k(j)}$ is the nested relative probe effect for the k -th probe in exon j (Purdom et al., 2008). High values of d_{ij} represent large discrepancies of sample i in exon j from the expected expression of this exon and are thus considered an indication for differential splicing. To avoid having to estimate the term d_{ij} explicitly and thus adding more noise to the parameter estimates, FIRMA instead formulates the detection of AS as an outlier detection problem: For each exon, a summary score is computed based on the residuals of the observed probe-level expression within that exon from the estimated expression produced by fitting the standard RMA model (20). A \log_2 FIRMA score of 0 indicates no departure from the model while extremely high or low scores can be indicative of exon inclusion or skipping, respectively. While there is no obvious threshold for calling an exon differentially spliced, it has been shown that high-scoring exons are tracking real AS events and that the difference between the FIRMA scores of two conditions allows for the detection of, e.g., cancer-specific splicing (Purdom et al., 2008). Interestingly, FIRMA has also been used for the detection of rhythmic AS in time-series microarray data from murine liver (McGlincy et al., 2012) (see subsection 1.3.2). However, FIRMA can only be used for the detection of cassette-type alternative exons which is the most common type of AS in mammals, but does not allow the reconstruction of more complex types of AS (**Figure 1-8**).

While FIRMA has originally been developed for Affymetrix exon arrays that contain only up to four probes per probe set ($\sim 10\%$ of probe sets have fewer than four probes due to probe selection region lengths and sequence constraints), it can also be applied to other splicing-sensitive arrays. For this thesis, splicing-sensitive arrays of the HTA type were used that contain a median of ten probes per exon (**Figure 1-9**). Derivatives of the FIRMA method include the Integrated Gene and Exon Model of Splicing (iGEMS) (Sood et al., 2016), the Random Effects for Identification of Differential Splicing (REIDS) model (Van Moerbeke et al., 2017), and FIRMAGene (Robinson and Speed, 2009), which was developed for the smaller Affymetrix Gene 1.0 ST array. While the gene array shares a large number of probes with the Affymetrix exon array, the design of the platform is not optimized for the detection of AS, necessitating adjustments in the algorithm's assumptions. Similar to the original FIRMA algorithm, FIRMAGene uses the RMA model to decompose probe-level microarray data into probe effects and expression levels and calculates probe-wise residuals from the RMA fit (Robinson and Speed, 2009). Based on the assumption that several adjacent poorly fitting probes for the same exon region that behave differently from the rest (residuals away from zero and in the same direction) can be evidence of potential AS, the algorithm scores the persistence of residuals from the RMA fit, yielding gene-level FIRMAGene scores (Robinson and Speed, 2009). In this thesis, both the original FIRMA algorithm as well as the FIRMAGene derivative are used to detect putative AS events in microarray data originating from different platforms. In particular, FIRMA/FIRMAGene scores are tested for circadian rhythms.

In recent years, high-throughput next-generation sequencing (NGS) techniques such as RNA-seq have largely replaced the use of microarrays in transcriptomic studies. RNA-seq enables the detection and quantification of distinct transcripts of a gene, making it possible to analyze differences in expression at the resolution of individual splice isoforms (Wang et al., 2009). RNA-seq datasets for transcriptomic

analyses commonly consist of tens to hundreds of millions of relatively short (<200 nt) single- or paired-end sequences (reads) that represent fragments of the original RNA molecules. To estimate the abundance of individual isoforms from these short sequences, the reads first need to be assigned to their original location on the reference genome or transcriptome in a process known as alignment. This is complicated by the fact that reads can contain sequencing errors and genomic variants that differ from the reference sequence, and the need to account for non-unique sequences and introns in the genome that lead to non-contiguous reads that span more than one exon (**Figure 1-10**). Various RNA-seq alignment methods have been developed that differ in speed, memory requirements, accuracy, and their ability to deal with spliced alignments. Following the alignment, reads covering a specific genomic feature are counted and normalized for sequencing depth and feature length to infer gene and/or transcript expression levels. Common RNA-seq metrics for expression levels are reads per kilobase million (RPKM) for single-end reads, fragments per kilobase million (FPKM) for paired-end reads, and transcripts per million (TPM) (Conesa et al., 2016).

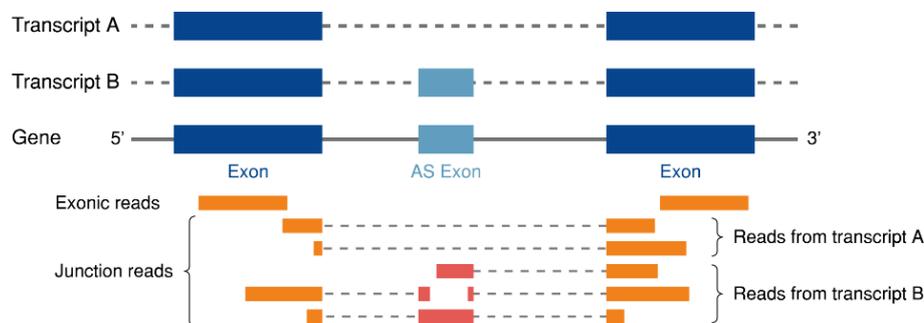


Figure 1-10: Types of RNA-seq reads. In contrast to exonic reads, junction reads are non-contiguous and span one or more exons. AS creates ambiguity concerning the origin of reads that do not map uniquely to a specific transcript. Adapted from Kim et al. (2015a).

In the following, two methods that were applied in this study for the mapping of RNA-seq reads and the subsequent quantification of transcript abundance are briefly described. The Spliced Transcripts Alignment to a Reference (STAR) aligner is one of the most widely used methods for splice-aware RNA-seq read alignment that is characterized by its high mapping speed and accuracy as well as its ability to discover novel splice junctions and chimeric alignments (Dobin et al., 2013). The STAR algorithm aligns non-contiguous read sequences directly to the reference genome in a seed searching step followed by a clustering-stitching-scoring procedure. The seed searching step is based on the concept of a sequential search for a maximal mappable prefix (MMP) between reads (or pairs of reads) and the reference genome (**Figure 1-11A**). For a read R , a read location i , a reference genome G , and a maximum mappable length (MML), the $MMP(R, i, G)$ is defined as the longest substring of the read ($R_i, R_{i+1}, \dots, R_{i+MML-1}$) that matches exactly one or more substrings of the genome. To enable a fast MMP search, the reference genome is converted to a suffix array prior to the mapping. In the following step, STAR clusters and stitches the previously aligned seeds together in accordance with a local alignment scoring scheme for (mis-)matches, indels, and splice junction gaps. For each read, the highest score determines which stitching combination is chosen as the best alignment, while in the case of multi-mapping reads,

all alignments with a score above a user-defined threshold are reported for subsequent analyses. After the generation of the alignment, further steps are needed for the quantification of gene abundance, e.g., counting tools such as HTSeq-count (Anders et al., 2015) or featureCounts (Liao et al., 2013). For the quantification of transcript (i.e., isoform) abundance based on a precomputed alignment, tools such as RSEM (Li and Dewey, 2011), RNA-eXpress (Forster et al., 2013), TIGAR (Nariai et al., 2013), and Salmon (Patro et al., 2017) can be used.

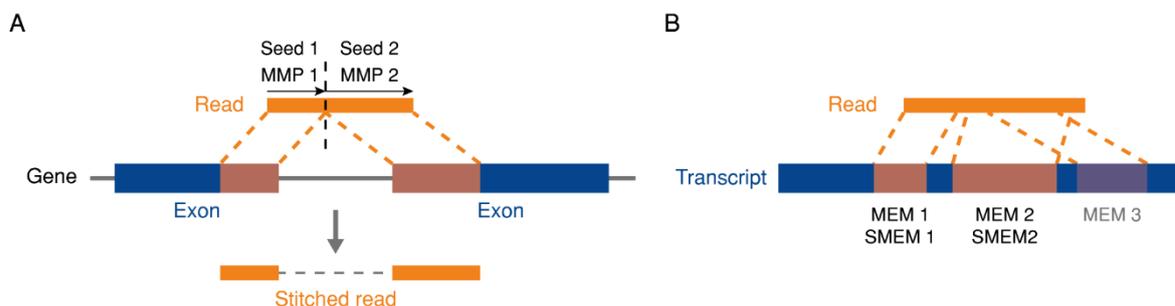


Figure 1-11: Schematic representation of the mapping strategies of STAR and Salmon. (A) STAR searches for MMPs between a read and the reference genome and stitches them together. Adapted from Dobin et al. (2013). (B) Salmon applies a quasi-mapping step where MEMs and SMEMs between a read and the transcripts are identified and chained. Adapted from Patro et al. (2016).

Salmon is a method for quantifying transcript abundance that has both an alignment-free, so called quasi-mapping mode, and an alignment-based mode (Patro et al., 2017). In the alignment-free mode, the quantification is performed directly with an index of the reference transcriptome and the raw reads as input. In the quasi-mapping step, the algorithm searches for maximal exact matches (MEM) between the reads (query) and the transcripts (reference) (Patro et al., 2016) (**Figure 1-11B**). A MEM between two sequences is defined as an exact match that cannot be extended in either direction without producing a mismatch. A super maximal exact match (SMEM) is a MEM that is not contained in any other MEM in either the query or the reference (Li, 2013). Salmon attempts to cover the read by forming chains of SMEMs and MEMs, allowing only small gaps between them. If the chain covers a user-defined fraction of the read, the transcript is considered as a possible locus of origin. Alternatively, Salmon can also be used in the alignment-based mode that requires a previously computed alignment to the transcriptome as input. In the following inference step, Salmon attempts to find nucleotide fractions (i.e., transcript abundances) that optimize the probability of a maximum likelihood model of the observed data while simultaneously learning parameters for a data-based auxiliary bias model to account for sample-specific parameters and biases (e.g., fragment length distributions). If a read maps to several transcriptomic positions, the algorithm considers all the projected positions and allocates it probabilistically by maximizing the joint likelihood of all the observed data.

Various strategies have been developed for the detection of differential splicing events between two conditions (reviewed in Adams et al. (2011)) which are not covered here because they do usually not consider time-series RNA-seq data. In recent years, there has been a growing interest in measuring temporal dynamics in gene expression on a genome-wide scale by high-throughput time-series experiments (Oh et al., 2014). Accordingly, several tools have been introduced in the last years that aim at

the detection of differential gene expression (Spies et al., 2017) and AS isoform switches (Guo et al., 2017; Huang, 2018; Nueda et al., 2017) in RNA-seq time-series data. However, none of these methods have been developed for the analysis of circadian datasets and thus usually require the repeated sampling of biological replicates at the same time point which is often not feasible in the experimental design of genome-wide studies on a circadian timescale. Under the assumption that AS events might be rhythmically regulated by the circadian clock (see subsection 1.3.2), new pipelines and methods for the genome-wide detection of putatively rhythmic changes in isoform expression need to be developed that take the specific challenges presented by circadian datasets into account. In this thesis, STAR and Salmon (in alignment-mode) are used jointly to quantify transcript-level expression for time-series data and thus enable the detection of rhythms in the transcription of individual splice variants.

1.3 Interplay between the Circadian Clock and mRNA Splicing

Circadian rhythmicity in gene expression was long assumed to be mainly due to the circadian TTFL, with particular emphasis on the rhythmic activation and inhibition of E-box-mediated transcription via the CLOCK:BMAL1 complex (see subsection 1.1.2). However, studies comparing circadian rhythmicity on nascent RNA- and mRNA-level in murine liver have revealed that rhythmic *de novo* transcription only accounts for about 22–30% of genes with rhythmic mRNA expression patterns, indicating that the majority of rhythmic mRNAs may instead result from a combination of diverse post-transcriptional events (Koike et al., 2012; Menet et al., 2012). Accordingly, RNA-based post-transcriptional mechanisms in clock regulation have increasingly moved into the focus of circadian research (Preußner and Heyd, 2016). Several post-transcriptional processes are known to be under clock control and to contribute to circadian gene expression and protein abundance, including alternative polyadenylation, mRNA degradation, translation, and mRNA splicing (Kojima and Green, 2014; Lim and Allada, 2013; Mermet et al., 2017).

1.3.1 Regulation of the Circadian Clock via Alternative Splicing

In recent years, AS has emerged as an important post-transcriptional mechanism responsible for the maintenance of the circadian clock and its response to environmental inputs in several circadian model systems (**Figure 1-12A**). Many core clock genes have more than one transcript, making them potential targets for AS mechanisms and post-transcriptional regulation, e.g., by producing clock isoforms with distinct functions or by controlling their expression levels via producing non-functional transcripts and/or inducing nonsense-mediated decay (NMD).

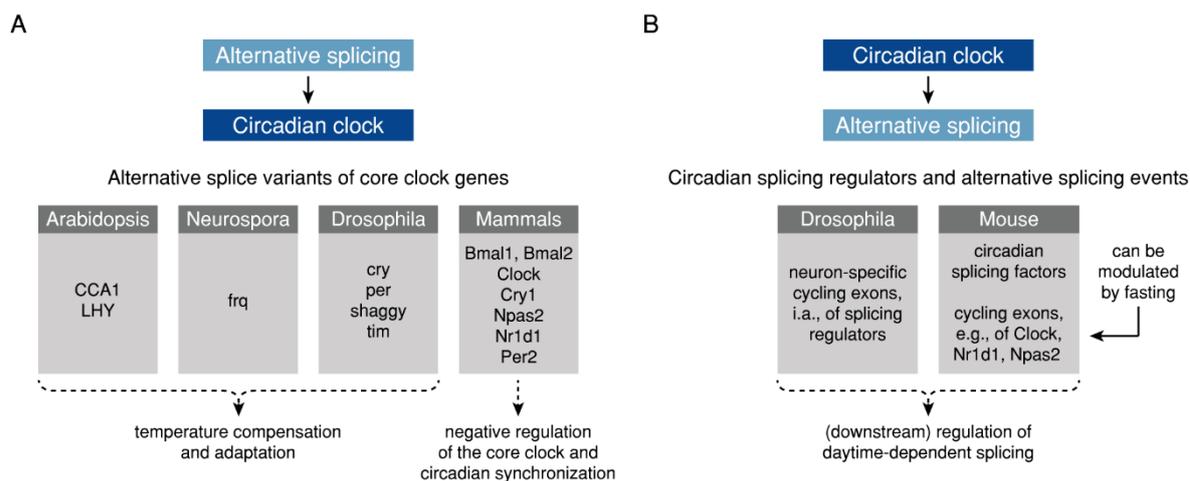


Figure 1-12: Interplay between the circadian clock and alternative splicing. (A) Regulation of the circadian clock via AS. Literature findings about alternative splice variants of core-clock genes and their potential functional consequences are represented for various organisms. (B) Regulation of AS via the circadian clock. Literature findings about circadian splicing regulators and circadian AS events and their potential functional consequences are represented for *Drosophila* and mouse.

Splice variants of clock components of the circadian model plant *Arabidopsis thaliana*, particularly of *LHY* and *CCA1*, have been extensively studied in the context of temperature adaptation and compensation. Temperature reduction leads to the retention of an *LHY* intron, likely resulting in NMD due to the introduction of a premature stop codon, thereby modulating the expression levels of the normally spliced *LHY* protein in response to cold (James et al., 2012; Seo et al., 2012). In contrast, cold temperatures have been found to prevent the intron retention of a *CCA1* transcript that acts as a dominant negative inhibitor of the normally spliced transcript (James et al., 2012). The reduced production of the alternative *CCA1* isoform likely contributes to freezing tolerance of plants (Seo et al., 2012). Moreover, several splicing-related genes have been identified that influence AS of clock genes in *Arabidopsis*, including SFs that play a conserved role in other species, e.g., *GEMIN2* (Schlaen et al., 2015), *LSM4* and *LSM5* (Perez-Santángelo et al., 2014), *PRMT5* (Sanchez et al., 2010), and *SR45* (Filichkin et al., 2015). In most cases, mutations of these SFs lead to a lengthening of the free-running circadian period, however, the underlying mechanisms of how changes in SF networks lead to the AS of clock genes and subsequently to changes in the circadian phenotype are still mostly unknown (Shakhmantsir and Sehgal, 2019).

Temperature-dependent AS events in clock genes are not limited to plants. In the fungal model organism *Neurospora crassa*, thermosensitive AS of the core clock gene *frq* determines the ratio of the long protein isoform l-FRQ to the short isoform s-FRQ that is responsible for fine-tuning the circadian period length and maintaining a robust rhythmicity in response to temperature changes (Diernfellner et al., 2007; Liu et al., 1997). The retention of an intron in the 3'-untranslated region of *Drosophila* core clock gene *per* is a thermosensitive splicing event that has been linked to the timing of evening activity in flies and likely plays a role in the adaptation to seasonal weather changes (Low et al., 2008; Majercak et al., 1999). Likewise, AS of *Drosophila* core clock gene *tim* is temperature-dependent and results in at

least four distinct isoforms of which two are cold-specific and which might contribute to the temperature compensation of the fly clock (Anduaga et al., 2019; Boothroyd et al., 2007; Evantal et al., 2018; Foley et al., 2019). Though not linked to temperature, rhythmic AS patterns of genes encoding for the clock proteins CRY and Shaggy have been identified in *Drosophila* circadian neurons (Wang et al., 2018).

AS also seems to control different aspects of the mammalian circadian clock, with many clock components displaying several isoforms, though the functional consequences of these AS events are often only poorly understood so far. Alternatively spliced variants of Bmal1 (Ikeda and Nomura, 1997; Lee et al., 2018; Yu et al., 1999) and Bmal2 (Schoenhard et al., 2002) have been described for humans, mice, and rats, where they often show divergent expression levels among tissues, indicating a functional relevance. For instance, the human-specific alternative isoform BMAL1a remains in the cytoplasm and does not enter the nucleus due to a loss of the N-terminal nuclear localization signal (Lee et al., 2018). However, BMAL1a still dimerizes with its canonical isoform or with CLOCK, thereby interfering with the formation of the CLOCK:BMAL1 heterodimer complex and thus acting as a negative regulator of the clock (Lee et al., 2018). PER2S, an alternative isoform of human PER2, has been found to be expressed both at mRNA- and protein-level in the nucleolus of human keratinocytes and likely plays a role in their circadian synchronization (Avitabile et al., 2014). An alternative *Cry1* isoform with an antiphasic expression pattern to the canonical isoform has been observed in the murine SCN (Pembroke et al., 2015). A study on rhythmic AS events in murine tissues detected by exon arrays further revealed circadian rhythms in exon inclusion of several other clock genes, including *Clock*, *Npas2* and *Nr1d1* (McGlincy et al., 2012). A well-studied example of an alternatively spliced isoform affecting a core component of the mammalian clock is the murine SF U2af26/U2af114. 24-h rhythmic skipping of exons 6 and 7 of *U2af26* results in a frameshift mutation that leads to the addition of a new C terminus to the protein which shows homology to the *Drosophila* clock protein TIM (Preußner et al., 2014). Analogous to TIM, the alternative isoform U2af26 Δ e67 interacts with Per1 and destabilizes it (Preußner et al., 2014). U2af26-deficient mice show near-arrhythmic Per1 levels and display increased phase advance adaptation following experimental jetlag, indicating a functional importance of U2af26 for correct circadian gene expression and the resetting of the clock (Preußner et al., 2014).

1.3.2 Regulation of Alternative Splicing via the Circadian Clock

As described in the previous subsection, several alternatively spliced isoforms have been reported to display 24-h rhythmic patterns at the mRNA and/or protein-level, indicating that the circadian clock might play a role in the regulation of AS (**Figure 1-12B**). Splicing is a rapid process that *in vivo* takes place in the range of seconds to minutes (Carmo-Fonseca and Kirchhausen, 2014; Huranová et al., 2010; Martin et al., 2013). The precise timing of splicing is assumed to be important for the dynamic control of AS decisions (Carmo-Fonseca and Kirchhausen, 2014). A temporal regulation of splicing via the circadian clock might both result in rhythmic patterns of AS, e.g., exon inclusion or intron retention, as well as in non-rhythmic splicing events, e.g., the production of an alternatively spliced isoform due

to changes in the oscillatory patterns of splicing-related genes. Indeed, SFs and other splicing-related genes have been observed to display rhythmic expression patterns with a circadian period in several organisms and can sometimes be linked to AS of possible target genes (Perez-Santángelo et al., 2012).

In a study exploring the circadian transcriptome of the *Drosophila* brain using RNA-seq libraries, the abundance of relatively few alternatively spliced isoforms was found to be regulated as a function of time of day, suggesting a limited influence of circadian AS regulation in the fly brain (Hughes et al., 2012a). However, many AS events were affected by loss of function of the clock gene *per* (Hughes et al., 2012a). In a more recent study, Wang et al. (2018) used a previously published time-series RNA-seq dataset to characterize AS profiles in subtypes of circadian *Drosophila* neurons. They discovered neuron-specific expression patterns of genes encoding for RBPs and rhythmic AS events, the majority of which did not cycle at the total mRNA level. In contrast, only few rhythmic AS events were identified in a negative control group of non-circadian neurons that do not express clock genes, reinforcing the hypothesis of a circadian regulation of rhythmic AS events that is possibly driven by rhythmic RBPs that are only active at specific times of the day (Wang et al., 2018). Among the genes with rhythmic AS events, several were encoding for splicing regulators and might thus be responsible for downstream daytime-dependent splicing changes.

An early example of mammalian circadian AS was reported in murine liver for the gene encoding for the signaling protein PSEN2 that shows oscillations of several shorter alternatively spliced isoforms in contrast to the non-rhythmic full-length isoform (Bélanger et al., 2006). The first genome-wide study on rhythmic AS in mammals was also conducted in murine liver (McGlinchy et al., 2012). Using time-series exon-microarrays, the authors identified several exons that were alternatively spliced in a circadian manner and that could be validated in other murine tissues. For the majority of these circadian exons, phase and amplitude of AS were found to be tissue-dependent and concurrent with circadian transcript abundance (McGlinchy et al., 2012), raising the idea that circadian AS occurs co-transcriptionally (Kojima and Green, 2014). Furthermore, it was found that hepatic circadian AS is at least partly controlled by the autonomous liver clock and that it could be modulated by fasting (McGlinchy et al., 2012). The authors also examined the hepatic mRNA expression of over 200 SFs listed by Grosso et al. (2008a) and identified 15 robustly circadian SFs, including *Srsf3*, *Srsf5*, *Sf3b1*, *Hnrnpdl*, *Cirbp*, and *Pcbp2*, further supporting the hypothesis that the expression of some SFs might be regulated by the circadian clock. In their multi-organ study of the murine circadian transcriptome, Zhang et al. (2014) found that circadian genes have more expressed splice isoforms than non-circadian genes (46% of protein-coding circadian genes expressed multiple isoforms) and that for circadian genes, the identity of the dominant isoform tended to differ more across tissues.

Due to being entrainable to light and persisting in constant darkness, rhythmic AS of the SF U2af26 (see subsection 1.3.1) had initially been termed circadian (Preußner et al., 2014). However, later experiments with young, not yet fully endothermic mice subjected to changes in the ambient temperature revealed that splicing of *U2af26* is instead controlled via circadian body temperature cycles that drive the rhythmic phosphorylation of SR proteins, e.g., of *Srsf2* and *Srsf7* (Preußner et al., 2017). The

authors found that low body temperature leads to SR protein phosphorylation, whereas high body temperature leads to their dephosphorylation. Other temperature-responsive, core clock-independent 24-h rhythmic AS events (also termed “endothermic circadian” by the authors) include the core transcription factor *Tbp*, as well as *Hnrnpa2b1* and *Ktn1*, but no core clock genes. Since most adult mammalian species are endothermic, the authors conclude that it is hard to discriminate between truly circadian and endothermic circadian rhythms, underlining the need for *in vivo* experimental setups that disrupt body temperature rhythms or *in vitro* cell culture experiments under constant temperature conditions (Preußner and Heyd, 2018).

While all of the aforementioned examples describe rhythmic changes in AS, components of the circadian clock can also influence the outcome of splicing without necessarily causing rhythms in isoform expression. In a recent study, a skeletal muscle-specific knockout of *Bmal1* was found to result in a shift in expression from a short to a long isoform of the largest known protein Titin in murine muscle tissues with potential implications for changes in muscle tension (Riley et al., 2018). Overall, there seems to be accumulating evidence for a reciprocal interplay between AS and the circadian circuitry with possible functional implications for diverse biological processes.

1.3.3 A Role for Clock-Regulated Splicing Events in Cancer?

As previously described, disruptions of the circadian clock (see subsection 1.1.3) and aberrant splicing (see subsection 1.2.3) are associated with various aspects of tumor biology. Several genes that have alternatively spliced isoforms with important functions in carcinogenesis- and tumor progression are themselves under control of the circadian clock or interact with clock components on the protein-level, including *MDM2* (Liu et al., 2018), *VEGFA* (Jensen and Cao, 2013), and *BCL-X* (Hua et al., 2006), indicating a possible circadian control of cancer-relevant AS. Additionally, several SFs with robustly circadian gene expression patterns in murine liver (McGlinchy et al., 2012) play a critical role in cancer: For instance, both SRSF3 and SRSF5 are upregulated in various cancers and have been identified as oncogenes responsible for the production of isoforms that promote cell proliferation and transformation (Jia et al., 2019; Yang et al., 2018). Changes in SF expression are often observed in cancer where they can affect the splicing of target genes and thus contribute to tumorigenesis (Sveen et al., 2016).

Therefore, it is conceivable that clock-controlled oscillations in the expression levels of SFs might be required for the correct production of a specific isoform at a specific time during the daily cycle. A clock-controlled timing of AS might serve as an additional temporal layer of proteome diversification. A disruption of this hypothesized time-of-day-dependent AS could interfere with molecular processes such as metabolism, the cell cycle, and proliferation. Cancer cells with dysregulated circadian clocks could potentially hijack this process by producing cancer-relevant isoforms that ultimately lead to carcinogenesis and tumor progression.

1.4 Aim and Structure of the Thesis

The question whether the circadian clock regulates the abundance of alternatively spliced isoforms has been named one of the outstanding open problems in the circadian field (Hughes et al., 2012a). Though comprehensive efforts have been made to uncover circadian rhythms in AS on the genome-scale in *Drosophila* (Hughes et al., 2012a; Wang et al., 2018) and in mice (McGlinchy et al., 2012), the mechanics and the potential functions of rhythmic mRNA splicing remain largely unknown. Emerging evidence points to an interplay between clock components and splicing-related genes that might affect mammalian cancer development and progression.

This thesis aims to investigate the hypothesized circadian control of splicing in mammals and its potential role in cancer by a computational analysis of time-series transcriptome data of human cancer cell lines and healthy mammalian tissues (**Figure 1-13**). In particular, it aims to determine

- whether the expression of genes encoding for spliceosome components and SFs is circadian and whether their expression patterns differ between healthy mammalian tissues and cells in different stages of cancer, and
- whether circadian rhythms in AS events can be observed in mammalian tissues and human cancer cells and whether the produced isoforms might be involved in the timing of cancer-relevant processes, such as metabolism, proliferation, apoptosis, angiogenesis, and metastasis.

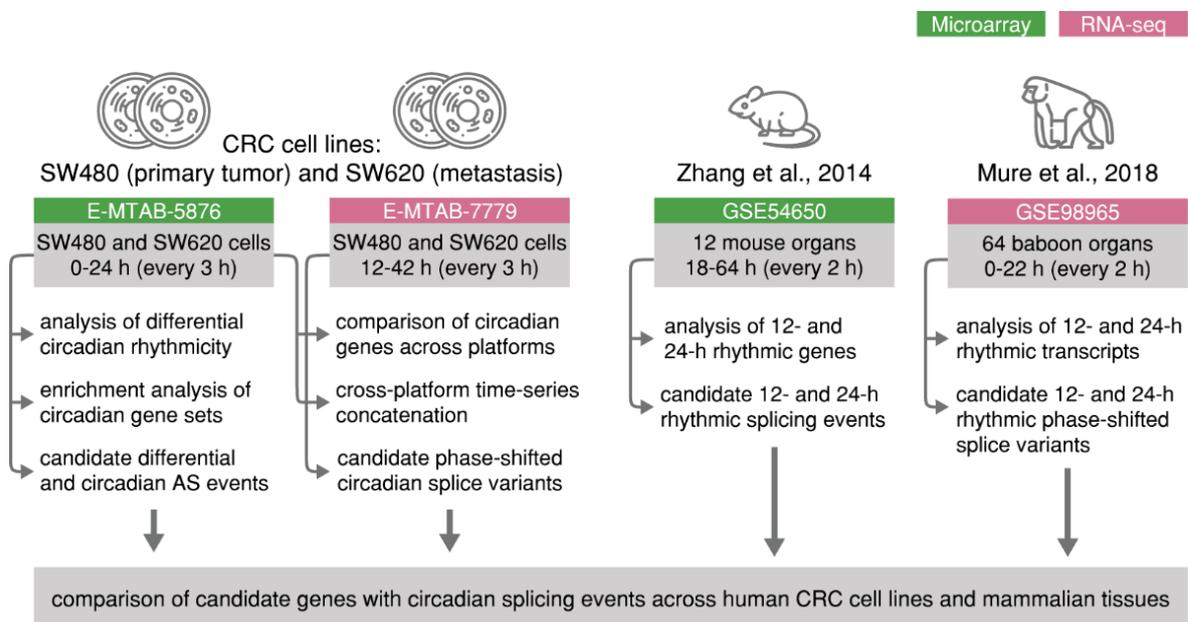


Figure 1-13: Structure of the thesis.

To this end, a computational analysis for the detection of candidate rhythmic splicing events in DNA microarray and RNA-seq time-series datasets was conducted. In the first part of the analysis, the circadian transcriptome of a model system of human CRC progression was investigated based on time-series microarray data from two cell lines that originate from a primary colorectal tumor (SW480) and a lymph node metastasis (SW620) of the same patient. 24-h rhythms in gene expression were identified

and explored to elucidate how transcriptional rhythmicity and the temporally regulated biological pathways might differ between CRC tumor stages. Differentially rhythmic genes were compared between the cell lines, with a particular focus on changes in oscillations of spliceosome components and SFs. Candidate differential AS events between SW480 and SW620 cells as well as putative circadian AS events were detected and their possible functional consequences for tumor progression were explored.

For the second part of the analysis, the circadian transcriptome of the same CRC progression model was profiled by RNA-seq. Being of different lengths and starting at different time points, the two datasets were normalized and concatenated to gain a longer time-series that allows for a more robust identification of circadian rhythms in gene expression. Gene expression and 24-h rhythmic gene sets were compared between the two platforms. Differentially rhythmic splice variants of the same gene were identified based on the transcript-level RNA-seq data and compared between the cell lines as well as with the candidate circadian AS events gained from the microarray data. In the third part of the analysis, the hypothesized 24-h rhythmicity of AS was further investigated in two previously published multi-organ circadian transcriptome datasets of a nocturnal rodent (Zhang et al., 2014) and a diurnal primate (Mure et al., 2018) that serve as examples for healthy mammalian tissues. In addition, the data was screened for potentially 12-h rhythmic genes and AS events, since the sampling resolution of the time-series was higher than that of the CRC cell line data. Rhythmic expression patterns of splicing-related genes were analyzed and compared between species, with a view on elucidating potential mechanisms that might be responsible for a rhythmic regulation of AS in mammalian tissues in health and disease.

2 Materials and Methods

2.1 Materials

2.1.1 Deposited Data

Resource	Author	Identifier/URL
Adapter sequences TruSeq3-PE-2.fa	Trimmomatic development team	https://github.com/timflutre/trimmomatic/blob/master/adapters/TruSeq3-PE-2.fa
Affymetrix mogene10 annotation data mogene10stranscriptcluster.db	MacDonald (2017)	http://bioconductor.org/packages/release/data/annotation/html/mogene10stranscriptcluster.db.html
CDF mogene10 annotation data MoGene-1_0-st-v1,mm9.cdf	FIRMAGene development team	http://bioinf.wehi.edu.au/folders/firmagene/
Circadian microarray data of 12 murine tissues	Zhang et al. (2014)	GEO: GSE54650
Circadian RNA-seq data of 12 murine tissues	Zhang et al. (2014)	GEO: GSE54651
Circadian RNA-seq data of 64 olive baboon tissues	Mure et al. (2018)	GEO: GSE98965
Circadian microarray data of cell lines SW480 and SW620	El-Athman et al. (2018)	ArrayExpress: E-MTAB-5876
Circadian RNA-seq data of cell lines SW480 and SW620	El-Athman et al. (2019)	ArrayExpress: E-MTAB-7779
Genome wide annotation for human org.Hs.eg.db	Marc Carlson	https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html
Genome wide annotation for mouse org.Mm.eg.db	Marc Carlson	https://bioconductor.org/packages/release/data/annotation/html/org.Mm.eg.db.html
Human reference genome Ensembl release 92, GRCh38	Genome Reference Consortium	https://www.ensembl.org/Homo_sapiens/Info/Index
Olive baboon (Papio Anubis) reference genome Ensembl release 93, Panu_3.0	Genome Reference Consortium	https://www.ensembl.org/Papio_anubis/Info/Index

2.1.2 Software and Algorithms

Resource	Version	Author	URL
AME	4.12.0	McLeay and Bailey (2010)	http://meme-suite.org/doc/ame.html
Aroma.Affymetrix	3.1.1	Bengtsson et al. (2008)	https://cran.rstudio.com/web/packages/aroma.affymetrix/index.html
Bedtools	2.26.0	Quinlan (2014)	https://github.com/arq5x/bedtools2
circular	0.4-93	Agostinelli and Lund (2017)	https://cran.r-project.org/web/packages/circular/index.html
clusterProfiler	3.10.1	Yu et al. (2012)	http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
DeltaCCD	0.0.0.9001	Shilts et al. (2018)	https://github.com/hughey/deltaccd
DODR	0.99.2	Thaben and Westermark (2016)	https://cran.r-project.org/web/packages/DODR/index.html
edgeR	3.20.9	Robinson et al. (2010)	https://bioconductor.org/packages/release/bioc/html/edgeR.html
eulerr	5.0.0	Larsson (2018)	https://cran.r-project.org/web/packages/eulerr/index.html
FastQC	0.11.7	Andrews (2010)	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
FIRMAGene	0.9.8	Robinson and Speed (2009)	https://rdrr.io/rforge/FIRMAGene/
GenomeGraphs	1.38.0	Durinck et al. (2009)	https://bioconductor.org/packages/release/bioc/html/GenomeGraphs.html
HarmonicRegression	1.91	Lück et al. (2014)	https://cran.r-project.org/web/packages/HarmonicRegression/index.html
Mfuzz	2.38.0	Kumar and Futschik (2007)	http://bioconductor.org/packages/release/bioc/html/Mfuzz.html
Oligo	1.42.0	Carvalho and Irizarry (2010)	https://bioconductor.org/packages/release/bioc/html/oligo.html
RAIN	1.12.0	Thaben and Westermark (2014)	https://www.bioconductor.org/packages/release/bioc/html/rain.html
Salmon	0.10.2	Patro et al. (2017)	https://github.com/COMBINE-lab/salmon
STAR	2.6.0a	Dobin et al. (2013)	https://github.com/alexdobin/STAR
Trimmomatic	0.38	Bolger et al. (2014)	http://www.usadel-lab.org/cms/?page=trimmomatic
Tximport	1.6.0	Soneson et al. (2015)	https://bioconductor.org/packages/release/bioc/html/tximport.html

2.2 Molecular and Cell Biology Methods

2.2.1 Sample Preparation

Unless stated otherwise, wet-lab experiments were carried out by Luise Fuhr from the AG Relógio. For this reason, only an overview of molecular and cell biology methods is given here. For details on materials and methods for the cell culture and the sample preparation, please refer to Fuhr et al. (2018), El-Athman et al. (2018), and El-Athman et al. (2019).

Cells from the human colorectal carcinoma cell lines SW480 and SW620 were seeded one day prior to the start of the experiment and kept in constant conditions at 37°C in a humidified atmosphere with 5% CO₂. Cell populations were synchronized by changing the culture medium. Time-series sampling (with every time point sampled from an individual plate of cells to avoid longitudinal batch effects) was either started immediately (for the microarrays) or 12 h after synchronization (for RNA-seq) (**Figure 2-1**). Samples were taken every 3 h for a time-series of either 24 h (for the microarrays) or 30 h (for RNA-seq) and prepared for RNA extraction. The microarray hybridization was carried out by the Labor für funktionelle Genomforschung of the Charité – Universitätsmedizin Berlin, DE. The CEL files containing raw intensities have been deposited in the ArrayExpress database at EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-5876. The mRNA libraries for RNA-seq were sequenced on an Illumina NextSeq 500 platform to an average depth of 74M (min. 49M, max. 129M) 75 bp paired-end reads at the EMBL GeneCore Facility, Heidelberg, DE. The raw RNA-seq read files have been deposited in the ArrayExpress database at the EMBL-EBI (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-7779.

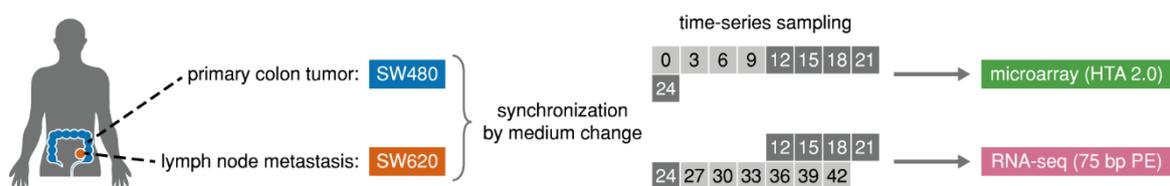


Figure 2-1: Sampling of a cellular model of human CRC progression. Gene expression of two CRC cell lines originating from the same patient were profiled over a time-series of 24 h and 30 h using microarrays and RNA-seq, respectively. The SW480 cell line is derived from the primary colon adenocarcinoma and the cell line SW620 is derived from a lymph node metastasis.

2.3 Bioinformatics and Statistical Methods

Unless stated otherwise, bioinformatics and statistical analyses were conducted using the R programming language (v3.4–3.5) in the RStudio integrated development environment.

2.3.1 Processing of Microarray Transcriptome Data

The raw microarray expression data of twelve murine organs (**Table S 5**) sampled every 2 h for two circadian cycles by Zhang et al. (2014) was downloaded from the Gene Expression Omnibus (GEO) database (GSE54650). The raw microarray expression data of the human CRC cell lines and the murine organs was preprocessed for all time points of each dataset as one batch using the RMA methodology (see subsection 1.2.4) (Bolstad et al., 2003; Irizarry et al., 2003a) as implemented in the oligo package (v1.42.0) (Carvalho and Irizarry, 2010). A transcript cluster consists of one or more exon clusters (on exon array) or probes (on gene arrays) and roughly corresponds to known or putative genes (Affymetrix). Transcript clusters of the human CRC cell lines SW480 and SW620 were annotated with NCBI Gene IDs (Entrez IDs) and Ensembl IDs using Affymetrix HTA 2.0 annotation data (hta20transcriptcluster.db, v8.7.0). For genes annotated by multiple transcript clusters, only the transcript cluster with the highest mean expression over all time points in both cell lines was retained to enable a comparison across cell lines. For the comparison of the microarray data with the RNA-seq data of the CRC cell lines, only the transcript cluster with the highest mean expression over all time points of the respective cell line was retained. For the murine data, transcript clusters were annotated with Ensembl IDs using Affymetrix mogene10 annotation data (mogene10sttranscriptcluster.db, v8.7.0). In the following, microarray transcript clusters will be referred to as “genes” to avoid confusion with features identified on transcript-level in the RNA-seq data.

2.3.2 Processing of RNA-seq Transcriptome Data

For the human CRC cell lines RNA-seq dataset, quality control was performed using FastQC (v0.11.7) (Andrews, 2010). Since the quality control revealed overrepresented sequences in several samples that might be adapters, they were cut using Trimmomatic (v0.38) (Bolger et al., 2014) with TruSeq3-PE-2.fa adapter sequences (default parameters). Only paired-end reads were retained for the downstream analyses. RNA-seq reads (paired-end, 100 bp) from 12 murine tissues sampled every 6 h for two circadian cycles (GSE54651) (Zhang et al., 2014) and RNA-seq reads (single-end, 50 bp) from 64 olive baboon tissues (**Table S 5**) sampled every 2 h for one circadian cycle (GSE98965) (Mure et al., 2018) were downloaded from the European Nucleotide Archive. Reads for the baboon iris (IRI) sampled at ZT22 were missing. Sequencing reads were aligned to the human genome (Homo_sapiens.GRCh38, Ensembl release 92), the murine genome (Mus_musculus.GRCm38, Ensembl release 95), and the olive baboon genome (Panu_3.0, Ensembl release 93), respectively, using the STAR software (v2.6.0a) (Dobin et al., 2013) (see subsection 1.2.4) with default parameters and the option --quantMode TranscriptomeSAM. The TranscriptomeSAM option generates both a genome and a transcriptome alignment by first aligning reads to the genome and then searching for concordance between the genomic alignment and annotated transcripts. If a genomic alignment is mapped to several transcriptomic coordinates, it is projected to all of them. The baboon amygdala (AMY) sample taken at ZT08 had a very low mapping rate (< 1%) and was therefore excluded from further analyses. The resulting STAR transcriptome alignments were saved as unsorted bam files. Based on the alignment, transcript-

level abundances were quantified in units of TPM using Salmon (v0.10.2) (Patro et al., 2017) in alignment-based mode (see subsection 1.2.4) with default parameters and the `--seqBias` option which enables the algorithm to learn and correct for sequence-specific biases in the input data.

The tximport package (v1.6.0) (Soneson et al., 2015) was used to import and scale the resulting transcript-level count tables by first multiplying TPM by feature length and then scaling up to the library size (lengthScaledTPM), resulting in both transcript-level (txOut=TRUE, based on Ensembl Gene IDs) and summarized gene-level (txOut=FALSE, based on Ensembl Transcript IDs) count estimates. For each sample, normalization factors to scale the raw library sizes were calculated using the trimmed mean of M-values (TMM) method and counts were \log_2 -transformed using the `cpm` function from the edgeR package (v3.20.9) (Robinson et al., 2010). For the murine and the baboon tissues, lowly expressed genes and transcripts were filtered out, retaining only features with at least 0.5 counts per million (CPM) on average over all time points per tissue. For the human CRC cell lines, all features with at least 0.5 CPM on average over all time points in at least one of the two cell lines were retained. For the comparison of the microarray data with the RNA-seq data of the CRC cell lines, all features with at least 0.5 CPM on average over all time points per cell line were retained. Subsequently, counts were renormalized using only the selected features.

2.3.3 Rhythmicity Analysis

To detect rhythmic features and parameters, the parametric harmonic regression method (see subsection 1.1.5) as implemented in the HarmonicRegression package (v1.91) (Lück et al., 2014) and the non-parametric RAIN algorithm (see subsection 1.1.5) as implemented in the RAIN package (v1.12.0) (Thaben and Westermarck, 2014) were applied, using successive filtering steps. Rhythmicity analysis was performed on the absolute RMA-preprocessed gene expression values for the array data and on the transcript- and gene-level CPM values for the RNA-seq data.

For the human CRC cell lines microarray dataset, 24-h rhythmicity was assessed using the RAIN algorithm with a period of 24 h. Phases and amplitudes were estimated by fitting a robust harmonic regression model with a period of 24 h to the time-series if the harmonic regression $p < 0.5$. For a harmonic regression $p \geq 0.5$, the phase as estimated by the RAIN algorithm was taken instead and the amplitude was calculated as the peak-to-trough ratio of the maximum and the minimum expression value. A mean expression threshold of 2.5 and a fold change (FC) amplitude threshold of 1.15 were used to filter for expressed gene with biologically significant rhythms. The RAIN p -values of the remaining genes were FDR-corrected using the BH procedure. Significance for 24-h rhythmic genes from the CRC cell lines microarray dataset was bounded by $q < 0.08$. For the differential rhythmicity analysis of the 24-h rhythmic transcriptome of the CRC cell lines microarray dataset, the robust DODR method from the DODR package (v0.99.2) (Thaben and Westermarck, 2016) was applied for all genes that were identified as robustly 24-h rhythmic in at least one of the two cell lines. DODR p -values were BH-adjusted for multiple testing with a DODR q -value cutoff of 0.05 determining genes with differential rhythmicity.

Genes were further divided in genes with amplitude gains in one of the two cell lines when their absolute \log_2 amplitude change in one cell line compared to the other was > 0.5 and in genes with a pure phase shift when their FC amplitude was > 1.15 in both cell lines, their absolute \log_2 amplitude change < 0.1 , and the phase shift between the cell lines > 1 h.

For the murine multi-organ microarray dataset, the baboon multi-organ RNA-seq dataset, and the CRC cell lines RNA-seq dataset, 12-h and 24-h rhythmicity was assessed using the RAIN algorithm with a period of 12 h and 24 h. In the case of multiple genes annotated with the same Ensembl Gene ID in the array data, only the gene with the lowest RAIN p -value per tissue was retained. Phase and relative amplitude were determined by fitting a robust harmonic regression model with a period of 12 h and 24 h, respectively. For both the mouse array and the baboon and CRC cell lines RNA-seq data, rhythmic features with a relative amplitude < 0.1 were excluded from the set. The p -values of the remaining genes and transcripts of the mouse array data were FDR-corrected using the BH method. Significance for 12-h and 24-h rhythmic features was bounded by $q < 0.05$ for the mouse array data, by $p < 0.005$ for the baboon RNA-seq data, and by $p < 0.05$ for the CRC cell lines RNA-seq data.

Circular mean and median of phases across several tissues were calculated with the R package `circular` (v0.4-93) (Agostinelli and Lund, 2017). Phase synchrony across tissues was tested by the Rayleigh test of uniformity as implemented in the R package `circular` (v0.4-93).

2.3.4 Correlation Analysis

For the human CRC cell lines microarray data, the relationship between the expression values of ten core clock genes (*BMAL1*, *NPAS2*, *CLOCK*, *CRY1*, *CRY2*, *NR1D1*, *NR1D2*, *PER1*, *PER2*, *PER3*) and two clock-regulated genes (*DBP*, *TEF*) was quantified by calculating the Spearman correlation between each pair of genes using the `DeltaCCD` package (v0.0.0.9001) (Shilts et al., 2018). For reference Spearman correlations, the samples from the murine multi-organ microarray dataset from Zhang et al. (2014) were used.

For the comparison of the human CRC cell lines microarray data with data from different RNA-seq pipelines, correlation of expression values of commonly expressed genes was conducted between all four pipeline methods (*microarray*, *STAR + featureCounts*, *STAR + Salmon*, *Salmon*) in a pairwise manner for samples of the same cell line taken at identical time points, considering only those time points that are shared between all methods (12–24 h). The resulting Pearson correlation coefficients were averaged over all time points for each method comparison. Circadian parameters (i.e., phases and relative amplitudes) of 24-h rhythmic genes were compared between the three datasets (microarray, RNA-seq, and the concatenation of both) in a pairwise manner. Only genes that were commonly identified to be 24-h rhythmic were taken into account for the correlation analyses. Circular Pearson correlation coefficients were computed for phases of 24-h rhythmic genes and statistical significance was tested using the function `core-circular` from the R package `circular` (v0.4-93) for circular statistics

(Agostinelli and Lund, 2017). For relative amplitudes, Pearson correlation coefficients were computed, and statistical significance was tested using the function `cor.test` from the R package `stats`.

2.3.5 Clustering Analysis

For the human CRC cell lines microarray data, noise-robust soft clustering of 24-h rhythmic gene expression patterns was performed with the `Mfuzz` package (v2.38.0) (Kumar and Futschik, 2007) using default parameters. The number of clusters was set to four. Prior to clustering, the \log_2 gene expression intensities were transformed to the standard normal distribution.

2.3.6 Detection of Alternatively Spliced Exons in Microarray Data

For the human CRC cell lines transcriptome array data, putative alternatively spliced exons were predicted using the FIRMA algorithm (Purdom et al., 2008) as implemented in the `aroma.affymetrix` package (v3.1.1) (Bengtsson et al., 2008) and a custom chip definition file (CDF) for HTA 2.0 arrays from Brainarray (v19) (Dai et al., 2005). For each cell line and each time point, FIRMA scores were computed for all probe sets and annotated with Ensembl exon IDs. Exons with a \log_2 expression < 2.5 were marked as absent. Exons that were absent in more than half of the samples of a cell line were excluded from the analysis, as were genes where at least half of the probe sets in more than half of the samples of a cell line were absent. Paired differences in FIRMA scores were computed as the \log_2 FC of two FIRMA scores for the same exon at the same time point. 24-h rhythmic changes in FIRMA scores over time were determined by RAIN ($p < 0.05$) (Figure 2-2A).

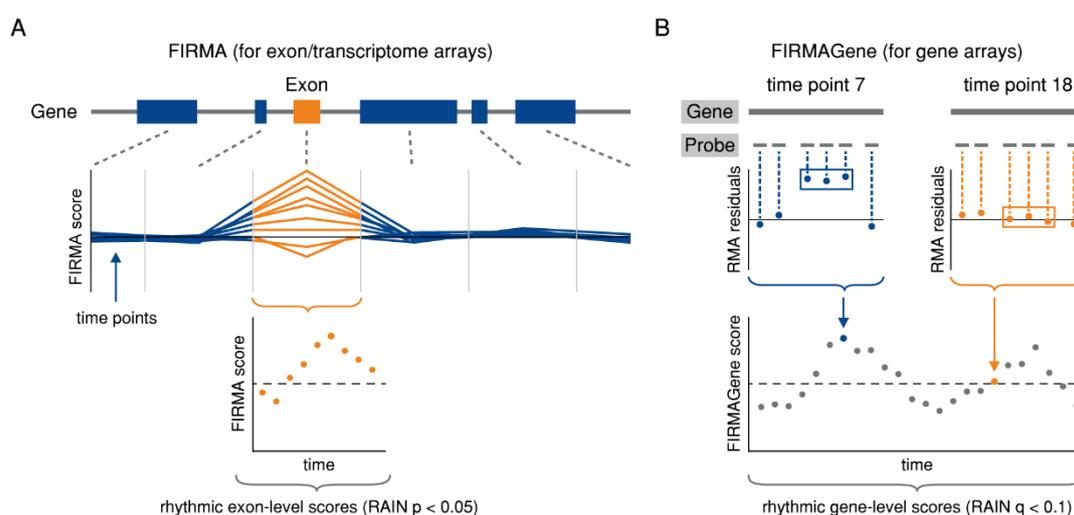


Figure 2-2: Schematic representation of the FIRMA and FIRMAGene analysis to identify candidate rhythmic splicing events in microarray data. (A) FIRMA detects exon-specific changes in expression levels derived from exon/transcriptome arrays and scores them depending on whether they deviate from the expected gene expression level by fitting the RMA model. (B) FIRMAGene scores the persistence of residuals of adjacent gene array probes for the same exon region from the RMA fit, yielding gene-level FIRMAGene scores for each individual time point. Rhythmic changes in FIRMA or FIRMAGene scores were identified using RAIN.

For the murine whole-transcript microarray data, candidate genes with putative rhythmic differential splicing events were predicted using the FIRMAGene method that was specifically developed for the Affymetrix Gene 1.0 ST platform (Robinson and Speed, 2009) (**Figure 2-2A**). Gene-level FIRMAGene scores were calculated for each time point using the FIRMAGene package (v0.9.8) in the Aroma platform as implemented in the aroma.affymetrix package (v3.1.1) and a CDF for MoGene arrays (MoGene-1_0-st-v1,mm9.cdf). Phases and relative amplitudes of the FIRMAGene scores were determined by fitting a robust harmonic regression using the HarmonicRegression package (v1.91) with a period of 12 h and 24 h, respectively. Scores with a relative amplitude < 0.1 were excluded. The RAIN p -values of amplitude-filtered scores were FDR-corrected using the BH method and significance for 12-h and 24-h rhythmic FIRMAGene scores was bounded by $q < 0.1$.

2.3.7 Detection of Differentially Rhythmic Splice Variants in RNA-seq Data

For the detection of differentially rhythmic splice variants in the RNA-seq data, the robustDODR method from the DODR package (v0.99.2) was applied to calculate the p -value of differential rhythmicity for each pair of rhythmic transcripts. For the baboon RNA-seq data, two rhythmic transcripts were defined as a pair if they had the same gene identifier and the same period in the same tissue and if one transcript had a RAIN $p < 0.005$ and the other RAIN $p < 0.05$. For the human CRC cell lines RNA-seq data, all transcripts pairs with a RAIN $p < 0.05$ were considered. The DODR p -values of the total number of 12-h and 24-h rhythmic splice variant pairs (SVPs) per tissue were FDR-corrected using the BH method and significance for differential rhythmicity was bounded by DODR $q < 0.05$. To specifically identify splice variants with similar amplitudes and a strong phase difference, only differentially rhythmic SVPs with a relative amplitude difference < 2 and a phase difference ≥ 2 h (for baboon tissues) for 12-h rhythmic transcripts and ≥ 4 h for 24 h rhythmic transcripts (for baboon tissues and human CRC cell lines) were retained for subsequent analyses.

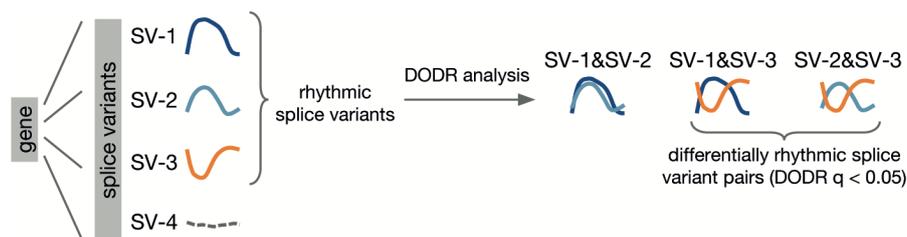


Figure 2-3: Schematic representation of the analysis of differentially rhythmic splice variant pairs in RNA-seq data. The rhythmic patterns of all expressed and rhythmic (with the same period) splice variants of the same gene in the same tissue are compared in a pairwise manner using the robust DODR method in order to identify differentially rhythmic SVPs with a phase shift in peak expression.

2.3.8 Prediction of Transcription Factor Binding Sites

For the three sets of differentially 24-h rhythmic genes identified based on the SW480 and SW620 microarray data (genes with a higher amplitude in SW480 cells, gene with a higher amplitude in SW620

cells and phase-shifted genes), promoter sequences were extracted $\pm 1,000$ bp of the RefSeq transcription start sites (TSSs). For the baboon RNA-seq dataset, promoter sequences were extracted for the two sets of splicing-related genes detected to be 24-h rhythmic in at least ten of the baboon tissues that showed an early and a late peak of expression, respectively, and for a set of arrhythmic splicing-related genes. For each consistently 24-h rhythmic splicing-related gene, all associated 24-h rhythmic transcripts were identified, while for the set of arrhythmic splicing-related genes, the associated transcripts were filtered for those expressed in at least one baboon tissue but not detected to be 24-h rhythmic ($p > 0.05$) in any tissue. Sequences of promoter regions ± 1000 bp of the TSSs from Ensembl BioMart (Ensembl Genes 95) (Zerbino et al., 2017) were extracted using bedtools (v2.26.0) (Quinlan, 2014). Duplicate sequences were removed. The enrichment of four potential clock transcription factor binding sites (D-box, E-box, E'-box, and RRE) in the promoter regions of the respective gene sets was tested using the Analysis of Motif Enrichment (AME) tool of the MEME Suite (v4.12.0) (McLeay and Bailey, 2010) based on position-specific scoring matrices from McGlincy et al. (2012), employing the average odds scoring method and one-tailed Fisher's exact test with shuffled input sequences as control.

2.3.9 Functional Enrichment Analysis

Enriched GO terms and KEGG pathways for 24-h rhythmic gene clusters from the SW480 and SW620 microarray data were computed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (v6.8) (Huang et al., 2009) against a background of all expressed genes in the respective cell line. The resulting annotations were grouped into annotation clusters based on common gene members. The enrichment score of each annotation cluster is defined as the geometric mean (in $-\log_{10}$ scale) of the individual annotations' p -values. DAVID was further used to compute enriched GO terms for the three sets of differentially rhythmic transcripts in SW480 and SW620 cells and for the candidate genes with differential AS events between the cell lines, based on the microarray data. Circadian pathways were determined by PSEA (v1.1) (Zhang et al., 2016) based on the sets of 24-h rhythmic genes identified in the SW480 and SW620 microarray data. Gene sets were downloaded from the Molecular Signatures database (MSigDB) C2 (KEGG gene sets) (v6.1) (Subramanian et al., 2005). Sets containing fewer than five 24-h rhythmic genes were excluded from the analysis. The Kuiper test was used to identify circadian gene sets by comparing the phases of all 24-h rhythmic genes (rounded to the full hour) belonging to each gene set to a uniform background distribution and by testing for non-uniformity ($q < 0.01$).

Gene Ontology (GO) enrichment analysis of gene sets identified based on the murine and the baboon data as well as the SW480 and SW620 RNA-seq data was performed using the clusterProfiler R package (3.10.1) (Yu et al., 2012). For the two sets of murine candidate genes with 12-h and 24-h rhythmic FIRMAGene scores, p -values were FDR-corrected using the BH procedure with significance bounded by $q < 0.01$. The baboon candidate genes with differentially rhythmic phase-shifted SVPs were mapped to orthologous human genes using Ensembl BioMart (Ensembl Genes 95). P -values were FDR-corrected using the BH procedure with significance bounded by $p < 0.002$ for genes with 24-h rhythmic

SVPs in SW480 and with $q < 0.05$ for genes with 24-h rhythmic SVPs in SW620 cells and by $q < 0.05$ for genes with 24-h rhythmic and by $p < 0.001$ for genes with 12-h rhythmic SVPs in baboon tissues.

2.3.10 Compilation of Lists of Splicing-related Genes

A list of 254 human spliceosome components and other splicing-related genes (**Table S 1**) was compiled from three sources: from a publication by Relógio et al. (2005) and from two public databases for spliceosome components, SpliceosomeDB (Cvitkovic and Jurica, 2012), and for human SFs, SpliceAid-F (Giulietti et al., 2012). Genes were grouped according to their recruitment to different complexes of the spliceosome (A, B, B_{act}, and C complex), different snRNPs (U11/12, U1, 17S U2, U5, and U4/U6 snRNP), and other spliceosomal complexes and proteins (PRP19 and RES complex, EJC/mRNP, and Sm/LSm proteins), as well as hnRNPs, SR proteins, and other splicing regulators. A larger list of 426 human spliceosome and splicing-related genes (**Table S 1**) was compiled by aggregating elements from the following sources: 1) the previous list of 254 spliceosome components and splicing regulators and 2) 404 splicing factor genes from Seiler et al. (2018). Mapping these genes to their orthologues using Ensembl BioMart (Ensembl Genes 95) (Zerbino et al., 2017) resulted in 429 splicing-related genes for olive baboon and 451 splicing-related genes for mouse.

2.3.11 Visualization

Unless otherwise stated, graphics were created in the R software environment using the ggplot2 package (v3.2.0) for data visualization (Wickham, 2016). For the human CRC cell lines microarray data, exon-level expression and FIRMA scores for individual genes were visualized at the genomic level using the GenomeGraphs package (v1.38.0) (Durinck et al., 2009). Tissue-wise area-proportional Venn diagrams of the logical relations between 12-h and 24-h rhythmic gene and transcript sets of the baboon RNA-seq dataset were visualized using the eulerr package (v5.0.0) (Larsson, 2018). All other area-proportional Venn diagrams were visualized using BioVenn (Hulsen et al., 2008). Graphics were refined and figures were assembled in Adobe Illustrator CC (v17.0.0).

3 Results

3.1 Analysis of the Circadian Transcriptome of Human CRC Cell Lines

To investigate the hypothesized influence of circadian regulated AS on cancer development, the CRC cell lines SW480 (derived from a primary carcinoma) and SW620 (derived from a lymph node metastasis from the same male patient) were chosen as a model system of tumor progression. The chapter starts with a comparison of the time-series expression of core clock genes and circadian transcription factors between the two cell lines, followed by a detailed analysis of differential rhythmicity on the transcriptome-level. In the second and third part, biological processes enriched for 24-h rhythmic genes are analyzed, with a particular focus on the spliceosome pathway and other splicing-related genes. The chapter closes with an analysis of candidate differential and circadian AS events in the cell line model and their possible implications for tumor progression. The results described in this chapter have for the most part been published in El-Athman et al. (2018). Text passages and figures of the publication have been adapted for this thesis.

3.1.1 Genome-wide Analysis of Differential Rhythmicity between SW480 and SW620 Cells

The circadian transcriptome of the CRC cell lines SW480 and SW620 was profiled using whole transcriptome microarrays. RNA was sampled every 3 h for a full circadian cycle (24 h) after synchronization of the cells by a change of the growth medium. The 14 core clock components *BMAL1*, *BMAL2*, *CLOCK*, *NPAS2*, *CRY1*, *CRY2*, *PER1*, *PER2*, *PER3*, *NR1D1*, *NR1D2*, *RORA*, *RORB*, and *RORC*, and the clock-regulated transcription factor genes *DBP* and *TEF* were all expressed with medium to high levels in SW480 and SW620 cells, with the exception of *RORB* that displayed only relatively low expression levels in both cell lines (mean RMA expression value < 4) (**Figure S 1**). As expected for a functioning core clock system, *BMAL1* was oscillating in anti-phase to *PER1*, *PER2*, *PER3*, *NR1D1*, and *NR1D2* in SW480 cells (**Figure 3-1A**). However, several of the robust clock gene oscillations in SW480 cells were severely diminished in the metastasis-derived cell line: *BMAL1*, *NR1D1*, *PER3*, and *DBP* were not oscillating in SW620 cells, whereas *NR1D2* and *CRY1* oscillated in a circadian manner in both cell lines but displayed lower amplitudes in SW620 cells, indicating a disruption of the core clock in SW620 cells. Overall, core clock genes tended to have lower expression levels in SW620 cells when compared to SW480 cells (**Figure S 1**). For genes whose oscillations were diminished or abolished in SW620 cells,

such as *BMAL1*, *NR1D1*, *NR1D2*, and *PER3*, the mean expression levels over time in SW620 cells corresponded to the troughs of the oscillation in SW480 cells.

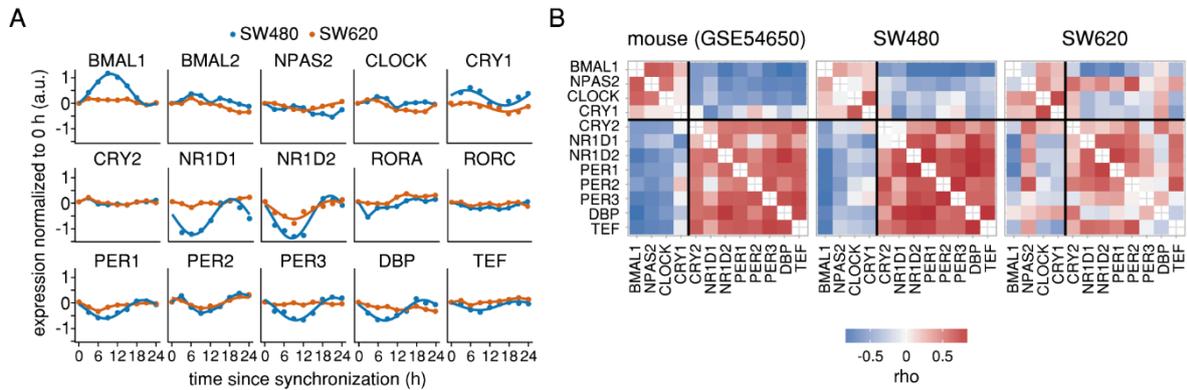


Figure 3-1: Core clock genes and circadian transcription factors in SW480 and SW620 cells (microarray data). (A) Time-series expression of core clock genes in primary tumor-derived SW480 cells (blue) and metastasis-derived SW620 cells (orange). A harmonic regression was fitted to genes that were identified as 24-h rhythmic (RAIN $q < 0.08$ after filtering for genes with a minimum FC amplitude of 1.15). (B) Heatmaps of Spearman correlation (ρ) between each pair of core clock genes for SW480 and SW620 cells in comparison to a healthy murine multi-organ reference set (GSE54650).

To further quantify the extent of circadian perturbation in the CRC cell lines, the Spearman correlation between the expression values of ten core clock genes and *DBP* and *TEF* was computed (Shilts et al., 2018) and compared to a reference signature of clock gene co-expression from healthy mouse organs (Zhang et al., 2014) (**Figure 3-1B**). Mammalian clock genes and proteins are known to be highly conserved, thus allowing for a comparison of correlation patterns across species. The correlation pattern of SW480 cells closely resembled the reference pattern: Two groups of clock genes (*BMAL1*, *NPAS2*, *CLOCK*, and *CRY1* versus *CRY2*, *NR1D1*, *NR1D2*, *PER1*, *PER2*, *PER3*, *DBP*, and *TEF*) were positively correlated with genes within their own group but negatively correlated with genes from the respective other group. In SW620 cells, the separation in two groups was less distinct and the genes *NPAS2* and *DBP* showed a negative correlation with genes within their own group, indicating a stronger dysregulation of the clock in the metastasis-derived cell line in comparison to cells derived from a primary CRC tumor.

In the next step, 24-h rhythmic gene expression was analyzed and compared between both cell lines at the whole transcriptome-level. 24-h rhythmic genes were determined by the RAIN algorithm and corrected for multiple testing after filtering for genes with a minimal FC amplitude of 1.15. The non-parametric RAIN algorithm was chosen for the analysis of rhythmicity because it allows for the detection of rhythms with arbitrary waveforms and only requires the period length (~ 24 h) as input, which can be inferred from previous bioluminescence measurements of core clock gene reporter activity in the cell lines (Fuhr et al., 2018). Given an FDR of 8%, 2282 ($\sim 9.5\%$) genes were identified in SW480 cells and 1921 ($\sim 8.0\%$) in SW620 cells (**Figure 3-2A**). The FDR cutoff of 8% was chosen because a more stringent cutoff of 5% yielded no 24-h rhythmic genes for SW620 cells (**Figure S 2A**). This seems improbable given that circadian oscillations of several core clock genes were observed in the cell line (**Figure 3-1A**) which are likely propagated to target clock-controlled genes. The resulting percentages

of 24-h rhythmic genes per cell line are comparable to previous results gained from mammalian tissues (Zhang et al., 2014). In both cell lines, the sets of 24-h rhythmic genes displayed bimodal phase distributions (**Figure 3-2B and C**). In SW480 cells, ~47% of the 24-h rhythmic genes peaked between 6 to 12 h (early peak) and ~23% between 19–24/0 h after synchronization (late peak). The late peak occurred ~12 h after the peak of rhythmic *BMAL1* expression (9.9 h), and the early peak occurs ~12 h after the peaks of *NR1D1* (18.2 h) and *NR1D2* (20.5 h), suggesting that the observed oscillations in SW480 cells might be regulated via rhythmic E-box and ROREs binding. In SW620 cells, ~58% of the 24-h rhythmic genes peak between 4–10 h (early peak) and ~18% between 16–21 h after synchronization (late peak). The phase shift of -2–3 h from SW480 to SW620 cells points to a shift of clock activity from primary tumor to metastasis, possibly resulting from the observed dysregulation of the core clock in SW620 cells.

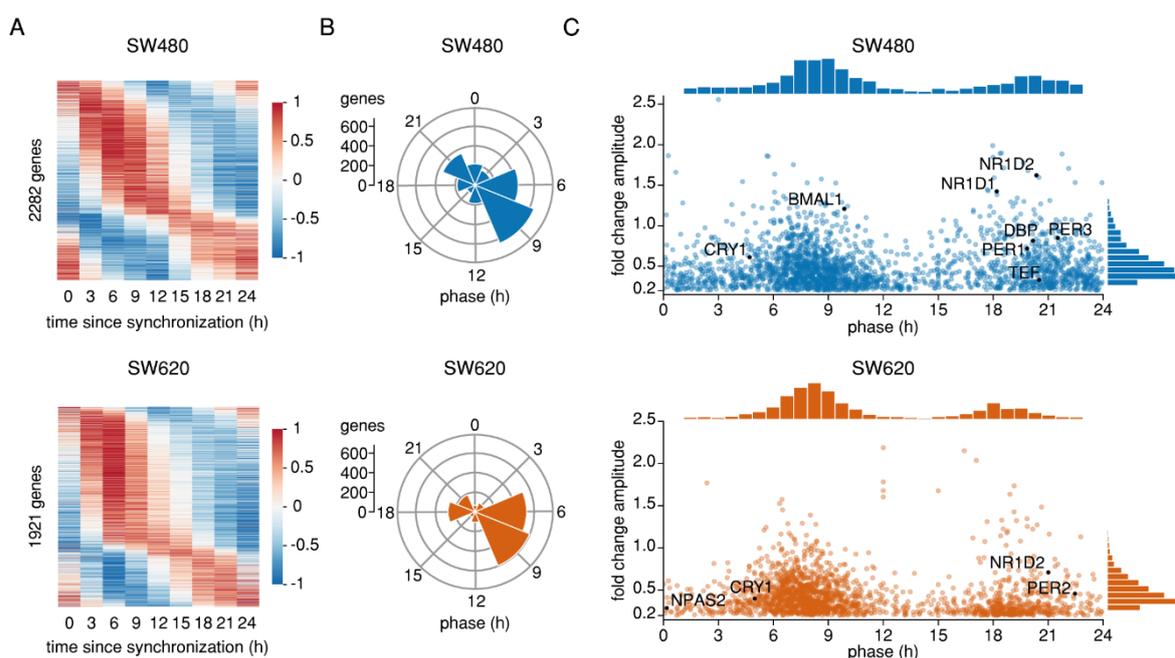


Figure 3-2: Transcriptome-level analysis of 24-h rhythmic genes in SW480 and SW620 cells (microarray data). (A) Median-normalized, phase-ordered expression heatmaps of genes that were identified as 24-h rhythmic in SW480 and SW620 cells (RAIN $q < 0.08$ after filtering for genes with a minimum FC amplitude of 1.15). Each row represents one gene. (B) 3-h phase bins of the 24-h rhythmic gene sets identified in SW480 SW620 cells. (C) Phase and amplitude distributions of the 24-h rhythmic genes identified in SW480 and SW620 cells. Each 24-h rhythmic gene is represented by a dot; core clock genes are highlighted in black.

Less than 600 genes were identified as 24-h rhythmic in both cell lines, yet many of the non-intersecting genes also showed oscillations in the respective other cell line (**Figure S 2B and C**). This observation demonstrates that even though a gene may be defined as rhythmic in one condition according to the chosen statistical threshold, while not being significantly rhythmic in another condition, it does not necessarily follow that they also have significantly different expression patterns. To attain a better understanding regarding the changes in circadian rhythmicity during tumor progression, the amplitudes and phases of individual genes were compared between both cell lines using the robust DODR method

for the detection of differential rhythmicity. The DODR method tests for significant differences in rhythmicity, i.e., in phase or amplitude, between two time-series by fitting them to sine curves with a fixed period and free phase and amplitude.

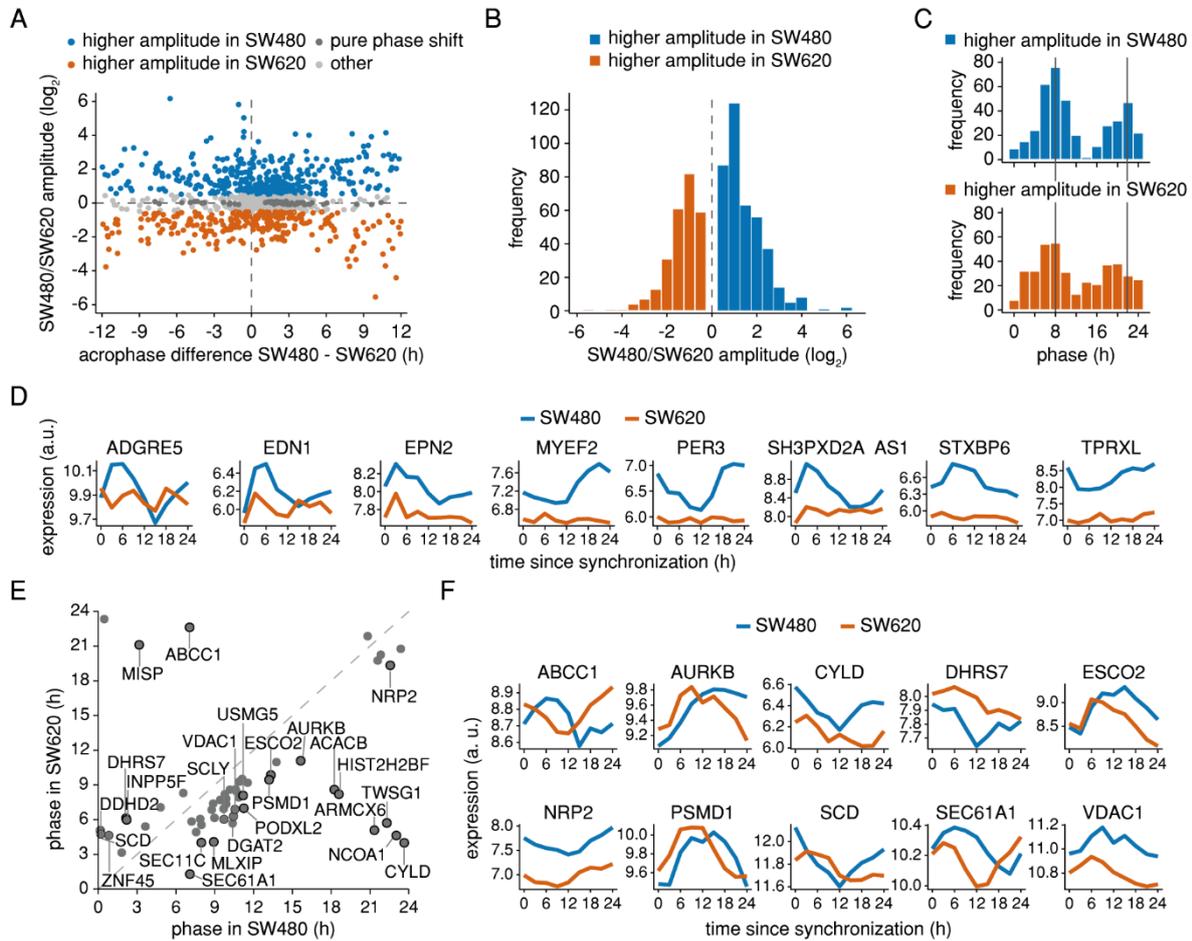


Figure 3-3: Differential rhythmicity analysis between rhythmic genes in SW480 and SW620 cells (microarray data). (A) Changes in rhythmicity (amplitudes and/or phases) from SW480 to SW620 cells of the genes that show 24-h rhythms in either one of the cell lines were estimated using the robust DODR method. The transcripts with BH-adjusted $p < 0.05$ were further divided in genes with a higher amplitude in SW480 cells (blue), genes with a higher amplitude in SW620 cells (orange), and genes with a pure phase shift (dark gray). (B) Log₂ amplitude changes from SW480 to SW620 cells for genes with a higher amplitude in either SW480 or SW620. (C) Phases for genes with higher amplitudes in SW480 cells (upper panel, blue) and SW620 cells (lower panel, orange). (D) Time-series expression in SW480 cells (blue) and SW620 cells (orange) for eight genes with the largest loss of oscillations in SW620 cells. (E) Phases in SW480 and SW620 cells for genes with pure phase shifts. Genes with absolute phase shifts > 3 h (black border) are labeled with the gene name. Selected transcripts represented in (F) are written in bold. (F) Time-series expression for selected phase-shifted genes in SW480 cells (blue) and SW620 cells (orange).

DODR p -values were computed for all 3606 pairs of 24-h rhythmic genes that oscillated in at least one of the two cell lines. The resulting 1005 genes with significant differential rhythmicity between the cell lines (BH-adjusted DODR $p < 0.05$) were further grouped in three sets (see subsection 2.3.3 for criteria): Genes with a higher amplitude in SW480 cells (39.5%), genes with a higher amplitude in SW620 cells (25.9%), and genes displaying a pure phase shift (5.6%) (**Figure 3-3A**). The remaining differentially rhythmic genes did not meet any of the criteria and were therefore not assigned to any group. In

line with the observed weaker oscillations of core clock genes in SW620 cells, a tendency for increased amplitudes in genes with higher amplitudes in SW480 cells compared to genes with higher amplitudes in SW620 cells could be observed (**Figure 3-3B**). Furthermore, genes with a higher amplitude in SW620 cells tended to have earlier phases than those with a higher amplitude in SW480 cells (**Figure 3-3C**), again suggesting a temporal shift in the activity of clock-controlled genes in the metastasis-derived cell line.

To gain mechanistic insights into possible causes for the altered rhythmicity between the circadian transcriptome of SW480 and SW620 cells, a search for enriched transcription factor binding sites among the differentially rhythmic genes sets was conducted. Promoter regions within 1000 bp of the TSSs were queried for overrepresented clock transcription factor binding sites (E-box, E'-box, D-box, and RORE). The E-box motif was found to be enriched for both the set of genes with amplitude gains in SW480 (adj. $p = 1.42e-8$) and SW620 cells (adj. $p = 0.0035$). Additionally, the RORE motif was enriched for genes with an amplitude gain in SW480 cells (adj. $p = 0.024$), raising the possibility that the loss of NR1D1 oscillations might in part be responsible for the diminished amplitudes in SW620 circadian gene expression. No clock transcription factor binding motif was found to be enriched for the set of phase-shifted genes.

Eight genes with an absolute \log_2 FC > 4 between the two cell lines were identified that all displayed amplitude gains and higher expression levels in SW480 cells (**Figure 3-3D**). The set includes the core-clock gene *PER3*, five other protein-coding genes (*ADGRE5*, *EDN1*, *EPN2*, *MYEF2*, *STXBP6*), and two non-coding RNA genes (*TPRXL*, *SH3PXD2A-AS1*). *MYEF2* is a paralog of the SF *HNRNPM* and acts as a suppressor factor in myelinating and erythroid cells (van Riel et al., 2012). *ADGRE5*, *EDN1*, and *STXBP6* are all involved in cell-cell adhesion (Davenport et al., 2016; Lenka et al., 2017; Sood et al., 2016), whereas *EPN2* is involved in clathrin-mediated endocytosis and acts as a suppressor of VEGF-mediated angiogenesis (Rahman et al., 2016). In a cancer context, angiogenesis describes the process of blood vessel growth that provides a tumor with oxygen and nutrients, allowing it to grow beyond a limited size and to metastasize to distant sites. The adhesive properties of cells are essential for the interaction of tumor cells with endothelial cells during angiogenesis (Farahani et al., 2014). The non-coding RNA genes *TPRXL* and *SH3PXD2A-AS1* have both been linked to cancer as well (Hao et al., 2018; Torres-Martín et al., 2015). It is thus conceivable that the combination of higher expression levels and stronger oscillations of angiogenesis- and other cancer-related genes contribute to differences in the phenotypes of primary tumor- and metastasis-derived CRC cells. 25 genes displayed an estimated pure phase shift > 3 h between the cell lines (**Figure 3-3E and F**). Several of them encode for membrane-related proteins (*DHRS7*, *NRP2*, *PSMD1*, *SCD*, *SEC61A1*) or are involved in transmembrane transport (*ABCC1*, *VDAC1*) while others are related to the cell cycle (*AURKB*, *CYLD*, *ESCO2*). *ABCC1* encodes for the protein MRP1 that plays an important role in conferring resistance to chemotherapeutic drugs in cancer cells (Kunická and Souček, 2014), making it a potential target for cancer stage-specific chronotherapeutical approaches that take the internal time of patients into account.

To further analyze whether the three types of differentially rhythmic genes between the cell lines are involved in biologically related processes or share molecular functions, a functional annotation analysis was conducted against the background of all 24-h rhythmic genes in either cell line (**Figure S 3**). Biological processes enriched for genes that oscillated in SW480 cells but lost their rhythmicity in the metastasis-derived cell line include chromatin silencing and modification, the regulation of mitotic nuclear division, and homophilic cell adhesion. Genes with higher amplitudes in the metastasis-derived cell line are associated with, i.e., the regulation of transcription from an RNA polymerase II promoter and the regulation of smooth muscle cell proliferation. Cell adhesion properties determine the cell's abilities to escape from its site of origin, invade other tissues, and ultimately metastasize and thus play an important role in cancer progression (Okegawa et al., 2004). Likewise, the smooth muscle layer is often lacking in tumor vessels and the apoptosis of vascular smooth muscle cells facilitates the intravasation of CRC cells into the bloodstream and promotes metastasis (Li et al., 2017). The loss of oscillations in the expression of genes involved in cell cycle regulation and cell adhesion and the gain of rhythmicity for genes regulating smooth muscle cell proliferation might enable CRC cells to develop properties that are favorable for proliferation and the formation of metastases. Biological processes enriched for the set of phase-shifted genes include replication fork processing, DNA binding, and retinoid X receptor (RXR) binding. Interestingly, the RXR element has previously been reported as a binding motif of circadian TFs (Westermarck and Herzog, 2013) and RXRs interact with several elements of the positive limb of the circadian TTFL. For instance, the interaction of RXR α with CLOCK and NPAS2 hinders the transcriptional activation of clock gene expression (McNamara et al., 2001). RXR α further modulates the canonical Wnt/ β -catenin signaling pathway, which is deregulated in CRC due to over-phosphorylation of RXR α that leads to hyperactivated β -catenin signaling (Ruan et al., 2017). A phase shift of oscillating RXR-binding genes might thus be responsible for a change in the temporal regulation of cancer-relevant downstream pathways.

Overall, the analysis of the circadian transcriptome of SW480 and SW620 cells demonstrates that the dysregulated core clock observed in SW620 cells likely leads to the differential rhythmicity of clock-controlled genes, several of which are associated with hallmarks of tumor progression. However, it remains to be elucidated how the loss or gain of rhythms in gene expression and the change of phases contribute to the cellular phenotype in comparison to overall changes in gene expression between the primary tumor- and the metastasis-derived cell line.

3.1.2 Functional Annotation of Circadian Genes in CRC Cells Reveals Phase-shifted Pathways

To explore the functional relevance of the complete circadian transcriptome of the CRC cell lines, 24-h rhythmic genes were grouped into four clusters, respectively, based on their temporal expression in SW480 (**Figure S 4A**) and in SW620 cells (**Figure S 5A**). The clusters contain 200–800 rhythmic genes each and exhibit different phases of peak expression. A functional annotation of the eight gene sets revealed various important pathways and biological processes known to be under circadian control,

such as the cell cycle (e.g., G1/S transition of mitotic cell cycle and DNA replication initiation) and diverse metabolic processes (e.g., canonical glycolysis, fructose and mannose metabolism, starch and sucrose metabolism) (**Figure S 4B** and **Figure S 5B**). Other biological processes enriched for the 24-h rhythmic gene sets are less well known to be controlled by the clock, e.g., cell-cell adhesion and mRNA splicing. Interestingly, some biological processes were identified as rhythmic in both cell lines, but the associated gene clusters showed expression peaks at different times of the circadian day. For instance, in SW480 cells, the spliceosome pathway is enriched for the gene set that peaks at ~9 h after synchronization, whereas in SW620 cells, it is associated with the gene set peaking at about 6 h after synchronization. Other processes such as cell-cell adhesion are enriched for more than one cluster in the same cell line, indicating that while these processes are enriched for 24-h rhythmic genes, the associated genes have diverse temporal expression patterns.

The latter case demonstrates that the binning of genes with continuous peak times into discrete clusters prior to the functional annotation does not necessarily allow for the detection of pathways that are exclusively enriched for genes with similar peak times. To specifically identify biologically related gene sets with temporally coordinated transcription, the periodicity of the data (i.e., the peak expression phases of the rhythmic genes) needs to be taken into account as part of the enrichment analysis. In analogy to gene set enrichment analysis where each set member contributes to the score of the entire set, PSEA makes use of the peak phases to identify biological processes enriched for sets of 24-h rhythmic genes with non-uniformly distributed phases and to estimate the circular mean phase of the entire set. Applying PSEA for the complete sets of 24-h rhythmic genes, 15 and 36 KEGG pathways were found to be significantly temporally coordinated ($q < 0.01$) in SW480 and SW620 cells, respectively (**Figure 3-4**). With the exception of the systemic lupus erythematosus pathway, all phase-clustered circadian pathways in SW480 cells had a circular mean phase between 4 and 12 h after synchronization. In SW620, the circular mean phases were even more closely clustered between ~5 and ~10 h after synchronization. Despite the bimodal distributions of peak phases of 24-h rhythmic genes (see subsection 3.1.1), no pathways with circular mean phases between 16 and 4 h were identified in either cell line. In the primary tumor cell line SW480, the analysis further revealed a significant temporal orchestration for important signaling pathways, including the Jak-STAT signaling pathway (4.9 h) and the Toll-like receptor pathway (6.8 h) that were not identified in the previous functional annotation of circadian gene clusters. In the metastatic SW620 cells, several cancer-associated pathways were found to be temporally coordinated, including small/non-small cell lung cancer (7.4 h/7.3 h), pancreatic cancer (7.4 h), and oxidative phosphorylation (8.8 h). Several temporally coordinated pathways are shared between both cell lines, including protein export (circular mean phase of 7.4/6.2 h after synchronization in SW480/SW620 cells), antigen processing and presentation (7.6/6.6 h), and the spliceosome (9.4/8.5 h), as well as nucleotide metabolic processes (purine metabolism: 9.8/8.3 h; pyrimidine metabolism: 10.0/8.6 h), DNA replication (11.1/9.2 h), several DNA repair-related pathways (base excision repair: 10.3/9.2 h; mismatch repair: 11.1/9.4 h, homologous recombination: 11.1/10.0 h), and the cell cycle (11.4/8.0 h).

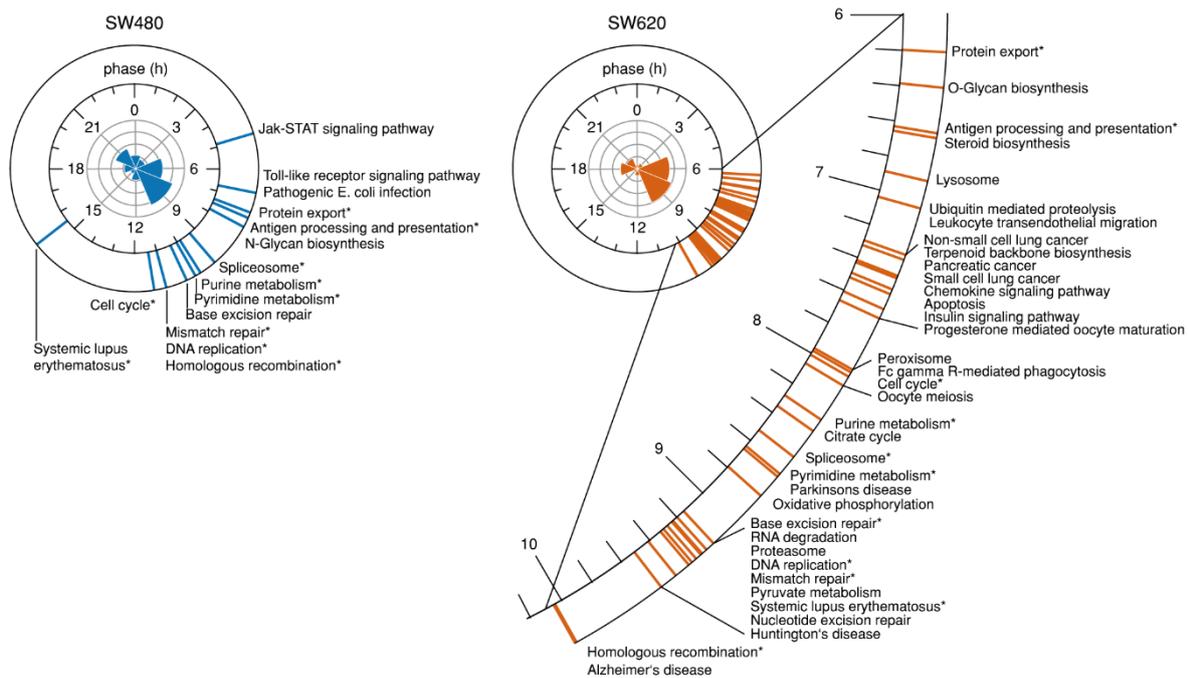


Figure 3-4: Phase-clustered circadian pathways in SW480 and SW620 cells (microarray data). The circular axis (blue lines: SW480; orange lines: SW620) represents the average phase of all 24-h rhythmic genes from a KEGG pathway. Pathways that are temporally coordinated in both cell lines are marked with an asterisk. The circular histograms show the 3-h phase bins of the 24-h rhythmic genes.

To further assess temporal shifts from SW480 to SW620 cells on the level of phase-clustered circadian pathways, PSEA was applied to identify processes with significant phase shifts between the cell lines. Based on the set of 597 genes that oscillated in both cell lines, 16 pathways were identified as significantly phase-shifted with shifts between 0.5 and 2.5 h (**Figure 3-5A**). Interestingly, the spliceosome was among the resulting phase-clustered pathways whose associated genes showed a small but significant shift of peak expression (0.76 h, $q < 0.01$) (**Figure 3-5A**). In both cell lines, the sets of the circadian spliceosome-associated genes were unimodally distributed, though the composition of the sets and the peak times of the genes differed (**Figure 3-5B**). Any phase shift smaller than 3 h has to be considered carefully because the small sampling resolution does not allow the exact determination of phases. However, since the result is not based on the expression of a single gene but rather on the peak phases of several genes, it can be taken as an indication that the rhythmic orchestration of the spliceosome undergoes subtle but significant changes during CRC metastasis.

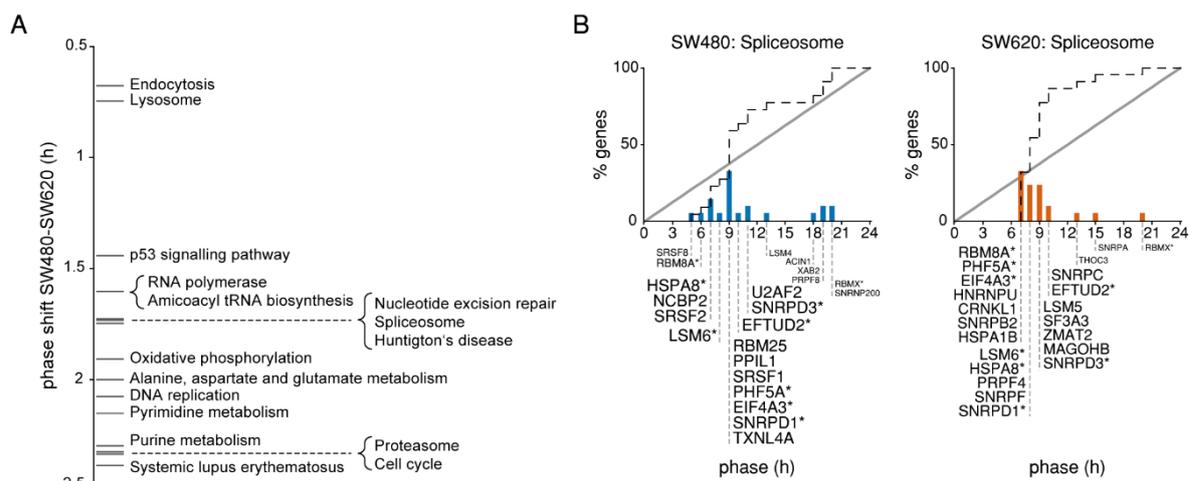


Figure 3-5: Spliceosome-associated genes are phase-shifted between SW480 and SW620 cells (microarray data). (A) Circadian pathways with a significant phase shift between SW480 and SW620 cells. (B) Phases of spliceosome-associated genes in SW480 cells (left panel) and SW620 cells (right panel). The dotted line depicts the empirical cumulative phase distribution of spliceosome-associated genes. The solid line depicts the uniform distribution. The height of the bars corresponds to the percentage of 24-h rhythmic genes in the spliceosome pathway that peak at a certain time after synchronization. A greater contribution of a gene to the overall phase clustering of the pathway is represented by a larger font size.

3.1.3 Detection of Differential Rhythmicity of Splicing-related Genes

In order to further investigate and compare the temporal expression patterns of genes involved in the splicing process between the two CRC cell lines, a list of 254 splicing-related genes was compiled (see subsection 2.3.10) (**Table S 1**). The list consists of components recruited at different complexes of the spliceosome, different snRNPs, other spliceosomal complexes and proteins (PRP19, RES complex, EJC/mRNP, and LSm proteins), as well as hnRNPs, SR proteins, and other splicing regulators. In SW480 cells, 46 (~18%) of the splicing-related genes from the list showed 24-h rhythms in expression (**Figure 3-6A**) compared to 35 genes (~14%) in SW620 cells (**Figure 3-6B**). The percentages are nearly twice the respective percentages of 24-h rhythmic genes identified on the whole transcriptome-level (see subsection 3.1.1). As expected from the results of the functional annotation and the circadian pathway analysis (see subsection 3.1.2), the phase distributions of the 24-h rhythmic splicing-related genes differed between SW480 and SW620 cells: 30 out of the 35 24-h rhythmic spliceosome components and splicing regulators in SW620 cells peaked between 6 and 10 h after synchronization whereas in SW480 cells, the phases were distributed throughout the circadian day with a peak of expression around 9 h after synchronization. The 24-h rhythmic splicing-related genes identified in SW480 and SW620 cells belong to diverse spliceosomal complexes and splicing-related protein classes (**Figure 3-6C and D**). However, neither phases nor amplitudes were clustered for specific spliceosome complexes or for splicing genes encoding for proteins from the same class or family, indicating a complex temporal regulation of splicing throughout the circadian day.

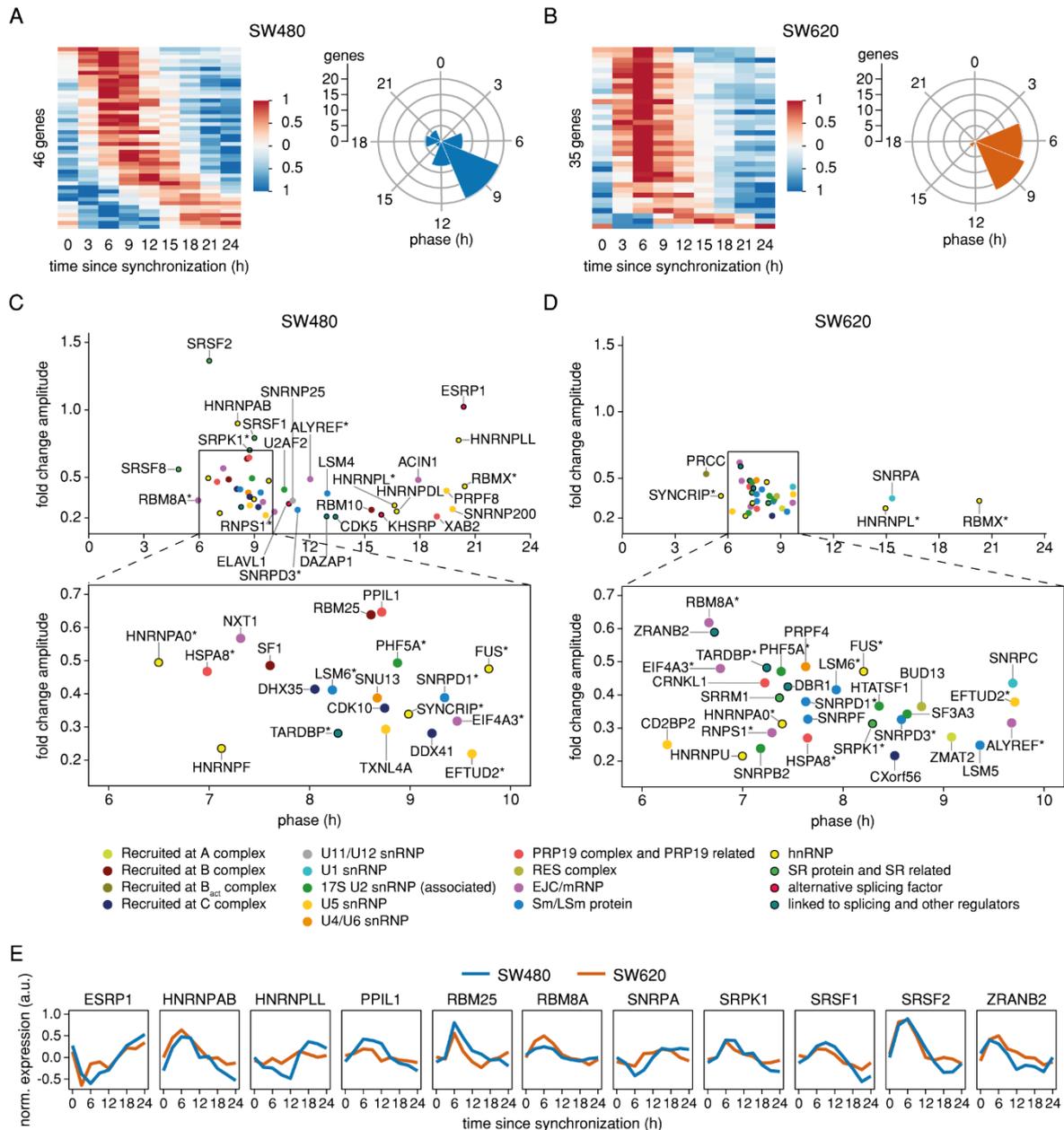


Figure 3-6: Splicing-related genes display 24-h rhythms in SW480 and SW620 cells (microarray data). Median-normalized, phase-ordered expression heatmap (left panel) and 3-h phase bins (right panel) of 24-h rhythmic splicing-related genes in (A) SW480 (blue) and (B) SW620 cells (orange). Phases and amplitudes of 24-h rhythmic splicing-related genes in (C) SW480 and (D) SW620 cells. Genes displaying 24-h rhythms in both cell lines are marked by an asterisk. The genes are color-coded according to their protein class/family or their recruitment to the spliceosome complex. (E) Median-normalized time-series expression of candidate robustly circadian splicing-related genes in SW480 cells (blue) and SW620 cells (orange). Shown are genes with a FC amplitude ≥ 1.5 that were identified to be 24-h rhythmic in at least one of the cell lines.

The rhythmic splicing-related genes identified in SW480 cells include several well characterized SFs (*SF1*, *SRSF1*, *SRSF2*, *SRSF8*, *ESRP1*, *U2AF2*) and hnRNP-encoding genes (*HNRNPA0*, *HNRNPAB*, *HNRNPDL*, *HNRNPF*, *HNRNPL*, *HNRNPLL*, *FUS*). 17 splicing-related genes were identified as 24-h rhythmic in both cell lines: five hnRNP-encoding genes (*FUS*, *HNRNPA0*, *HNRNPL*, *RBMX*, *SYNCRIP*), one SR protein encoding gene (*SRPK1*), three genes encoding for Sm and LSm proteins (*LSM6*, *SNRPD1*, *SNRPD3*), four

EJC/mRNP components (*ALYREF*, *EIF4A3*, *RBM8A*, *RNPS1*), one element of the PRP19 complex (*HSPA8*), as well as three other spliceosome components and splicing regulators (*EFTUD2*, *PHF5A*, *TARDBP*). Eleven genes had an estimated FC amplitude ≥ 1.5 in at least one of the cell lines and thus represent good candidates for robustly circadian spliceosome components and SFs (**Figure 3-6E**). They encode for three SR proteins (*SRSF1*, *SRSF2*, *SRPK1*), two hnRNPs (*HNRNPAB*, *HNRNPLL*), a component of the U1 snRNP (*SNRPA*), two RNA binding motif (RBM) proteins (*RBM25*, *RBM8A*), a PRP19 complex-related protein (*PPIL1*) and two SFs (*ESRP1*, *ZRANB2*). When comparing the time-series expression of splicing-related genes that oscillated in at least one of the cell lines, it becomes apparent that many of them had higher expression levels in SW620 cells (**Figure S 6**).

16 out of the 64 genes that oscillated in either one or both of the cell lines were identified as significantly differentially rhythmic (DODR adj. $p < 0.05$) (see subsection 3.1.1). Among them, there were four genes with higher amplitudes in SW480 cells (*ACIN1*, *ALYREF*, *NXT1*, *RBM10*), five genes with amplitude gains in SW620 cells (*CXorf56*, *DBR*, *EIF4A3*, *HNRNPU*, *PRCC*), and two phase-shifted genes (*FUS*, *PHF5A*). However, several of the genes that were not identified as differentially rhythmic also showed strong differences in (rhythmic) expression. For instance, *HNRNPLL* and *NXT1* displayed robust oscillations in SW480 cells but none in SW620 cells, and *PPIL1* and *SRSF1* had diminished amplitudes in SW620 cells compared to SW480 cells. Moreover, many splicing-related genes showed similar but slightly shifted rhythms in SW480 and SW620 cells. There seems to be a tendency for rhythmic splicing-related genes peaking 2–3 h earlier in SW620 cells than in SW480 cells, e.g., for *ESRP1*, *FUS*, *SF3A3*, and *SNRNP25*. However, these putative phase shifts in circadian transcription are hard to quantify with confidence due to the relatively low sampling resolution of 3 h. Nonetheless, it can be concluded that several splicing-related genes show robust oscillations in CRC cell lines that tend to differ between primary tumor and metastasis-derived cells.

3.1.4 Identification of Differential and Rhythmic Splicing Events in SW480 and SW620 Cells

Differences in the expression levels of SFs are known to cause changes in AS of their target genes. To investigate whether the observed differential rhythmicity of splicing-related genes leads to differences in the outcome of AS between the two CRC cell lines, the FIRMA method was applied (Purdom et al., 2008). The algorithm allows for the detection of candidate cassette-type alternative exons for single samples without replicates (see subsection 1.2.4). FIRMA scores are calculated on exon-level based on the residuals of the observed probe-level expression from the estimated expression of the RMA model. A high positive FIRMA score is an indication for exon inclusion whereas a low negative FIRMA score indicates exon skipping.

The first part of the analysis aimed at the identification of “static” changes of AS between the cells lines that showed high differences across the pairwise comparison of all time points. For this purpose, the paired differences in FIRMA scores between SW480 and SW620 cells were computed for identical time points. Of 401,839 expressed exons, 299 exons (0.0007%) of 256 genes had significantly different (BH-adjusted $p < 0.001$) FIRMA scores with high differences (mean absolute FIRMA score \log_2 FC ≥ 1) and

were selected as candidate exons for differential AS events between the two cell lines (**Figure 3-7A**, **Table S 2**). 158 (52.8%) of the candidate exons had higher mean FIRMA scores in SW480 cells and 141 (47.2%) had higher scores in SW620 cells, indicating that cassette exons were included or excluded to approximately equal shares between the cell lines. A functional annotation analysis revealed various GO biological processes to be enriched for the set of candidate alternatively spliced genes ($p < 0.05$) (**Table S 3**). Many of the processes are associated with tumor progression and metastasis, such as the regulation of cell proliferation and apoptotic processes, response to drugs, cell adhesion, and angiogenesis (**Figure 3-7B**). Interestingly, axon guidance, the most enriched term for the candidate gene set, has recently been reported as a potential mediator of colon cancer metastasis (Rokavec et al., 2017). The top 30 candidate alternatively spliced genes (mean absolute FIRMA score $\log_2 FC \geq 1.35$) have 19 exons whose FIRMA scores were higher in SW480 cells and 11 exons whose FIRMA scores were higher in SW620 cells (**Figure 3-7A**). Several of the candidate genes have well-characterized AS variants that play a role in invasion and metastasis, e.g., *NCAM1* (**Figure S 7A**) which encodes for a neural adhesion molecule capable of heterotypic and homotypic binding and has been described as a tumor suppressor (Roesler et al., 1997). The *NCAM1* gene consists of at least 19 exons and has three known major isoforms (Huerta et al., 2001). A splicing defect between exons 12 and 13 that leads to the production of truncated transcripts and the loss of long isoform NCAM1-L-180 has previously been reported for SW620 but not for SW480 cells and is linked to clinically aggressive CRC (Huerta et al., 2001; Roesler et al., 1997). The candidate exon identified in the FIRMA analysis is likely part of the short NCAM1-204 transcript that contains a retained intron and is annotated as a processed transcript without an open reading frame in the Ensembl database. It is currently unclear how its loss might affect cellular adhesion properties in SW620 cells and whether the observed differential AS event corresponds to the splicing defect previously found to be responsible for the lost expression of NCAM1-L-180 between SW480 and SW620 cells. Nonetheless, the identification of *NCAM1* as a candidate gene for differential splicing events between SW480 and SW620 cells seems to be well supported by literature. Further candidate genes with AS events that play a well-studied role in cancer include *CD44* (**Figure S 7B**) and *FGFR2* (**Figure S 7C**) that are both involved in EMT (see subsection 1.2.3). The candidate differential splicing event identified for *CD44* overlaps with the genetic region of the variant *CD44* exons v1 – v10 and likely corresponds to v4/exon 9 (**Figure S 7B**). The FIRMA score is higher in SW480 compared to SW620 cells, indicating a loss of the exon in the metastasis-derived cell line which is the case for several *CD44* isoforms and might impact on the attachment properties of the resulting molecule. Two candidate differential AS events were detected for *FGFR2* (**Figure S 7C**), one with a higher FIRMA score in SW480 cells and the other in SW620 cells. The latter event might correspond to the variable inclusion of the *FGFR2* cassette exons III3b and IIIc which is responsible for determining the epithelial or mesenchymal character of the cell.

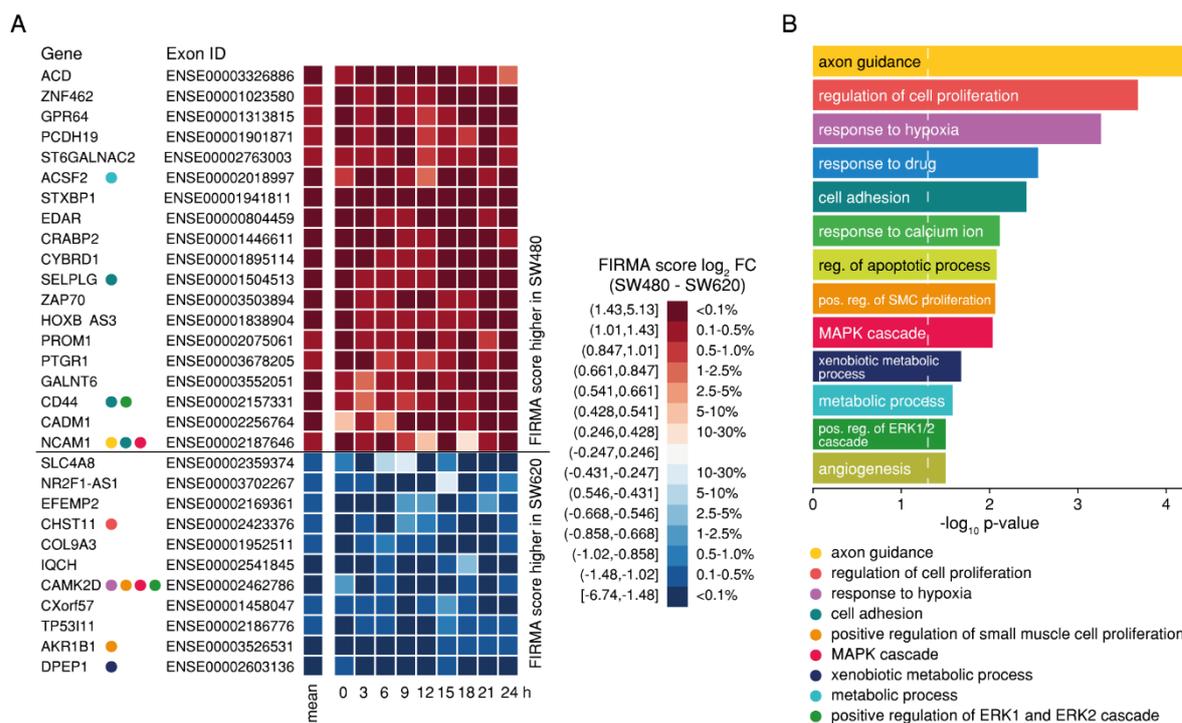


Figure 3-7: Candidate differential AS events between SW480 and SW620 cells (microarray data) are associated with cancer-relevant processes. (A) Paired differences in FIRMA scores of the top 30 candidate exons with differential AS events in SW480 and SW620 cells. The color scale is not evenly spaced but based on the percentiles of all FIRMA scores for all probesets and samples after filtering. (B) Selected GO terms (biological processes) enriched for the 256 candidate genes with differential AS events.

The second part of the analysis aimed at the detection of putative circadian AS in the CRC cell lines. Exons with 24-h rhythmic variations in FIRMA scores were identified individually for each cell line (RAIN $p < 0.05$ and peak-trough-ratio of FIRMA score ≥ 2), resulting in 59 candidate rhythmic alternative exons in SW480 cells (59 genes) (**Figure S 8A**) and half as many in SW620 cells (29 exons, 29 genes) (**Figure S 8B**). Peaks of FIRMA scores could be observed across the whole circadian cycle for both cell lines, indicating that the putative circadian regulation of AS is not limited to a specific time-of-day which is in line with the diverse phases identified for 24-h rhythmic splicing related-genes (**Figure 3-6**). Many of the genes with candidate rhythmic exons also oscillated on gene-level in one or both of the cell lines, suggesting that circadian AS is often concurrent with circadian transcription. Only for two of the candidate genes, the calcium/calmodulin-dependent protein kinase encoding *CAMK2G* and the DNA damage gene *NBR1*, 24-h rhythmic AS were detected in both cell lines. In addition to *CAMK2G* and *NBR1*, seven other candidates were selected for a closer inspection due to their interesting roles in the circadian system, splicing, and cancer (**Table 1**).

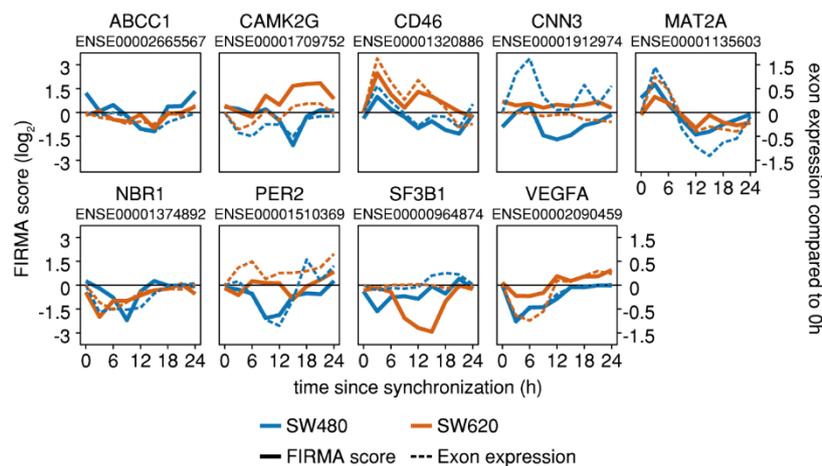


Figure 3-8: Comparison of selected candidate circadian AS events between SW480 and SW620 cells (microarray data). (A) Circadian FIRMA scores (solid line) and exon expression profiles (dotted line) for selected exons with 24-h rhythmic AS events in SW480 (blue) and SW620 cells (orange).

Six of the genes had candidate circadian AS events in SW480 cells (*ABCC1*, *CD46*, *CNN3*, *MAT2A*, *PER2*, and *VEGFA*) and one gene (*SF3B1*) in SW620 cells. In most cases, the rise and fall of the rhythmic FIRMA score was similar to variations in the exon expression (**Figure 3-8**). Most FIRMA scores also showed variations over time in the respective other cell line, with the exception of the *CNN3* exon which had a 24-h rhythmic FIRMA score in SW480 cells and a constant score in SW620 cells. Interestingly, the circadian FIRMA score profiles of the exons from *CAMK2G* and *NBR1* peaked at different times, suggesting a phase-shift of rhythmic AS between SW480 and SW620 cells for these genes. *NBR1* has previously been reported as a putative CLOCK target gene in CRC (Alhopuro et al., 2010). *CD46*, *MAT2A*, and *CNN3* have been found to be overexpressed in CRC and *CNN3* has furthermore been proposed as a marker for CRC lymph node metastasis (Chen et al., 2007; Cho et al., 2016; Nakarai et al., 2015). Though not much is known concerning the consequences of AS of these genes, it is conceivable that the production of functionally distinct isoforms at a specific time-of-day might play a part in the progression of cancer cells, e.g., by evading circadian immune system signals.

Different splice variants of the candidate gene *ABCC1* have previously been characterized (Grant et al., 1997), some of which have been found to confer a drug resistance phenotype in ovarian cancer (He et al., 2004). Moreover, circadian transcriptional oscillations of *ABCC1* have been observed in Caco-2 cells (Ballesta et al., 2011) and diverse tissues in mouse, rat, and monkey (Ozturk et al., 2017). In the present study, *ABCC1* was found to display phase-shifted 24-h rhythmic expression patterns between SW480 and SW620 cells (**Figure 3-3F**). The dysregulation of the core clock system in SW620 cells could be the reason for the observed shift in the oscillatory expression and the putative circadian AS of *ABCC1* which might impact on drug resistance in colon cancer patients (Hu et al., 2016). While circadian AS has previously been reported for core clock genes in mice, the candidates did not include any member of the *Per* gene family (McGlinicy et al., 2012). The candidate circadian AS exon of *PER2* identified in SW480 cells is part of the C-terminal region of the PER2S splicing variant that was found to be preferentially localized in the nucleolus (Avitabile et al., 2014) (see subsection 1.3.1). Interestingly, the disturbance of nucleolar function in response to cellular stress is linked to cancer via regulation of the p53 pathway:

The disruption of the nucleolar structure triggers the release of the tumor suppressor ARF and several nucleolar riboproteins into the nucleoplasm where they bind to MDM2, thereby preventing MDM2-mediated p53 degradation, leading to cell cycle arrest or apoptosis (Suzuki et al., 2012). The loss of *PER2S* circadian AS could thus potentially lead to alterations in the nucleolar function due to disrupted circadian timing of its components which may impact on p53 activation and ultimately promote carcinogenesis.

Table 1: Overview of candidate genes with circadian AS events in SW480 and SW620 cells (microarray data).

Gene	Role in cancer	Effects of AS	Observations in SW480/SW620
ABCC1	overexpression confers resistance to chemotherapeutic drugs in cancer cell lines (He et al., 2004)	higher frequency of AS in ovarian tumors than in matched normal tissues; splice variants confer drug resistance (He et al., 2004)	phase-shifted 24-h rhythmic gene expression; 24-h rhythmic AS in SW480
CAMK2G	stabilization of MYC protein via phosphorylation; promotion of T cell lymphoma (Gu et al., 2017)	aberrant splicing in patients with myotonic dystrophy (Perfetti et al., 2014)	24-h rhythmic AS of the same exon in both cell lines (phase-shifted)
CD46	overexpressed in CRC (Cho et al., 2016)	unknown	24-h rhythmic gene expression in SW620; 24-h rhythmic AS in SW480; similar pattern in SW620
CNN3	overexpressed in CRC; marker for CRC lymph node metastasis (Nakarai et al., 2015)	unknown	24-h rhythmic AS in SW480
MAT2A	overexpressed in CRC; promotes growth of colon cancer cell lines (Chen et al., 2007)	unknown	24-h rhythmic gene expression in SW480; 24-h rhythmic AS in SW480
NBR1	mediator of migration via autophagy-dependent focal adhesion turnover (Kenific et al., 2016); CLOCK target gene in CRC (Alhopuro et al., 2010)	unknown	24-h rhythmic gene expression in SW480; 24-h rhythmic AS of the same exon in both cell lines (phase-shifted)
PER2	overexpression induces apoptotic cell death in murine and human cancer cells via downregulation of <i>c-Myc</i> , <i>Bcl-XL</i> , and <i>Bcl-2</i> and upregulation of <i>p53</i> and its target gene <i>Bax</i> (Hua et al., 2006; Oda et al., 2009)	nucleolar localization of PER2S splicing variant (Avitabile et al., 2014)	diminished amplitude of oscillations in SW620 cells; 24-h rhythmic AS in SW480
SF3B1	cancer-associated mutations promote alternative branch-point usage of target genes (Alsafadi et al., 2016)	unknown	24-h rhythmic AS in SW620
VEGFA	key mediator of angiogenesis (the formation of new blood vessels) to supply tumors with oxygen and nutrients, allowing them to grow and metastasize (Carmeliet, 2005)	overexpression of pro-angiogenic VEGFxxx isoforms promote angiogenesis and downregulation of anti-angiogenic VEGFxxxb isoforms in cancer (Kaida et al., 2012)	24-h rhythmic AS in SW480; similar pattern in SW620

SF3B1 is the only SF for which a candidate AS event was identified. The affected exon is located in the HEAT repeat domain 8 which encodes for part of the canonical isoform of the gene. While only little is known regarding the effects of AS of *SF3B1*, its knockdown in human myeloid cell lines has been reported to result in over 500 differentially regulated splicing variants of nearly 400 genes, including

TP53 (Dolatshad et al., 2015). A similar downstream effect could potentially result from the changes in circadian AS of *SF3B1* observed in this study. However, further experiments are necessary to elucidate a possible functional outcome of the proposed mechanism. *VEGFA* is one of the most well-studied examples of how AS can affect cancer-relevant processes. Depending on the splice site choice, *VEGFA* isoforms can act either as anti-angiogenic or as a pro-angiogenic proteins (see subsection 1.2.3). Furthermore, *vegfa* has been identified as a direct transcriptional target of *Bmal1* in zebrafish where VEGF-dependent developmental angiogenesis is under control of the circadian clock (Jensen and Cao, 2013). So far, however, there have been no reports of circadian AS of *VEGFA*. The candidate circadian *VEGFA* exon detected in this study is part of the transcripts *VEGFA-224* and *VEGFA-225* which are both annotated as protein-coding in the Ensembl database. However, it is unclear whether the encoded proteins act in pro- or anti-angiogenic way. Alternative splice forms of *VEGFA* are regulated by SR proteins, e.g., *SRSF1*, which has been reported to act as an oncogene and which displays an altered rhythmic phenotype in the metastasis-derived cell line *SW620* when compared to *SW480*.

Overall, the analysis of both static and circadian AS events in the CRC cell line model yielded several candidate genes with interesting implications in tumor-related biological processes, supporting the hypothesis that the circadian clock might be involved in the regulation of AS in cancer progression. However, it still remains to be elucidated whether the observed changes are truly caused by elements of the clock and whether they are functional on the protein-level.

3.2 Comparison of Circadian Microarray and RNA-Seq Data of CRC Cell Lines

For the second part of the study, the circadian transcriptomes of the human CRC cell lines *SW480* and *SW620* were additionally profiled by RNA-seq. In the first two parts of the chapter, the RNA-seq data is compared to the microarray data to investigate how well 24-h rhythmic genes and associated circadian parameters are replicated across platforms. In addition, the two time-series are normalized and concatenated to form a longer time-series. A manuscript based on the results described in these chapters is currently in preparation. In the third part of the chapter, differentially rhythmic splice variants of the same gene are detected, and the resulting candidate circadian AS events are compared to those gained from the microarray data. The results described in the third part of the chapter have for the most part been published in El-Athman et al. (2019). Text passages and figures of the publication have been updated and adapted for this thesis.

3.2.1 Correlation of Gene Expression between Microarray and RNA-seq Data

The time-series microarray data (E-MTAB-5876) and RNA-seq data (E-MTAB-7779) of the cell lines *SW480* and *SW620* have been produced under identical conditions in the same lab with a sampling resolution of 3 h. However, the two datasets do not cover the same time range: The microarray dataset consists of nine samples that were taken from 0 to 24 h after synchronization of the cells by medium

change, whereas the eleven RNA-seq samples were taken from 12 to 42 h after synchronization. Accordingly, the datasets share five sampling time points (12, 15, 18, 21, and 24 h). The microarray dataset covers a full circadian cycle with one additional time replicate of the first time point (0 h = 24 h) and the RNA-seq datasets covers 1.25 circadian cycles with time replicates of three time points (12 h = 36 h; 15 h = 39 h; 18 h = 42 h). For this reason, it is not only of interest to investigate how circadian gene sets might differ between the two technological platforms, but also how the datasets could be normalized and concatenated to form a longer time-series.

The correlation between gene expression values gained from the different platforms was compared between datasets and different preprocessing methods. While the RMA method is commonly used for the preprocessing of microarray data, no standard method for the preprocessing of circadian RNA-seq data has yet been established (Li et al., 2015). For this reason, a tripartite pipeline was used for the alignment and quantification of the raw RNA-seq reads: Two of the chosen pipelines employed the STAR aligner (Dobin et al., 2013) for the alignment of the raw reads to the human genome. Additionally, the STAR genome coordinates were transformed to the transcriptome space to get a transcriptome alignment. Subsequently, genome-aligned reads were quantified by featureCounts (*STAR + featureCounts*), resulting in gene counts, whereas transcriptome-aligned reads were quantified by Salmon in alignment-based mode (*STAR + Salmon*), resulting in transcript counts. In the third pipeline, reads were directly quantified by Salmon in quasi-mapping-based mode without a prior alignment step (*Salmon*). R packages tximport (Soneson et al., 2015) and edgeR (Robinson et al., 2010) were used to normalize the Salmon transcript counts and to summarize them to gene-level abundances.

The average mapping rates of STAR ($81.7\% \pm 2.55$) and Salmon ($87.1\% \pm 0.95$) were comparable and consistent across time points and cell lines (**Figure S 9A**). An expression cutoff of at least 0.5 CPM per gene on average over all time points resulted in 13904 to 16478 expressed genes in SW480 cells and 13495 to 15933 expressed genes in SW620 cells, depending on the preprocessing method (**Figure S 9B**). For both cell lines, *STAR + featureCounts* yielded the highest number of expressed genes, followed by *STAR + Salmon*. Independent of the chosen method, the number of expressed genes in SW620 was lower than in SW480 cells. Analogously, different cutoffs for expressed genes were compared for the microarray data, resulting in 21798 expressed genes with a minimum mean \log_2 RMA-preprocessed intensity value of 3 across all time points in SW480 cells and 21764 genes in SW620 cells (**Figure S 9C**). Since there are no universal cutoffs to exclude genes with low expression values either for microarray intensity values or RNA-seq counts, the chosen cutoffs are arbitrary and could be adjusted to gain expressed gene sets of comparable sizes for the different platforms. For the present study, this is not necessary because subsequent correlation and rhythmicity analyses were focused on genes that were found to be commonly expressed across platforms.

After the preprocessing, the resulting RNA-seq expression values of commonly expressed genes were compared to each other and to the microarray expression values in a gene-wise correlation analysis. For each cell line, gene expression was compared in a pairwise manner for samples taken at identical

time points, considering only those time points that were shared between all methods (12–24 h) (**Figure 3-9**). Expression values determined by the three RNA-seq analysis pipelines had high mean Pearson correlation coefficients among each other (> 0.9) which were similar across cell lines. The highest correlation was observed between *STAR + Salmon* and *Salmon* for both cell lines, indicating that the alignment/quasi-mapping step had a smaller impact on the resulting gene expression values than the subsequent quantification of the mapped reads. The mean correlation coefficients between the three RNA-seq pipelines and the microarray data were very similar across methods and cell lines, ranging from 0.72 to 0.74. Thus, it can be concluded that the choice of the RNA-seq analysis pipeline did not have a large influence on the correlation coefficients, which were primarily determined by the chosen platform. Nonetheless, since the highest correlation of the three pipelines with the microarray data was observed for *STAR + Salmon* for both cell lines, the *STAR + Salmon* gene expression data was considered the best approximation to the microarray data and was used for subsequent analyses (hereafter denoted as “RNA-seq data”).

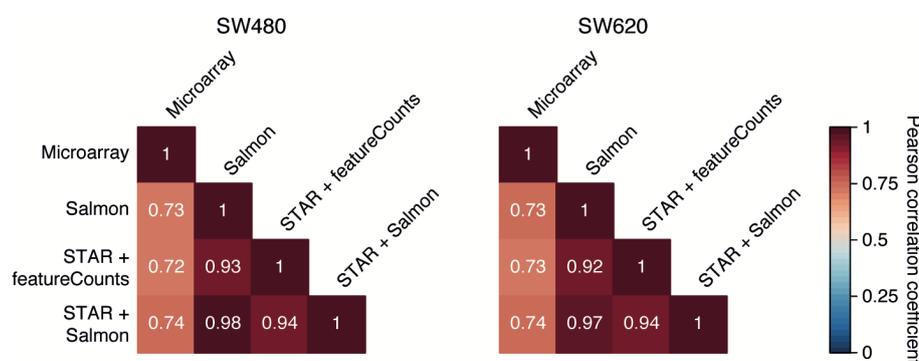


Figure 3-9: Correlation of gene expression between circadian microarray and RNA-seq data of human CRC cell lines. Mean Pearson correlation coefficients between expression values determined by different platforms and different downstream methods for SW480 cells (left panel) and SW620 cells (right panel). Expression values were determined based on microarray data and RNA-seq data (*Salmon*, *STAR + featureCounts*, *STAR + Salmon*). Shown are the mean Pearson correlation coefficients between gene expression values determined for identical time points (12–24 h since synchronization).

3.2.2 Normalization and Cross-Platform Concatenation of Circadian Datasets

Since the microarray and RNA-seq datasets contain overlapping but not identical time-series, a concatenation of the data could yield a longer time-series of nearly two circadian cycles (42 h) with a consistent sampling resolution. However, independent of the correlation of gene expression, the range of the abundance measure of the two methods is different: While microarrays measure fluorescence intensities and yielded RMA-preprocessed expression values in a \log_2 -scaled range from 0 to 13.5, RNA-seq quantified gene expression in CPM in a greater dynamic \log_2 -range from -8 to 15 (**Figure 3-10A**). To circumvent this dataset shift and make expression values comparable across platforms, TDM-normalization of the RNA-seq data was applied (Thompson et al., 2016). The TDM method transforms an RNA-seq dataset to have a similar distribution as a microarray dataset while keeping inter-observation

dependencies intact. The distribution of the TDM-transformed RNA-seq expression data closely resembled that of the microarray data for both cell lines (**Figure 3-10A**). However, when comparing mean expression levels over the complete time-series of individual genes, a tendency for a lower mean expression for lowly expressed genes and a higher mean expression for highly expressed genes was observed for the RNA-seq data which is likely due to its higher dynamic range (**Figure 3-10B**). Nonetheless, for the bulk of genes, the gene expression values of both platforms were comparable after the TDM-transformation of the RNA-seq data, thus allowing for a concatenation of the datasets.

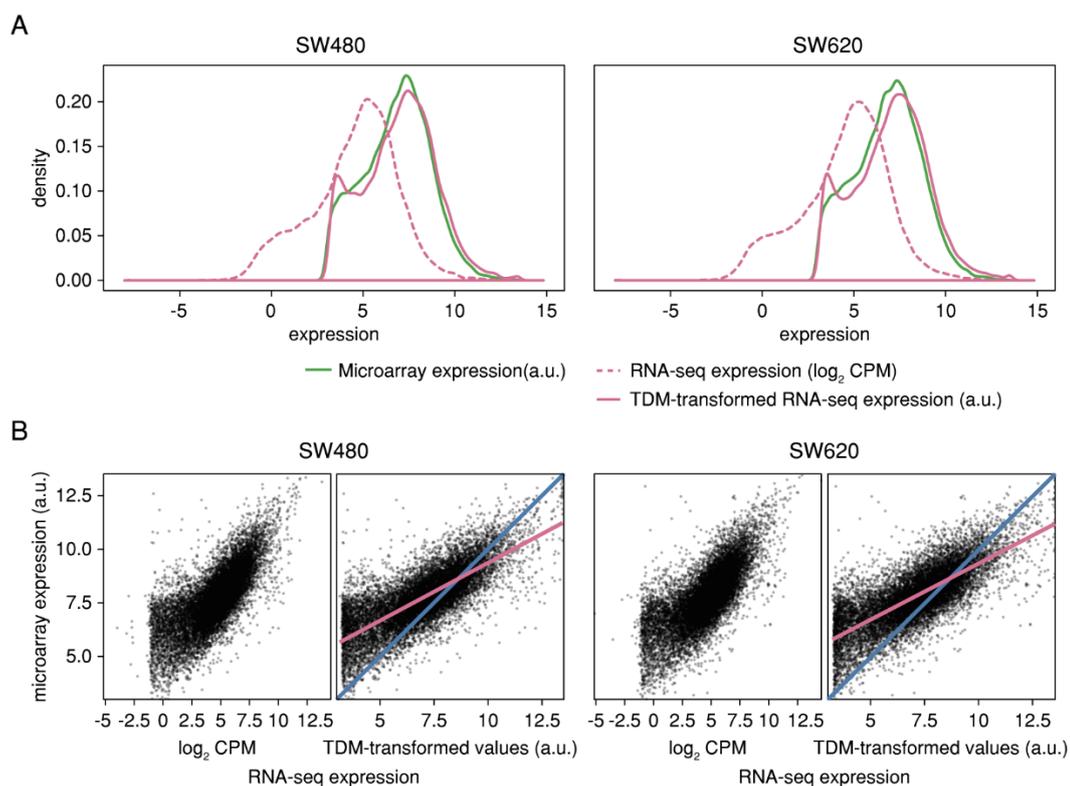


Figure 3-10: RNA-seq expression data of CRC cell lines before and after TDM-normalization. Density of gene expression values in SW480 cells (left panel) and SW620 cells (right panel) of microarray data (green line), and RNA-seq data before (dashed pink line) and after TDM-transformation (solid pink line). (B) Scatterplots comparing mean gene expression across all time points between the microarray (y-axis) and the RNA-seq dataset (x-axis) before (respective left panel) and after TDM-transformation (respective right panel) for SW480 cells (two leftmost panels) and SW620 (two rightmost panels). The pink line represents the linear regression fitted to the data and the blue line represents the ideal regression.

In the next step, 24-h rhythmic gene sets for all three datasets were identified using both the parametric harmonic regression method and the non-parametric RAIN algorithm with an input period of 24 h (**Figure S 10**). For the concatenated datasets, shared time points were treated as replicate measurements of the same time-series. For a q -value cutoff of 0.05 and a minimum relative amplitude of 0.1, harmonic regression yielded more 24-h rhythmic genes than RAIN for both cell lines and all three datasets, with the exception of the SW620 microarray dataset. Using RAIN, a higher number of rhythmic genes was identified based on the microarray data than based on the concatenated data, whereas the trend was reversed when using harmonic regression, suggesting that the latter method is better suited

for the detection of rhythmicity in the concatenated datasets. For this reason, only the rhythmic gene sets identified by harmonic regression were used for further analyses.

When comparing the overlaps between the 24-h rhythmic gene sets identified for either of the two different platforms and the concatenation of both time-series, it becomes apparent that not only the number but also the identity of rhythmic genes and their circadian parameters differed greatly between the methods (**Figure 3-11**). The intersection of all three datasets only amounted to 110 (2.3%) genes commonly identified as 24-h rhythmic in SW480 cells and 52 (1.5%) in SW620 cells (**Figure 3-11A**). For both cell lines, only very few genes were commonly identified as rhythmic by the microarray and RNA-seq data but not by the concatenated data, indicating that the concatenation of the two time-series successfully captured the circadian transcriptome identified by either one of the platforms. Additionally, many genes were identified as significantly 24-h rhythmic in the concatenated data that were not detected in either one of the two shorter time-series, possibly capturing false negative circadian genes. Contrary to expectations and previous results gained from the microarray data (see subsection 3.1.1), more 24-h rhythmic genes were identified for SW620 cells (728) than for SW480 cells (564) based on the RNA-seq data. This is particularly surprising because the time-series expression of core clock genes in the RNA-seq data closely resembled that of the previously analyzed microarrays (**Figure 3-1A**). Core clock genes had smaller amplitudes or were non-rhythmic in SW620 cells compared to SW480 cells and their mean expression tended to be lower, with the exception of *NPAS2* and *PER2* (**Figure S 11**). Based on these observations which indicate a dysregulated core clock system in the metastasis-derived cell line, one would expect fewer clock-controlled genes on the transcriptome-level. In line with this assumption, a higher number of 24-h rhythmic genes were detected in SW480 (3426) than in SW620 cells (1892) based on the concatenated dataset.

The comparison of the phase distributions of the 24-h rhythmic gene sets revealed another discrepancy between the microarray and the RNA-seq data: While the microarray data for SW480 cells showed a bimodal peak distribution with one phase peak at ~8 h and the second at ~21 h, the RNA-seq data displayed a small phase peak at ~1 h and a large peak at ~16 h, indicating a shift of ~-5 h in overall circadian transcription measured for the same cell line using two different platforms (**Figure 3-11B**). Interestingly, the concatenation of the data led to phase peaks at ~10 h and ~22 h, which resembled the distribution of the microarray data but were shifted in the opposite direction. For SW620 cells, the RNA-seq phase peaks were likewise shifted, but only by about ~-3 h and the concatenation of the data resulted in a phase distribution that resembled a superposition of both distributions (**Figure 3-11B**). The distributions of the relative amplitudes of 24-h rhythmic genes detected for the different platforms and cell lines were similar, though there were small but significant differences between all of the sets of each cell line ($p < 0.05$ determined by two-sided Mann-Whitney test adjusted for multiple comparisons using the BH method) (**Figure 3-11C**). For both cell lines, the RNA-seq data set yielded slightly higher mean relative amplitudes (SW480: 0.197, SW620: 0.209) than the microarray (SW480: 0.178, SW620: 0.17) or the concatenated data (SW480: 0.174, SW620: 0.161), which can probably be ascribed to the higher dynamic range of the RNA-seq data.

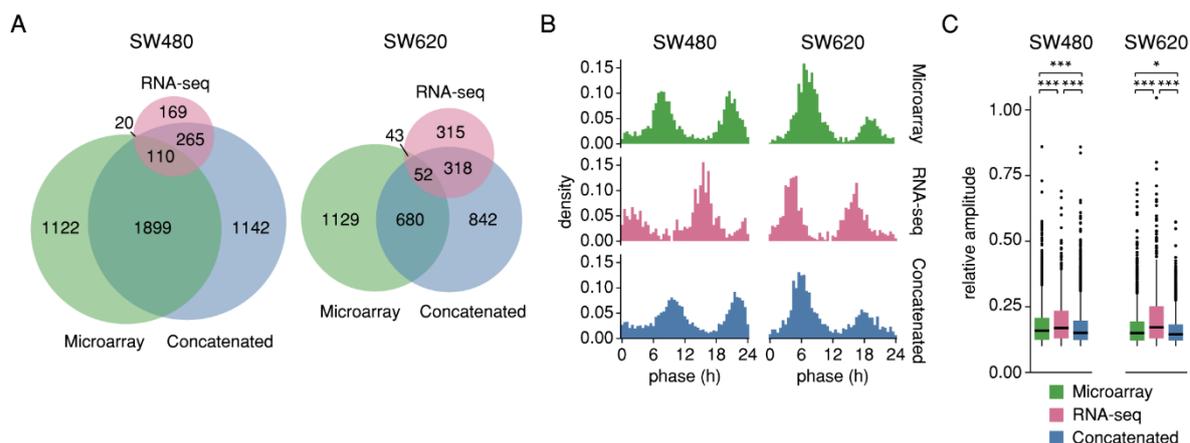


Figure 3-11: Comparison of 24-h rhythmic gene sets of SW480 and SW620 cells identified based on microarray and RNA-seq data. (A) Venn diagrams representing the intersections between the 24-h rhythmic gene sets identified based on microarray data (green), RNA-seq data (pink) and the concatenated expression of both platforms (blue) in SW480 cells (left panel) and SW620 cells (right panel). (B) Phase distributions of the 24-h rhythmic genes sets identified based on microarray data (green), RNA-seq data (pink) and the concatenated expression of both platforms (blue) in SW480 cells (left panel) and SW620 cells (right panel). (C) Boxplots of the relative amplitudes of the 24-h rhythmic genes sets identified based on microarray data (green), RNA-seq data (pink) and the concatenated expression of both platforms (blue) in SW480 cells (left panel) and SW620 cells (right panel). Significance was determined by a two-sided Mann-Whitney test adjusted for multiple comparisons using the BH method (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$). Rhythmic genes and their parameters were determined by harmonic regression ($q < 0.05$ and relative amplitude ≥ 0.1).

In the next step, circadian parameters were compared in a gene-wise manner for all genes that were identified to be 24-h rhythmic in the same cell line in at least two of the datasets. The phases estimated based on microarray and RNA-seq data of SW620 cells had a lower absolute circular Pearson correlation coefficient (0.05) than those estimated for SW480 cells (0.26) and were not significantly correlated ($p = 0.58$) (**Figure 3-12A**). The correlation between the phases estimated based on the concatenated data and the microarray data was much higher (circular Pearson correlation coefficient (r) = 0.99 for SW480 cells and 0.85 for SW620 cells) and significant ($p < 0.05$) (**Figure 3-12B**), as was the correlation between the concatenated phases and the RNA-seq phases (circular $r = 0.95$ for SW480 cells and 0.87 for SW620 cells) (**Figure 3-12C**). The gene-wise correlation of relative amplitude values was highest between the microarray and the concatenated data ($r = 0.83$ for SW480 cells and 0.77 for SW620 cells) (**Figure 3-12D**). The correlation was considerably lower between RNA-seq data and concatenated data ($r = 0.32$ for SW480 cells and 0.48 for SW620 cells), where a trend for higher amplitudes could be observed for the RNA-seq data (**Figure 3-12F**). The lowest correlation was observed between amplitudes of the RNA-seq and the microarray data for both cell lines ($r = 0.28$ for SW480 cells and 0.20 for SW620 cells) (**Figure 3-12E**).

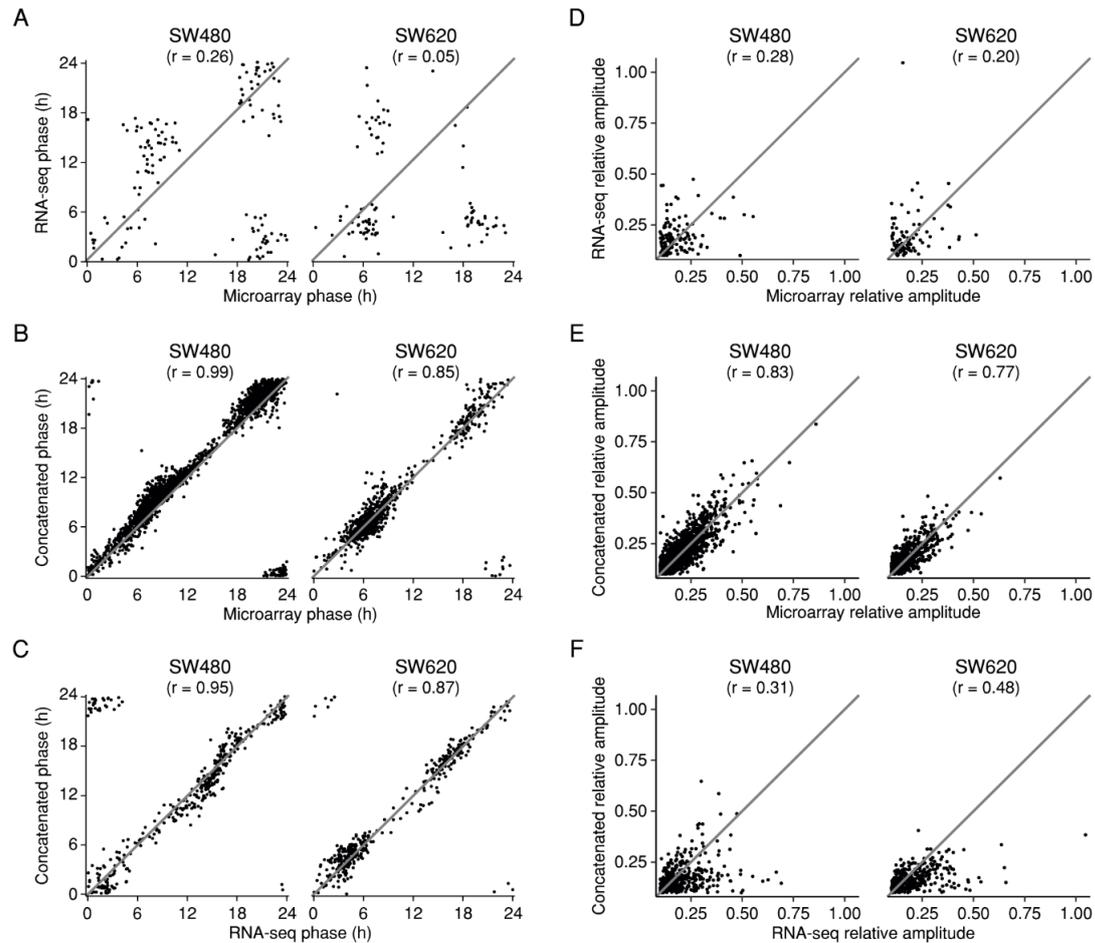


Figure 3-12: Gene-wise comparison of circadian parameters of 24-h rhythmic genes identified based on microarray and RNA-seq data of SW480 and SW620 cells. Scatterplots and circular Pearson correlation coefficients of phases estimated based on (A) microarray data and RNA-seq data, (B) microarray data and concatenated data, and (C) RNA-seq and concatenated data for SW480 cells (left panels) and SW620 cells (right panels). Scatterplots and Pearson correlation coefficients of amplitudes estimated based on (D) microarray data and RNA-seq data, (E) microarray data and concatenated data, and (F) RNA-seq and concatenated data for SW480 cells (left panels) and SW620 cells (right panels).

The preceding comparisons of circadian parameters were conducted only between genes that were commonly identified as 24-h rhythmic between the different methods. In particular between microarray and RNA-seq data, these gene intersections were quite small (**Figure 3-11A**). As previously observed when comparing differential rhythmicity between SW480 and SW620 cells based on the microarray data, genes can have similar oscillatory expression despite not being identified as significantly rhythmic in one of the conditions (**Figure S 2C**). For this reason, the time-series expression of all genes identified as rhythmic in the respective other platform was compared by expression heatmaps (**Figure S 12**). While many of the genes identified to be 24-h rhythmic in SW480 cells based on the microarray data had similar expression patterns in the RNA-seq data for the shared time points (12–24 h), they lacked a distinct second peak or trough of expression in later time points. In the opposite case (microarray expression of genes identified as 24-h rhythmic in RNA-seq data) similar expression patterns could again be observed for the shared time points for SW480 cells, whereas earlier time points displayed a disordered pattern. For SW620 cells, expression patterns of rhythmic gene sets were even

more dissimilar between platforms, suggesting that many of the genes identified as 24-h rhythmic for the metastatic cell line are not truly circadian. In contrast, the expression heatmaps ordered according to the phases estimated for the 24-h rhythmic genes of the concatenated data showed circadian patterns for both platforms and cell lines. Moreover, five out of the ten 24-h rhythmic genes with the lowest p -values in SW480 cells identified based on the concatenated data were core clock genes (*BMAL1*, *NR1D1*, *NR1D2*, *PER2*, and *PER3*). Overall, core clock genes displayed robust oscillations with similar phases and amplitudes across the whole time-series (**Figure 3-13**), indicating that the normalization and concatenation of the data is a suitable method for the detection of circadian rhythms in transcription across platforms.

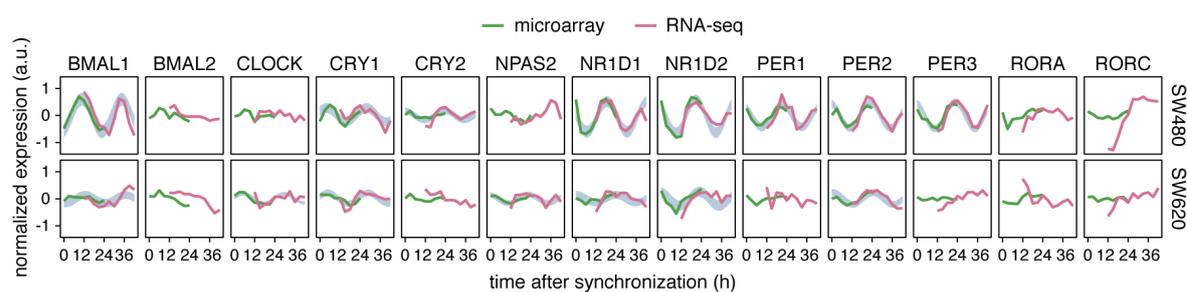


Figure 3-13: Expression of core clock genes in the concatenated microarray and RNA-seq time-series of SW480 and SW620 cells. Mean-normalized time-series expression of core clock genes in SW480 cells (top row) and SW620 cells (bottom row) based on the concatenated data. Microarray expression values are represented by green lines and RNA-seq expression values by pink lines. The blue area marks the confidence area of the harmonic regression fitted to the concatenated data for 24-h rhythmic genes ($q < 0.05$ and relative amplitude ≥ 0.1).

3.2.3 Detection of Differentially Rhythmic Phase-shifted Splice Variants in CRC Cell Lines

In the next step, the RNA-seq data was used for the detection of putative differentially rhythmic splice variants. Transcript-level expression values were estimated using STAR for the alignment and Salmon in alignment-mode for the quantification as previously described (*STAR + Salmon*). In the following, splice variants detected on transcript-level are referred to as ‘transcripts’, whereas transcribed genes detected on summarized gene-level are referred to as ‘genes’. Applying a RAIN p -value cutoff of 0.05 and a relative amplitude ≥ 0.1 , 1,343 24-h rhythmic genes and 6,003 24-h rhythmic transcripts were identified in SW480 cells and 1,934 24-h rhythmic genes and 5,395 24-h rhythmic transcripts in SW620 cells (**Figure S 13A**). Interestingly, the ratio of rhythmic transcripts to rhythmic genes was higher for SW480 cells, indicating that despite the higher number of genes detected as rhythmic in SW620, the number of splice variants was higher in the primary tumor-derived cell line (**Figure 3-14A**). On average, a 24-h rhythmic gene had 1.02 rhythmic transcripts in SW480 cells and 0.93 in SW620 cells. In both cell lines, about 70% of 24-h rhythmic genes displayed at least one rhythmic transcript (**Figure S 13B**). However, there were also many genes identified as 24-h rhythmic on transcript-level only, possibly indicating a rhythmic regulation of AS that leads to discrepancies between rhythmic features identified on gene- and transcript-level.

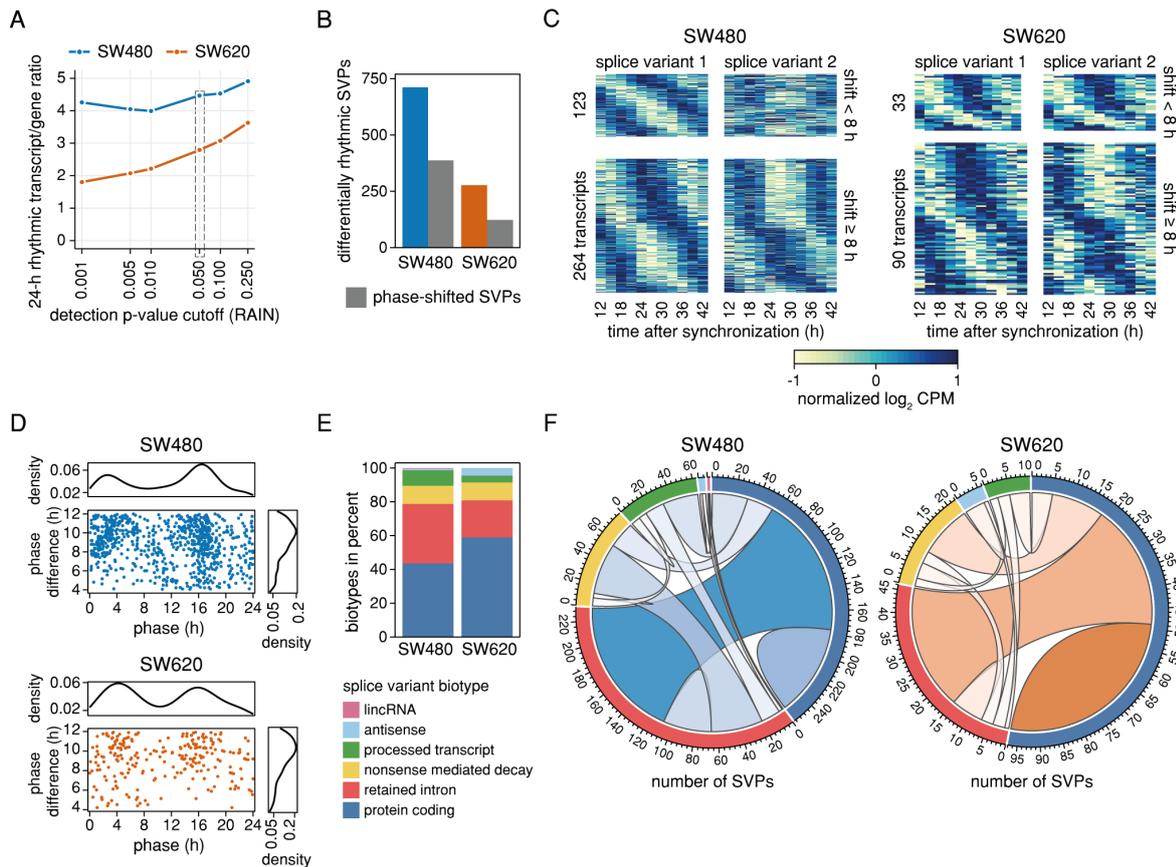


Figure 3-14: Differentially 24-h rhythmic phase-shifted splice variants of the same gene in human CRC cell lines. (A) Ratio of 24-h rhythmic transcripts and genes for SW480 (blue) and SW620 (orange) for different RAIN p -value cutoff levels. (B) Number of 24-h rhythmic differentially rhythmic SVPs in the CRC cell lines. The proportion of phase-shifted SVPs is marked in gray. (C) Phase-sorted expression heatmaps of phase-shifted SVPs in SW480 (left panel) and SW620 (right panel). Each row represents one SVP sorted according to the phases of splice variant 1. The heatmaps are further subdivided according to the length of the phase shift between the splice variants in a pair. (D) Phase distribution and pairwise phase-difference of phase-shifted differentially 24-h rhythmic SVPs in SW480 cells (upper panel, blue) and SW620 cells (lower panel, orange). (E) Splice variant biotypes of 24-h phase-shifted SVPs in percent. (F) Chord diagrams representing the biotypes of SVPs for SW480 (left panel) and SW620 (right panel).

An analysis pipeline was developed to identify splice variants of the same gene that displayed phase-shifted rhythms in expression on the transcript-level in the same cell line and might thus represent distinct splice isoforms that are produced at different times of the circadian cycle. To detect these cases of differential rhythmicity, the DODR method was applied. Each 24-h rhythmic transcript was compared to all other 24-h rhythmic transcripts of the same gene in the same cell line in a pairwise manner (see subsection 2.3.7). For SW480 cells, 711 differentially rhythmic splice variant pairs (SVPs) were identified, compared to 277 pairs in SW620 cells (**Figure 3-14B**). Differential rhythmicity detected by DODR can be due to changes in either amplitude or phase of the rhythmic features, or both. Assuming that the observed differences in splice variants of the same gene are due to circadian-regulated AS, they should be shifted in phase but have similar amplitudes. Accordingly, the differentially rhythmic SVPs were filtered for pairs with an amplitude ratio < 2 that were phase-shifted by at least 4 h. The phase shift between two variants was defined as the smallest difference between their two phases (< 12 h).

The filtering resulted in 387 phase-shifted 24-h rhythmic SVPs (belonging to 280 genes) in SW480 cells and 123 (belonging to 100 genes) in SW620 cells (**External data file 1**). Of these SVPs, more than two thirds had a phase shift greater than 8 h. When representing the expression of the candidate transcripts of an SVP in heatmaps ordered according to the phase of one of the splice variants, it becomes apparent that the expression of the second variant tended to be in anti-phase for SVPs with phase shift greater than 8 h (**Figure 3-14C**). In SW480 cells, there was a peak of expression for phase-shifted splice variants at ~16 h and a smaller one at ~3 h, whereas the distribution in SW620 cells was bimodal with peaks at ~4 h and ~16 h (**Figure 3-14D**). Only six genes with phase-shifted SVPs (*ASPSCR1*, *HDAC6*, *HNRNPH1*, *LRRC75A-AS1*, *NDUFV1*, and *PAX8-AS1*) were shared between SW480 and SW620 cells, suggesting that the rhythmic regulation of splice variant expression might be cancer stage-specific.

About 44% of the splice variants included in SVPs in SW480 cells and 59% in SW620 cells were annotated as protein-coding (**Figure 3-14E**). In SW480 cells, 31.8% of the SVPs consisted of a protein-coding transcript and a transcript with a retained intron, followed by SVPs of two protein-coding transcripts (20.7%) (**Figure 3-14F**). In SW620 cells, the largest set were SVPs of two protein-coding transcripts (39%), followed by SVPs consisting of a protein-coding transcript and a transcript with a retained intron (25.5%) (**Figure 3-14F**). For many human alternative splice variants, it is not yet known whether they are functional and whether their putative function differs from that of the canonical isoform. Accordingly, most gene enrichment and functional annotation tools are currently only possible for genes and not for individual splice variants. For this reason, a GO enrichment analysis was conducted on gene-level for all candidate genes with phase-shifted SVPs (**Table S 4**). Enriched biological processes in SW620 cells included RNA splicing, post-transcriptional regulation of gene-expression, and the regulation of translation (BH-adjusted $p < 0.05$) (**Figure 3-15B**). For SW480 cells, the analysis revealed no significantly enriched processes when applying the same cutoff. However, for a more relaxed cutoff of $p < 0.002$, lipoprotein-related processes, the regulation of the androgen receptor pathway, and RNA splicing were found among the biological processes enriched for genes with phase-shifted SVPs (**Figure 3-15A**). Interestingly, membrane androgen receptors expressed in colon cancer cells have been reported to induce strong apoptotic responses via inhibition of pro-survival signals and blocking migration (Gu et al., 2011). Constitutively active splice variants of androgen receptors contribute the development of resistance to androgen deprivation therapy in human prostate tumors (Liu et al., 2014).

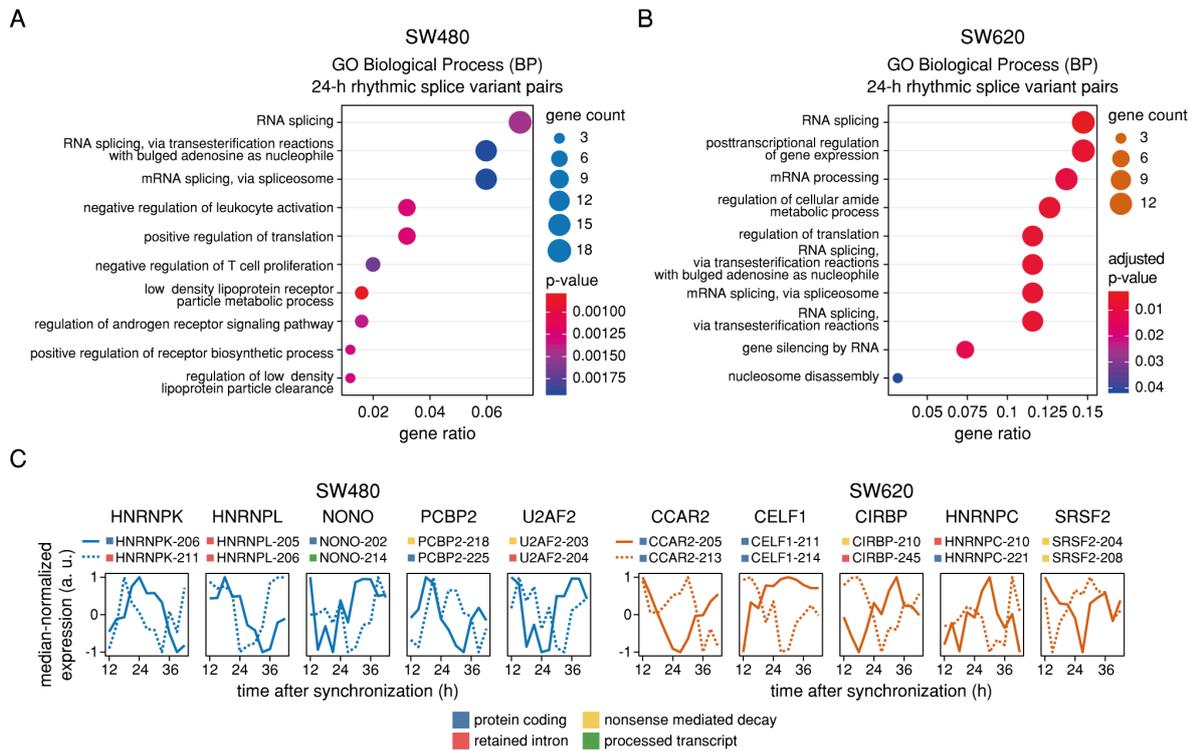


Figure 3-15: Phase-shifted splice variants in SW480 and SW620 cells are linked with the splicing process. Enriched GO terms (Biological Process – BP) for the sets of genes that exhibited 24-h rhythmic phase-shifted SVPs in (A) SW480 and (B) in SW620 cells. For visualization, GO terms were manually curated to remove redundant categories. The complete lists of results can be found in **Table S 4**. Median-normalized expression (in a range from -1 to 1) of 24-h rhythmic phase-shifted SVPs from selected splicing-related genes in SW480 (blue) and SW620 (orange).

The prevalence of splicing processes for both cell lines suggest that splicing-related genes themselves might be spliced in a time-of-day-dependent way. To find out which splicing-related genes could be affected, the previous list of 254 spliceosome components and splicing regulators (see subsection 3.1.3) was joined with a list of 404 human SF genes from Seiler et al. (2018), resulting in an extended list of 426 human splicing-related genes (**Table S 1**). 23 genes from the extended list were among the candidates in SW480 cells (*CLK4*, *DDX39B*, *DDX3X*, *DDX5*, *FMR1*, *GRSF1*, *HNRNPH1*, *HNRNPK*, *HNRNPL*, *INTS6L*, *MOV10*, *NONO*, *NXF1*, *PCBP2*, *RBM3*, *RBM39*, *SRRT*, *SRSF11*, *TAF15*, *TTC14*, *U2AF2*, *U2SURP*, and *ZNF131*) and nine were found for SW620 cells (*CCAR2*, *CELF1*, *CIRBP*, *DDX17*, *HNRNPC*, *HNRNPH1*, *PABPC1*, *RBM4*, *SRSF2*) (**Figure 3-15C**). Only four of the splicing-related genes with phase-shifted SVPs in SW480 cells (*RBM39*, *SRRT*, *SRSF11*, and *U2AF2*) (**Figure S 14A**) and none in SW620 cells were found to be oscillating on gene-level in the RNA-seq data (RAIN $p < 0.05$ and relative amplitude ≥ 0.1) (**Figure S 14B**). Likewise, none of them were found to 24-h rhythmic in the concatenated data (harmonic regression $q < 0.05$ and relative amplitude ≥ 0.1) (**Figure S 14C**), indicating that their rhythmic expression might be masked by the phase-shifted expression of their individual transcripts.

In the next step, the sets of genes with phase-shifted SVPs were compared to the set of candidate genes with circadian AS events identified based on the microarray data (see subsection 3.1.4). Only two candidate genes, *CAMK2G* and *PPP2R5C*, were shared across platforms and analysis methods for SW480 cells and none for SW620 cells. Alternatively spliced forms of *CAMK2G* have been identified in patients

with myotonic dystrophy (Perfetti et al., 2014) (**Table 1**). Based on the microarray data, *CAMK2G* has previously been identified as a candidate gene with circadian AS events for both SW480 and SW620 cells but with phase-shifted FIRMA score profiles (see subsection 3.1.4). The PPP2R5C phosphatase is involved in the modulation of liver metabolism (Cheng et al., 2015) and further acts as a tumor suppressor via dephosphorylation of p53 (Nobumori et al., 2013). Five different splice variants of human *PPP2R5C* have been detected both in samples from healthy individuals and from leukemia patients, though not much is known concerning their individual function (Zheng et al., 2011). Another gene, *INTS11* (also known as *CPSF3L*), that encodes for a part of the Integrator complex, was found to be a shared candidate between the SW480 microarray data and the SW620 RNA-seq data. Conversely, *SH3D19* was identified as a shared candidate between the SW620 microarray data and the SW480 RNA-seq data.

In light of the important role of AS in cancer, the candidate phase-shifted SVPs were further searched for known cancer-relevant isoforms. Interesting candidates were found for *ANKHD1* in SW480 and *MYO1C* in SW620 (**Figure 3-16**). *ANKHD1* is an ankyrin repeat and KH domain-containing protein for which two functional AS variants have been characterized: In contrast to the canonical isoform, VBARP-L (11 exons) and VBARP-S (9 exons) both contain only a single ankyrin repeat motif and lack the signature KH domain. (Miles et al., 2005). The 24-h rhythmic transcript ANKHD1-205 in SW480 encodes for VBARP-L, whereas the ~10.5 h phase-shifted transcript ANKHD1-210 consists of eight exons that encode for the KH domain and do not overlap with the exons of ANKHD1-205. *ANKHD1* is overexpressed in acute leukemia (Traina et al., 2006) and the VBARP isoforms have been reported to have an anti-apoptotic effect and to play a role in cell survival pathways (Miles et al., 2005). *MYO1C* is a member of the unconventional myosin gene family and produces three AS isoforms through N-terminal splicing, which differ in their functions and their nucleo-cytoplasmic partitioning (Zattelman et al., 2017). The 24-h rhythmic transcript MYO1C-202 in SW620 encodes for the isoform MYO1C^C, which localizes predominantly to the plasma-membrane and is involved in cell migration and signal transduction. The ~7.4 h phase-shifted transcript MYO1C-203 encodes for the isoform MYO1C^B, also known as nuclear myosin I or MYO1C¹⁶, which localizes mostly to the nucleus where it is involved in transcription, mRNA maturation, and chromatin remodeling (Zattelman et al., 2017). Interestingly, the third isoform of MYO1C, MYO1C^A, is involved in prostate cancer cell migration (Maly et al., 2017) and it has been speculated that alterations of MYO1C at the plasma membrane might affect the metastatic potential of tumor cells (Venit et al., 2016). Notably, the expression patterns of *ANKHD1* and *MYO1C* splice variants differ between the cell lines (**Figure 3-16**), indicating that they might also be specific for different stages of tumor progression.

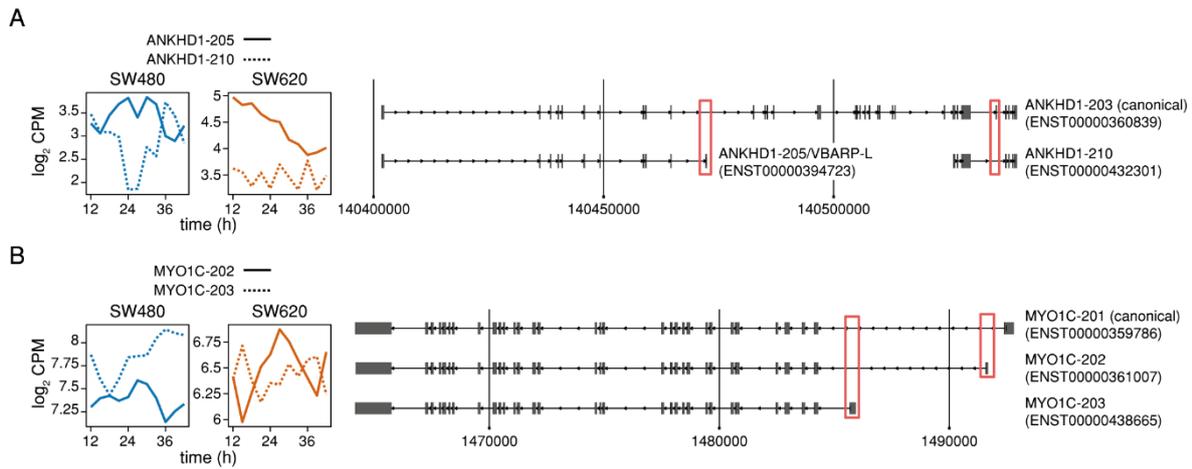


Figure 3-16: Candidate genes with phase-shifted splice variants are linked with apoptosis and migration. Expression (\log_2 CPM) and visualization of the genomic structure of (A) the transcripts ANKHD1-205 and ANKHD1-210 and (B) MYO1C-202 and MYO1C-203 in SW480 (blue) and SW620 (orange). Differences in the exonic composition are marked with red rectangles.

Altogether, the search for phase shifted SVPs has revealed several candidate genes which might be alternatively spliced in a time-of-day-dependent way in human CRC cell lines, some of which are associated with cancer-relevant processes, such as apoptosis and migration. Thus, it is conceivable that a temporal regulation of functionally distinct isoforms might enable CRC cells to produce protein isoforms at specific times of the day that further promote their malignant transformation.

3.3 Analysis of Rhythmic Splicing Events in Mammalian Tissues

In the third chapter, potential circadian rhythms in AS are analyzed in healthy mammalian tissues to find out whether the hypothesized rhythmic regulation of AS is conserved across species and to further elucidate its biological functions. As a representative model, two previously published mammalian circadian transcriptome datasets were chosen: A microarray dataset of twelve organs from male mice, sampled every 2 h for two circadian cycles (Zhang et al., 2014) and an RNA-seq dataset of 64 tissues and brain regions from olive baboons (**Table S 5**), sampled every 2 h for one circadian cycle (Mure et al., 2018). The higher sampling resolution of the datasets additionally allows for the analysis of ultradian rhythms with a period of 12 h (see subsection 1.1.4). In the first part of the chapter, 12-h and 24-h rhythms in gene expression and putative rhythmic changes in AS are identified for murine organs. In the second and third part, transcript-level oscillations in expression are compared to gene-level oscillations, followed by the detection of phase-shifted splice variants of the same gene in baboon tissues. In the last part of the chapter, circadian oscillations in the expression of splicing-related genes are compared between the two species. The results described in this chapter have for the most part been published in El-Athman et al. (2019). Text passages and figures of the publication have been updated and adapted for this thesis.

3.3.1 Identification of Putative Ultradian and Circadian AS Events in Murine Organs

The multi-organ mouse microarray dataset from Zhang et al. (2014) was re-analyzed for this study, starting with preprocessing of the raw data to gain expression values of genes for all individual time points, followed by an analysis of rhythmicity to identify rhythmic features and their rhythmic parameters. Gene-level expression values for the murine microarray data were calculated based on RMA-preprocessed intensity values for transcript clusters. A harmonic regression analysis for a period range of 5–30 h and a significance cutoff $p < 0.01$ revealed not only clusters of rhythmic genes for periods of circadian length (~24 h) but also for ultradian periods at the second harmonic of circadian rhythmicity (~12 h) for the murine tissues (**Figure S 15A**). The latter were particularly prominent for the three types of muscle-related tissue, aorta (AOR), heart (HEA), and skeletal muscle (MUS), but were also observed for the adrenal gland (ADG) and brown and white adipose tissue (BAT and WAT). The RAIN algorithm was used to identify rhythmic features with 24-h or 12-h periods for each of the twelve murine tissues separately. Amplitudes and phases of rhythmic genes were estimated by fitting a harmonic regression with the same period.

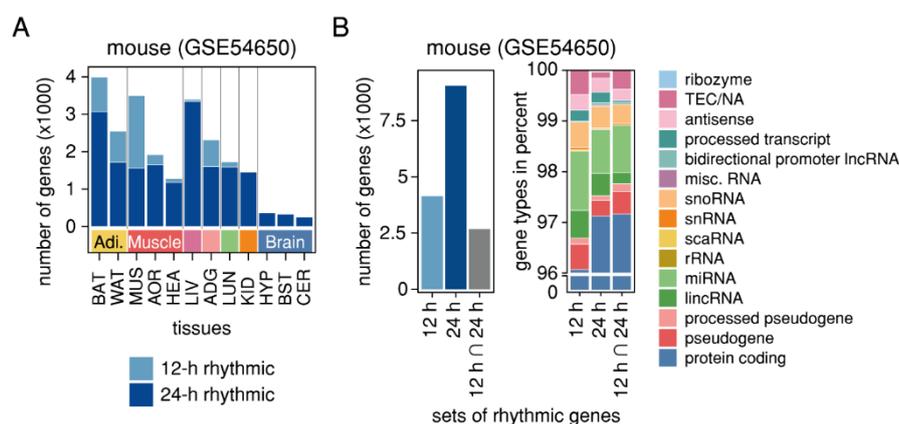


Figure 3-17: Re-analysis of the circadian transcriptome of twelve murine tissues from Zhang et al. (2014) (microarray data, GSE54650). (A) Number of 12-h rhythmic (light blue) and 24-h rhythmic (dark blue) genes in twelve murine tissues. See **Table S 5** for the meaning of tissue abbreviations. Rhythmic genes were determined by RAIN ($q < 0.05$ after filtering for genes with a relative amplitude ≥ 0.1) (B) Number of genes (left panel) and gene types in percent (right panel) for the total sets of 12-h and 24-h rhythmic genes across all murine tissues and their intersection.

Given an FDR of 5% (after pre-filtering of RAIN p -values for genes with relative amplitude ≥ 0.1), 24-h and to a lesser extent 12-h rhythmic genes were identified for all of the tissues (**Figure 3-17A** and **Figure S 16A**). As already reported in the original publication (Zhang et al., 2014), the number of circadian genes differs between the organs, with the highest number identified in the liver (LIV) (3205 genes) and the lowest in the cerebellum (CER) (220 genes). In total, 9,045 genes were identified as 24 h rhythmic and 4,140 genes as 12-h rhythmic in at least one of the murine tissues, including 2673 genes with both 24-h and 12-h rhythms, depending on the tissue (**Figure 3-17B**). The complete set of genes displaying 12-h rhythmicity in at least one murine tissue included fewer protein-coding genes than the set of genes detected as 24-h rhythmic (**Figure 3-17B**). Interestingly, not only the number of 24-h rhythmic genes but also the prevalence of 12-h rhythmic genes varied greatly throughout the murine

tissues. The median percentage of 12-h rhythmic genes in the collective sets of identified rhythmic genes (12-h and 24-h rhythmic) was 8% which is comparable to previous studies on sub-circadian transcriptional rhythms in mouse liver (Hughes et al., 2009). However, MUS (55%), ADG (31%), WAT (32%), and BAT (23%) constituted outliers which displayed higher percentages of 12-h rhythmic genes.

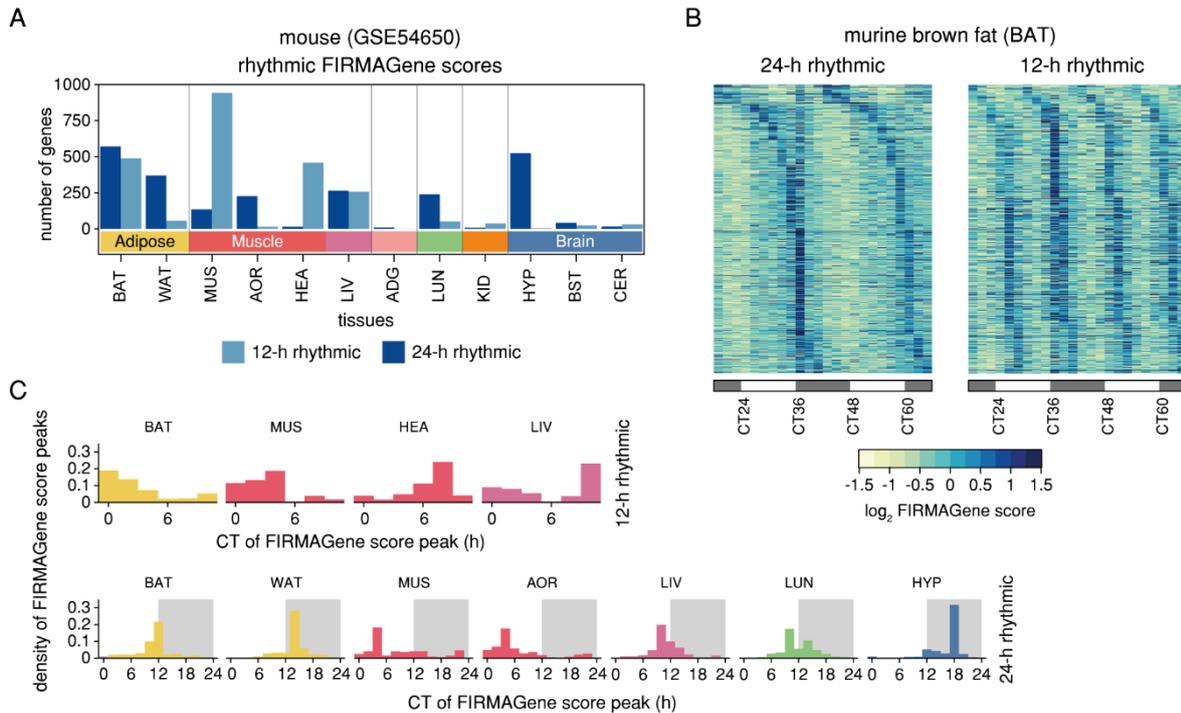


Figure 3-18: Analysis of whole-transcript microarray data (GSE54650) reveals candidate genes with 24-h and 12-h rhythmic AS events in murine tissues. (A) Number of genes with 24-h rhythmic (dark blue) and 12-h (light blue) candidate AS events in the twelve murine tissues (RAIN $q < 0.1$ and relative amplitude ≥ 0.1). (B) Exemplary phase-sorted heatmaps of 24-h rhythmic (left panel) and 12-h rhythmic (right panel) FIRMAGene scores in BAT. Phases were determined by RAIN. (C) Distribution of peak times of 12-h and 24-h rhythmic FIRMAGene scores for tissues with at least 100 genes with rhythmic candidate AS events.

While the design of whole-transcript (or “Gene”) microarrays (Figure 1-9) does not allow for the quantification of transcript- or exon-level expression, it can nevertheless be used to identify candidate genes with potentially rhythmic AS events by applying the FIRMAGene method. FIRMAGene scores potential splicing events based on the persistence of probe-wise residuals from the RMA fit (see subsection 1.2.4). In accordance with the previous analysis of rhythmic splicing in microarray data (see subsection 3.1.4), the non-parametric RAIN algorithm was used to identify genes whose FIRMAGene scores vary over time with underlying 24-h or 12-h rhythms, thus indicating a circadian or ultradian regulation of AS. Overall, 2,417 unique genes with 24-h rhythmic and 2,368 unique genes with 12-h rhythmic putative AS events were identified across all murine tissues (RAIN $q < 0.1$ and relative amplitude ≥ 0.1) (Figure 3-18A and External data file 2). The highest number of genes with putative rhythmic AS events was identified for MUS with a period of 12 h (943 genes), followed by events with a 24-h period in BAT (571 genes) (Figure 3-18B) and the hypothalamus (HYP, 524 genes). The high number of genes with putative circadian AS events in HYP is remarkable because only few genes show 24-h rhythms in

transcription in the same tissue (**Figure 3-17A**). Surprisingly, 12-h rhythmic splicing events did not only make up a large proportion of all identified rhythmic events in tissues with many 12-h rhythmic genes like MUS (87%), and BAT (46%) (**Figure 3-18B**), but also in HEA (97%) and LIV (49%) where only few 12-h rhythmic genes have been identified (**Figure 3-17A**). In adipose tissues as well as in LIV and lung (LUN), the 24-h rhythmic FIRMAGene scores peaked predominantly at the transition from the light to the dark phase, while in muscle tissues, they peaked in the middle of the subjective day, and in HYP, there was a strong peak in the middle of the subjective night (**Figure 3-18C**). Peak times for 12-h rhythmic putative AS events were less distinct and differed even between tissues of a similar type, e.g., MUS and HEA.

The number of putative AS events is dependent on the amplitude cutoff and the FDR. For an FDR of 5%, only few genes were found to display candidate AS events in more than one tissue (data not shown). However, it seems likely, that a circadian regulation of splicing of a gene is not unique to a single tissue but also occurs in other tissues of a similar type. Indeed, identical circadian AS events have previously been detected for several mammalian tissues (McGlincy et al., 2012). Thus, a gene with putative rhythmic AS events in several tissues is considered a better candidate than a gene for which only a single AS was identified, though its rhythmicity q -value might be lower. For an FDR cutoff of 10%, putative 24-h rhythmic AS events were detected in up to six tissues for one gene, e.g., for the core-clock genes *Bmal2* (two tissues) and *Nr1d1* (six tissues, including LIV), the clock-controlled gene *Ciart* (two tissues), and for the SF gene *Mbnl2* (two tissues) (**Figure S 17A**). Curiously, the set of genes with 12-h rhythmic FIRMAGene scores in more than one tissue also included several components of the core-clock and clock-controlled genes, such as *Bmal1* (three tissues), *Nr1d2* (two tissues), *Npas2* (four tissues), *Dbp* (four tissues) and *Ciart* (three tissues) (**Figure S 17B**). The putative AS event of *Ciart* in MUS was identified to be significantly rhythmic for a period of both 24-h and 12-h which shows that it can be difficult to distinguish between circadian rhythms in AS and their second harmonics, i.e., ultradian 12-h rhythms. A 12-h rhythmic putative AS event was also identified for the transcription factor gene *Xbp1* in murine BAT. Both the splicing efficiency of *Xbp1* and the abundance of the spliced isoform 12-h rhythmic have previously been found to be 12-h rhythmic *in vivo* and *in vitro* (Cretenet et al., 2010; Zhu et al., 2017) (see subsection 1.1.4). In contrast, the SF gene *U2af26/U2af114* whose splicing is controlled via circadian body temperature cycles in mouse (Preußner et al., 2017) (see subsections 1.3.2 and 1.3.3) was neither among the candidates for genes with 24-h or 12-h rhythmic AS events. Interestingly, another gene with a temperature-responsive 24-h rhythmic AS event reported by Preußner et al. (2017), *Tbp*, had a candidate 12-h rhythmic AS event in MUS.

In the next step, the results derived from the murine microarray data were further compared to RNA-seq data from the same biological material (Zhang et al., 2014). The RNA-seq data was sampled for the same time range of two circadian cycles but with a large sampling resolution of 6 h. For this reason, the RNA-seq dataset was only used to identify expressed splice variants instead of rhythmic ones. For many genes with candidate rhythmic FIRMAGene scores, two or more expressed splice variants were identified in the respective tissue (**Figure S 17C**), indicating that an AS event does indeed take place, though it does not have to be rhythmic. However, depending on the tissue, for ~18–57% of the genes

with 24-h rhythmic scores and ~11–40% of the genes with 12-h rhythmic scores, only one expressed splice variant was identified for the chosen cutoff in the respective tissue. Accordingly, the putative rhythmic AS events gained from the FIRMAGene analysis should be considered with reservations because they are likely to include many false positives.

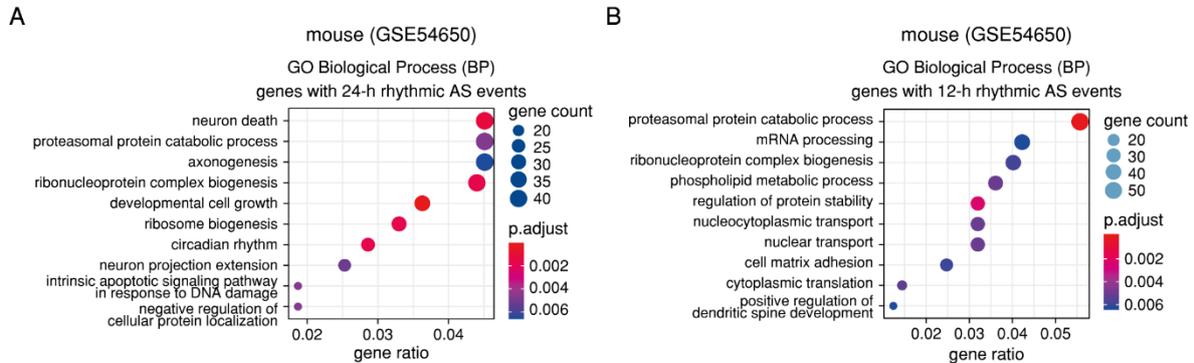


Figure 3-19: Enriched processes for candidate genes with rhythmic AS events in murine tissues (microarray data). Enriched GO terms (Biological Process – BP) for the sets of genes with (A) 24-h rhythmic and (B) 12-h rhythmic candidate AS events in the murine tissues (GSE54650) that had at least two expressed splice variants in the RNA-seq data (GSE54651). GO terms were manually curated to remove redundant categories. Only the first ten GO terms are shown. The complete lists of results can be found in **Table S 6**.

A GO enrichment analysis was performed only for the sets of candidate genes with two or more expressed splice variants (**Table S 6**). The biological processes enriched for the 945 genes with 24-h rhythmic FIRMAGene scores (BH-adjusted $p < 0.01$) include developmental cell growth, neuron death, circadian rhythm, ribosome biogenesis, and the intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator (**Figure 3-19A**). The detection of circadian rhythm among the enriched processes further reinforces the hypothesized reciprocal interplay between rhythmic AS and the circadian clock. For the set of 995 genes with 12-h rhythmic FIRMAGene scores, protein catabolic-related processes and metabolic processes were found to be enriched (BH-adjusted $p < 0.01$), in addition to nuclear and nucleocytoplasmic transport, the regulation of protein stability, cytoplasmic translation, and cell-matrix adhesion (**Figure 3-19B**). The prevalence of metabolic and catabolic processes suggests that 12-h rhythmic AS events might be involved in the hypothesized metabolic role of ultradian rhythmicity in mammals. Still, further experiments are necessary to clarify whether 12-h rhythms in isoform expression truly exist and which functional role they might play.

3.3.2 Comparison of Transcript-level and Gene-level Expression in Baboon Tissues

The second circadian multi-organ mammalian dataset used in this study consists of RNA-seq reads from 64 different tissues resected from olive baboons for one circadian day (Mure et al., 2018). The dataset was re-analyzed to gain expression values of genes and transcripts for all individual time points following the same RNA-seq preprocessing pipeline as previously described for the human CRC cell line data (see subsection 3.2.1). A harmonic regression analysis for a period range of 6–26 h and a significance cutoff $p < 0.01$ revealed clusters of rhythmic genes for periods of approximately circadian

length (~24 h) for the majority of tissues, as well as for shorter periods of ~12 h for some tissues, e.g., for the prostate (PRO), the smooth muscle (SMM), and bone marrow (BOM) (**Figure S 15B**). Many of the period distributions in baboon tissues did not center exactly around 24 h but slightly earlier which might be attributed to the relatively short time-series of a single circadian cycle. Rhythmic features with 24-h or 12-h periods were identified using the RAIN algorithm and amplitudes and phases of rhythmic features were estimated by fitting a harmonic regression with the same period. Instead of correcting the RAIN p -values for multiple testing, a low p -value cutoff was chosen to get rhythmic features for all tissues (**Figure S 16A**). Applying the criteria RAIN $p < 0.005$ and relative amplitude ≥ 0.1 , both 24-h and 12-h rhythmic genes and transcripts could be detected across all tissues (**Figure 3-20A**). The highest number of 24-h rhythmic genes was found in the thyroid (THY, 3807 genes), followed by the kidney cortex (KIC, 3513), the stomach fundus (STF, 3357 genes), and omental fat (OMF, 3038 genes). These number are comparable to those identified by the original publication where THY and STF were also identified as the tissues with the highest number (> 3000) of genes with 24-h rhythms of expression (Mure et al., 2018).

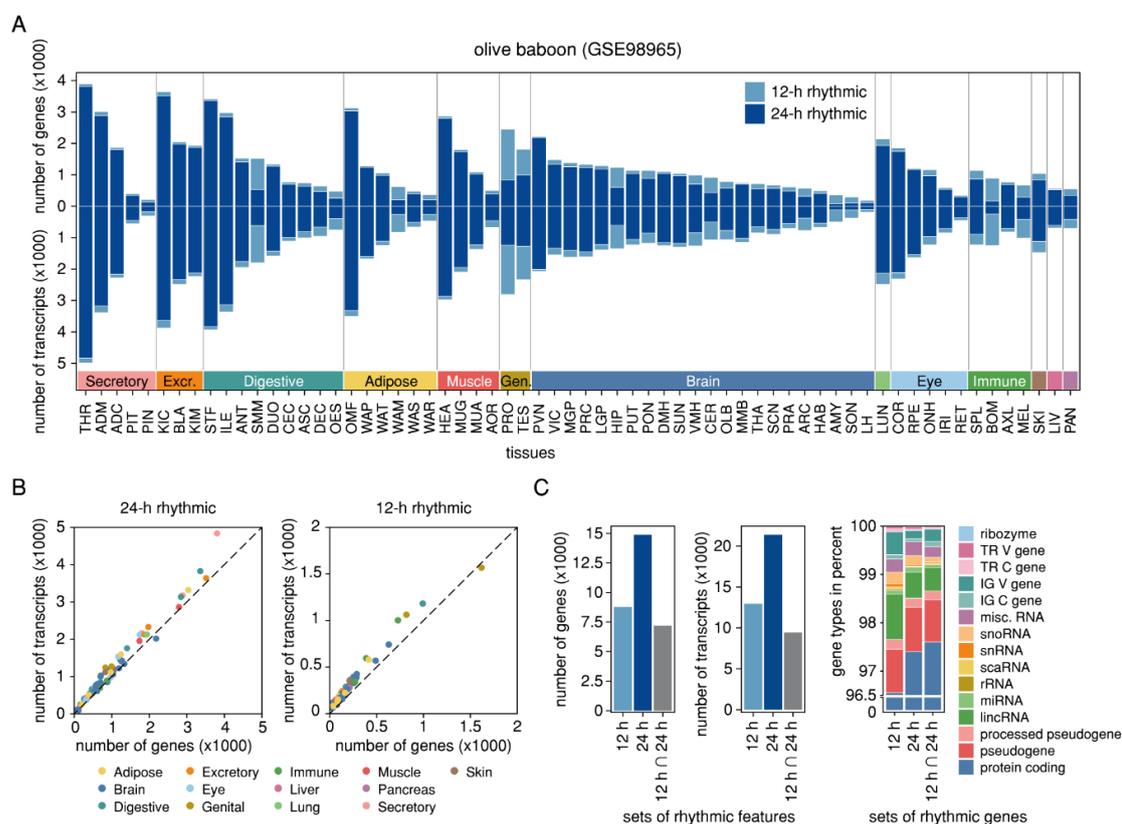


Figure 3-20: Re-analysis of the circadian transcriptome of 64 olive baboon tissues from Mure et al. (2018) on gene- and transcript-level (RNA-seq data, GSE98965). (A) Number of 12-h rhythmic (light blue) and 24-h rhythmic (dark blue) genes (upper panel) and transcripts (lower panel) in 64 baboon tissues (GSE98965). See **Table S 5** for the meaning of tissue abbreviations. Rhythmic features were determined by RAIN ($p < 0.005$ and relative amplitude ≥ 0.1) (B) Number of rhythmic genes (x -axis) versus number of rhythmic transcripts (y -axis) with a period of 24 h (left panel) and 12 h (right panel) for each tissue. Tissues are color-coded according to tissue type. (C) Number of genes (left panel), transcripts (center panel) and gene types in percent (right panel) for the total sets of 12-h and 24-h rhythmic features across all tissues and their intersection.

In addition, several tissues also displayed large proportions of 12-h rhythmic genes: While, the median percentage of 12-h rhythmic genes in the collective sets of rhythmic genes was 11%, there were some baboon tissue where the majority of rhythmic genes was detected to be 12-h rhythmic, including BOM (81%, 1001 genes), mesenteric white adipose tissue (WAM, 67%, 579 genes), PRO (66%, 1568 genes), and SMM (65%, 1568 genes). In total, 14,881 genes were identified as 24-h rhythmic and 8,763 genes as 12-h rhythmic in at least one of the baboon tissues (**Figure 3-20C**). The genes detected to be 12-h rhythmic included fewer protein-coding genes than the set of 24-h rhythmic genes. However, as already observed for the murine multi-organ dataset, there was a large intersection between the two sets of rhythmic genes, with ca. 82% of the 12-h rhythmic genes identified as 24-h rhythmic in at least one other tissue. Applying the same criteria for the transcript-level counts of the baboon data, 21,379 transcripts were identified as 24-h and 12,941 transcripts as 12-h rhythmic (**Figure 3-20C**), resulting in a roughly linear correlation between the number of rhythmic genes and the number of rhythmic transcripts (**Figure 3-20B**).

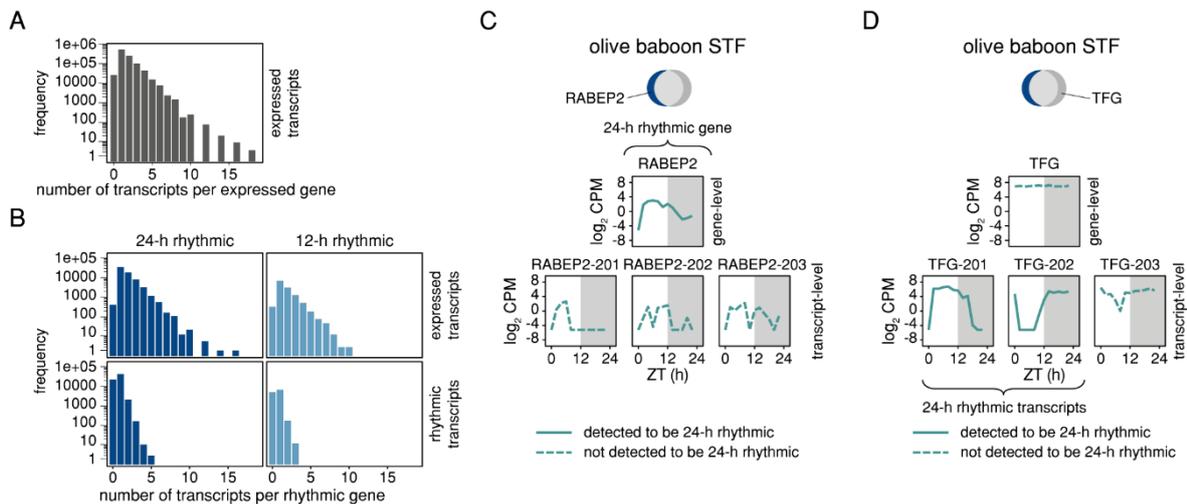


Figure 3-21: Discrepancies between time-series expression identified on gene- and transcript-level in baboon tissues (RNA-seq data, GSE98965) (A) Frequency barplot of the number of expressed transcripts per expressed genes in all baboon tissues. (B) Frequency barplots of the number of expressed transcripts per 24-h rhythmic gene (dark blue, upper left panel) and per 12-h rhythmic gene (light blue, upper right panel), and of the number of 24-h rhythmic transcripts per 24-h rhythmic gene (dark blue, lower left panel) and of 12-h rhythmic transcripts per 24-h rhythmic gene (light blue, lower right panel) in all baboon tissues. (C) Example of a gene (*RABEP2*) detected to be 24-h rhythmic in baboon stomach fundus (STF) on gene-level but not on transcript-level. (D) Example of a gene (*TFG*) detected to be 24-h rhythmic in the stomach fundus (STF) on transcript-level but not on gene-level.

In the next step, the identity and the rhythmic parameters of the rhythmic genes detected on gene- and on transcript-level were compared for each baboon tissue. Surprisingly, the intersection of the two gene sets only amounted to an average of 40% of the set union for the 24-h rhythmic genes (**Figure S 18A**) and to an average of 25% of the set union for the 12-h rhythmic genes (**Figure S 18B**). Interestingly, the tissue-wise median of RAIN *p*-values of genes from the intersection (identified as rhythmic on both transcript- and summarized gene-level) was lower than the *p*-value median of the genes identified as rhythmic on gene-level for the majority of tissues (**Figure S 19A**). The phase distributions of

rhythmic features were similar between gene- and transcript-level, but differed between tissues, in particular for features with 12-h rhythms in expression (**Figure S 19B**). Both 24-h and 12-h rhythmic transcript sets tended to have higher relative amplitudes than their counterparts on gene-level (**Figure S 19C**), which is likely due to the transcript counts summarization step of the RNA-seq preprocessing pipeline. Across all baboon tissues, an expressed gene had an average of 1.7 expressed transcripts. Some genes had up to 18 expressed transcripts (**Figure 3-21A**), e.g., *EPB41L2*, *RACK1*, and *ENSPANG00000011737*, a novel orthologous gene to human gene *DHRS4*. A similar ratio was observed for the average number of expressed transcripts per rhythmic gene (24-h rhythmic: 1.80; 12-h rhythmic: 1.68) (**Figure 3-21B**). The average number of rhythmic transcripts per rhythmic gene was lower (24-h rhythmic: 0.71; 12-h rhythmic: 0.60), the maximum being five 24-h rhythmic transcripts for *MT3* and *RACK1* (**Figure 3-21B**).

For genes with several expressed transcripts, discrepancies between the time-series expression of the gene and the time-series expression of its individual transcripts could be observed. For instance, the gene *RABEP2* was 24-h rhythmic in baboon STF, whereas its three expressed transcripts were not (**Figure 3-21C**). The opposite case was detected for the gene *TFG*: Two out of the three expressed *TFG* transcripts in STF were identified as 24-h rhythmic, albeit with different phases, while the summarized expression on gene-level was constant (**Figure 3-21C**). In the first case, the summarization of transcript-level counts to gene-level count estimates led to the gene being defined as rhythmic despite the individual transcripts being arrhythmic, making *RABEP2* a likely false positive candidate circadian gene. In the second case, the summarization led to a masking of the phase-shifted rhythmicity of the two 24-h rhythmic transcripts, making *TFG* a likely false negative. The latter case could be the result of a potential circadian regulation of AS: Two splice variants of the same gene are rhythmically transcribed with their expression peaking at different times of the circadian day. Hence, a rhythmicity analysis on transcript-level is likely to yield new results, including candidate circadian and/or ultradian alternatively spliced genes, that would be overlooked in a rhythmicity analysis conducted on gene-level only.

3.3.3 Detection of Differentially Rhythmic Phase-shifted Splice Variants in Baboon Tissues

Following the previously described analysis pipeline for the detection of putative rhythmic AS in circadian RNA-seq data (see subsection 3.2.3), differentially rhythmic SVPs were identified in 21 and 16 baboon tissues, respectively (**Figure 3-22A**). As previously described for the human RNA-seq data (see subsection 3.2.3), the differentially rhythmic SVPs were further filtered for pairs with an amplitude ratio < 2 that were phase-shifted by at least a sixth of their period length in the same tissue (**Figure 3-22B**). This filtering step resulted in a total of 2,278 (1,620 unique SVPs and 1,250 unique genes) and 501 (483 unique SVPs and 443 unique genes) phase-shifted SVPs with 24-h and 12-h rhythms in expression, respectively (**Figure 3-22C** and **External data file 3**).

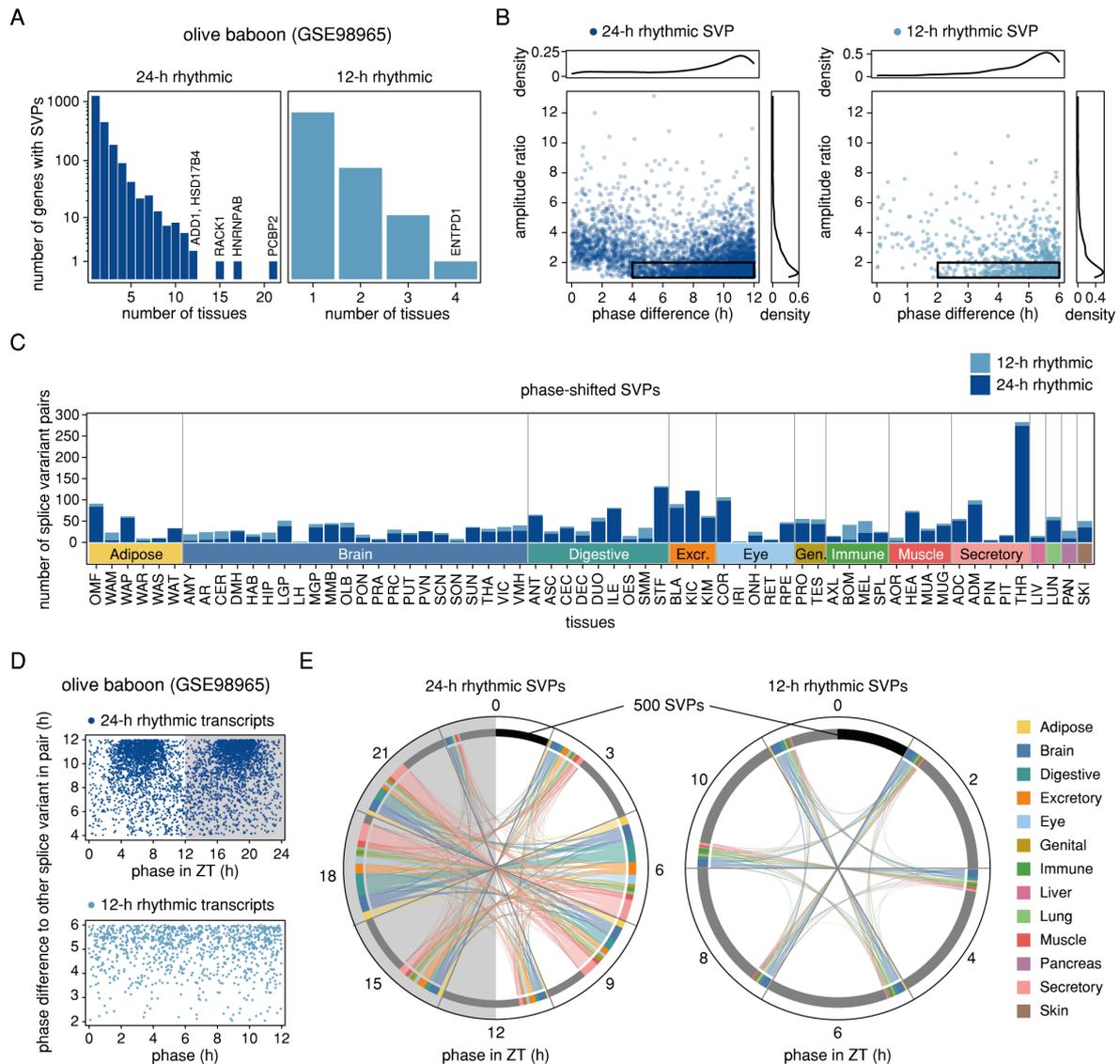


Figure 3-22: Analysis of differentially rhythmic splice variant pairs of the same gene in the same baboon tissue (RNA-seq data, GSE98965). (A) Distribution of the number of genes with differentially 24-h rhythmic (upper panel, dark blue) and 12-h rhythmic (lower panel, light blue) SVPs. (B) Amplitude ratio and phase difference of SVPs found to be differentially rhythmic with a period of 24 h (left panel) and 12 h (right panel). Phase-shifted SVPs (amplitude ratio < 2 and phase-shift \geq period/6) are marked by black rectangles. (C) Number of 24-h (dark blue) and 12-h rhythmic (light blue) phase-shifted SVPs in the 64 baboon tissues. (D) Phase distribution and pairwise phase difference of 24-h rhythmic (upper panel, dark blue) and 12-h rhythmic (lower panel, light blue) phase-shifted SVPs for all baboon tissues. (E) Chord diagrams representing the relations between the phases of 24-h rhythmic (left panel) and 12-h rhythmic phase-shifted SVPs. Peak-phases are clustered in bins of 3 h (left panel) and 2 h (right panel) and colored according to the type of tissue in which they were detected.

An analysis of the phase distribution and the phase differences of all phase-shifted 24-h rhythmic splice variants pairs revealed two clusters of splice variants that exhibited a phase-shift greater than 10 h with two peaks of expression: one between ZT4–ZT10 and the other between ZT15–ZT21 (**Figure 3-22D**). When further comparing the relations between the two peak times of phase-shifted SVPs in chord diagrams, it becomes apparent that the majority of 24-h rhythmic phase-shifted SVPs was expressed at opposing times of the daily cycle, with one transcript peaking in the middle of the subjective

day and the other in the middle of the subjective night (**Figure 3-22E**). For the smaller set of 12-h rhythmic SVPs, a tendency for phase-shifts greater than 4 h but no clear peak times over the 12-h cycle were observed (**Figure 3-22D and E**). 364 (22.5%) of the 24-h rhythmic phase-shifted SVPs and 439 (35.1%) of the affected genes appeared in more than one tissue, suggesting that the putative circadian regulation of AS is prevalent across baboon tissues of different organ types.

To learn more about possible physiological functions of rhythmic AS, a GO enrichment analysis was conducted on gene-level for all candidate genes with phase-shifted SVPs. Biological processes enriched for genes with 24-h rhythmic SVPs included RNA processing, RNA splicing processes, Golgi vesicle transport, and the regulation of mRNA metabolic processes (adjusted $p < 0.05$) (**Figure 3-23A and Table S 5**). The prevalence of RNA splicing-related processes among the enriched terms suggests that splicing-related genes are among the targets of the putative circadian regulation of AS, as previously observed for the human CRC cell lines SW480 and SW620 (see subsection 3.2.3). No significantly enriched processes (BH-adjusted $p < 0.05$) were identified for the smaller set of genes with 12-h rhythmic phase-shifted SVPs, though some GO terms were found for a less stringent cutoff ($p < 0.001$), including Golgi vesicle, cytosolic, and microtubule-based transport (**Figure 3-23B and Table S 5**).

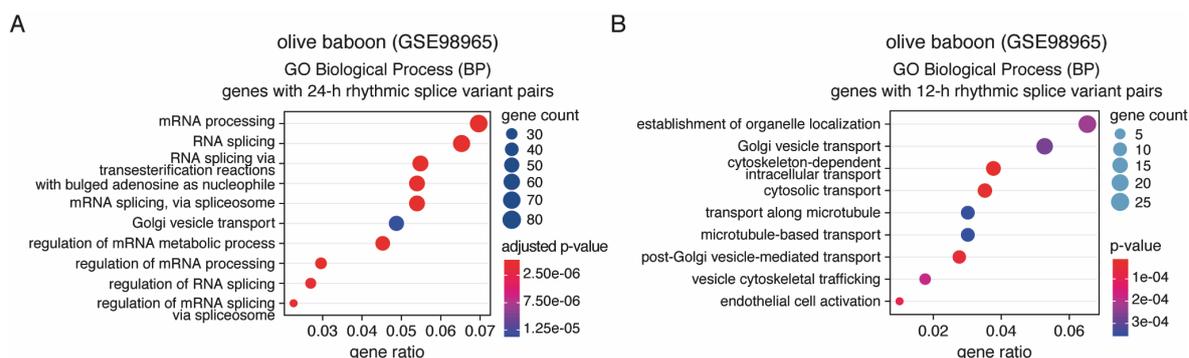


Figure 3-23: Enriched biological processes for genes with rhythmic phase-shifted splice variants in baboon tissues (RNA-seq data, GSE98965). Enriched GO terms (Biological Process – BP) for the sets of genes that exhibited (A) 24-h rhythmic and (B) 12-h rhythmic phase-shifted SVPs in at least on baboon tissue. Shown are the top ten enriched processes. The complete lists of results can be found in **Table S 5**.

Assuming that the phase-shifted SVPs represent true rhythmic splicing events, one would expect some of the candidates to appear with similar phases in more than one tissue of the same or possibly even of different organ types. For the set of 12-h rhythmic phase-shifted splice variants, 16 SVPs were detected in two tissues (*ATP13A1*, *C2orf49*, *CTNNB1*, *ENSPANG00000032468* (an orthologous gene to human *PARL*), *ENTPD1*, *GTDC1*, *KPNB1*, *NBAS*, *NDUFB2*, *PI4KB*, *PLEKHA7*, *TBC1D23*, *TPM1*, *UBAP1*, *VPS13A*, and *ZNF691*) and a single SVP in three tissues (*DNAJC13*). The small number of occurrences prevents testing for phase synchrony of the individual splice variants across tissues. In contrast, there were 22 phase-shifted 24-h rhythmic SVPs that were detected in six or more baboon tissues and therefore constitute good candidates for robustly circadian AS events in baboon (**Figure 3-24 and Table S 8**). The top candidate SVPs included 43 different splice variants that originate from 21 genes: *ADD1*, *CCNL2*, *CPNE1*, *DRG1*, *EIF3H*, *EXOC3*, *HNRNPAB*, *HSD17B4*, *MCFD2*, *MOB4*, *NELFE*, *PAF1*, *PCBP2*, *PEX6*,

For the majority of the candidates, not much is known concerning the expression and/or function of alternative splice variants in mammals, with the exception of *CCNL2*, *HSD17B4*, *SEPHS1*, and *ZNF207*. *CCNL2* is a member of the cyclin family that has been found to be involved in splicing and to inhibit growth and induce apoptosis of tumor cells, possibly via regulating RNA processing of apoptosis-related factors such as p53, Bax, and Bcl-2 (Yang et al., 2004). Three protein isoforms encoded by *CCNL2* have been reported to be localized in distinct cellular compartments and to bind to SR proteins (Berke et al., 2001; Loyer et al., 2008). *HSD17B4* encodes for a member of the HSD17B family of hydroxysteroid dehydrogenases that have various enzymatic functions in steroidogenesis and steroid metabolism. Five protein-coding splice variants of *HSD17B4* have been described in humans and the loss of the only androgen-inactivating *HSD17B4* splice variant has been found to promote the development of castration-resistant prostate cancer (CRPC) (Ko et al., 2018). Interestingly, one of the phase-shifted SVPs of *HSD17B4* has been identified in baboon PRO, indicating that there might be an isoform-dependent circadian inactivation of androgen in the prostate. Moreover, a second member of the HSD17B family, *HSD17B10*, displayed phase-shifted 24-h rhythmic SVPs in five different baboon tissues. Higher levels of *HSD17B10* mRNA have been detected in CRPC bone metastases compared to non-malignant prostate and primary prostate tumor tissue (Jernberg et al., 2013). The protein encoded by *SEPHS1* plays a role in maintaining cellular redox homeostasis. Disruption of murine *Sephs1* has been reported to lead to an inhibition of cell proliferation (Na et al., 2018). For human *SEPHS1*, five splice variants with different subcellular locations and expression patterns have been described that vary depending on cell cycle stage (Kim et al., 2010). Two isoforms of human *ZNF207* play distinct roles during cell differentiation of embryonic stem cells: While *ZNF207* isoform C is present in embryonic stem cells where it controls the transcription of genes responsible for self-renewal and pluripotency, *ZNF207* isoform B can be found in differentiated cells and cancer cells where it regulates mitotic chromosome alignment (Fang et al., 2018).

The list of genes with candidate differentially rhythmic SVPs was further searched for genes with experimentally validated functionally distinct splice isoforms in healthy mammalian tissues (Bhuiyan et al., 2018). Of the 23 human genes meeting the criteria of Bhuiyan and colleagues, ten orthologous olive baboon genes with differentially 24-h rhythmic SVPs in at least one tissue were found (*CFLAR*, *EIF4G1* and *EIF4G2*, *KLF6*, *MADD*, *MST1R*, *PML*, *PRMT5*, *STIM2*, and *SUN1*). Out of these, phase-shifted SVPs were found for the eukaryotic initiation factor gene *EIF4G2* (three tissues), the splicing regulator gene *PRMT5* (two tissues), *SUN1* (one tissue) and for the tumor suppressor genes *KLF6* (one tissue) and *PML* (two tissues). Human *SUN1* isoforms differ in their tissue distribution and have opposing roles in the regulation of cell migration as shown by isoform-specific knockdown experiments (Nishioka et al., 2016). AS of *KLF6* can result in the dominant negative splice isoform *KLF6-SV1*, which has been identified as a key driver of breast cancer and prostate cancer metastasis in humans by promoting cell survival, migration, and invasion (Hatami et al., 2013; Narla et al., 2008).

To find out whether genes with 24-h rhythmic phase-shifted SVPs are shared across species, the baboon candidates were compared to the human candidate genes identified based on the RNA-seq dataset of the CRC cell lines SW480 and SW620 (see subsection 3.2.3), revealing 74 shared candidate

genes. For 19.8% of the human candidates, an orthologous candidate gene could be identified in at least one baboon tissue, suggesting that genes affected by circadian AS are at least partly conserved between diurnal primate species. The 74 candidates included eleven splicing-related genes (*CIRBP*, *DDX5*, *DDX17*, *HNRNPH1*, *MIR3064*, *NONO*, *NXF1*, *PCBP2*, *RBM3*, *RBM39*, *SRRT*, and *TAF15*), among them tumor suppressors and oncogenes. The cold-inducible protein CIRBP is known as a mediator of cancer-related inflammation and has been reported to act both as a tumor suppressor and a tumor promoter, depending on the cell type and cancer stage (Lujan et al., 2018). The SF RBM39, also known as CAPER- α , has been found to alter the ratio of VEGF isoforms in breast cancer and Ewing's sarcoma cells, thereby influencing tumor growth and vessel density (Dowhan et al., 2005; Huang et al., 2012). NONO constitutes an important coupling element between the mammalian circadian clock and the cell cycle (see subsection 1.1.3) and PCBP2 is overexpressed in several cancers and has recently been described as an oncogenic SF (Guo and Jia, 2019). Further common candidates include the genes *HDAC6* and *PRMT2*, that are both involved in the regulation of the androgen receptor signaling pathway. Together with the two members of the HSD17B family identified among the top candidates with phase-shifted SVPs in baboon tissues, this finding further points to a possible role of mammalian circadian AS in androgen signaling. Four shared candidate genes (*ARFGAP1*, *DCTN3*, *GOSR2*, and *SPTAN1*) are associated with ER to Golgi vesicle-mediated transport, a process that was found to be enriched for genes with phase-shifted SVPs in baboon tissues (**Figure 3-23A**). When additionally taking the candidate genes with rhythmic AS from the microarray analysis of SW480 and SW620 cells into account, further robust candidate genes emerged: Five genes were shared between baboon tissues and SW480 cells (*CAMK2G*, *ILK*, *MIR6727*, *MSH6*, *NBR1*, *PPP2R5C*, and *UBR4*) and four genes were shared with SW620 cells (*CAMK2G*, *MFF*, *NBR1*, and *ST5*). Both *CAMK2G* and *PPP2R5C* were previously identified as candidate genes for SW480 cells based on both the microarray and the RNA-seq data (see subsection 3.2.3). Moreover, *CAMK2G* and *NBR1* were the only candidate genes with circadian AS events that were found to be shared across cell lines in the microarray data (**Figure S 8**). The fact that these genes were likely also alternatively spliced in a rhythmic manner in baboon tissues further reinforces the previous findings for the CRC cell lines. The baboon candidate genes were also compared to the murine candidate genes with putative rhythmic AS events (identified based on microarray data) and more than two expressed splice variants (identified based on RNA-seq data) (see subsection 3.1.1). Prior to the comparison, genes of both species were mapped to orthologous human genes. For candidates with 12-h rhythmic AS events, 50 genes were found to be shared between mouse and baboon, compared to 99 candidate genes with 24-h rhythmic events. The intersection of 24-h rhythmic candidate genes from all three different species independent of the platform revealed a total of five common candidates between murine and baboon tissues and human cell lines, including two SFs (*NONO* and *RBM39*) and three other genes (*DCTN3*, *POLR2H* and *SMARCE1*).

Overall, phase-shifted differentially 24-h and to a lesser extent 12-h rhythmic SVPs were identified in a variety of baboon tissues, suggesting a time-of-day dependent regulation of mRNA splicing in diurnal primates. Several 24-h rhythmic SVPs were detected in more than one baboon tissue, making them robust candidates for circadian AS events. Functional annotation of the candidates revealed splicing-

related processes to be enriched, indicating a reciprocal interplay between circadian AS and genes involved in splicing. In line with this observation, the intersection of candidate genes derived from the human CRC cell lines and the murine tissues included several splicing-related genes that are likely spliced in a circadian manner in different mammalian species. However, it still remains to be elucidated whether the detected rhythmic splice variants have a functional effect and whether the specific isoforms and their putative distinct functions are conserved across species.

3.3.4 Conserved Rhythmicity of Splicing-related Genes across Mammalian Tissues

Following the assumption that AS is to some extent a rhythmic process, one would expect a subset of splicing-related genes to be consistently rhythmic across tissues and possibly also across species. Oscillating spliceosome components and SFs have previously been identified both on gene- and transcript-level for the human CRC cell line pair SW480 and SW620 (see subsections 3.1.3 and 3.2.3). A search was conducted for splicing-related genes identified as either 12-h or 24-h rhythmic in at least a quarter of the tissues from mouse and baboon (three out of twelve tissues in mouse and 16 out of 64 tissues in baboon) (**Figure 3-25A and B**). The 426 human genes from the extended list of splicing-related genes were mapped to 451 orthologous murine genes and to 429 orthologous baboon genes of which 409 and 408 were expressed in at least one tissue, respectively. Of the expressed splicing-related genes, ~38% were detected to be either 12-h or 24-h rhythmic in at least one of the murine tissues and ~95% were detected to be either 12-h or 24-h rhythmic in at least one of the baboon tissues. As expected, most of the genes showed 24-h oscillations in transcription whereas 12-h oscillations constituted an exception. In mouse, some splicing-related genes were detected to be rhythmic for both a 24-h and a 12-h period which might be attributed to the difficulty of separating true ultradian rhythms from coexpressed circadian rhythms and their harmonics (van der Veen and Gerkema, 2016).

The topmost five consistently rhythmic splicing-related genes in mouse tissues were the hnRNP-encoding genes *Cirbp* and *Fus*, and the heat shock protein encoding genes *Hspa1b*, *Hspa5*, and *Hspb1* (**Figure 3-25B**). While the heat shock protein encoding genes might be taken as an indication of a regulation via diurnal changes in body temperature, *HSPA1B* and *HSPA5* have also been found to be 24-h rhythmic in the CRC cell lines that were sampled while being kept in constant temperature conditions (**Figure S 14C**). *Fus* is a neurodegeneration-associated protein that binds to distinct RNA sites of its target genes to regulate the inclusion of alternative exons and has recently been found to be transcriptionally regulated by core-clock component *Nr1d1* and to modulate the expression of core-clock genes *Per2* and *Cry1* (Jiang et al., 2018; Rogelj et al., 2012). *FUS* was also detected as 24-h rhythmic in both SW480 and SW620 cells (**Figure 3-6C and D**). In baboon, the topmost five consistently rhythmic splicing-related genes were the hnRNP encoding gene *HNRNPDL*, *ENSPANG00000032278* (a novel orthologous gene to murine genes *Hspa1a* and *Hspa1b*), the integrator complex subunit-coding gene *INTS1*, the RBP encoding gene *RNF40*, and the SF gene *XAB2* (**Figure 3-25A**). *HNRNPDL* and *XAB2* have also been identified as 24-h rhythmic in SW480 cells (**Figure 3-6C**). In addition, several splicing-related genes were found to be consistently rhythmic in both mouse and baboon tissues, including

Hspa1a/Hspa1b/ENSPANG00000032278, *Hspb1/HSPB1*, *Rbm45/RBM45*, and *Srsf5/SRSF5*. *Hspb1* and *Srsf5* have previously been reported to be regulated via temperature variations: While *Hspb1* expression has been associated with the efficiency of splicing recovery after heat shock (Arrigo and Gibert, 2013), transcript and protein levels of the splicing factor *Srsf5* have been shown to be cold-inducible in mammalian cells cultured at 32 °C (Fujita et al., 2017). Interestingly, human SRSF5 also displayed 24-h rhythms in transcription in SW480 cells kept under constant conditions, indicating that oscillations of the gene are not solely due to changes in the internal body temperature of animals. With the exception of *HSPA5* in baboon and *Hnrnpdl* in mouse, the phases of all genes were significantly unimodally distributed across tissues in both species ($p < 0.05$ determined by Rayleigh test) (Table 2).

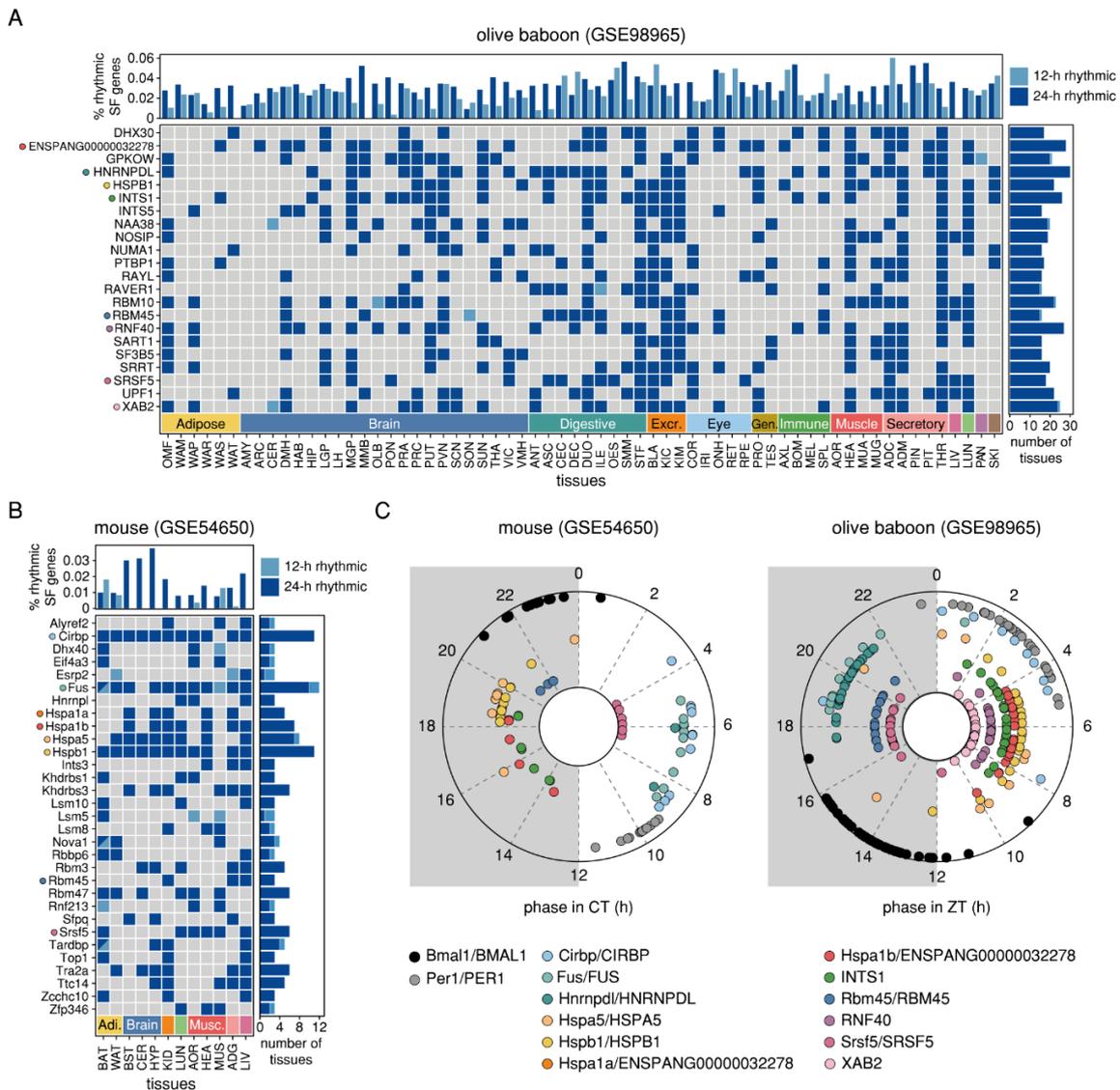


Figure 3-25: Conserved 24-h rhythmicity of splicing-related genes in mouse and baboon. Categorical heatmaps for splicing-related genes detected as rhythmic (24-h rhythmic: dark blue; 12-h rhythmic: light blue) in at least (A) 16 baboon tissues and (B) three murine tissues. The respective upper panels show the percentage of all rhythmic splicing-related genes in the set of all rhythmic genes per tissue and the respective right panels show the number of tissues where the splicing-related genes were detected as rhythmic. The topmost five consistently rhythmic genes of each species are marked by a colored dot. (C) Peak phases of consistently 24-h rhythmic splicing-related genes and the core clock genes *Bmal1/BMAL1* and *Per1/PER1* in mouse tissues (left panel) and baboon tissues (right panel).

The comparison of circular mean phases of orthologous splicing-related genes between mouse and baboon revealed most of them to be phase-shifted by about 12 h (**Table 2**), as expected for nocturnal and diurnal mammals (Mure et al., 2018). A notable exception to the observed ~12-h shift between species were the genes *Rbm45/RBM45* and *Cirbp/CIRBP* whose mean circular phases were only shifted by about 3 h. In mouse, *Cirbp* peaked in the middle of the subjective day, coinciding with a trough in the core body temperature of male C57Bl/6 mice (Sanchez-Alavez et al., 2011) which supports the hypothesis of a temperature-dependent control of *Cirbp* (Gotic et al., 2016; Gotic and Schibler, 2017). Yet, the morning peak of expression of *CIRBP* at ~ZT 3 in baboon overlapped with the time at which the body temperature of the animals was reported to reach its peak (~37.8°C) (Mure et al., 2018), indicating that *Cirbp/CIRBP* control mechanisms might differ between diurnal and nocturnal mammals. An alternative explanation could be a circadian regulation of *CIRBP* AS that results in the phase-shifted expression of different isoforms, as previously detected in SW620 cells (**Figure 3-15C**) and diverse baboon tissues (see subsection 3.3.3).

Table 2: Topmost consistently 24-h rhythmic splicing-related genes in mouse and baboon tissues.

Gene (mouse/baboon)	Circular mean phase		Rayleigh test <i>p</i> -value		Phase difference of circular mean phases (CT-ZT)
	mouse (CT)	baboon (ZT)	mouse	baboon	
<i>Cirbp/CIRBP</i>	6.41	2.93	0	0.0032	3.48
<i>Fus/FUS</i>	6.70	20.16	0	1.25e-05	13.46
<i>Hnrnpdl/HNRNPDL</i>	7.32	19.74	---	7.37e-12	12.42
<i>Hspa1a/ENSPANG00000032278</i>	16.70	6.34	0.0110	1.37e-10	10.36
<i>Hspa1b/ENSPANG00000032278</i>	16.20	6.34	0.0012	1.37e-10	9.86
<i>Hspa5/HSPA5</i>	19.43	7.61	0.0023	0.409	11.82
<i>Hspb1/HSPB1</i>	19.26	6.17	0	6.90e-08	13.09
<i>Rbm45/RBM45</i>	21.52	18.72	0.0362	7.11e-07	2.80
<i>Srsf5/SRSF5</i>	17.54	5.44	3.97e-04	3.05e-07	12.10

In order to gain a broader view of the peak phase distribution of the 24-h rhythmic splicing-related genes in baboon, the candidate set was extended to include all genes detected to be rhythmic with a circadian period in at least ten tissues. The resulting 83 genes displayed a bimodal distribution of peak phases: The first cluster contains about two thirds of the set (54 genes) and peaks during the subjective day (ZT6.1 ± 0.77) and the second cluster contains about one third of the set (29 genes) and peaks during the subjective night (ZT18.6 ± 1.1) of the animals (**Figure 3-26A**). For all genes, the phases across tissues were significantly unimodally distributed ($p < 0.05$ determined by Rayleigh test). Since promoter regions of clock-controlled genes are known to be GC-rich (Bozek et al., 2009), the GC content of the promoter sequences (TSS ± 1,000 bp) of the two clusters of consistently 24-h rhythmic splicing-related genes was compared to that of the 18 arrhythmic splicing-related genes not detected to be 24-h rhythmic in any baboon tissue. The promoters of the set of daytime-peaking splicing-related genes were found to be significantly more GC-rich than those of the arrhythmic splicing-related genes

($p = 1.058e-06$ determined by Wilcoxon rank sum test) and the nighttime-peaking splicing-related genes ($p = 0.002101$ determined by Wilcoxon rank sum test) (**Figure 3-26B**). An enrichment test for clock transcription factor binding sites in the promoter regions of the two clusters further revealed the D-box motif to be significantly enriched for the set of daytime-peaking splicing-related genes (Bonferroni-adjusted $p = 0.0298$). The observed oscillations might thus be due to a direct regulation via first-order clock-controlled transcription factors that bind to D-boxes in a phase-specific manner, such as *DBP* and *TEF* (Bozek et al., 2009). *DBP* and *TEF* were among the top consistently 24-h rhythmic genes in baboon (24-h rhythmic in 25 and 26 tissues, respectively), suggesting that they are important factors for the downstream propagation of circadian rhythms in baboon. Thus, they might contribute to the regulation of splicing processes by binding to splicing-related genes. According to the TRANSFAC database, both the daytime-peaking gene *DHX30* and the nighttime-peaking gene *SYNCRIP* are targets of *DBP* (Matys et al., 2003; Rouillard et al., 2016). In contrast, no enrichment for clock transcription factor binding sites was found for the set of nighttime-peaking splicing-related genes. This could be due to the smaller set size or the length of the chosen promoter sequences. Alternatively, the genes might be regulated via first-order clock-controlled genes, e.g., by circadian transcription factors that show 24-h rhythmic expression patterns but do not constitute core-clock components themselves, or through secondary circadian effects such as diurnal changes in core body temperature.

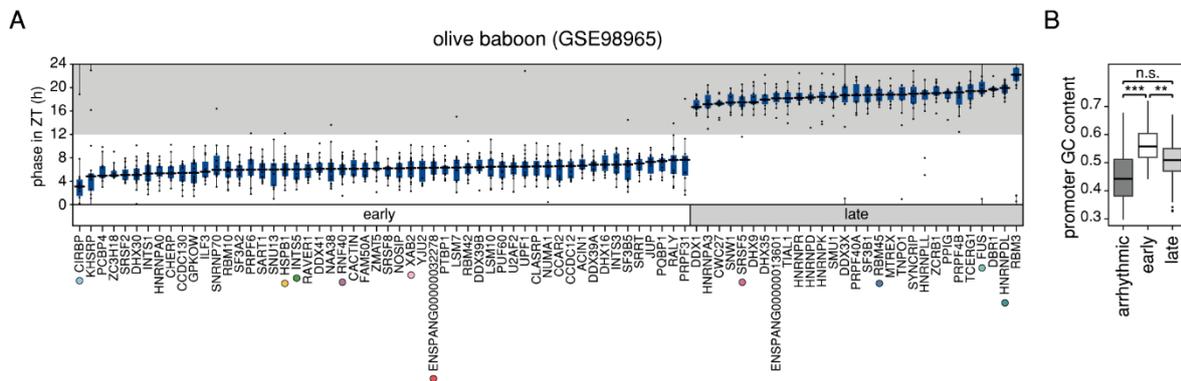


Figure 3-26: Phase distribution of robustly oscillating splicing-related genes across baboon tissues. (A) Circular boxplots of peak phases of splicing-related genes that were detected as 24-h rhythmic in at least 10 baboon tissues. Circular boxplots show the circular median and inter-quartile range (IQR). IQR is extended with whiskers to the largest and smallest value, respectively, and can cross the 0/24 h divide, but no further than 1.5x IQR from hinges. The topmost five consistently rhythmic genes of each species are marked by a colored dot. (B) Promoter GC content in sets of arrhythmic, daytime-peaking and nighttime-peaking 24-h rhythmic splicing-related genes calculated for TSS \pm 1,000 bp. The Wilcoxon rank sum was used to check for significant differences (n.s.: not significant, **: $p < 0.01$ ***: $p < 0.001$).

In conclusion, many splicing-related genes showed coordinated 24-h rhythms in expression across diverse tissues from both mouse and baboon, suggesting a circadian control of the spliceosome machinery, possibly via core clock genes and circadian transcription factors. When comparing peak phases of the most frequently oscillating splicing-related genes, a \sim 12-h phase-shift between diurnal and nocturnal animals could be observed for most genes, indicating that the putative circadian regulation of mRNA splicing is conserved in mammals.

4 Discussion

4.1 A CRC Cell Line Model Reveals Disruption of Clock Genes during Tumor Progression

Cancer cell lines display a variety of circadian phenotypes, ranging from strong oscillators with robust circadian rhythms of core clock genes to weak oscillators with a near-complete loss of rhythmicity (Relógio et al., 2014). To investigate a possible link between the circadian clock and AS in a cancer context, the CRC cell line pair SW480/SW620 was chosen as a model system of human tumor progression. Both cell lines were derived from the same male CRC patient, with SW480 established from a primary colorectal carcinoma (classified as Dukes' type B) and SW620 established from a lymph node metastasis (classified as Dukes' type C). The cell lines display identical mutations in the CRC-critical genes *p53* and *KRAS*, and have been classified as being microsatellite stable and having a negative CpG island methylator phenotype (Berg et al., 2017). Alterations in AS during CRC development are common (Bisognin et al., 2014) and a differential isoform usage between SW480 and SW620 cells has previously been reported (Huerta et al., 2001). The circadian phenotypes of both cell lines have been determined by time-series bioluminescence measurements of the promoter activity of the core clock genes *BMAL1* and *PER2*, revealing SW480 to be a robustly oscillating cell line with a circadian period of ~23.8–24.6 h, whereas SW620 displays only weak oscillations (Fuhr et al., 2018). Hence, the cell lines were considered as a suitable *in vitro* model for studying differences in the circadian transcriptome and in AS during tumor progression.

To get a comprehensive view of changes in clock phenotypes during CRC progression, the circadian transcriptome of both cell lines was profiled twice for at least one day, using both microarrays and RNA-seq. The results of both platforms revealed stable oscillations with a period of ~24 h in the expression of core clock genes in SW480 cells, with antiphasic rhythms detected between *BMAL1* and several of its target genes. In contrast, core clock genes in SW620 showed either diminished oscillations with lower amplitudes or arrhythmic expression patterns with little variation over time. These findings are in line with previous reports of core clock gene phenotypes observed in SW480 and SW620 cells, as determined by measurements of promoter activity by bioluminescence and gene expression by time-series qPCR (Fuhr et al., 2018; Relógio et al., 2014). Accordingly, it can be concluded that the circadian transcription of core clock genes is disrupted in SW620 cells in comparison to SW480 cells. While the robust oscillations in SW480 cells can be taken as an indication for a functioning core clock

system in the primary tumor-derived cell line, this needs to be further validated by a comparison to healthy reference data. Recent studies have shown that the co-expression of core-clock genes is perturbed in many solid human cancers when compared to signatures of normal tissues of the same organ type (Shilts et al., 2018). To date, there is no healthy circadian transcriptome data of human colon cells available. However, due to the high conservation of components of the circadian systems in mammals, the co-expression of core clock genes in the CRC cell lines can instead be compared to that of healthy cells from another species. The correlation signature of core clock genes in SW480 cells closely resembled a reference signature derived from healthy mouse organs (Zhang et al., 2014), whereas the signature for SW620 cells differed from both. Hence, it can be inferred that SW480 cells are in possession of a functioning core clock system that is comparable to that of healthy cells in terms of core clock gene co-expression. With further progression of the cancer, the circadian clock appears to get dysregulated, as evidenced by the diminished or lost rhythmicity in the expression of its core components in the metastatic cell line SW620. Similar findings concerning the disruption of core clock gene rhythmicity have recently been made for the breast cancer cell line MCF7 in comparison to the non-tumorigenic breast epithelial cell line MCF10A (Ye et al., 2018). These results support the hypothesis of the circadian clock acting as a tumor suppressor (Fu and Lee, 2003) and dysregulated circadian rhythmicity as a potential hallmark of cancer (El-Athman and Relógio, 2018).

4.2 Circadian Rhythmicity Differs between Primary Tumor- and Metastasis-derived Cells

In recent years, the advent of NGS techniques such as RNA-seq has gradually replaced microarrays as the method of choice for gene expression profiling, including the transcriptional profiling of circadian rhythms (Li et al., 2015). RNA-seq offers several advantages in comparison to microarrays, including a higher dynamic range of expression and the ability to measure alternative splice variants and other forms of RNA processing (Ozsolak and Milos, 2011). Nonetheless, the wealth of published microarray data in the circadian domain is still valuable for various meta-analysis studies, e.g., for benchmarking new algorithms for rhythm detection (Hutchison et al., 2018) or as training data for machine learning approaches (Hughey et al., 2016). Accordingly, comparative studies for different methods of gene expression profiling are common, however, they usually compare gene expression values, differentially expressed genes, and enriched pathways across platforms for matched single-time point datasets (Nookaew et al., 2012; Wolff et al., 2018; Zhao et al., 2014b) and do not consider circadian datasets.

In mammalian cells, oscillations of the core clock system are propagated to the expression of hundreds to thousands of clock-controlled target genes in a tissue-specific manner. A transcriptome-wide analysis of circadian rhythmicity in the CRC cell lines revealed that rhythmic gene sets and associated rhythmic parameters were highly dependent on the underlying platform used for profiling gene expression. Independent of using a parametric or a non-parametric algorithm for the detection of rhythmicity, considerably smaller sets of 24-h rhythmic genes were identified based on the RNA-seq data in comparison to the microarray data. Based on the microarray data, a higher number of 24-h rhythmic genes was

detected for SW480 cells than for SW620 cells, as expected from the diminished oscillations in core clock gene expression in the metastatic cell line. Surprisingly, the opposite case was observed for the RNA-seq data, where more rhythmic genes were detected in SW620 cells, despite the perturbed oscillations of core clock genes in this cell line. Moreover, the intersections between 24-h rhythmic gene sets identified for the same cell line based on the different platforms were very small. When comparing circadian parameters of 24-h rhythmic gene sets, further discrepancies between the platforms emerged. Even for genes that were commonly identified as rhythmic based on expression values of both platforms, the correlation between estimated amplitudes and phases was low, in particular for rhythmic genes detected in SW620 cells. The reasons for these discrepancies are currently unknown. A possible explanation could be the fact that the microarray and the RNA-seq datasets were not sampled during the same time frame. The microarray data was sampled immediately after synchronization of the cells for one circadian cycle, whereas sampling for the RNA-seq data started 12 h later for 1.25 circadian cycles. For many genes that were identified as rhythmic based on the microarray data, the expression at the overlapping time points of both datasets was similar, yet no second peak or trough was observed for later time points in the RNA-seq data. It is conceivable that the microarray sets contain true circadian genes whose amplitudes decrease dependent on how much time has passed since synchronization of the cells. Accordingly, a second peak or trough would not be as obvious as the first and could erroneously be perceived as constant expression. Bioluminescence measurements of *BMAL1* promoter activity in SW480 cells have shown that amplitudes of oscillations decrease over time due to a loss of synchronicity between cultured cells, but that rhythmicity is nonetheless robust for at least five days (Fuhr et al., 2018). In line with this observation, the oscillations of core clock genes as determined by RNA-seq data are robust for the whole time-series. Alternatively, the 24-h rhythmic microarray gene sets could also contain false positive circadian genes that show transient effects upon medium change, i.e., burst and decay in the expression of immediate early genes. Such transient effects can resemble circadian oscillations during the first 24 h but do not persist in a rhythmic manner when the synchronizing stimulus is absent (Hughes et al., 2017). It is also possible that both time-series taken individually are too short for the reliable identification of 24-h rhythmic genes. Simulations with synthetic time-series have shown that sampling periods shorter than two full circadian cycles make the analysis sensitive to outliers in the data and can lead to an increase of false negatives (Hughes et al., 2017). In order to gain more robust results that take the information from both microarray and RNA-seq measurements into account, the circadian datasets were normalized and concatenated across platforms, yielding a longer time-series with two replicate measurements for overlapping time points. For both cell lines, rhythmicity analysis of the concatenated dataset resulted in a greater number of 24-h rhythmic genes than for either of the platforms separately. Rhythmic parameters estimated based on the concatenated time-series were well correlated with parameters identified based either on the microarray or the RNA-seq data. It follows, that the normalization and concatenation of time-series gene expression data from two different platforms might prove a viable method for gaining robustly circadian genes.

In addition to the cross-platform analysis, rhythmic genes were also compared in detail between the two cell lines (based on the microarray data) to uncover changes in the circadian transcriptome upon tumor progression. SW480 and SW620 cells displayed distinct sets of 24-h rhythmic genes that were comparable in size but had only a small intersection. Moreover, the bimodal phase distribution of 24-h rhythmic genes differed between the cell lines, with slightly earlier peak times observed in SW620 cells compared to SW480 cells. Apparently, the observed disruption of the core clock system in SW620 cells led to a change in the identity and timing of clock-controlled target genes. However, it has to be taken into account that the comparison of sets of rhythmic features via set operations can be misleading due to an overstatement of the differences between the sets (Hughes et al., 2017). In line with this observation, it was shown that despite only having been identified as rhythmic in one of the CRC cell lines, many genes displayed similar variations in gene expression over time in the respective other cell line. To detect significant changes in the circadian transcriptome, a statistical analysis of differential rhythmicity was conducted between the two cell lines that directly compared the expression of 24-h rhythmic genes detected in either SW480 or SW620 cells, or both. For both cell lines, genes with a loss or gain of oscillations were identified, as well as commonly rhythmic genes whose peak phases were shifted between SW480 and SW620 cells. Interestingly, genes with significantly different rhythmic expression patterns between the two cancer stages were enriched for biological processes that are important for enhancing the metastatic potential of tumor cells, such as the regulation of smooth muscle cell proliferation and cell adhesion properties (Li et al., 2017; Okegawa et al., 2004). A phase-set enrichment analysis for 24-h rhythmic genes sets revealed additional cancer-associated pathways to be temporally coordinated, some of which displayed phase-shifts between SW480 and SW620 cells. In view of these findings, it can be assumed that tumor cells in different progression stages benefit from a dysregulated circadian system via the phase-shifted activation of cancer-relevant rhythmic pathways that might help them to, e.g., evade circadian immune surveillance (Sulli et al., 2019). However, due to the relatively low resolution of the sampling time, no definite conclusion concerning the effect of small phase-shifts in the timing of rhythmic biological pathways in tumor and metastatic cells can be drawn at this point.

4.3 Rhythmic AS Events in CRC are likely Regulated by Circadian Splicing Genes

Aberrant splicing events are a common characteristic of carcinogenesis and tumor progression. Various alternatively spliced isoforms detected in tumor cells constitute drivers of cancer hallmarks, such as resisting apoptosis, inducing invasion and metastasis, and regulating angiogenesis (El Marabti and Younis, 2018). Many of these cancer-specific alterations in splicing can be traced back to malfunctions in the splicing machinery and splicing regulatory networks (David and Manley, 2010; Grosso et al., 2008b). Interestingly, the spliceosome was among the temporally coordinated biological pathways found to be enriched for genes with a subtle but significant phase-shift between the CRC cell lines. In

view of the differences of the circadian transcriptome observed between SW480 and SW620 cells, the changes in rhythmicity of the spliceosome pathway might be the result of a malfunctioning clock.

Based on time-series microarray and RNA-seq data as well as a concatenation of both datasets, robust 24-h rhythms were detected in the expression of genes encoding for diverse spliceosomal components and SFs, including hnRNPs, SR proteins, and RBM proteins. Circadian rhythms in the expression of 15 SF genes have previously been identified in murine liver (McGlinchy et al., 2012). Some of the candidate circadian SFs from McGlinchy et al. (2012) were also found to be rhythmic in the CRC cell lines, including *DDX46*, *DHX9*, *HNRNPDL*, *KHDRBS1*, *SF3B1*, *SRSF3*, and *SRSF5*, suggesting that a circadian regulation of splicing might be conserved between different mammalian species. While some of the splicing-related genes were only rhythmically expressed in one of the CRC cell lines, others were oscillating in both, but displayed changes in amplitude and phase. For instance, the SF encoding genes *ESRP1*, *FUS*, and *SF3A3* tended to peak 2-3 h earlier in SW620 cells compared to SW480 cells. However, due to the relatively low resolution of the time-series, these subtle differences in phase were hard to quantify statistically. An experimental validation of their rhythmic transcription in SW480 and SW620 cells via high-resolution time-series qPCR could shed further light on putative differences in the circadian expression of splicing-related genes in CRC progression.

Changes in the expression levels of SFs have been observed in various human cancers and can influence the outcome of AS, even in the absence of mutations (Anczuków and Krainer, 2016). Thus, it seems likely that changes in the rhythmic expression of splicing-related genes in cancer might likewise be responsible for altering the AS outcome of their target genes. The majority of alterations of splicing events during CRC development and progression occur during the transition from normal to tumor cells, while other changes take place during the progression from primary tumor to metastatic cells (Bisognin et al., 2014). In the absence of healthy human reference data, the CRC cell line model aimed at the detection of the latter set of AS events. A differential splicing analysis of the transcriptome arrays revealed over two hundred genes with candidate differentially spliced exons between SW480 and SW620 cells. Many of the predicted AS events were involved in processes relevant to tumor progression, migration, and invasion. Prominent candidates included the neural cell adhesion molecule-encoding gene *NCAM1* and the EMT-associated genes *CD44* and *FGFR2*. Differential *NCAM1* isoform usage has previously been reported for SW480 and SW620 cells, and is linked to clinically aggressive CRC (Huerta et al., 2001). Defective AS events in malignant CRC are hypothesized to lead to impaired intracellular adhesion between colonocytes that enable the tumor to evolve to more aggressive stages (Huerta et al., 2001). AS of both *CD44* and *FGFR2* is regulated by the epithelial SFs *ESRP1* and *ESRP2* and has been associated with cell proliferation and invasion via EMT (Warzecha et al., 2009). As previously stated, rhythmic *ESRP1* expression displayed a shift in phase between SW480 and SW620 cells. A similar shift of peak expression between cell lines was observed for *ESRP2*, which was additionally expressed in antiphase to *ESRP1*. It is conceivable that the temporal shift in the peak expression of *ESRP1* and *ESRP2* during tumor progression leads to changes in AS of their target genes and subsequently to the production of isoforms that promote malignant transformation by enabling EMT. However, the overall expression of both *ESRP1* and *ESRP2* was higher in SW620 compared to SW480 cells

which makes it difficult to discriminate between effects that are caused by circadian phase shifts and those caused by increased expression levels.

In addition to the above described differential splicing events between different tumor stages, the time-series data of the CRC cell lines was also probed for 24-h rhythmic changes in AS. For the transcriptome array data, candidate exons with oscillating FIRMA scores were identified for each cell line separately. A similar analysis based on time-series exon arrays has previously been conducted to detect putative circadian AS in murine liver (McGlincy et al., 2012). In contrast to microarrays, RNA-seq allows for the quantification of expression at the resolution of individual splice variants (Wang et al., 2009). Following the assumption that circadian AS results in at least two different isoforms that are expressed at different times of the day, the analysis aimed at the detection of differentially rhythmic pairs of splice variants of the same gene that do not oscillate with the same phase. Both methods yielded a higher number of putative rhythmic AS events for SW480 cells with only a small number of candidate genes shared between the CRC cell lines. This might be taken as an indication that the circadian regulation of AS is altered in human tumor progression. Most exons that exhibited circadian rhythms in splicing in SW480 cells based on the microarray data, such as *ABCC1*, *PER2*, and *VEGFA*, lost them in the metastasis-derived cell line. Alternatively spliced isoforms of *ABCC1* play an important role in conferring drug resistance (He et al., 2004), while *VEGFA* isoforms can act as either pro- or anti-angiogenic factors, thus regulating the growth and metastatic ability of tumor cells (Kaida et al., 2012). A disruption in the rhythmic regulation of AS of these genes might enable tumor cells to constitutively express isoforms that favor the progression of cancer cells to a more malignant state. In contrast, the alternatively spliced candidate exon of the SF gene *SF3B1*, started to exhibit rhythmic splicing behavior in SW620 cells, suggesting a change in the regulation of SF expression levels during tumor progression via AS. In line with this finding, the differential rhythmicity analysis of the RNA-seq data revealed many phase-shifted pairs of splice variants for splicing-related genes in both cell lines. This opens up interesting possibilities for the mechanistic control of circadian regulated splicing in mammalian cells. Under the assumption that the circadian expression of splicing-related genes is translated to the protein level, circadian rhythms in the abundance of spliceosomal components and splicing factors might be responsible for the subsequent rhythmic AS of their own pre-mRNA in a self-regulatory system. Many SFs, hnRNPs, and core spliceosome components are known to autoregulate their own expression via AS and NMD pathways (McGlincy and Smith, 2008; Pervouchine et al., 2019). For instance, conserved sequences in SR genes that contain premature stop codons cause the production of mRNAs that are degraded by NMD in a process termed unproductive splicing (Lareau et al., 2007). The functional importance of alternative and unproductive splicing of SFs in cancer has been shown for the SR gene *SRSF1* which is involved in the regulation of EMT in SW480 cells via the production of a constitutively active splice variant of the proto-oncogene *RON* (Valacca et al., 2010). While *SRSF1* was among the robustly 24-h rhythmic genes in SW480 cells (with a decreased amplitude in SW620 cells), it was not among the candidates with phase-shifted splice variants. The observed fine-tuned changes in the expression levels of SFs and spliceosome components might in turn be required for the correct production of isoforms at a specific time-of-day. A similar downstream regulation of daytime-dependent AS

has previously been suggested based on rhythmic splicing events of splicing regulators observed in *Drosophila* neurons (Wang et al., 2018). Changes in the system, e.g., loss of oscillations or shifts in phase in the expression of SFs, might lead to the loss of isoforms that prevent and the gain of isoforms that favor the malignant development of cells. In line with this hypothesis, several interesting cancer-relevant phase-shifted isoforms were identified in the CRC cell lines, including transcripts from *ANKHD1* and *MYO1C*. Alternatively spliced isoforms of both genes are associated with cancer-related biological processes, namely cell survival and migration (Maly et al., 2017; Miles et al., 2005), again hinting at a possible role of rhythmic AS in CRC tumor progression. The daytime-specific production of cancer-relevant isoforms might enable cancer cells to avoid circadian immune system control. Isoform-specific knockdown or knockout experiments of selected candidate genes are likely to reveal more about the functional role of rhythmic AS in cancer.

Only two genes with candidate circadian AS events were identified for the same cell line by both platforms: *CAMK2G* and *PPP2R5C*. The surprisingly small number of common candidate genes might in part be ascribed to inherent differences of the technologies and the subsequent analysis methods. While the transcriptome array used for profiling gene expression of the CRC cell lines is splicing-sensitive and has been designed for the detection of different types of AS, the FIRMA algorithm only allows for the detection of cassette-type AS events (Purdom et al., 2008). Thus, it is possible that the analysis of the RNA-seq data has revealed more complex types of circadian AS events that were missed by the microarray analysis. Moreover, thresholds of statistical significance for circadian AS events are not directly comparable between methods, which might explain why so few microarray candidates were validated by the RNA-seq analysis. Nonetheless, the small number of shared candidate genes between the two platforms underlines the need for future validation experiments of candidate circadian AS events. To date, not much is known concerning possibly distinct isoform functions of *CAMK2G* and *PPP2R5C*, though alternatively spliced forms of *CAMK2G* have been identified in patients with myotonic dystrophy (Perfetti et al., 2014). Interestingly, *CAMK2G* was also one of two genes for which a candidate circadian AS event was detected for both SW480 and SW620 cells, making it a robust candidate for circadian AS in human colon tissues. Since *PPP2R5C* is known to act as a tumor suppressor via dephosphorylation of p53 (Nobumori et al., 2013), it can be surmised that an alternatively spliced isoform might function as a dominant negative inhibitor of the normally spliced transcript and thus play a role in tumor progression.

Taken together, the analysis of circadian microarray and RNA-seq data revealed robust 24-h rhythms in the expression of genes encoding for spliceosome components and SF in an *in vitro* model of CRC progression. Differences in rhythmicity between the two tumor stages were associated with changes in static and possibly circadian AS of target genes. The production of different isoforms at specific times of the daily cycle might in turn be responsible for the gain of oncogenic properties during tumor progression.

4.4 Rhythmic Splicing Genes and AS Events are Widespread in Mammalian Tissues

The previously discussed studies have revealed 24-h rhythms in the expression of splicing-related genes as well as in candidate AS events of human CRC cell lines, pointing at a circadian regulation of splicing that is altered during cancer progression. Following the analysis of circadian AS events in an *in vitro* model of a matched primary tumor/metastasis cell line pair, it was of interest to investigate whether a rhythmic regulation of splicing can also be observed *in vivo* in healthy tissues from other mammalian species. For this reason, a meta-analysis of two multi-organ circadian datasets from mouse (Zhang et al., 2014) and baboon (Mure et al., 2018) was conducted. The chosen datasets represent the most comprehensive circadian transcriptome datasets from diurnal and nocturnal mammals currently available. Since both time-series feature a relatively high sampling resolution (2 h), an additional search was conducted for putative ultradian rhythms in gene expression and AS with a period of 12 h.

For both species, circadian and to a lesser extent ultradian rhythms in gene expression were detected across tissues from all organ types. On average, ~8% of the identified oscillating genes per tissue in mice and 11% in baboon had a period of 12 h, which is a slightly higher percentage than reported for murine liver (Hughes et al., 2009). Conversely, in some tissues, 12-h rhythms in transcription were dominating, e.g., in murine adrenal gland, adipose and skeletal muscle tissues and baboon bone marrow, smooth muscle, adipose and genital tissues. These findings indicate that the prevalence of 12-h rhythms in transcription is heterogeneous across tissue types, and that ultradian rhythms might even be more prevalent than circadian rhythms in some mammalian organs. Since ultradian rhythms in gene expression were not the focus of this thesis, functional implications of the associated gene sets were not further investigated.

Several consistently 24-h rhythmic splicing-related genes were identified in both murine and baboon tissues. As expected, rhythms between diurnal and nocturnal species were shifted by approximately 12 h. Conserved 24-h rhythmic splicing-related genes included *FUS* and *SRSF5*, that both peaked during the respective active phase of the animals, as well as several heat shock protein-coding genes that peaked in the respective rest phase. While *FUS* has previously been reported to be transcriptionally regulated by the core clock protein NR1D1 (Jiang et al., 2018; Rogelj et al., 2012), expression of *SRSF5* is cold-inducible (Fujita et al., 2017), suggesting a body temperature-dependent control of splicing (Preußner et al., 2017). However, *SRSF5* and several heat shock protein-coding genes also displayed 24-h rhythmic patterns in transcription in the human CRC cells kept under constant temperature conditions, indicating that its oscillations are not solely due to temperature changes. Based on the baboon data, a set of more than 80 consistently 24-h rhythmic splicing-related genes was identified that displayed oscillations in expression with unimodally distributed phases in at least ten tissues. For all consistently rhythmic splicing-related genes taken together, a bimodal distribution of peak phases was observed, with two thirds of the genes peaking during the subjective daytime and one third of the genes peaking during subjective nighttime. These findings are in line with previous results of McGlinchy et al.

(2012) who also reported two clear clusters of peak phases for murine circadian SFs during the subjective day and night of the animals. Promoter regions of the set of daytime-peaking splicing-related baboon genes were not only significantly more GC-rich than those of arrhythmic splicing-related genes, but also enriched for the D-box motif, a clock transcription factor binding site. GC-rich motifs have previously been reported to be overrepresented in promoter regions of clock-controlled genes (Bozek et al., 2009), pointing to a possible regulation of the daytime-peaking splicing-related genes via first-order D-box binding clock-controlled transcription factors such as *DBP* and *TEF*.

The whole-transcript arrays used by Zhang et al. (2014) for profiling the circadian transcriptome in murine organs are not splicing-sensitive and thus only allow for the identification of putative AS events at gene- and not at exon-level. While the study by Zhang et al. (2014) also contains matched RNA-seq data for the same organs, the low sampling resolution of 6 h does not suffice to identify 12-h or 24-h rhythmic features with confidence and was therefore only used to determine the number of expressed splice variants per candidate gene. Several core clock and clock-controlled genes were among the candidates for which potential 24-h rhythmic AS events were detected in more than one murine tissue, including *Bmal2*, *Nr1d1*, and *Ciart*. A circadian regulation of splicing of an *Nr1d1* exon has previously been reported and experimentally validated in murine liver (McGlincy et al., 2012). The prevalence of clock-related genes among the candidates further supports the hypothesis of a reciprocal interplay between the circadian clock and AS in mammals. Genes with putative 12-h rhythmic AS events likewise included several clock-related genes, as well as *Xbp1*, a previously reported case for ultradian isoform production (Cretenet et al., 2010; Zhu et al., 2017). The detection of several known targets of rhythmically regulated AS among the candidate genes suggests that microarray analysis might be of use for the identification of putative AS events in the absence of better data. However, it has to be considered that the candidates detected by applying the FIRMAGene method do not always represent true AS events but can also be due to technical reasons, such as cross-hybridizing probes or probes having different intensity ranges (Robinson and Speed, 2009). Accordingly, candidate genes with rhythmic AS events identified based on the murine microarray data should be considered with reservations and must be independently validated in future studies.

In contrast to whole-transcript arrays, RNA-seq data allows for the quantification of individual isoforms (Zhang et al., 2017). Interesting discoveries were made when comparing the results of the rhythmicity analysis made on gene- and on transcript-level for the baboon tissues. While the correlation between the number of rhythmic genes and the number of rhythmic transcripts per tissue was roughly linear, the identity of rhythmic features differed depending on whether they were identified based on transcript counts or based on summarized gene counts. Some genes were classified as rhythmic without any of their expressed splice variants being likewise identified as rhythmic. Conversely, some genes were not defined as rhythmic, despite having oscillating splice variants. Overall, genes identified as rhythmic on both transcript- and gene-level had lower *p*-values according to non-parametric rhythmicity analysis than genes identified on gene-level only. Moreover, the sets of rhythmic transcripts displayed higher estimated amplitudes than the sets of rhythmic genes. Therefore, a rhythmicity analysis that takes transcript-level information into account could not only be of use for the

detection of differentially rhythmic isoforms but also for the detection of robustly oscillating genes. The transcript-level rhythmicity analysis of the baboon RNA-seq data revealed a multitude of genes with differentially 24-h rhythmic pairs of splice variants that had similar amplitudes but different phases. The majority of pairs peaked at opposing times of the daily cycle with a phase difference of more than 8 h, either in the middle of the subjective day or the middle of the subjective night. The observed bimodal distribution of splice variant phases indicates a possibly distinct functional role of the involved isoforms that depends on their expression level at a specific time-of-day. In contrast, genes with 12-h rhythmic phase shifted splice variants were far less prevalent and displayed no clear phase distribution across baboon tissues. Hence, it remains unclear whether they represent true rhythmically regulated AS events or whether they are simply the result of arrhythmic variations in transcription. In line with the results gained from the analysis of the CRC cell lines, splicing-related processes were found to be significantly enriched for genes with 24-h rhythmic phase-shifted splice variants. A subset of genes displayed identical pairs of phase-shifted splice variants across several baboon tissues. Phases of the individual splice variants were unimodally distributed across tissues of different organ types, pointing to a tightly controlled regulation of rhythmic isoform expression. Among the most frequently identified pairs of splice variants, there were various splicing-related genes, as well as two genes associated with functions in androgen activation, suggesting a possible role of circadian AS in steroid metabolism. Interestingly, androgen-receptor signaling was also one of the processes that was found to be enriched for 24-h rhythmic candidate AS events in SW480 cells. Several of the identified candidate genes with phase-shifted splice variants in baboon tissues are known to produce functionally distinct isoforms in human cells, some of which are associated with cancer-relevant processes such as cell survival, migration, and invasion. However, it still remains to be elucidated whether the specific isoforms and their functions are conserved between human and baboon.

In conclusion, phase-shifted splice variants of the same gene with a period of approximately 24 h are widespread across healthy mammalian tissues, supporting the hypothesis of a circadian regulation of AS that contributes to the temporal diversification of the proteome. Several of the candidate alternatively spliced genes encode for functionally distinct protein isoforms that might influence cellular processes in a time-of-day-dependent way, including the splicing process itself.

4.5 Limitations of the Study

The computational analyses conducted in the framework of this thesis provide solid evidence for a circadian regulation of mammalian splicing in both health and disease. Nevertheless, there are several limitations to the results and their extrapolation to other systems, which dictate caution in the biological interpretation and underline the need for further experimental validation.

In vitro cell line models are essential for the understanding of tumor biology and the development of possible treatment strategies. Large parts of the computational analyses performed in this study are based on experimental measurements of gene expression in two human CRC cell lines that were originally derived from the same patient. Due to their shared genetic background and their different tumor

stages, the cell lines SW480 and SW620 are a commonly used model system for the investigation of primary tumor and metastatic behavior and CRC progression (Chen et al., 2014a; Ma et al., 2014; Yeh et al., 2008). Yet, the cell line pair represents only the characteristics of the tumor and metastasis of one type of cancer of a single male patient. This is of particular importance in view of the strong sex-dependency of optimal treatment schedules observed for the chronomodulated delivery of chemotherapy in CRC patients (Giacchetti et al., 2006). Moreover, immortalized cancer cell lines that have been established several decades ago, have often accumulated additional genetic and epigenetic changes during cultivation passages (Boot et al., 2016; Lange et al., 2014). Accordingly, the observations made in the model system do not necessarily translate to general characteristics of CRC tumor progression in humans.

The interpretation of the differential rhythmicity observed between SW480 cells and SW620 cells is based on the assumption that the cell populations respond equally well to synchronizing cues. It is conceivable that instead of harboring a disrupted core clock, SW620 cells simply differ in their ability to be entrained via medium change. Preliminary results of clock protein measurements conducted in the group indicate that variations in the circadian phenotype also persist on single-cell level, suggesting that the observed differences in the circadian transcriptome are indeed due to inherent differences in the molecular core clock (El-Athman et al., 2018). However, further single-cell measurements are necessary to exclude the alternative. It would also be of interest to investigate how an entrainment of the cells by rhythmic temperature changes would influence the circadian transcriptome. Further limitations of the study can be traced back to the sampling schemes of the experimental data. Hughes et al. (2017) propose a sampling resolution of 2 h for two circadian cycles in order to identify robustly circadian genes with at least two peaks and/or troughs in expression. Due to financial limitations, the CRC cell lines were sampled with a resolution of 3 h for one full circadian cycle (microarray data) and for 1.25 circadian cycles (RNA-seq data). Discrepancies in the circadian transcriptomes assayed by microarrays and RNA-seq have already been discussed in detail (see subsection 4.2) and could partially be mitigated by the merging of both datasets. Nonetheless, a longer time-series profiled by RNA-seq would likely further reduce the false negative and false positive rates. The same problem also affects the analysis of candidate circadian features in the baboon tissues that were sampled for 24 h. Moreover, the sacrificed baboons were not of the same lineage, making it likely that many of the observed variations in gene expression are due to inter-individual differences, thereby masking true circadian genes.

Technical aspects of the platforms and the associated computational analysis likewise introduce a number of challenges to the study. Platform-specific limitations concerning the detection of AS events based on different types of microarrays have already been discussed (see subsection 4.3 and 4.4). In the last years, RNA-seq has largely replaced microarrays as the method of choice for the detection of AS. Nevertheless, the accurate detection and quantification of alternative splice variants in RNA-seq data remains a hard problem that requires a combination of sufficiently long reads, a high read coverage, and an appropriate number of replicates per condition. Simulation experiments have shown that the identification of the correct isoform becomes the more difficult, the higher the number of possible

alternative splice variants of a gene (Hayer et al., 2015). The design of high-throughput circadian transcriptome studies usually does not include replicates of the same time points (Sefer et al., 2016), making the detection of rhythmic splice variants a particularly challenging problem (Li et al., 2015). While the CRC cell line RNA-seq data produced for this thesis has a relatively high coverage (average of ~75 million reads per sample), the coverage of the baboon data is rather low (average of ~20 million reads per sample). For the baboon data, this problem was partially mitigated by focusing on candidate genes for which phase-shifted isoform pairs were identified in more than one tissue. The repeated occurrence of the same rhythmic splice variants increases the confidence in the validity of the candidate circadian AS event. However, AS is known to be tissue-specific (Baralle and Giudice, 2017), a fact that is likely also true for circadian AS (McGlinchy et al., 2012). Thus, the experimental design of future transcriptome studies on circadian AS events should implement further measures for the accurate detection of alternative splice isoforms, e.g., by including additional time point replicates and increasing coverage and/or RNA-seq read length (Li et al., 2015). Alternatively, one could aim to identify circadian AS on exon-level only, without considering the detection of specific splice variants (Wang et al., 2018).

Further ambiguities concerning the validity of the results are due to the nature of the statistical tests and filtering steps that have been conducted as part of the downstream computational analyses. As true for most bioinformatics analyses, the choice of algorithms, FDRs, and other cutoffs has a decisive influence on the ultimate results of the studies, e.g., on the number of candidate circadian genes, potential AS events, and enriched pathways. Accordingly, candidate rhythmic AS events should be experimentally validated in follow-up time-series studies, e.g., by isoform-specific RT-qPCR (Londoño and Philipp, 2016) or Western blot analysis with isoform-specific antibodies. Even if the existence of a candidate circadian AS event has been verified using independent technical approaches, information concerning the functional impact of specific splice variants is still sparse to date (Li et al., 2016). This is especially true for less-extensively studied model organisms such as the olive baboon, thus preventing a functional annotation on isoform-level. Future experimental studies that aim at targeted isoform depletion of candidate circadian splice variants might yield more information concerning the functional effects of rhythmic AS.

5 Conclusion and Outlook

The circadian clock controls the daily timing of a multitude of biological processes via transcriptional, posttranscriptional, translational, and posttranslational mechanisms that induce 24-h rhythms in the abundance of gene products. A circadian regulation of mRNA splicing has previously been postulated for several organisms, including mammals (McGlinchy et al., 2012; Wang et al., 2018). Since both dysregulations of the circadian clock and aberrant splicing are implicated in cancer development and progression, it is conceivable that the hypothesized circadian control of AS events likewise plays a role in cancer-relevant processes.

This thesis provides novel evidence for a widespread circadian regulation of AS in human cancer cell lines and healthy mammalian tissues. A computational analysis of circadian transcriptome data was conducted based on an *in vitro* model system of a matched tumor/metastasis-derived pair of human CRC cell lines as well as previously published reference datasets from healthy mammalian tissues. The study revealed differences in the circadian system between the CRC cell lines, both for the core clock and the overall circadian transcriptome. While the primary tumor-derived cell line SW480 displayed robust circadian rhythms in the expression of core clock genes, oscillations were diminished or lost in the metastasis-derived cell line SW620, suggesting a perturbation of the clock that is associated with tumor progression. On transcriptome level, circadian gene sets and parameters likewise differed between both cell lines. A functional enrichment analysis of circadian gene sets revealed changes in the identity and timing of rhythmically regulated biological pathways between the cell lines, including several cancer-relevant processes, as well as the spliceosome pathway. Across both CRC cell lines as well as in the reference mammalian tissues, 24-h rhythms were detected in the expression of genes encoding for a variety of spliceosome components, SFs, and other splicing-related proteins. Several of the candidate circadian regulated splicing-related genes were not only identified as robustly oscillating with similar phases across tissues of the same organism but were also found to be conserved across species, with an approximate phase shift of 12 h between nocturnal and diurnal mammals. In the CRC progression model, oscillations of splicing-related genes differed between primary tumor- and metastasis-derived cells, suggesting a subtle but significant phase shift in the postulated rhythmic activity of the spliceosome pathway.

In view of these findings and the known importance of SF expression patterns for splicing decisions, it seems likely that oscillations of splicing-related genes might be propagated to the protein-level and ultimately result in a rhythmic regulation of AS. Candidate 24-h rhythmic AS events were detected across all mammalian tissues and cell lines, providing further evidence for a circadian regulation of AS.

While there were some indications for ultradian splicing events with a period of 12 h in mammalian tissues, the sampling resolution of the experiments made it difficult to determine whether they truly exist or whether they are simply a result of biological and technical noise. Interestingly, for both the set of healthy primate tissues and the CRC cell lines, many of the candidate 24-h rhythmic AS events were identified for splicing-related genes. This finding suggests a partially autoregulatory mechanism behind the postulated circadian regulation of AS: Components of the core clock and circadian transcription factors could be responsible for oscillations in the expression of clock-controlled splicing-related genes. Oscillating SFs could in turn regulate rhythmic AS events of their own mRNAs, as well as of other splicing-related genes, resulting in the production of functionally distinct SF isoforms that influence the daytime-dependent splicing of downstream target genes. The functional implications of a temporal regulation of AS are manifold. Functionally distinct isoforms of the same gene could be produced at different times of the circadian cycle, localize in different cellular compartments, and participate in different biological pathways. Rhythmic AS that results in unproductive splicing at one time of the day could control the circadian expression of a protein-coding variant of the same gene, independent of circadian transcription. The anti-phasic production of truncated isoforms might, e.g., be responsible for occupying binding partners of canonical isoforms and thus interfere with complex formation in a circadian manner, acting as a time-dependent negative regulator. In addition to an independent experimental validation of the candidate circadian AS events, it still remains to ascertain whether the affected splice variants are indeed functionally relevant and thus contribute to the postulated temporal diversity of the proteome. Both in the CRC cell lines and the primate mammalian tissues, candidate circadian AS events were identified for a number of genes whose isoforms are associated with cancer-relevant processes, such as angiogenesis, migration, and cell survival, but also androgen signaling. From a translational perspective, the observed circadian regulation of AS in cancer opens up interesting possibilities concerning new strategies of cancer treatment that combine chronotherapeutical approaches with therapies that target cancer-specific splicing isoforms as potential drug targets.

Nevertheless, several open questions remain, before the initial hypothesis of a cancer-relevant circadian regulation of AS can be confirmed or rejected. Are circadian AS events characteristic to specific cancer types and stages? Do the produced isoforms have a functional effect or are they simply the by-product of an increasingly disrupted circadian system during cancer progression that leads to defects in splicing control? Further experiments in other human cancer cell lines could reveal whether the association between oscillating splicing genes and potentially cancer-relevant rhythmic AS events hold true for other *in vitro* model systems as well. Possible experimental models for such experiments are the human CRC cell line HCT116 which displays robust circadian oscillations in the promoter activity of *BMAL1* (Relógio et al., 2014), or the matched primary-metastasis triplet consisting of the cell lines Isreco1, Isreco2, and Isreco3 that are derived from a primary colon carcinoma and its corresponding liver and peritoneal metastases, respectively (Cojot et al., 1997). When following up the hypothesis that circadian regulated AS might be involved in androgen signaling, the prostate cell lines RC-77N/E and RC-77T/E, a matched pair of non-malignant and malignant tumor-derived cell lines from the same patient (Theodore et al., 2010), could prove a useful model system for future studies.

The proposed mechanism of circadian regulated AS likewise needs to be further explored. In order to determine whether clock components or circadian transcription factors are truly responsible for the observed rhythms in the expression of splicing-related genes and AS events, knockdown or knockout experiments of clock or clock-controlled genes are required. It would also be of interest to find out more about the downstream regulation driving circadian AS. A computational analysis of splicing-related RBP target genes using publicly available data from cross-linking immunoprecipitation experiments (Van Nostrand et al., 2018) could reveal whether the expression of circadian RBPs and the AS patterns of their target genes are correlated. In this way, candidate circadian regulated SFs could be linked to potential target genes. It also has to be considered that the results gained in this thesis are based on data that was produced in bulk experiments of synchronized cell populations and thus precludes the ability to analyze transcription at the level of single cells. Hence, it would be of interest to investigate whether different splice variants of the same gene are expressed at different times of the day in the same cell or whether the observed candidate circadian AS events are due to heterogeneity among cells. Technologically, single-cell RNA-seq for the detection of individual splice variants is hardly feasible at the moment (Arzalluz-Luque and Conesa, 2018), in particular when considering the additional challenge of reconstructing cycling pseudo-time-series patterns (Liu et al., 2017b). However, future developments in single-cell sequencing techniques and algorithms might make it possible to tackle this challenging problem.

In conclusion, this thesis provides novel and compelling evidence for a circadian regulation of alternative mRNA splicing in mammals via clock-controlled oscillations of splicing regulatory factors. It further highlights the potential implications of this temporal diversification of the transcriptome for the daytime-dependent control of biological pathways and illustrates how failures in the system might enable cells to acquire oncogenic properties that contribute to the development and progression of cancer.

6 Bibliography

- Adams, D.R., Ron, D., and Kiely, P.A. (2011). RACK1, A multifaceted scaffolding protein: Structure and function. *Cell communication and signaling* 9, 22.
- Affymetrix. Affymetrix Exon and Gene Array Glossary.
- Agostinelli, C., and Lund, U. (2017). R package circular: Circular Statistics (version 0.4–93), CA: Department of Environmental Sciences, Informatics and Statistics, Ca'Foscari University, Venice, Italy. UL: Department of Statistics, California Polytechnic State University, San Luis Obispo, California, USA.
- Alhopuro, P., Björklund, M., Sammalkorpi, H., Turunen, M., Tuupanen, S., Biström, M., Niittymäki, I., Lehtonen, H.J., Kivioja, T., and Launonen, V. (2010). Mutations in the circadian gene CLOCK in colorectal cancer. *Molecular Cancer Research* 8, 952-960.
- Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., and Roman-Roman, S. (2016). Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nature communications* 7, 10615.
- Anczuków, O., and Krainer, A.R. (2016). Splicing-factor alterations in cancers. *Rna* 22, 1285-1301.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166-169.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
- Anduaga, A.M., Evantal, N., Patop, I.L., Bartok, O., Weiss, R., and Kadener, S. (2019). Thermosensitive alternative splicing senses and mediates temperature adaptation in *Drosophila*. *eLife* 8, e44642.
- Arrigo, A.-P., and Gibert, B. (2013). Protein interactomes of three stress inducible small heat shock proteins: HspB1, HspB5 and HspB8. *International Journal of Hyperthermia* 29, 409-422.
- Arzalluz-Luque, Á., and Conesa, A. (2018). Single-cell RNAseq for the study of isoforms—how is that possible? *Genome biology* 19, 110.
- Aschoff, J. (1965). Circadian rhythms in man. *Science* 148, 1427-1432.
- Aschoff, J. (1981). Freerunning and entrained circadian rhythms. In *Biological rhythms* (Springer), pp. 81-93.
- Aschoff, J. (1983). Circadian control of body temperature. *Journal of thermal Biology* 8, 143-147.
- Aschoff, J., and Wever, R. (1962). Spontanperiodik des Menschen bei Ausschluss aller Zeitgeber. *Naturwissenschaften* 49, 337-342.
- Avitabile, D., Genovese, L., Ponti, D., Ranieri, D., Raffa, S., Calogero, A., and Torrisi, M.R. (2014). Nucleolar localization and circadian regulation of Per2S, a novel splicing variant of the Period 2 gene. *Cellular and molecular life sciences* 71, 2547-2559.
- Ballesta, A., Dulong, S., Abbara, C., Cohen, B., Okyar, A., Clairambault, J., and Levi, F. (2011). A combined experimental and mathematical approach for molecular-based optimization of irinotecan circadian delivery. *PLoS computational biology* 7, e1002143.
- Baralle, F.E., and Giudice, J. (2017). Alternative splicing as a regulator of development and tissue identity. *Nature Reviews Molecular Cell Biology* 18, 437.
- Bargiello, T.A., Jackson, F.R., and Young, M.W. (1984). Restoration of circadian behavioural rhythms by gene transfer in *Drosophila*. *Nature* 312, 752.
- Beersma, D.G., and Gordijn, M.C. (2007). Circadian control of the sleep–wake cycle. *Physiology & behavior* 90, 190-195.
- Bélanger, V., Picard, N., and Cermakian, N. (2006). The circadian regulation of Presenilin-2 gene expression. *Chronobiology international* 23, 747-766.

- Bell-Pedersen, D., Cassone, V.M., Earnest, D.J., Golden, S.S., Hardin, P.E., Thomas, T.L., and Zoran, M.J. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature Reviews Genetics* 6, 544-556.
- Bengtsson, H., Simpson, K., Bullard, J., and Hansen, K. (2008). aroma. affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory (tech report).
- Benna, C., Helfrich-Förster, C., Rajendran, S., Monticelli, H., Pilati, P., Nitti, D., and Mocellin, S. (2017). Genetic variation of clock genes and cancer risk: a field synopsis and meta-analysis. *Oncotarget* 8, 23978.
- Berg, K.C., Eide, P.W., Eilertsen, I.A., Johannessen, B., Bruun, J., Danielsen, S.A., Bjørnslett, M., Meza-Zepeda, L.A., Eknæs, M., and Lind, G.E. (2017). Multi-omics of 34 colorectal cancer cell lines—a resource for biomedical studies. *Molecular cancer* 16, 116.
- Berke, J.D., Sgambato, V., Zhu, P.-P., Lavoie, B., Vincent, M., Krause, M., and Hyman, S.E. (2001). Dopamine and glutamate induce distinct striatal splice forms of Ania-6, an RNA polymerase II-associated cyclin. *Neuron* 32, 277-287.
- Bhuiyan, S.A., Ly, S., Phan, M., Huntington, B., Hogan, E., Liu, C.C., Liu, J., and Pavlidis, P. (2018). Systematic evaluation of isoform function in literature reports of alternative splicing. *BMC genomics* 19, 637.
- Bieler, J., Cannavo, R., Gustafson, K., Gobet, C., Gatfield, D., and Naef, F. (2014). Robust synchronization of coupled circadian and cell cycle oscillators in single mammalian cells. *Molecular systems biology* 10, 739.
- Bishehsari, F., Levi, F., Turek, F.W., and Keshavarzian, A. (2016). Circadian rhythms in gastrointestinal health and diseases. *Gastroenterology* 151, e1-e5.
- Bisognin, A., Pizzini, S., Perilli, L., Esposito, G., Mocellin, S., Nitti, D., Zanovello, P., Bortoluzzi, S., and Mandruzzato, S. (2014). An integrative framework identifies alternative splicing events in colorectal cancer development. *Molecular oncology* 8, 129-141.
- Blencowe, B.J. (2006). Alternative splicing: new insights from global analyses. *Cell* 126, 37-47.
- Blencowe, B.J. (2017). The relationship between alternative splicing and proteomic complexity. *Trends in biochemical sciences* 42, 407-408.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.
- Bolstad, B.M., Irizarry, R.A., Åstrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185-193.
- Bonnal, S., Vigevani, L., and Valcárcel, J. (2012). The spliceosome as a target of novel antitumour drugs. *Nature reviews Drug discovery* 11, 847-859.
- Boot, A., van Eendenburg, J., Crobach, S., Ruano, D., Speetjens, F., Calame, J., Oosting, J., Morreau, H., and van Wezel, T. (2016). Characterization of novel low passage primary and metastatic colorectal cancer cell lines. *Oncotarget* 7, 14499.
- Boothroyd, C.E., Wijnen, H., Naef, F., Saez, L., and Young, M.W. (2007). Integration of light and temperature in the regulation of circadian gene expression in *Drosophila*. *PLoS genetics* 3, e54.
- Bourgon, R., Gentleman, R., and Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences* 107, 9546-9551.
- Bozek, K., Relógio, A., Kielbasa, S.M., Heine, M., Dame, C., Kramer, A., and Herzog, H. (2009). Regulation of clock-controlled genes in mammals. *PLoS one* 4, e4882.
- Braun, R., Kath, W.L., Iwanaszko, M., Kula-Eversole, E., Abbott, S.M., Reid, K.J., Zee, P.C., and Allada, R. (2018). Universal method for robust detection of circadian state from gene expression. *Proceedings of the National Academy of Sciences* 115, E9247-E9256.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* 68, 394-424.
- Brooks, A.N., Choi, P.S., de Waal, L., Sharifnia, T., Imielinski, M., Saksena, G., Peadarallu, C.S., Sivachenko, A., Rosenberg, M., and Chmielecki, J. (2014). A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. *PLoS one* 9, e87361.
- Brown, S.A. (2014). Circadian clock-mediated control of stem cell division and differentiation: beyond night and day. *Development* 141, 3105-3111.
- Brown, S.A., Ripperger, J., Kadener, S., Fleury-Olela, F., Vilbois, F., Rosbash, M., and Schibler, U. (2005). PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator. *Science* 308, 693-696.
- Bünning, E. (1932). Über die Erbllichkeit der Tagesperiodizität bei den Phaseolus Blättern. *Jb wiss Bot* 81, 411-418.
- Carmeliet, P. (2005). VEGF as a key mediator of angiogenesis in cancer. *Oncology* 69, 4-10.

- Carmo-Fonseca, M., and Kirchhausen, T. (2014). The timing of pre-mRNA splicing visualized in real-time. *Nucleus* 5, 11-14.
- Carvalho, B.S., and Irizarry, R.A. (2010). A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26, 2363-2367.
- Cederroth, C.R., Albrecht, U., Bass, J., Brown, S.A., Dyhrfeld-Johnsen, J., Gachon, F., Green, C.B., Hastings, M.H., Helfrich-Förster, C., and Hogenesch, J.B. (2019). Medicine in the Fourth Dimension. *Cell Metabolism* 30, 238-250.
- Chacko, E., and Ranganathan, S. (2009). Comprehensive splicing graph analysis of alternative splicing patterns in chicken, compared to human and mouse. *BMC genomics* 10, S5.
- Chen, B., Liu, Y., Jin, X., Lu, W., Liu, J., Xia, Z., Yuan, Q., Zhao, X., Xu, N., and Liang, S. (2014a). MicroRNA-26a regulates glucose metabolism by direct targeting PDHX in colorectal cancer cells. *BMC Cancer* 14, 443.
- Chen, H., Xia, M., Lin, M., Yang, H., Kuhlenkamp, J., Li, T., Sodir, N.M., Chen, Y.H., Josef-Lenz, H., and Laird, P.W. (2007). Role of methionine adenosyltransferase 2A and S-adenosylmethionine in mitogen-induced growth of human colon cancer cells. *Gastroenterology* 133, 207-218.
- Chen, L., Bush, S.J., Tovar-Corona, J.M., Castillo-Morales, A., and Urrutia, A.O. (2014b). Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Molecular biology and evolution* 31, 1402-1413.
- Chen, S.-T., Choo, K.-B., Hou, M.-F., Yeh, K.-T., Kuo, S.-J., and Chang, J.-G. (2005). Deregulated expression of the PER1, PER2 and PER3 genes in breast cancers. *Carcinogenesis* 26, 1241-1246.
- Cheng, Y.-S., Seibert, O., Klötting, N., Dietrich, A., Straßburger, K., Fernández-Veledo, S., Vendrell, J.J., Zorzano, A., Blüher, M., and Herzog, S. (2015). PPP2R5C couples hepatic glucose and lipid homeostasis. *PLoS genetics* 11, e1005561.
- Cho, Y.-S., Do, M.-H., Kwon, S.-Y., Moon, C., Kim, K., Lee, K., Lee, S.-J., Hemmi, S., Joo, Y.-E., and Kim, M.S. (2016). Efficacy of CD46-targeting chimeric Ad5/35 adenoviral gene therapy for colorectal cancers. *Oncotarget* 7, 38210.
- Climente-Gonzalez, H., Porta-Pardo, E., Godzik, A., and Eyra, E. (2017). The functional impact of alternative splicing in cancer. *Cell reports* 20, 2215-2226.
- Cohen, S.E., and Golden, S.S. (2015). Circadian rhythms in cyanobacteria. *Microbiology and Molecular Biology Reviews* 79, 373-385.
- Cojot, J.-F., Sordat, I., Silvestre, T., and Sordat, B. (1997). Differential display cloning identifies motility-related protein (MRP1/CD9) as highly expressed in primary compared to metastatic human colon carcinoma cells. *Cancer Research* 57, 2593-2597.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szczesniak, M.W., Gaffney, D.J., Elo, L.L., and Zhang, X. (2016). A survey of best practices for RNA-seq data analysis. *Genome biology* 17, 13.
- Cooper, D.L., and Dougherty, G.J. (1995). To metastasize or not? Selection of CD44 splice sites. *Nature medicine* 1, 635.
- Cornelissen, G. (2014). Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling* 11, 16.
- Cretenet, G., Le Clech, M., and Gachon, F. (2010). Circadian clock-coordinated 12 hr period rhythmic activation of the IRE1 α pathway controls lipid metabolism in mouse liver. *Cell metabolism* 11, 47-57.
- Crnko, S., Du Pré, B.C., Sluijter, J.P., and Van Laake, L.W. (2019). Circadian rhythms and the molecular clock in cardiovascular biology and disease. *Nature* 41569, 019-0167.
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I.M., Bennett, R., Bhai, J., Billis, K., and Boddu, S. (2018). Ensembl 2019. *Nucleic acids research* 47, D745-D751.
- Cvitkovic, I., and Jurica, M.S. (2012). Spliceosome database: a tool for tracking components of the spliceosome. *Nucleic acids research* 41, D132-D141.
- Dai, M., Wang, P., Boyd, A.D., Kostov, G., Athey, B., Jones, E.G., Bunney, W.E., Myers, R.M., Speed, T.P., and Akil, H. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic acids research* 33, e175-e175.
- Dallmann, R., Okyar, A., and Lévi, F. (2016). Dosing-time makes the poison: circadian regulation and pharmacotherapy. *Trends in molecular medicine* 22, 430-445.
- Davenport, A.P., Hyndman, K.A., Dhaun, N., Southan, C., Kohan, D.E., Pollock, J.S., Pollock, D.M., Webb, D.J., and Maguire, J.J. (2016). Endothelin. *Pharmacological reviews* 68, 357-418.
- David, C.J., and Manley, J.L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. *Genes & development* 24, 2343-2364.
- Davis, K., Roden, L.C., Leaner, V.D., and van der Watt, P.J. (2019). The tumour suppressing role of the circadian clock. *IUBMB life*.

- Davison, A.C. (2003). *Statistical models*, Vol 11 (Cambridge University Press).
- Davuluri, R.V., Suzuki, Y., Sugano, S., Plass, C., and Huang, T.H.-M. (2008). The functional consequences of alternative promoter use in mammalian genomes. *Trends in Genetics* 24, 167-177.
- de Candolle, A.P. (1832). *Physiologie végétale, ou Exposition des forces et des fonctions vitales des végétaux*, Vol 3 (Béchet jeune).
- de Mairan, J.-J. (1729). *Observation botanique. Histoire de l'Académie Royale des Sciences Paris*.
- Dibner, C., Schibler, U., and Albrecht, U. (2010). The mammalian circadian timing system: organization and coordination of central and peripheral clocks. *Annual review of physiology* 72, 517-549.
- Diernfellner, A., Colot, H.V., Dintsis, O., Loros, J.J., Dunlap, J.C., and Brunner, M. (2007). Long and short isoforms of *Neurospora* clock protein FRQ support temperature-compensated circadian rhythms. *FEBS letters* 581, 5759-5764.
- 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.
- Dolatshad, H., Pellagatti, A., Fernandez-Mercado, M., Yip, B.H., Malcovati, L., Attwood, M., Przychodzen, B., Sahgal, N., Kanapin, A.A., and Lockstone, H. (2015). Disruption of SF3B1 results in deregulated expression and splicing of key genes and pathways in myelodysplastic syndrome hematopoietic stem and progenitor cells. *Leukemia* 29, 1092.
- Dorman, S.N., Viner, C., and Rogan, P.K. (2014). Splicing mutation analysis reveals previously unrecognized pathways in lymph node-invasive breast cancer. *Scientific reports* 4.
- Dowhan, D.H., Hong, E.P., Auboeuf, D., Dennis, A.P., Wilson, M.M., Berget, S.M., and O'Malley, B.W. (2005). Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPER α and CAPER β . *Molecular cell* 17, 429-439.
- Duez, H., Van Der Veen, J.N., Duhem, C., Pourcet, B., Touvier, T., Fontaine, C., Derudas, B., Baugé, E., Havinga, R., and Bloks, V.W. (2008). Regulation of bile acid synthesis by the nuclear receptor Rev-erb α . *Gastroenterology* 135, 689-698. e685.
- Durinck, S., Bullard, J., Spellman, P.T., and Dudoit, S. (2009). GenomeGraphs: integrated genomic data visualization with R. *BMC bioinformatics* 10, 2.
- Eckel-Mahan, K.L., and Storm, D.R. (2009). Circadian rhythms and memory: not so simple as cogs and gears. *EMBO reports* 10, 584-591.
- El Marabti, E., and Younis, I. (2018). The Cancer Spliceome: reprogramming of alternative splicing in cancer. *Frontiers in molecular biosciences* 5.
- El-Athman, R., Fuhr, L., and Relógio, A. (2018). A Systems-Level Analysis Reveals Circadian Regulation of Splicing in Colorectal Cancer. *EBioMedicine*.
- El-Athman, R., Genov, N.N., Mazuch, J., Zhang, K., Yu, Y., Fuhr, L., Abreu, M., Li, Y., Wallach, T., and Kramer, A. (2017). The Ink4a/Arf locus operates as a regulator of the circadian clock modulating RAS activity. *PLoS biology* 15, e2002940.
- El-Athman, R., Knezevic, D., Fuhr, L., and Relógio, A. (2019). A Computational Analysis of Alternative Splicing across Mammalian Tissues Reveals Circadian and Ultradian Rhythms in Splicing Events. *International journal of molecular sciences* 20, 3977.
- El-Athman, R., and Relógio, A. (2018). Escaping circadian regulation: an emerging Hallmark of Cancer? *Cell systems* 6, 266-267.
- Evantal, N., Anduaga, A.M., Bartok, O., Patop, I.L., Weiss, R., and Kadener, S. (2018). Thermosensitive alternative splicing senses and mediates temperature adaptation in *Drosophila*. *bioRxiv*, 503409.
- Fang, F., Xia, N., Angulo, B., Carey, J., Cady, Z., Durruthy-Durruthy, J., Bennett, T., Sebastiano, V., and Pera, R.A.R. (2018). A distinct isoform of ZNF207 controls self-renewal and pluripotency of human embryonic stem cells. *Nature communications* 9.
- Farahani, E., Patra, H.K., Jangamreddy, J.R., Rashedi, I., Kawalec, M., Rao Pariti, R.K., Batakis, P., and Wiechec, E. (2014). Cell adhesion molecules and their relation to (cancer) cell stemness. *Carcinogenesis* 35, 747-759.
- Feillet, C., Krusche, P., Tamanini, F., Janssens, R.C., Downey, M.J., Martin, P., Teboul, M., Saito, S., Lévi, F.A., and Bretschneider, T. (2014). Phase locking and multiple oscillating attractors for the coupled mammalian clock and cell cycle. *Proceedings of the National Academy of Sciences* 111, 9828-9833.
- Feillet, C., Van Der Horst, G.T., Levi, F., Rand, D.A., and Delaunay, F. (2015). Coupling between the circadian clock and cell cycle oscillators: implication for healthy cells and malignant growth. *Frontiers in neurology* 6, 96.
- Filichkin, S.A., Cumbie, J.S., Dharmawardhana, P., Jaiswal, P., Chang, J.H., Palusa, S.G., Reddy, A., Megraw, M., and Mockler, T.C. (2015). Environmental stresses modulate abundance and timing of alternatively spliced circadian transcripts in *Arabidopsis*. *Molecular plant* 8, 207-227.

- Filipski, E., King, V.M., Li, X., Granda, T.G., Mormont, M.-C., Liu, X., Claustrat, B., Hastings, M.H., and Lévi, F. (2002). Host circadian clock as a control point in tumor progression. *Journal of the National Cancer Institute* 94, 690-697.
- Filipski, E., Subramanian, P., Carriere, J., Guettier, C., Barbason, H., and Levi, F. (2009). Circadian disruption accelerates liver carcinogenesis in mice. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* 680, 95-105.
- Foley, L., Ling, J., Joshi, R., Evantal, N., Kadener, S., and Emery, P. (2019). *Drosophila* PSI controls circadian period and the phase of circadian behavior under temperature cycle via tim splicing. *eLife* 8, e50063.
- Forster, S.C., Finkel, A.M., Gould, J.A., and Hertzog, P.J. (2013). RNA-eXpress annotates novel transcript features in RNA-seq data. *Bioinformatics* 29, 810-812.
- Fu, L., and Kettner, N.M. (2013). The circadian clock in cancer development and therapy. *Progress in molecular biology and translational science* 119, 221.
- Fu, L., and Lee, C.C. (2003). The circadian clock: pacemaker and tumour suppressor. *Nature reviews Cancer* 3, 350.
- Fuhr, L., El-Athman, R., Scrima, R., Cela, O., Carbone, A., Knoop, H., Li, Y., Hoffmann, K., Laukkanen, M.O., and Corcione, F. (2018). The Circadian Clock Regulates Metabolic Phenotype Rewiring Via HKDC1 and Modulates Tumor Progression and Drug Response in Colorectal Cancer. *EBioMedicine* 33, 105-121.
- Fujita, T., Higashitsuji, H., Higashitsuji, H., Liu, Y., Itoh, K., Sakurai, T., Kojima, T., Kandori, S., Nishiyama, H., and Fukumoto, M. (2017). TRPV4-dependent induction of a novel mammalian cold-inducible protein SRSF5 as well as CIRP and RBM3. *Scientific reports* 7, 2295.
- Gallego, M., and Virshup, D.M. (2007). Post-translational modifications regulate the ticking of the circadian clock. *Nature reviews Molecular cell biology* 8, 139.
- Gallego-Paez, L., Bordone, M., Leote, A., Saraiva-Agostinho, N., Ascensão-Ferreira, M., and Barbosa-Morais, N. (2017). Alternative splicing: the pledge, the turn, and the prestige. *Human Genetics*, 1-28.
- Gaucher, J., Montellier, E., and Sassone-Corsi, P. (2018). Molecular cogs: interplay between circadian clock and cell cycle. *Trends in cell biology* 28, 368-379.
- Gery, S., Komatsu, N., Baldjyan, L., Yu, A., Koo, D., and Koeffler, H.P. (2006). The circadian gene *per1* plays an important role in cell growth and DNA damage control in human cancer cells. *Molecular cell* 22, 375-382.
- Ghigna, C., Valacca, C., and Biamonti, G. (2008). Alternative splicing and tumor progression. *Current genomics* 9, 556-570.
- Giacchetti, S., Bjarnason, G., Garufi, C., Genet, D., Iacobelli, S., Tampellini, M., Smaaland, R., Focan, C., Coudert, B., and Humblet, Y. (2006). Phase III trial comparing 4-day chronomodulated therapy versus 2-day conventional delivery of fluorouracil, leucovorin, and oxaliplatin as first-line chemotherapy of metastatic colorectal cancer: the European Organisation for Research and Treatment of Cancer Chronotherapy Group. *Journal of Clinical Oncology* 24, 3562-3569.
- Giles, T.D. (2006). Circadian rhythm of blood pressure and the relation to cardiovascular events. *Journal of Hypertension* 24, S11-S16.
- Giulietti, M., Piva, F., D'Antonio, M., D'Onorio De Meo, P., Paoletti, D., Castrignano, T., D'Erchia, A.M., Picardi, E., Zambelli, F., and Principato, G. (2012). SpliceAid-F: a database of human splicing factors and their RNA-binding sites. *Nucleic acids research* 41, D125-D131.
- Gotic, I., Omid, S., Fleury-Olela, F., Molina, N., Naef, F., and Schibler, U. (2016). Temperature regulates splicing efficiency of the cold-inducible RNA-binding protein gene *Cirbp*. *Genes & development* 30, 2005-2017.
- Gotic, I., and Schibler, U. (2017). Posttranscriptional mechanisms controlling diurnal gene expression cycles by body temperature rhythms. *RNA biology* 14, 1294-1298.
- Gotoh, T., Kim, J.K., Liu, J., Vila-Caballer, M., Stauffer, P.E., Tyson, J.J., and Finkielstein, C.V. (2016). Model-driven experimental approach reveals the complex regulatory distribution of p53 by the circadian factor *Period 2*. *Proceedings of the National Academy of Sciences* 113, 13516-13521.
- Grandin, L.D., Alloy, L.B., and Abramson, L.Y. (2006). The social zeitgeber theory, circadian rhythms, and mood disorders: review and evaluation. *Clinical psychology review* 26, 679-694.
- Grant, C.E., Kurz, E.U., Cole, S.P., and Deeley, R.G. (1997). Analysis of the intron-exon organization of the human multidrug-resistance protein gene (MRP) and alternative splicing of its mRNA. *Genomics* 45, 368-378.
- Gréchez-Cassiau, A., Rayet, B., Guillaumond, F., Teboul, M., and Delaunay, F. (2008). The circadian clock component *BMAL1* is a critical regulator of *p21WAF1/CIP1* expression and hepatocyte proliferation. *Journal of Biological Chemistry* 283, 4535-4542.
- Greenham, K., and McClung, C.R. (2015). Integrating circadian dynamics with physiological processes in plants. *Nature Reviews Genetics* 16, 598.

- Grosso, A.R., Gomes, A.Q., Barbosa-Morais, N.L., Caldeira, S., Thorne, N.P., Grech, G., Von Lindern, M., and Carmo-Fonseca, M. (2008a). Tissue-specific splicing factor gene expression signatures. *Nucleic acids research* 36, 4823-4832.
- Grosso, A.R., Martins, S., and Carmo-Fonseca, M. (2008b). The emerging role of splicing factors in cancer. *EMBO reports* 9, 1087-1093.
- Gu, S., Papadopoulou, N., Nasir, O., Föllner, M., Alevizopoulos, K., Lang, F., and Stournaras, C. (2011). Activation of membrane androgen receptors in colon cancer inhibits the prosurvival signals Akt/bad in vitro and in vivo and blocks migration via vinculin/actin signaling. *Molecular Medicine* 17, 48-58.
- Gu, Y., Zhang, J., Ma, X., Kim, B.-w., Wang, H., Li, J., Pan, Y., Xu, Y., Ding, L., and Yang, L. (2017). Stabilization of the c-Myc protein by CAMKII γ promotes T cell lymphoma. *Cancer cell* 32, 115-128. e117.
- Guo, J., and Jia, R. (2019). Splicing factor poly (rC)-binding protein 1 is a novel and distinctive tumor suppressor. *Journal of cellular physiology* 234, 33-41.
- Guo, W., Calixto, C.P., Brown, J.W., and Zhang, R. (2017). TSIS: an R package to infer alternative splicing isoform switches for time-series data. *Bioinformatics* 33, 3308-3310.
- Halberg, F., Halberg, E., Barnum, C., and Bittner, J.J. (1959). Physiologic 24-hour periodicity in human beings and mice, the lighting regimen and daily routine. *Photoperiodism and related phenomena in plants and animals* 55, 803-878.
- Halberg, F., Tong, Y.L., and Johnson, E. (1967). Circadian system phase—an aspect of temporal morphology; procedures and illustrative examples. In *The cellular aspects of biorhythms* (Springer), pp. 20-48.
- Hamilton, T. (1969). Influence of environmental light and melatonin upon mammary tumour induction. *British Journal of Surgery* 56, 764-766.
- Han, S.P., Tang, Y.H., and Smith, R. (2010). Functional diversity of the hnRNPs: past, present and perspectives. *Biochemical Journal* 430, 379-392.
- Hansen, J. (2001). Increased breast cancer risk among women who work predominantly at night. *Epidemiology* 12, 74-77.
- Hao, S., Yao, L., Huang, J., He, H., Yang, F., Di, Y., Jin, C., and Fu, D. (2018). Genome-Wide Analysis Identified a Number of Dysregulated Long Noncoding RNA (lncRNA) in Human Pancreatic Ductal Adenocarcinoma. *Technology in cancer research & treatment* 17, 1533034617748429.
- Hardin, P.E., Hall, J.C., and Rosbash, M. (1990). Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. *Nature* 343, 536.
- Hastings, M., O'Neill, J.S., and Maywood, E.S. (2007). Circadian clocks: regulators of endocrine and metabolic rhythms. *Journal of Endocrinology* 195, 187-198.
- Hatami, R., Sieuwerts, A.M., Izadmehr, S., Yao, Z., Qiao, R.F., Papa, L., Look, M.P., Smid, M., Ohlssen, J., and Levine, A.C. (2013). KLF6-SV1 drives breast cancer metastasis and is associated with poor survival. *Science translational medicine* 5, 169ra112-169ra112.
- Hayer, K.E., Pizarro, A., Lahens, N.F., Hogenesch, J.B., and Grant, G.R. (2015). Benchmark analysis of algorithms for determining and quantifying full-length mRNA splice forms from RNA-seq data. *Bioinformatics* 31, 3938-3945.
- He, X., Ee, P.R., Coon, J.S., and Beck, W.T. (2004). Alternative splicing of the multidrug resistance protein 1/ATP binding cassette transporter subfamily gene in ovarian cancer creates functional splice variants and is associated with increased expression of the splicing factors PTB and SRp20. *Clinical cancer research* 10, 4652-4660.
- Hu, T., Li, Z., Gao, C.-Y., and Cho, C.H. (2016). Mechanisms of drug resistance in colon cancer and its therapeutic strategies. *World journal of gastroenterology* 22, 6876.
- Hua, H., Wang, Y., Wan, C., Liu, Y., Zhu, B., Yang, C., Wang, X., Wang, Z., Cornelissen-Guillaume, G., and Halberg, F. (2006). Circadian gene *mPer2* overexpression induces cancer cell apoptosis. *Cancer science* 97, 589-596.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37, 1-13.
- Huang, G., Zhou, Z., Wang, H., and Kleinerman, E.S. (2012). CAPER- α alternative splicing regulates the expression of vascular endothelial growth factor165 in Ewing sarcoma cells. *Cancer* 118, 2106-2116.
- Huang, Y. (2018). Structured Bayesian methods for splicing analysis in RNA-seq data.
- Huber, A.-L., Papp, S.J., Chan, A.B., Henriksson, E., Jordan, S.D., Kriebs, A., Nguyen, M., Wallace, M., Li, Z., and Metallo, C.M. (2016). CRY2 and FBXL3 cooperatively degrade c-MYC. *Molecular cell* 64, 774-789.
- Huerta, S., Srivatsan, E.S., Venkatesan, N., Peters, J., Moatamed, F., Renner, S., and Livingston, E.H. (2001). Alternative mRNA splicing in colon cancer causes loss of expression of neural cell adhesion molecule. *Surgery* 130, 834-843.

- Hughes, M.E., Abruzzi, K.C., Allada, R., Anafi, R., Arpat, A.B., Asher, G., Baldi, P., De Bekker, C., Bell-Pedersen, D., and Blau, J. (2017). Guidelines for genome-scale analysis of biological rhythms. *Journal of biological rhythms* 32, 380-393.
- Hughes, M.E., DiTacchio, L., Hayes, K.R., Vollmers, C., Pulivarthy, S., Baggs, J.E., Panda, S., and Hogenesch, J.B. (2009). Harmonics of circadian gene transcription in mammals. *PLoS genetics* 5, e1000442.
- Hughes, M.E., Grant, G.R., Paquin, C., Qian, J., and Nitabach, M.N. (2012a). Deep sequencing the circadian and diurnal transcriptome of *Drosophila* brain. *Genome research* 22, 1266-1281.
- Hughes, M.E., Hogenesch, J.B., and Kornacker, K. (2010). JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of biological rhythms* 25, 372-380.
- Hughes, M.E., Hong, H.-K., Chong, J.L., Indacochea, A.A., Lee, S.S., Han, M., Takahashi, J.S., and Hogenesch, J.B. (2012b). Brain-specific rescue of Clock reveals system-driven transcriptional rhythms in peripheral tissue. *PLoS genetics* 8, e1002835.
- Hughey, J.J., Hastie, T., and Butte, A.J. (2016). ZeitZeiger: supervised learning for high-dimensional data from an oscillatory system. *Nucleic acids research* 44, e80-e80.
- Hulsen, T., de Vlieg, J., and Alkema, W. (2008). BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC genomics* 9, 488.
- Huranová, M., Ivani, I., Benda, A., Poser, I., Brody, Y., Hof, M., Shav-Tal, Y., Neugebauer, K.M., and Staněk, D. (2010). The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. *The Journal of cell biology* 191, 75-86.
- Hutchison, A.L., Allada, R., and Dinner, A.R. (2018). Bootstrapping and empirical Bayes methods improve rhythm detection in sparsely sampled data. *Journal of biological rhythms* 33, 339-349.
- Hutchison, A.L., Maienschein-Cline, M., Chiang, A.H., Tabei, S.A., Gudjonson, H., Bahroos, N., Allada, R., and Dinner, A.R. (2015). Improved statistical methods enable greater sensitivity in rhythm detection for genome-wide data. *PLoS computational biology* 11, e1004094.
- IARC (2007). IARC Monographs Programme finds cancer hazards associated with shiftwork, painting and firefighting. World Health Organization, Lyon.
- Ikeda, M., and Nomura, M. (1997). cDNA cloning and tissue-specific expression of a novel basic helix-loop-helix/PAS protein (BMAL1) and identification of alternatively spliced variants with alternative translation initiation site usage. *Biochemical and biophysical research communications* 233, 258-264.
- Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., and Sivachenko, A. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150, 1107-1120.
- Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003a). Summaries of Affymetrix GeneChip probe level data. *Nucleic acids research* 31, e15-e15.
- Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003b). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249-264.
- James, A.B., Syed, N.H., Bordage, S., Marshall, J., Nimmo, G.A., Jenkins, G.I., Herzyk, P., Brown, J.W., and Nimmo, H.G. (2012). Alternative splicing mediates responses of the Arabidopsis circadian clock to temperature changes. *The Plant Cell* 24, 961-981.
- Jensen, L., and Cao, Y. (2013). Clock controls angiogenesis. *Cell Cycle* 12, 405-408.
- Jernberg, E., Thysell, E., Ylitalo, E.B., Rudolfsson, S., Crnalic, S., Widmark, A., Bergh, A., and Wikström, P. (2013). Characterization of prostate cancer bone metastases according to expression levels of steroidogenic enzymes and androgen receptor splice variants. *PloS one* 8, e77407.
- Jeyaraj, S., O'Brien, D.M., and Chandler, D.S. (2008). MDM2 and MDM4 splicing: an integral part of the cancer spliceome. *Frontiers in bioscience (Landmark edition)* 14, 2647-2656.
- Jia, R., Ajiro, M., Yu, L., McCoy, P., and Zheng, Z.-M. (2019). Oncogenic splicing factor SRSF3 regulates ILF3 alternative splicing to promote cancer cell proliferation and transformation. *RNA* 25, 630-644.
- Jiang, W., Zhao, S., Jiang, X., Zhang, E., Hu, G., Hu, B., Zheng, P., Xiao, J., Lu, Z., and Lu, Y. (2016). The circadian clock gene Bmal1 acts as a potential anti-oncogene in pancreatic cancer by activating the p53 tumor suppressor pathway. *Cancer letters* 371, 314-325.
- Jiang, X., Zhang, T., Wang, H., Wang, T., Qin, M., Bao, P., Wang, R., Liu, Y., Chang, H.-C., and Yan, J. (2018). Neurodegeneration-associated FUS is a novel regulator of circadian gene expression. *Translational neurodegeneration* 7, 24.
- Jöchl, W. (1964). Trends in photophysiological concepts. *Annals of the New York Academy of Sciences* 117, 88-104.

- Jonckheere, A.R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika* 41, 133-145.
- Kaida, D., Schneider-Poetsch, T., and Yoshida, M. (2012). Splicing in oncogenesis and tumor suppression. *Cancer science* 103, 1611-1616.
- Kalluri, R., and Weinberg, R.A. (2009). The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation* 119, 1420-1428.
- Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature structural & molecular biology* 14, 185-193.
- Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. *Gene* 514, 1-30.
- Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika* 30, 81-93.
- Kenific, C.M., Stehbens, S.J., Goldsmith, J., Leidal, A.M., Faure, N., Ye, J., Wittmann, T., and Debnath, J. (2016). NBR1 enables autophagy-dependent focal adhesion turnover. *J Cell Biol* 212, 577-590.
- Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nature Reviews Genetics* 11, 345-355.
- Kim, D., Langmead, B., and Salzberg, S.L. (2015a). HISAT: a fast spliced aligner with low memory requirements. *Nature methods* 12, 357.
- Kim, E., Ilagan, J.O., Liang, Y., Daubner, G.M., Lee, S.C.-W., Ramakrishnan, A., Li, Y., Chung, Y.R., Micol, J.-B., and Murphy, M.E. (2015b). SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell* 27, 617-630.
- Kim, E., Magen, A., and Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic acids research* 35, 125-131.
- Kim, J.Y., Lee, K.H., Shim, M.S., Shin, H., Xu, X.-M., Carlson, B.A., Hatfield, D.L., and Lee, B.J. (2010). Human selenophosphate synthetase 1 has five splice variants with unique interactions, subcellular localizations and expression patterns. *Biochemical and biophysical research communications* 397, 53-58.
- Knutsson, A. (2003). Health disorders of shift workers. *Occupational medicine* 53, 103-108.
- Ko, H.-K., Berk, M., Chung, Y.-M., Willard, B., Bareja, R., Rubin, M., Sboner, A., and Sharifi, N. (2018). Loss of an androgen-inactivating and isoform-specific HSD17B4 splice form enables emergence of castration-resistant prostate cancer. *Cell reports* 22, 809-819.
- Koike, N., Yoo, S.-H., Huang, H.-C., Kumar, V., Lee, C., Kim, T.-K., and Takahashi, J.S. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* 338, 349-354.
- Kojima, S., and Green, C.B. (2014). Circadian genomics reveal a role for post-transcriptional regulation in mammals. *Biochemistry* 54, 124-133.
- Konopka, R.J., and Benzer, S. (1971). Clock mutants of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* 68, 2112-2116.
- Kowalska, E., Ripperger, J.A., Hoegger, D.C., Bruegger, P., Buch, T., Birchler, T., Mueller, A., Albrecht, U., Contaldo, C., and Brown, S.A. (2013). NONO couples the circadian clock to the cell cycle. *Proceedings of the National Academy of Sciences* 110, 1592-1599.
- Krishnaiah, S.Y., Wu, G., Altman, B.J., Growe, J., Rhoades, S.D., Coldren, F., Venkataraman, A., Orlarerin-George, A.O., Francey, L.J., and Mukherjee, S. (2017). Clock regulation of metabolites reveals coupling between transcription and metabolism. *Cell metabolism* 25, 961-974. e964.
- Kumar, L., and Futschik, M.E. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2, 5.
- Kunická, T., and Souček, P. (2014). Importance of ABCC1 for cancer therapy and prognosis. *Drug metabolism reviews* 46, 325-342.
- Ladomery, M. (2013). Aberrant alternative splicing is another hallmark of cancer. *International journal of cell biology* 2013.
- Laing, E.E., Möller-Levet, C.S., Poh, N., Santhi, N., Archer, S.N., and Dijk, D.-J. (2017). Blood transcriptome based biomarkers for human circadian phase. *Elife* 6, e20214.
- Lange, F., Franz, B., Maletzki, C., Linnebacher, M., Hühns, M., and Jaster, R. (2014). Biological and molecular effects of small molecule kinase inhibitors on low-passage human colorectal cancer cell lines. *BioMed research international* 2014.
- Lareau, L.F., Inada, M., Green, R.E., Wengrod, J.C., and Brenner, S.E. (2007). Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446, 926.
- Larsson, J. (2018). eulerr: area-proportional Euler and Venn diagrams with ellipses. R Package Version 3.

- Le, K.-q., Prabhakar, B.S., Hong, W.-j., and Li, L.-c. (2015). Alternative splicing as a biomarker and potential target for drug discovery. *Acta Pharmacologica Sinica* 36, 1212.
- Lee, J., Park, E., Kim, G.H., Kwon, I., and Kim, K. (2018). A splice variant of human Bmal1 acts as a negative regulator of the molecular circadian clock. *Experimental & molecular medicine* 50, 159.
- Lee Phillips, M. (2005). What makes life tick: Taking apart the living clock. *BioScience* 55, 928-933.
- Lee, S., Donehower, L.A., Herron, A.J., Moore, D.D., and Fu, L. (2010). Disrupting circadian homeostasis of sympathetic signaling promotes tumor development in mice. *PLoS one* 5, e10995.
- Lenka, G., Tsai, M.-H., Lin, H.-C., Hsiao, J.-H., Lee, Y.-C., Lu, T.-P., Lee, J.-M., Hsu, C.-P., Lai, L.-C., and Chuang, E.Y. (2017). Identification of methylation-driven, differentially expressed STXBP6 as a novel biomarker in lung adenocarcinoma. *Scientific Reports* 7, 42573.
- Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC bioinformatics* 12, 323.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:13033997*.
- Li, J., Grant, G.R., Hogenesch, J.B., and Hughes, M.E. (2015). Considerations for RNA-seq analysis of circadian rhythms. In *Methods in enzymology* (Elsevier), pp. 349-367.
- Li, W., Liu, C.-C., Kang, S., Li, J.-R., Tseng, Y.-T., and Zhou, X.J. (2016). Pushing the annotation of cellular activities to a higher resolution: predicting functions at the isoform level. *Methods* 93, 110-118.
- Li, W.-W., Wang, H.-Y., Nie, X., Liu, Y.-B., Han, M., and Li, B.-H. (2017). Human colorectal cancer cells induce vascular smooth muscle cell apoptosis in an exocrine manner. *Oncotarget* 8, 62049.
- Liao, Y., Smyth, G.K., and Shi, W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923-930.
- Lim, C., and Allada, R. (2013). Emerging roles for post-transcriptional regulation in circadian clocks. *Nature neuroscience* 16, 1544-1550.
- Liu, J., Zou, X., Gotoh, T., Brown, A.M., Jiang, L., Wisdom, E.L., Kim, J.K., and Finkielstein, C.V. (2018). Distinct control of PERIOD2 degradation and circadian rhythms by the oncoprotein and ubiquitin ligase MDM2. *Sci Signal* 11, eaau0715.
- Liu, L., Xie, N., Sun, S., Plymate, S., Mostaghel, E., and Dong, X. (2014). Mechanisms of the androgen receptor splicing in prostate cancer cells. *Oncogene* 33, 3140.
- Liu, Y., and Bell-Pedersen, D. (2006). Circadian rhythms in *Neurospora crassa* and other filamentous fungi. *Eukaryotic cell* 5, 1184-1193.
- Liu, Y., Garceau, N.Y., Loros, J.J., and Dunlap, J.C. (1997). Thermally regulated translational control of FRQ mediates aspects of temperature responses in the *Neurospora* circadian clock. *Cell* 89, 477-486.
- Liu, Y., González-Porta, M., Santos, S., Brazma, A., Marioni, J.C., Aebersold, R., Venkitaraman, A.R., and Wickramasinghe, V.O. (2017a). Impact of alternative splicing on the human proteome. *Cell reports* 20, 1229-1241.
- Liu, Z., Lou, H., Xie, K., Wang, H., Chen, N., Aparicio, O.M., Zhang, M.Q., Jiang, R., and Chen, T. (2017b). Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. *Nature communications* 8, 22.
- Londoño, J.C., and Philipp, S.E. (2016). A reliable method for quantification of splice variants using RT-qPCR. *BMC molecular biology* 17, 8.
- Long, J.C., and Caceres, J.F. (2009). The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal* 417, 15-27.
- Lopez, A.J. (1998). Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annual review of genetics* 32, 279-305.
- Low, K.H., Lim, C., Ko, H.W., and Edery, I. (2008). Natural variation in the splice site strength of a clock gene and species-specific thermal adaptation. *Neuron* 60, 1054-1067.
- Loyer, P., Trembley, J.H., Grenet, J.A., Busson, A., Corlu, A., Zhao, W., Kocak, M., Kidd, V.J., and Lahti, J.M. (2008). Characterization of cyclin L1 and L2 interactions with CDK11 and splicing factors influence of cyclin L isoforms on splice site selection. *Journal of biological chemistry* 283, 7721-7732.
- Lück, S., Thurley, K., Thaben, P.F., and Westermarck, P.O. (2014). Rhythmic degradation explains and unifies circadian transcriptome and proteome data. *Cell reports* 9, 741-751.
- Lujan, D.A., Ochoa, J.L., and Hartley, R.S. (2018). Cold-inducible RNA binding protein in cancer and inflammation. *Wiley Interdisciplinary Reviews: RNA* 9, e1462.
- Lutz, C.S. (2008). Alternative polyadenylation: a twist on mRNA 3' end formation. *ACS chemical biology* 3, 609-617.

- Ma, Y.S., Yang, I.P., Tsai, H.L., Huang, C.W., Juo, S.H., and Wang, J.Y. (2014). High glucose modulates antiproliferative effect and cytotoxicity of 5-fluorouracil in human colon cancer cells. *DNA Cell Biol* 33, 64-72.
- Mack, G.A., and Wolfe, D.A. (1981). K-sample rank tests for umbrella alternatives. *Journal of the American Statistical Association* 76, 175-181.
- Majercak, J., Sidote, D., Hardin, P.E., and Edery, I. (1999). How a circadian clock adapts to seasonal decreases in temperature and day length. *Neuron* 24, 219-230.
- Maly, I.V., Domaradzki, T.M., Gosity, V.A., and Hofmann, W.A. (2017). Myosin isoform expressed in metastatic prostate cancer stimulates cell invasion. *Scientific reports* 7, 8476.
- Martin, R.M., Rino, J., Carvalho, C., Kirchhausen, T., and Carmo-Fonseca, M. (2013). Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell reports* 4, 1144-1155.
- Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fiset, J.-F., Revil, T., and Chabot, B. (2008). hnRNP Proteins and Splicing Control. In *Advances in Experimental Medicine & Biology*, p. 123.
- Matsuo, T., Yamaguchi, S., Mitsui, S., Emi, A., Shimoda, F., and Okamura, H. (2003). Control mechanism of the circadian clock for timing of cell division in vivo. *Science* 302, 255-259.
- Matys, V., Fricke, E., Geffers, R., Gößling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., and Kel-Margoulis, O.V. (2003). TRANSFAC®: transcriptional regulation, from patterns to profiles. *Nucleic acids research* 31, 374-378.
- Mazzoccoli, G., Panza, A., Valvano, M., Palumbo, O., Carella, M., Paziienza, V., Biscaglia, G., Tavano, F., Di Sebastiano, P., and Andriulli, A. (2011). Clock gene expression levels and relationship with clinical and pathological features in colorectal cancer patients. *Chronobiology international* 28, 841-851.
- Mazzoccoli, G., Vinciguerra, M., Papa, G., and Piepoli, A. (2014). Circadian clock circuitry in colorectal cancer. *World Journal of Gastroenterology: WJG* 20, 4197.
- McGlincy, N.J., and Smith, C.W. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in biochemical sciences* 33, 385-393.
- McGlincy, N.J., Valomon, A., Chesham, J.E., Maywood, E.S., Hastings, M.H., and Ule, J. (2012). Regulation of alternative splicing by the circadian clock and food related cues. *Genome biology* 13, 1.
- McLeay, R.C., and Bailey, T.L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC bioinformatics* 11, 165.
- McNamara, P., Seo, S.-b., Rudic, R.D., Sehgal, A., Chakravarti, D., and FitzGerald, G.A. (2001). Regulation of CLOCK and MOP4 by nuclear hormone receptors in the vasculature: a humoral mechanism to reset a peripheral clock. *Cell* 105, 877-889.
- Meijer, J.H., and Schwartz, W.J. (2003). In search of the pathways for light-induced pacemaker resetting in the suprachiasmatic nucleus. *Journal of biological rhythms* 18, 235-249.
- Melamud, E., and Moul, J. (2009). Stochastic noise in splicing machinery. *Nucleic acids research* 37, 4873-4886.
- 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.
- Mermet, J., Yeung, J., and Naef, F. (2017). Systems chronobiology: global analysis of gene regulation in a 24-hour periodic world. *Cold Spring Harbor perspectives in biology* 9, a028720.
- Miki, T., Matsumoto, T., Zhao, Z., and Lee, C.C. (2013). p53 regulates Period2 expression and the circadian clock. *Nature communications* 4, 2444.
- Miles, M.C., Janket, M.L., Wheeler, E.D., Chattopadhyay, A., Majumder, B., DeRicco, J., Schafer, E.A., and Ayyavoo, V. (2005). Molecular and functional characterization of a novel splice variant of ANKHD1 that lacks the KH domain and its role in cell survival and apoptosis. *The FEBS journal* 272, 4091-4102.
- Mitsui, A., Kumazawa, S., Takahashi, A., Ikemoto, H., Cao, S., and Arai, T. (1986). Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically. *Nature* 323, 720.
- Mitsui, S., Yamaguchi, S., Matsuo, T., Ishida, Y., and Okamura, H. (2001). Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism. *Genes & development* 15, 995-1006.
- Mohawk, J.A., Green, C.B., and Takahashi, J.S. (2012). Central and peripheral circadian clocks in mammals. *Annual review of neuroscience* 35, 445.
- Mollet, I., Barbosa-Morais, N.L., Andrade, J., and Carmo-Fonseca, M. (2006). Diversity of human U2AF splicing factors: Based on the EMBO Lecture delivered on 7 July 2005 at the 30th FEBS Congress in Budapest. *The FEBS journal* 273, 4807-4816.
- Moore, R.Y., and Eichler, V.B. (1972). Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. *Brain research*.

- Mormont, M.-C., Waterhouse, J., Bleuzen, P., Giacchetti, S., Jami, A., Bogdan, A., Lellouch, J., Misset, J.-L., Touitou, Y., and Lévi, F. (2000). Marked 24-h rest/activity rhythms are associated with better quality of life, better response, and longer survival in patients with metastatic colorectal cancer and good performance status. *Clinical Cancer Research* 6, 3038-3045.
- Mteyrek, A., Filipinski, E., Guettier, C., Oklejewicz, M., van der Horst, G.T., Okyar, A., and Lévi, F. (2017). Critical cholangiocarcinogenesis control by cryptochrome clock genes. *International journal of cancer* 140, 2473-2483.
- Mure, L.S., Le, H.D., Benegiamo, G., Chang, M.W., Rios, L., Jillani, N., Ngotho, M., Kariuki, T., Dkhissi-Benyahya, O., and Cooper, H.M. (2018). Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science*, eaao0318.
- Na, J., Jung, J., Bang, J., Lu, Q., Carlson, B.A., Guo, X., Gladyshev, V.N., Kim, J., Hatfield, D.L., and Lee, B.J. (2018). Selenophosphate synthetase 1 and its role in redox homeostasis, defense and proliferation. *Free Radical Biology and Medicine* 127, 190-197.
- Nagoshi, E., Saini, C., Bauer, C., Laroche, T., Naef, F., and Schibler, U. (2004). Circadian gene expression in individual fibroblasts: cell-autonomous and self-sustained oscillators pass time to daughter cells. *Cell* 119, 693-705.
- Nakarai, C., Osawa, K., Akiyama, M., Matsubara, N., Ikeuchi, H., Yamano, T., Hirota, S., Tomita, N., Usami, M., and Kido, Y. (2015). Expression of AKR1C3 and CNN3 as markers for detection of lymph node metastases in colorectal cancer. *Clinical and experimental medicine* 15, 333-341.
- Nariai, N., Hirose, O., Kojima, K., and Nagasaki, M. (2013). TIGAR: transcript isoform abundance estimation method with gapped alignment of RNA-Seq data by variational Bayesian inference. *Bioinformatics* 29, 2292-2299.
- Narla, G., DiFeo, A., Fernandez, Y., Dhanasekaran, S., Huang, F., Sangodkar, J., Hod, E., Leake, D., Friedman, S.L., and Hall, S.J. (2008). KLF6-SV1 overexpression accelerates human and mouse prostate cancer progression and metastasis. *The Journal of clinical investigation* 118, 2711-2721.
- Nguyen, T.H.D., Galej, W.P., Fica, S.M., Lin, P.-C., Newman, A.J., and Nagai, K. (2016). CryoEM structures of two spliceosomal complexes: starter and dessert at the spliceosome feast. *Current opinion in structural biology* 36, 48-57.
- Nishioka, Y., Imaizumi, H., Imada, J., Katahira, J., Matsuura, N., and Hieda, M. (2016). SUN1 splice variants, SUN1_888, SUN1_785, and predominant SUN1_916, variably function in directional cell migration. *Nucleus* 7, 572-584.
- Nobumori, Y., Shouse, G.P., Wu, Y., Lee, K.J., Shen, B., and Liu, X. (2013). B56 γ tumor-associated mutations provide new mechanisms for B56 γ -PP2A tumor suppressor activity. *Molecular Cancer Research* 11, 995-1003.
- Nookaew, I., Papini, M., Pornputtapong, N., Scalcinati, G., Fagerberg, L., Uhlén, M., and Nielsen, J. (2012). A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic acids research* 40, 10084-10097.
- Noone, A., Howlader, N., Krapcho, M., Miller, D., Brest, A., Yu, M., Ruhl, J., Tatalovich, Z., Mariotto, A., and Lewis, D. (2018). SEER cancer statistics review, 1975–2015. Bethesda, MD: National Cancer Institute.
- Nowak, D.G., Amin, E.M., Rennel, E.S., Hoareau-Aveilla, C., Gammons, M., Damodoran, G., Hagiwara, M., Harper, S.J., Woolard, J., and Lodomery, M.R. (2010). Regulation of vascular endothelial growth factor (VEGF) splicing from pro-angiogenic to anti-angiogenic isoforms a novel therapeutic strategy for angiogenesis. *Journal of Biological Chemistry* 285, 5532-5540.
- Nueda, M.J., Martorell-Marugan, J., Martí, C., Tarazona, S., and Conesa, A. (2017). Identification and visualization of differential isoform expression in RNA-seq time series. *Bioinformatics* 34, 524-526.
- Oda, A., Katayose, Y., Yabuuchi, S., Yamamoto, K., Mizuma, M., Shirasou, S., Onogawa, T., Ohtsuka, H., Yoshida, H., and Hayashi, H. (2009). Clock gene mouse period2 overexpression inhibits growth of human pancreatic cancer cells and has synergistic effect with cisplatin. *Anticancer research* 29, 1201-1209.
- Oh, S., Song, S., Dasgupta, N., and Grabowski, G. (2014). The analytical landscape of static and temporal dynamics in transcriptome data. *Frontiers in genetics* 5, 35.
- Okegawa, T., Pong, R.-C., Li, Y., and Hsieh, J.-T. (2004). The role of cell adhesion molecule in cancer progression and its application in cancer therapy. *Acta Biochimica Polonica-English Edition*- 51, 445-458.
- Oltean, S., and Bates, D. (2014). Hallmarks of alternative splicing in cancer. *Oncogene* 33, 5311.
- Oltean, S., Sorg, B.S., Albrecht, T., Bonano, V.I., Brazas, R.M., Dewhirst, M.W., and Garcia-Blanco, M.A. (2006). Alternative inclusion of fibroblast growth factor receptor 2 exon IIIc in Dunning prostate tumors reveals unexpected epithelial mesenchymal plasticity. *Proceedings of the National Academy of Sciences* 103, 14116-14121.
- Oshima, T., Takenoshita, S., Akaike, M., Kunisaki, C., Fujii, S., Nozaki, A., Numata, K., Shiozawa, M., Rino, Y., and Tanaka, K. (2011). Expression of circadian genes correlates with liver metastasis and outcomes in colorectal cancer. *Oncology reports* 25, 1439-1446.

- Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics* *12*, 87.
- Ozturk, N., Lee, J.H., Gaddameedhi, S., and Sancar, A. (2009). Loss of cryptochrome reduces cancer risk in p53 mutant mice. *Proceedings of the National Academy of Sciences* *106*, 2841-2846.
- Ozturk, N., Ozturk, D., Kavakli, I.H., and Okyar, A. (2017). Molecular Aspects of Circadian Pharmacology and Relevance for Cancer Chronotherapy. *International journal of molecular sciences* *18*, 2168.
- Palmer, J.D. (1995). Review of the dual-clock control of tidal rhythms and the hypothesis that the same clock governs both circatidal and circadian rhythms. *Chronobiology International* *12*, 299-310.
- Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics* *40*, 1413-1415.
- Partch, C.L., Green, C.B., and Takahashi, J.S. (2014). Molecular architecture of the mammalian circadian clock. *Trends in cell biology* *24*, 90-99.
- Pastushenko, I., and Blanpain, C. (2018). EMT transition states during tumor progression and metastasis. *Trends in cell biology*.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2016). Salmon provides accurate, fast, and bias-aware transcript expression estimates using dual-phase inference. *bioRxiv*, 021592.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* *14*, 417.
- Patton, A.P., and Hastings, M.H. (2018). The suprachiasmatic nucleus. *Current Biology* *28*, R816-R822.
- Pembroke, W.G., Babbs, A., Davies, K.E., Ponting, C.P., and Oliver, P.L. (2015). Temporal transcriptomics suggest that twin-peaking genes reset the clock. *Elife* *4*, e10518.
- Perez-Santángelo, S., Mancini, E., Francey, L.J., Schlaen, R.G., Chernomoretz, A., Hogenesch, J.B., and Yanovsky, M.J. (2014). Role for LSM genes in the regulation of circadian rhythms. *Proceedings of the National Academy of Sciences* *111*, 15166-15171.
- Perez-Santángelo, S., Schlaen, R.G., and Yanovsky, M.J. (2012). Genomic analysis reveals novel connections between alternative splicing and circadian regulatory networks. *Briefings in functional genomics* *12*, 13-24.
- Perfetti, A., Greco, S., Fasanaro, P., Bugiardini, E., Cardani, R., Manteiga, J.M.G., Riba, M., Cittaro, D., Stupka, E., and Meola, G. (2014). Genome wide identification of aberrant alternative splicing events in myotonic dystrophy type 2. *PLoS one* *9*, e93983.
- Pervouchine, D., Popov, Y., Berry, A., Borsari, B., Frankish, A., and Guigó, R. (2019). Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. *Nucleic acids research* *47*, 5293-5306.
- Pickrell, J.K., Pai, A.A., Gilad, Y., and Pritchard, J.K. (2010). Noisy splicing drives mRNA isoform diversity in human cells. *PLoS genetics* *6*, e1001236.
- Pittendrigh, C.S. (1954). On temperature independence in the clock system controlling emergence time in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* *40*, 1018.
- Pittendrigh, C.S. (1960). Circadian rhythms and the circadian organization of living systems. Paper presented at: Cold Spring Harbor symposia on quantitative biology (Cold Spring Harbor Laboratory Press).
- Pittendrigh, C.S. (1993). Temporal organization: reflections of a Darwinian clock-watcher. *Annual review of physiology* *55*, 17-54.
- Preitner, N., Damiola, F., Zakany, J., Duboule, D., Albrecht, U., and Schibler, U. (2002). The orphan nuclear receptor REV-ERB α controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* *110*, 251-260.
- Prendergast, B.J., and Zucker, I. (2016). Ultradian rhythms in mammalian physiology and behavior. *Current opinion in neurobiology* *40*, 150-154.
- Preußner, M., Goldammer, G., Neumann, A., Haltenhof, T., Rautenstrauch, P., Müller-McNicoll, M., and Heyd, F. (2017). Body temperature cycles control rhythmic alternative splicing in mammals. *Molecular cell* *67*, 433-446. e434.
- Preußner, M., and Heyd, F. (2016). Post-transcriptional control of the mammalian circadian clock: implications for health and disease. *Pflügers Archiv-European Journal of Physiology*, 1-9.
- Preußner, M., and Heyd, F. (2018). Temperature-controlled Rhythmic Gene Expression in Endothermic Mammals: All Diurnal Rhythms are Equal, but Some are Circadian. *BioEssays* *40*, 1700216.
- Preußner, M., Wilhelmi, I., Schultz, A.-S., Finkernagel, F., Michel, M., Möröy, T., and Heyd, F. (2014). Rhythmic U2af26 alternative splicing controls PERIOD1 stability and the circadian clock in mice. *Molecular cell* *54*, 651-662.

- Purdom, E., Simpson, K.M., Robinson, M.D., Conboy, J., Lapuk, A., and Speed, T.P. (2008). FIRMA: a method for detection of alternative splicing from exon array data. *Bioinformatics* 24, 1707-1714.
- Quinlan, A.R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. *Current protocols in bioinformatics* 47, 11.12. 11-11.12. 34.
- Rahman, H.A., Wu, H., Dong, Y., Pasula, S., Wen, A., Sun, Y., Brophy, M.L., Tessneer, K.L., Cai, X., and McManus, J. (2016). Selective Targeting of a Novel Epsin-VEGFR2 Interaction Promotes VEGF-Mediated Angiogenesis Novelty and Significance. *Circulation research* 118, 957-969.
- Reeb, S.G. (2002). Plasticity of diel and circadian activity rhythms in fishes. *Reviews in Fish Biology and Fisheries* 12, 349-371.
- Refinetti, R. (2010). *Circadian physiology* (CRC press).
- Reinke, H., and Asher, G. (2019). Crosstalk between metabolism and circadian clocks. *Nature Reviews Molecular Cell Biology*, 1.
- Relógio, A., Ben-Dov, C., Baum, M., Ruggiu, M., Gemund, C., Benes, V., Darnell, R.B., and Valcárcel, J. (2005). Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *Journal of Biological Chemistry* 280, 4779-4784.
- Relógio, A., Thomas, P., Medina-Perez, P., Reischl, S., Bervoets, S., Gloc, E., Riemer, P., Mang-Fatehi, S., Maier, B., and Schäfer, R. (2014). Ras-mediated deregulation of the circadian clock in cancer. *PLoS genetics* 10, e1004338.
- Riley, L.A., Mijares, J.R., Zhang, X., and Esser, K.A. (2018). The Skeletal Muscle Molecular Clock Regulates Titin Splicing and Protein Expression. *Biophysical Journal* 114, 137a.
- 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.
- Robinson, M.D., and Speed, T.P. (2009). Differential splicing using whole-transcript microarrays. *BMC bioinformatics* 10, 156.
- Roenneberg, T., and Mellow, M. (2005). Circadian clocks—the fall and rise of physiology. *Nature Reviews Molecular Cell Biology* 6, 965-971.
- Roenneberg, T., and Mellow, M. (2016). The Circadian Clock and Human Health. *Current Biology* 26, R432-R443.
- Roenneberg, T., Wirz-Justice, A., and Mellow, M. (2003). Life between clocks: daily temporal patterns of human chronotypes. *Journal of biological rhythms* 18, 80-90.
- Roesler, J., Srivatsan, E., Moatamed, F., Peters, J., and Livingston, E.H. (1997). Tumor suppressor activity of neural cell adhesion molecule in colon carcinoma. *The American journal of surgery* 174, 251-257.
- Rogelj, B., Easton, L.E., Bogu, G.K., Stanton, L.W., Rot, G., Curk, T., Zupan, B., Sugimoto, Y., Modic, M., and Haberman, N. (2012). Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Scientific reports* 2, 603.
- Rokavec, M., Horst, D., and Hermeking, H. (2017). Cellular model of colon cancer progression reveals signatures of mRNAs, miRNA, lncRNAs, and epigenetic modifications associated with metastasis. *Cancer research* 77, 1854-1867.
- Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016.
- Ruan, H., Zhan, Y., Hou, J., Xu, B., Chen, B., Tian, Y., Wu, D., Zhao, Y., Zhang, Y., and Chen, X. (2017). Berberine binds RXR α to suppress β -catenin signaling in colon cancer cells. *Oncogene* 36, 6906.
- Sanchez, S.E., Petrillo, E., Beckwith, E.J., Zhang, X., Rugnone, M.L., Hernando, C.E., Cuevas, J.C., Herz, M.A.G., Depetris-Chauvin, A., and Simpson, C.G. (2010). A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature* 468, 112-116.
- Sanchez-Alavez, M., Alboni, S., and Conti, B. (2011). Sex- and age-specific differences in core body temperature of C57Bl/6 mice. *Age* 33, 89-99.
- Sato, T.K., Panda, S., Miraglia, L.J., Reyes, T.M., Rudic, R.D., McNamara, P., Naik, K.A., FitzGerald, G.A., Kay, S.A., and Hogenesch, J.B. (2004). A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron* 43, 527-537.
- Scheiermann, C., Kunisaki, Y., and Frenette, P.S. (2013). Circadian control of the immune system. *Nature Reviews Immunology* 13, 190.
- Schernhammer, E.S., Laden, F., Speizer, F.E., Willett, W.C., Hunter, D.J., Kawachi, I., and Colditz, G.A. (2001). Rotating night shifts and risk of breast cancer in women participating in the nurses' health study. *Journal of the National Cancer Institute* 93, 1563-1568.

- Schernhammer, E.S., Laden, F., Speizer, F.E., Willett, W.C., Hunter, D.J., Kawachi, I., Fuchs, C.S., and Colditz, G.A. (2003). Night-shift work and risk of colorectal cancer in the nurses' health study. *Journal of the National Cancer Institute* 95, 825-828.
- Scheving, L.A. (2000). Biological clocks and the digestive system. *Gastroenterology* 119, 536-549.
- Schlaen, R.G., Mancini, E., Sanchez, S.E., Perez-Santángelo, S., Rugnone, M.L., Simpson, C.G., Brown, J.W., Zhang, X., Chernomoretz, A., and Yanovsky, M.J. (2015). The spliceosome assembly factor GEMIN2 attenuates the effects of temperature on alternative splicing and circadian rhythms. *Proceedings of the National Academy of Sciences* 112, 9382-9387.
- Schoenhard, J.A., Eren, M., Johnson, C.H., and Vaughan, D.E. (2002). Alternative splicing yields novel BMAL2 variants: tissue distribution and functional characterization. *American Journal of Physiology-Cell Physiology* 283, C103-C114.
- Sebestyén, E., Singh, B., Miñana, B., Pagès, A., Mateo, F., Pujana, M.A., Valcárcel, J., and Eyra, E. (2016). Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome research* 26, 732-744.
- Sefer, E., Kleyman, M., and Bar-Joseph, Z. (2016). Tradeoffs between dense and replicate sampling strategies for high-throughput time series experiments. *Cell systems* 3, 35-42.
- Seiler, M., Peng, S., Agrawal, A.A., Palacino, J., Teng, T., Zhu, P., Smith, P.G., Caesar-Johnson, S.J., Demchok, J.A., and Felau, I. (2018). Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. *Cell reports* 23, 282-296. e284.
- Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., and Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF 65 and SF1/mBBP. *Molecular cell* 11, 965-976.
- Seo, P.J., Park, M.-J., Lim, M.-H., Kim, S.-G., Lee, M., Baldwin, I.T., and Park, C.-M. (2012). A self-regulatory circuit of CIRCADIAN CLOCK-ASSOCIATED1 underlies the circadian clock regulation of temperature responses in Arabidopsis. *The Plant Cell* 24, 2427-2442.
- Sephton, S.E., Sapolsky, R.M., Kraemer, H.C., and Spiegel, D. (2000). Diurnal cortisol rhythm as a predictor of breast cancer survival. *Journal of the National Cancer Institute* 92, 994-1000.
- Shakhmantsir, I., and Sehgal, A. (2019). Splicing the Clock to Maintain and Entrain Circadian Rhythms. *Journal of biological rhythms*, 0748730419868136.
- Shearman, L.P., Sriram, S., Weaver, D.R., Maywood, E.S., Chaves, I., Zheng, B., Kume, K., Lee, C.C., Hastings, M.H., and Reppert, S.M. (2000). Interacting molecular loops in the mammalian circadian clock. *Science* 288, 1013-1019.
- Shepard, P.J., and Hertel, K.J. (2009). The SR protein family. *Genome biology* 10, 1.
- Shi, Y. (2017). Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nature Reviews Molecular Cell Biology* 18, 655.
- Shilts, J., Chen, G., and Hughey, J.J. (2018). Evidence for widespread dysregulation of circadian clock progression in human cancer. *PeerJ* 6, e4327.
- Shostak, A. (2017). Circadian Clock, Cell Division, and Cancer: From Molecules to Organism. *International journal of molecular sciences* 18, 873.
- Smith, C.W., and Valcárcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in biochemical sciences* 25, 381-388.
- Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* 4.
- Sood, S., Szkop, K.J., Nakhuda, A., Gallagher, I.J., Murie, C., Brogan, R.J., Kaprio, J., Kainulainen, H., Atherton, P.J., and Kujala, U.M. (2016). iGEMS: an integrated model for identification of alternative exon usage events. *Nucleic acids research* 44, e109-e109.
- Spies, D., Renz, P.F., Beyer, T.A., and Ciaudo, C. (2017). Comparative analysis of differential gene expression tools for RNA sequencing time course data. *Briefings in bioinformatics*.
- Stephan, F.K., and Zucker, I. (1972). Circadian rhythms in drinking behavior and locomotor activity of rats are eliminated by hypothalamic lesions. *Proceedings of the National Academy of Sciences* 69, 1583-1586.
- Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 100, 9440-9445.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., and Lander, E.S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* 102, 15545-15550.

- Sugnet, C.W., Kent, W.J., Ares, M., and Haussler, D. (2004). Transcriptome and genome conservation of alternative splicing events in humans and mice. Paper presented at: Pacific Symposium on Biocomputing.
- Sulli, G., Lam, M.T.Y., and Panda, S. (2019). Interplay between circadian clock and cancer: new frontiers for cancer treatment. *Trends in cancer*.
- Suzuki, A., Kogo, R., Kawahara, K., Sasaki, M., Nishio, M., Maehama, T., Sasaki, T., Mimori, K., and Mori, M. (2012). A new PICTURE of nucleolar stress. *Cancer science* *103*, 632-637.
- Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R., and Skotheim, R.I. (2016). Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* *35*, 2413.
- Takahashi, J.S. (2017). Transcriptional architecture of the mammalian circadian clock. *Nature Reviews Genetics* *18*, 164.
- Terpstra, T.J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Math* *14*, 327-333.
- Thaben, P.F., and Westermark, P.O. (2014). Detecting rhythms in time series with RAIN. *Journal of biological rhythms*, 0748730414553029.
- Thaben, P.F., and Westermark, P.O. (2016). Differential rhythmicity: detecting altered rhythmicity in biological data. *Bioinformatics* *32*, 2800-2808.
- Theodore, S., Sharp, S., Zhou, J., Turner, T., Li, H., Miki, J., Ji, Y., Patel, V., Yates, C., and Rhim, J.S. (2010). Establishment and characterization of a pair of non-malignant and malignant tumor derived cell lines from an African American prostate cancer patient. *International journal of oncology* *37*, 1477-1482.
- Thompson, J.A., Tan, J., and Greene, C.S. (2016). Cross-platform normalization of microarray and RNA-seq data for machine learning applications. *PeerJ* *4*, e1621.
- Toh, K.L., Jones, C.R., He, Y., Eide, E.J., Hinze, W.A., Virshup, D.M., Ptáček, L.J., and Fu, Y.-H. (2001). An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* *291*, 1040-1043.
- Tomioka, K., and Matsumoto, A. (2010). A comparative view of insect circadian clock systems. *Cellular and Molecular Life Sciences* *67*, 1397-1406.
- Torres, M., Becquet, D., Franc, J.L., and François-Bellan, A.M. (2018). Circadian processes in the RNA life cycle. *Wiley Interdisciplinary Reviews: RNA* *9*, e1467.
- Torres-Martín, M., Kusak, M.E., Isla, A., Burbano, R.R., Pinto, G.R., Melendez, B., Castresana, J.S., and Rey, J.A. (2015). Whole exome sequencing in a case of sporadic multiple meningioma reveals shared NF2, FAM109B, and TPRXL mutations, together with unique SMARCB1 alterations in a subset of tumor nodules. *Cancer genetics* *208*, 327-332.
- Traina, F., Favaro, P.M., de Souza Medina, S., Duarte, A.d.S.S., Winnischofer, S.M.B., Costa, F.F., and Saad, S.T. (2006). ANKHD1, ankyrin repeat and KH domain containing 1, is overexpressed in acute leukemias and is associated with SHP2 in K562 cells. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* *1762*, 828-834.
- Tress, M.L., Abascal, F., and Valencia, A. (2017a). Alternative splicing may not be the key to proteome complexity. *Trends in biochemical sciences* *42*, 98-110.
- Tress, M.L., Abascal, F., and Valencia, A. (2017b). Most alternative isoforms are not functionally important. *Trends in biochemical sciences* *42*, 408.
- Triqueneaux, G., Thenot, S., Kakizawa, T., Antoch, M.P., Safi, R., Takahashi, J.S., Delaunay, F., and Laudet, V. (2004). The orphan receptor Rev-erb α gene is a target of the circadian clock pacemaker. *Journal of molecular endocrinology* *33*, 585-608.
- Tynes, T., Hannevik, M., Andersen, A., Vistnes, A.I., and Haldorsen, T. (1996). Incidence of breast cancer in Norwegian female radio and telegraph operators. *Cancer Causes & Control* *7*, 197-204.
- Valacca, C., Bonomi, S., Buratti, E., Pedrotti, S., Baralle, F.E., Sette, C., Ghigna, C., and Biamonti, G. (2010). Sam68 regulates EMT through alternative splicing-activated nonsense-mediated mRNA decay of the SF2/ASF proto-oncogene. *The Journal of cell biology* *191*, 87-99.
- van der Veen, D.R., and Gerkema, M.P. (2016). Unmasking ultradian rhythms in gene expression. *The FASEB Journal* *31*, 743-750.
- Van Moerbeke, M., Kasim, A., Talloen, W., Reumers, J., Göhlmann, H.W., and Shkedy, Z. (2017). REIDS: Random Effects for the Identification of Differential Splicing Using Exon and HTA Arrays.
- Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Blue, S.M., Dominguez, D., Cody, N.A., Olson, S., and Sundaraman, B. (2018). A large-scale binding and functional map of human RNA binding proteins. *bioRxiv*, 179648.

- van Riel, B., Pakozdi, T., Brouwer, R., Monteiro, R., Tuladhar, K., Franke, V., Bryne, J.C., Jorna, R., Rijkers, E.-J., and van Ijcken, W. (2012). A novel complex, RUNX1-MYEF2, represses hematopoietic genes in erythroid cells. *Molecular and cellular biology* 32, 3814-3822.
- Venit, T., Kalendová, A., Petr, M., Dzajak, R., Pastorek, L., Rohožková, J., Malohlava, J., and Hozák, P. (2016). Nuclear myosin I regulates cell membrane tension. *Scientific reports* 6, 30864.
- Ventre, S., Indrieri, A., Fracassi, C., Franco, B., Conte, I., Cardone, L., and Di Bernardo, D. (2015). Metabolic regulation of the ultradian oscillator Hes1 by reactive oxygen species. *Journal of molecular biology* 427, 1887-1902.
- Vollmers, C., Gill, S., DiTacchio, L., Pulivarthy, S.R., Le, H.D., and Panda, S. (2009). Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proceedings of the National Academy of Sciences* 106, 21453-21458.
- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701-718.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470-476.
- Wang, Q., Abruzzi, K.C., Rosbash, M., and Rio, D.C. (2018). Striking circadian neuron diversity and cycling of *Drosophila* alternative splicing. *Elife* 7, e35618.
- Wang, X., Yan, D., Teng, M., Fan, J., Zhou, C., Li, D., Qiu, G., Sun, X., Li, T., and Xing, T. (2012). Reduced expression of PER3 is associated with incidence and development of colon cancer. *Annals of surgical oncology* 19, 3081-3088.
- Wang, Y., Hua, L., Lu, C., and Chen, Z. (2011). Expression of circadian clock gene human Period2 (hPer2) in human colorectal carcinoma. *World journal of surgical oncology* 9, 166.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics* 10, 57.
- Warzecha, C.C., and Carstens, R.P. (2012). Complex changes in alternative pre-mRNA splicing play a central role in the epithelial-to-mesenchymal transition (EMT). Paper presented at: Seminars in cancer biology (Elsevier).
- Warzecha, C.C., Sato, T.K., Nabet, B., Hogenesch, J.B., and Carstens, R.P. (2009). ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. *Molecular cell* 33, 591-601.
- Welsh, D.K., Takahashi, J.S., and Kay, S.A. (2010). Suprachiasmatic nucleus: cell autonomy and network properties. *Annual review of physiology* 72, 551-577.
- Westermarck, P.O., and Herzog, H. (2013). Mechanism for 12 hr rhythm generation by the circadian clock. *Cell reports* 3, 1228-1238.
- Wickham, H. (2016). *ggplot2: elegant graphics for data analysis* (Springer).
- Wilcockson, D., and Zhang, L. (2008). Circatidal clocks. *Current Biology* 18, R753-R755.
- Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* 3, a003707.
- Wittenbrink, N., Ananthasubramanian, B., Münch, M., Koller, B., Maier, B., Weschke, C., Bes, F., de Zeeuw, J., Nowozin, C., and Wahnschaffe, A. (2018). High-accuracy determination of internal circadian time from a single blood sample. *The Journal of clinical investigation*.
- Wolff, A., Bayerlová, M., Gaedcke, J., Kube, D., and Beißbarth, T. (2018). A comparative study of RNA-Seq and microarray data analysis on the two examples of rectal-cancer patients and Burkitt Lymphoma cells. *PloS one* 13, e0197162.
- Xu, W., Seok, J., Mindrinos, M.N., Schweitzer, A.C., Jiang, H., Wilhelmy, J., Clark, T.A., Kapur, K., Xing, Y., and Faham, M. (2011). Human transcriptome array for high-throughput clinical studies. *Proceedings of the National Academy of Sciences* 108, 3707-3712.
- Yamaguchi, S., Mitsui, S., Yan, L., Yagita, K., Miyake, S., and Okamura, H. (2000). Role of DBP in the circadian oscillatory mechanism. *Molecular and cellular biology* 20, 4773-4781.
- Yan, G., Fukabori, Y., McBride, G., Nikolaropoulos, S., and McKeenan, W. (1993). Exon switching and activation of stromal and embryonic fibroblast growth factor (FGF)-FGF receptor genes in prostate epithelial cells accompany stromal independence and malignancy. *Molecular and cellular biology* 13, 4513-4522.
- Yang, L., Li, N., Wang, C., Yu, Y., Yuan, L., Zhang, M., and Cao, X. (2004). Cyclin L2, a novel RNA polymerase II-associated cyclin, is involved in pre-mRNA splicing and induces apoptosis of human hepatocellular carcinoma cells. *Journal of Biological Chemistry* 279, 11639-11648.
- Yang, Q., Pando, B.F., Dong, G., Golden, S.S., and van Oudenaarden, A. (2010). Circadian gating of the cell cycle revealed in single cyanobacterial cells. *Science* 327, 1522-1526.

- Yang, S., Jia, R., and Bian, Z. (2018). SRSF5 functions as a novel oncogenic splicing factor and is upregulated by oncogene SRSF3 in oral squamous cell carcinoma. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1865, 1161-1172.
- Ye, Y., Xiang, Y., Ozguc, F.M., Kim, Y., Liu, C.-J., Park, P.K., Hu, Q., Diao, L., Lou, Y., and Lin, C. (2018). The Genomic Landscape and Pharmacogenomic Interactions of Clock Genes in Cancer Chronotherapy. *Cell systems*.
- Yeh, C.S., Wang, J.Y., Chung, F.Y., Lee, S.C., Huang, M.Y., Kuo, C.W., Yang, M.J., and Lin, S.R. (2008). Significance of the glycolytic pathway and glycolysis related-genes in tumorigenesis of human colorectal cancers. *Oncol Rep* 19, 81-91.
- Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., and Nagasaki, M. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478, 64.
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OmicS: a journal of integrative biology* 16, 284-287.
- Yu, W., Ikeda, M., Abe, H., Honma, S., Ebisawa, T., Yamauchi, T., Honma, K.-i., and Nomura, M. (1999). Characterization of three splice variants and genomic organization of the mouse BMAL1 gene. *Biochemical and biophysical research communications* 260, 760-767.
- Zattelman, L., Regev, R., Ušaj, M., Reinke, P.Y., Giese, S., Samson, A.O., Taft, M.H., Manstein, D.J., and Henn, A. (2017). N-terminal splicing extensions of the human MYO1C gene fine-tune the kinetics of the three full-length myosin IC isoforms. *Journal of Biological Chemistry* 292, 17804-17818.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., and Girón, C.G. (2017). Ensembl 2018. *Nucleic acids research* 46, D754-D761.
- Zhang, C., Zhang, B., Lin, L.-L., and Zhao, S. (2017). Evaluation and comparison of computational tools for RNA-seq isoform quantification. *BMC genomics* 18, 583.
- Zhang, L., Hastings, M.H., Green, E.W., Tauber, E., Sladek, M., Webster, S.G., Kyriacou, C.P., and Wilcockson, D.C. (2013). Dissociation of circadian and circatidal timekeeping in the marine crustacean *Eurydice pulchra*. *Current Biology* 23, 1863-1873.
- Zhang, R., Lahens, N.F., Ballance, H.I., Hughes, M.E., and Hogenesch, J.B. (2014). A circadian gene expression atlas in mammals: implications for biology and medicine. *Proceedings of the National Academy of Sciences* 111, 16219-16224.
- Zhang, R., Podtelezchnikov, A.A., Hogenesch, J.B., and Anafi, R.C. (2016). Discovering biology in periodic data through phase set enrichment analysis (PSEA). *Journal of biological rhythms* 31, 244-257.
- Zhao, H., Zeng, Z.-L., Yang, J., Jin, Y., Qiu, M.-Z., Hu, X.-Y., Han, J., Liu, K.-Y., Liao, J.-W., and Xu, R.-H. (2014a). Prognostic relevance of Period1 (Per1) and Period2 (Per2) expression in human gastric cancer. *International journal of clinical and experimental pathology* 7, 619.
- Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X. (2014b). Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PloS one* 9, e78644.
- Zheng, H., Chen, Y., Chen, S., Niu, Y., Yang, L., Li, B., Lu, Y., Geng, S., Du, X., and Li, Y. (2011). Expression and distribution of PPP2R5C gene in leukemia. *Journal of hematology & oncology* 4, 21.
- Zheng, T., Wang, J., Zhao, Y., Zhang, C., Lin, M., Wang, X., Yu, H., Liu, L., Feng, Z., and Hu, W. (2013). Spliced MDM2 isoforms promote mutant p53 accumulation and gain-of-function in tumorigenesis. *Nature communications* 4, 2996.
- Zhu, B., Dacso, C.C., and O'Malley, B.W. (2018). Unveiling “Musica Universalis” of the Cell: A Brief History of Biological 12-Hour Rhythms. *Journal of the Endocrine Society* 2, 727-752.
- Zhu, B., Zhang, Q., Pan, Y., Mace, E.M., York, B., Antoulas, A.C., Dacso, C.C., and O'Malley, B.W. (2017). A cell-autonomous mammalian 12 hr clock coordinates metabolic and stress rhythms. *Cell metabolism* 25, 1305-1319. e1309.
- Zhu, J., Mayeda, A., and Krainer, A.R. (2001). Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Molecular cell* 8, 1351-1361.

7 Appendix

7.1 Supplementary Material

7.1.1 Supplementary Figures

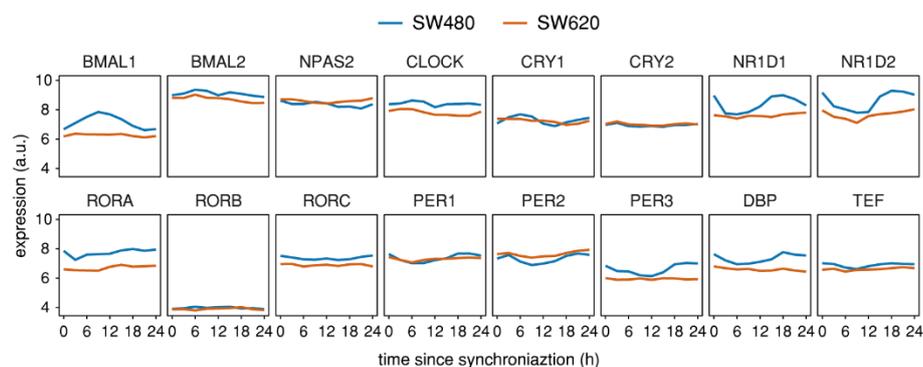


Figure S 1: Time-series expression of core clock genes and circadian transcription factors in SW480 and SW620 cells (microarray data). RMA-preprocessed time-series expression (\log_2) is shown for the primary tumor-derived SW480 cells (blue) and the metastasis-derived SW620 cells (orange).

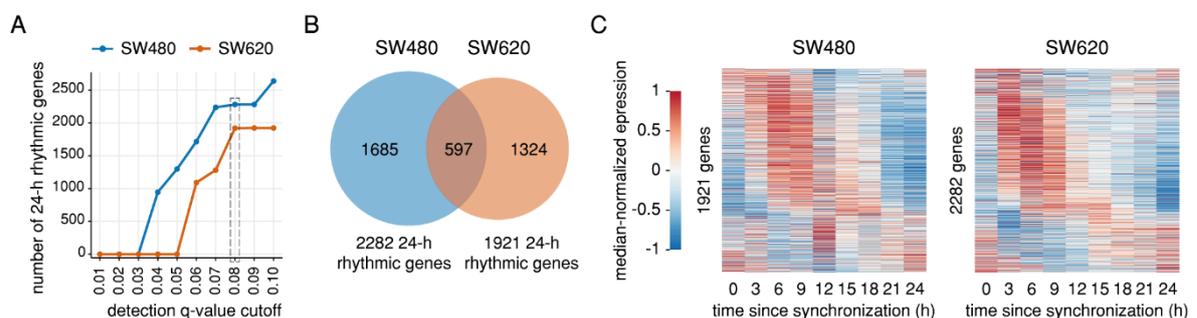


Figure S 2: Comparison of the circadian transcriptome between SW480 and SW620 cells (microarray data). (A) Number of 24-h rhythmic genes in SW480 cells (blue) and SW620 cells (orange) for different RAIN q -value cutoffs after filtering for genes with a FC amplitude > 1.15 . The gray rectangle marks the cutoff chosen to identify 24-h rhythmic genes for subsequent analyses. (B) Venn diagram of the intersection between 24-h rhythmic genes identified in SW480 cells (blue) and SW620 cells (orange). (C) Median-normalized, phase-ordered expression heatmaps of genes that were identified as 24-h rhythmic in the respective other cell line for SW480 (left) and SW620 cells (right). Each row represents one gene. The phase-ordering is based on the phase estimated for the respective other cell line.

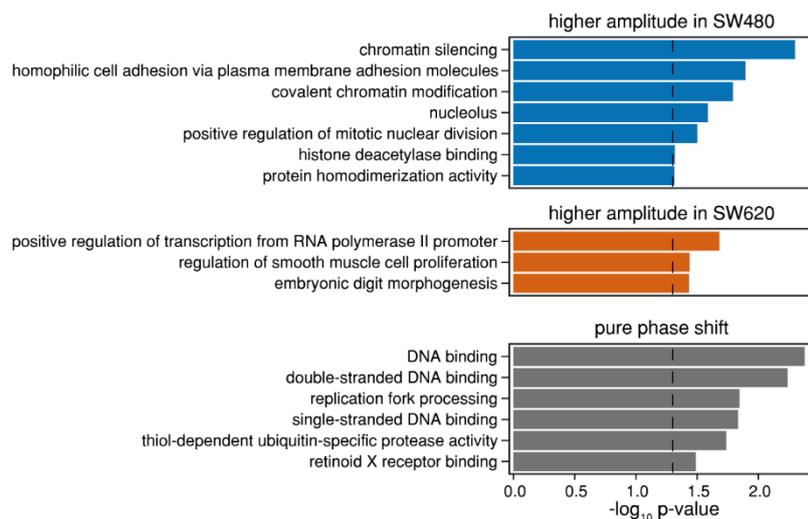


Figure S 3: Functional annotation for the sets of differentially rhythmic genes between SW480 and SW620 cells (microarray data). Enriched GO terms for the set of 24-h rhythmic genes with a higher amplitude in SW480 cells (blue), a higher amplitude in SW620 cells (orange) and with a pure phase-shift (gray).

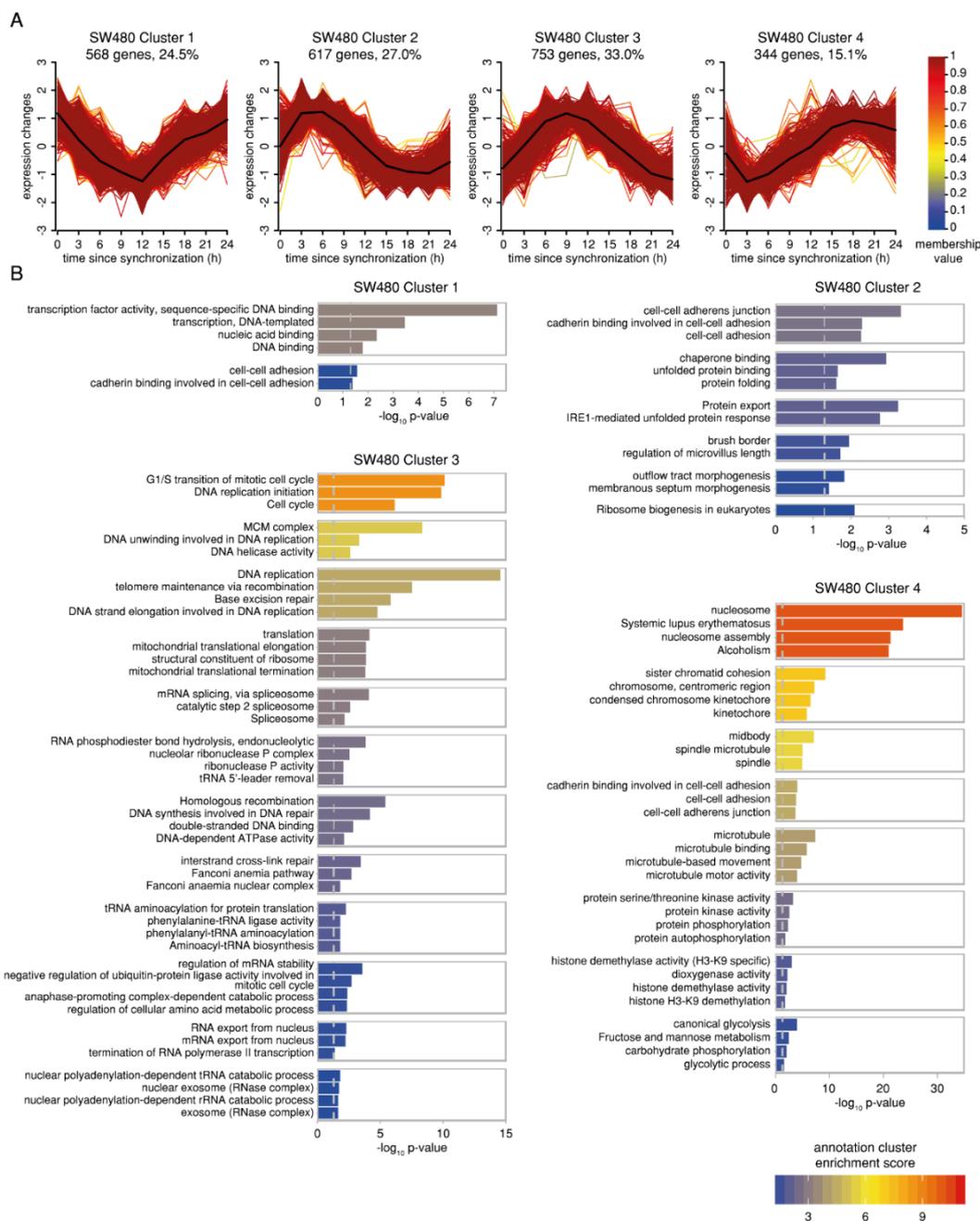


Figure S 4: Functional annotation of clusters of 24-h rhythmic genes in SW480 cells (microarray data). (A) Temporal expression clusters of 24-h rhythmic genes in SW480 cells. Each colored line represents the expression (transformed to the standard normal distribution) of a gene. The line colors indicate gradual membership values reflecting the strength of a gene's association with the cluster (red: high membership value, blue: low membership value). The black line indicates the cluster center. (B) Annotation clusters of enriched GO terms and KEGG pathways for the four clusters of 24-h rhythmic genes in SW480 cells. Shown are up to four highest ranking terms with $p < 0.05$ for all annotation clusters with an enrichment score ≥ 1.3 .

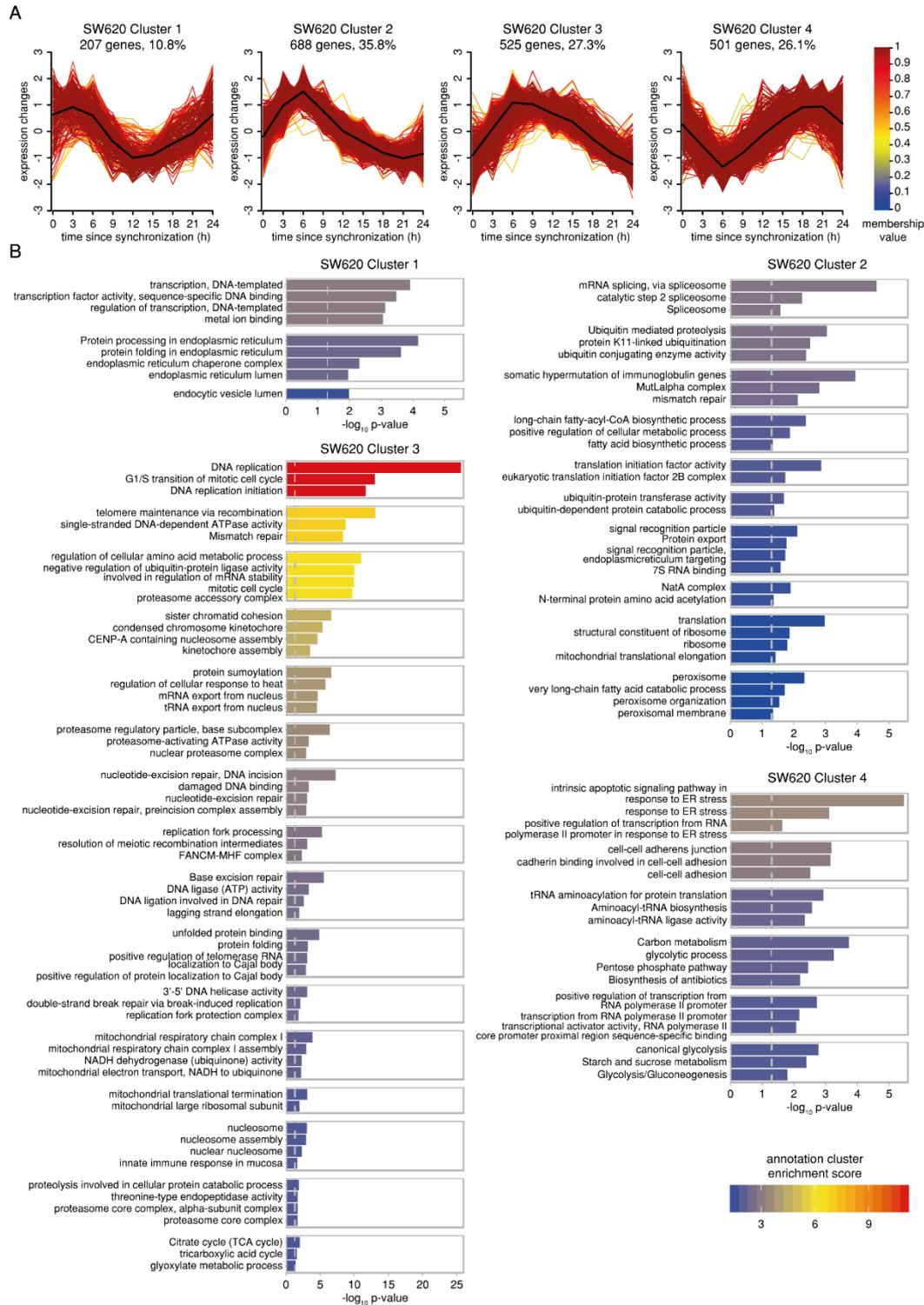


Figure S 5: Functional annotation of clusters of 24-h rhythmic genes in SW620 cells (microarray data). (A) Temporal expression clusters of 24-h rhythmic genes in SW620 cells. Each colored line represents the expression (transformed to the standard normal distribution) of a gene. The line colors indicate gradual membership values reflecting the strength of a gene's association with the cluster (red: high membership value, blue: low membership value). The black line indicates the cluster center. (B) Annotation clusters of enriched GO terms and KEGG pathways for the four clusters of 24-h rhythmic genes in SW620 cells. Shown are up to four highest ranking terms with $p < 0.05$ for all annotation clusters with an enrichment score ≥ 1.3 .

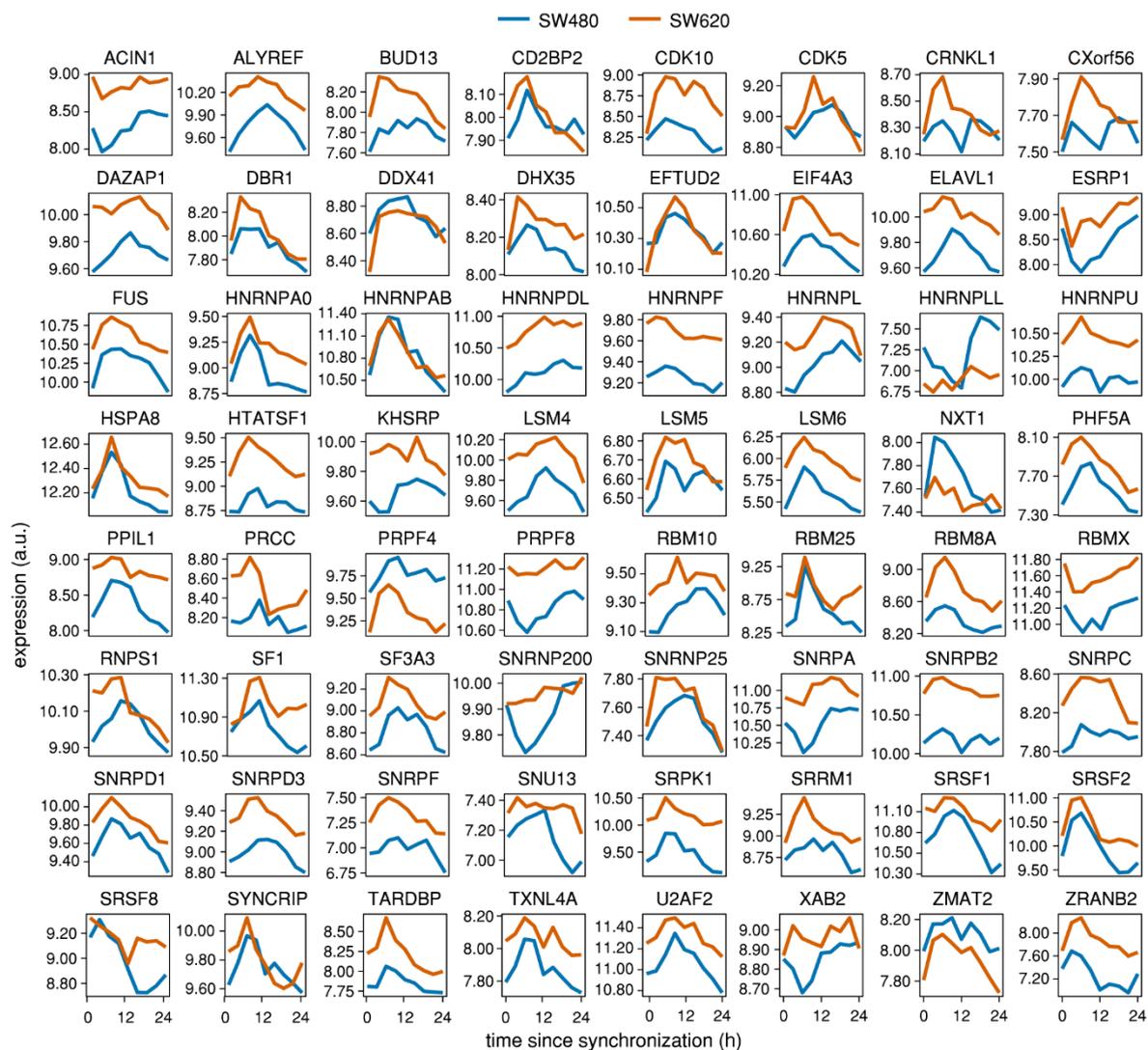


Figure S 6: Time-series expression of 24-h rhythmic splicing-related genes in CRC cell lines (microarray data). Shown are 64 splicing-related genes that were identified to be rhythmic in SW480 cells (blue) and/or SW620 cells (orange).

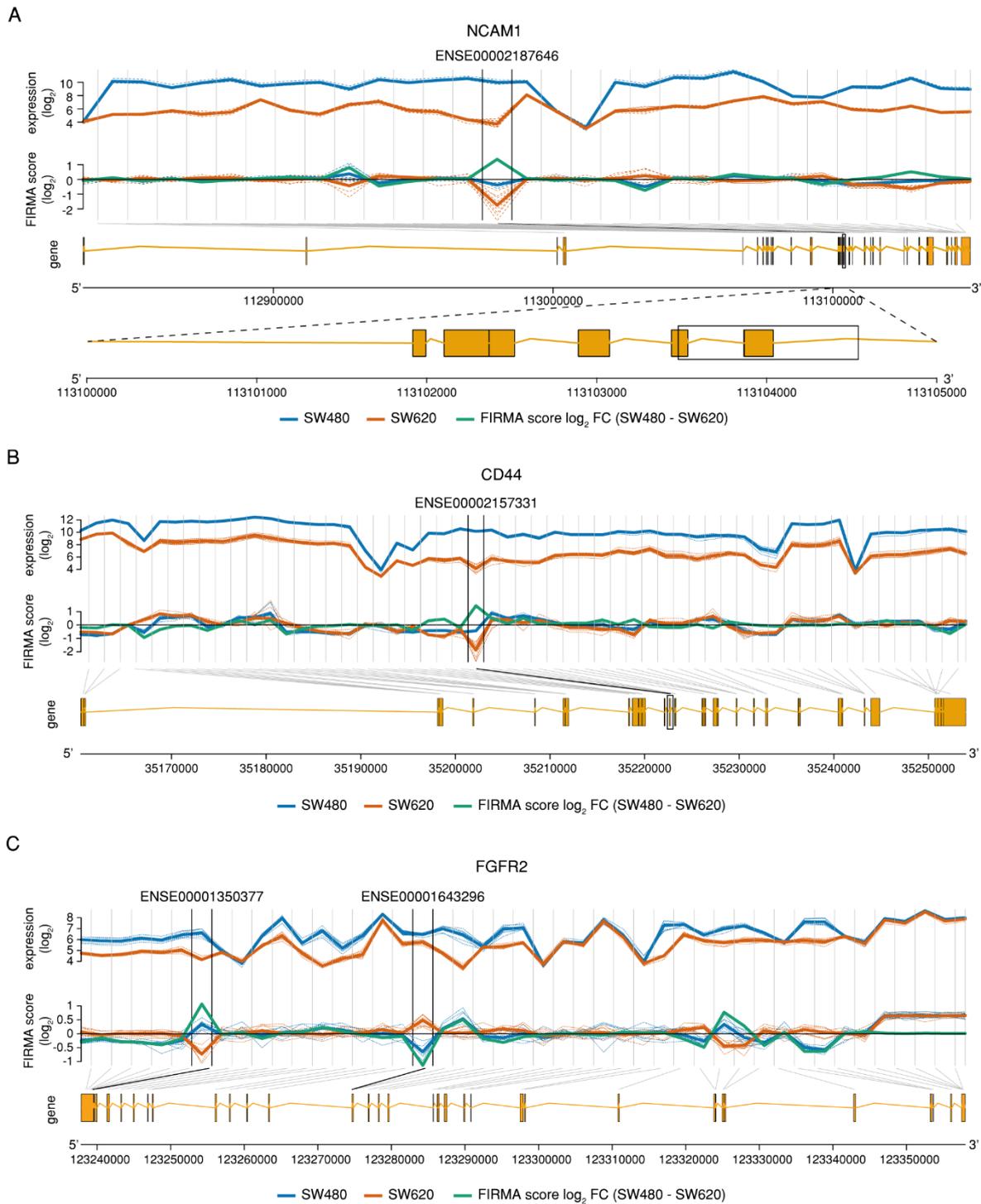


Figure S 7: Differential AS events in *NCAM1* and the EMT-associated genes *CD44* and *FGFR2* (microarray data). Exon-level expression (upper panel), FIRMA scores (middle panel), and genomic representation (lower panel) of the candidate genes (A) *NCAM1*, (B) *CD44* (B), and (C) *FGFR2*. Vertical lines separate the individual probesets covering the genes. Gray diagonal lines indicate the localization of the probesets within the genome. For each probeset, the exon-level expression and the FIRMA scores of the individual time points are depicted by a dotted line and the respective mean values are depicted by a solid line for SW480 (blue) and SW620 (orange), respectively. The mean \log_2 FC of the FIRMA score between cell lines (SW480-SW620) is represented by a green line.

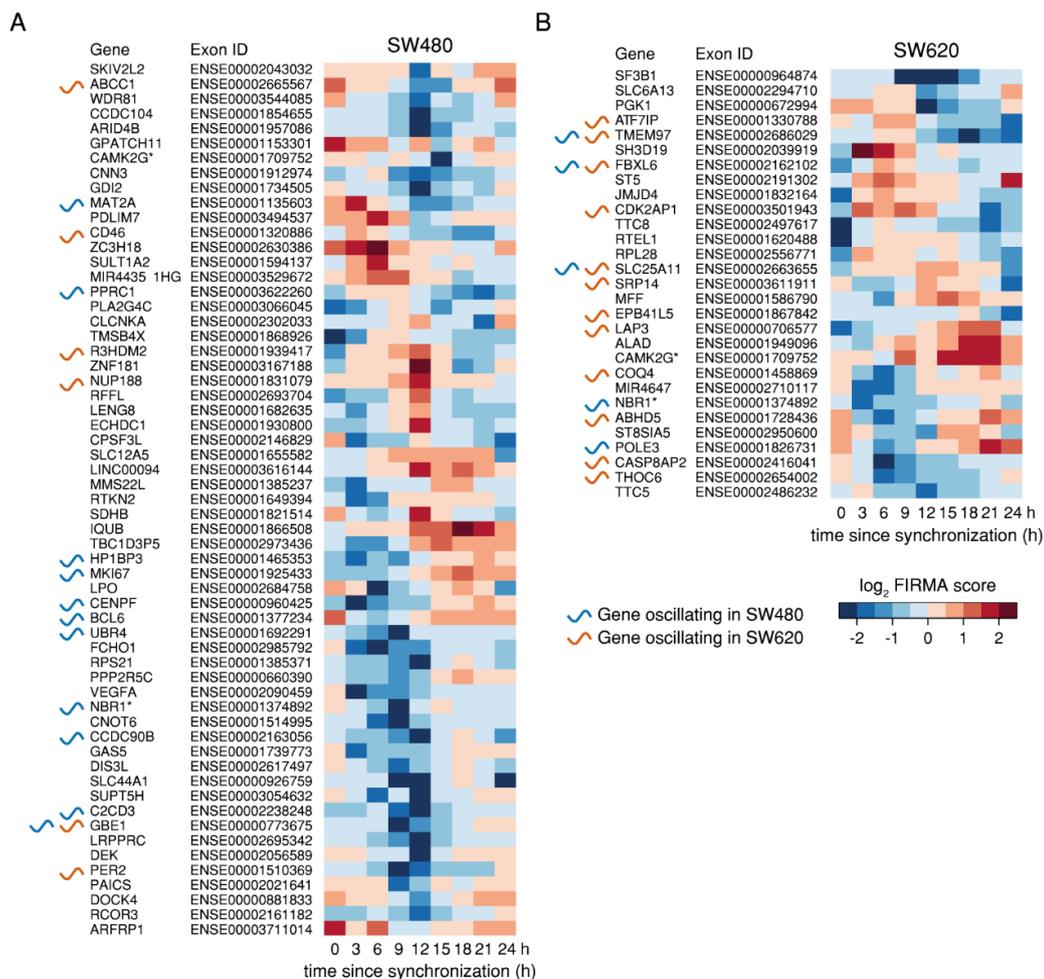


Figure S 8: Candidate circadian AS events in SW480 and SW620 cells (microarray data). Phase-ordered heatmaps of exons with 24-h rhythmic FIRMA scores in (A) SW480 and (B) SW620 cells. Exons displaying oscillating AS events in both cell lines are marked by an asterisk. The respective genes are marked with an oscillatory curve if they show 24-h transcriptional rhythms in SW480 (blue) or SW620 cells (orange).

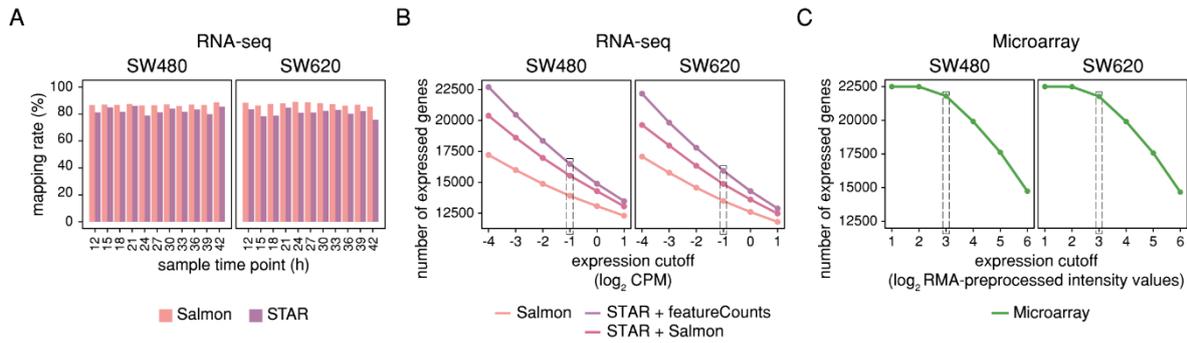


Figure S9: RNA-seq mapping rates and number of expressed genes in the microarray and RNA-seq datasets for different analysis methods and cutoffs. (A) Mapping rates of RNA-seq reads for time point samples from SW480 cells (left panel) and SW620 cells (right panel) using Salmon (percentage of mapped reads) and STAR (percentage of uniquely mapped reads). (B) Number of expressed genes for RNA-seq samples of SW480 cells (left panel) and SW620 cells (right panel) for different expression cutoffs and methods (*Salmon*, *STAR + featureCounts*, *STAR + Salmon*). The gray rectangle marks the cutoff that was chosen to identify expressed gene for subsequent analyses. (C) Number of expressed genes for microarray samples of SW480 cells (left panel) and SW620 cells (right panel) for different expression cutoffs. The gray rectangle marks the cutoff chosen to determine expressed gene for subsequent analyses.

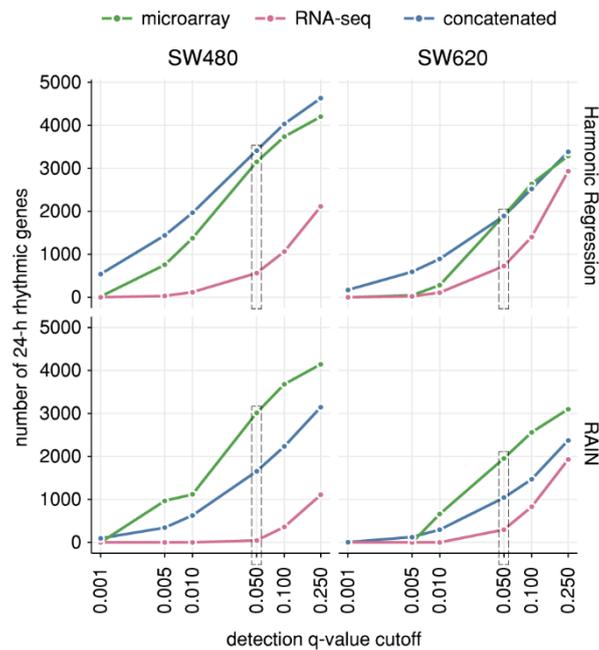


Figure S10: Rhythmic gene sets identified in SW480 and SW620 cells (microarray and RNA-seq data). Number of 24-h rhythmic genes in SW480 cells (left) and SW620 cells (right) for the three different datasets (microarray: green; RNA-seq: pink; concatenated: blue) and different q -value cutoffs determined either by harmonic regression (upper panels) or RAIN (lower panels). Additionally, 24-h rhythmic genes have a relative amplitude cutoff of 0.1, as estimated by harmonic regression. RAIN p -values were corrected after filtering for genes with a relative amplitude ≥ 0.1 . The gray rectangle marks the cutoff chosen to determine 24-h rhythmic genes for subsequent analyses.

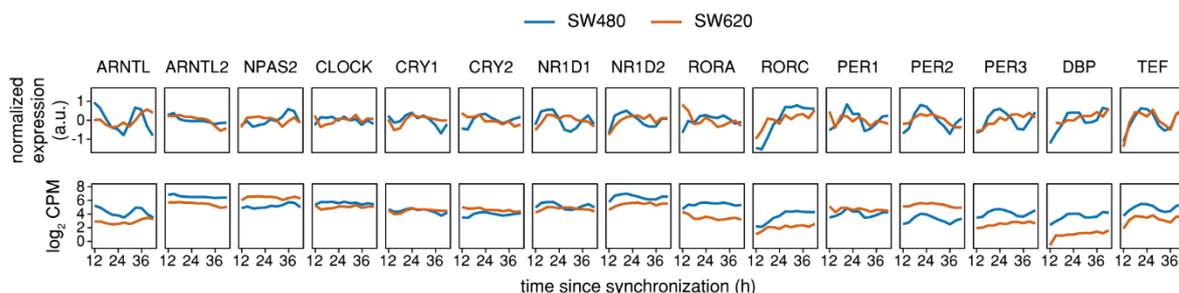


Figure S 11: Time-series expression data of core clock genes and circadian transcription factors in SW480 and SW620 cells (RNA-seq data). The upper panel shows \log_2 CPM normalized to the mean of all time points and the lower panel shows the \log_2 CPM for SW480 cells (blue line) and SW620 cells (orange line).

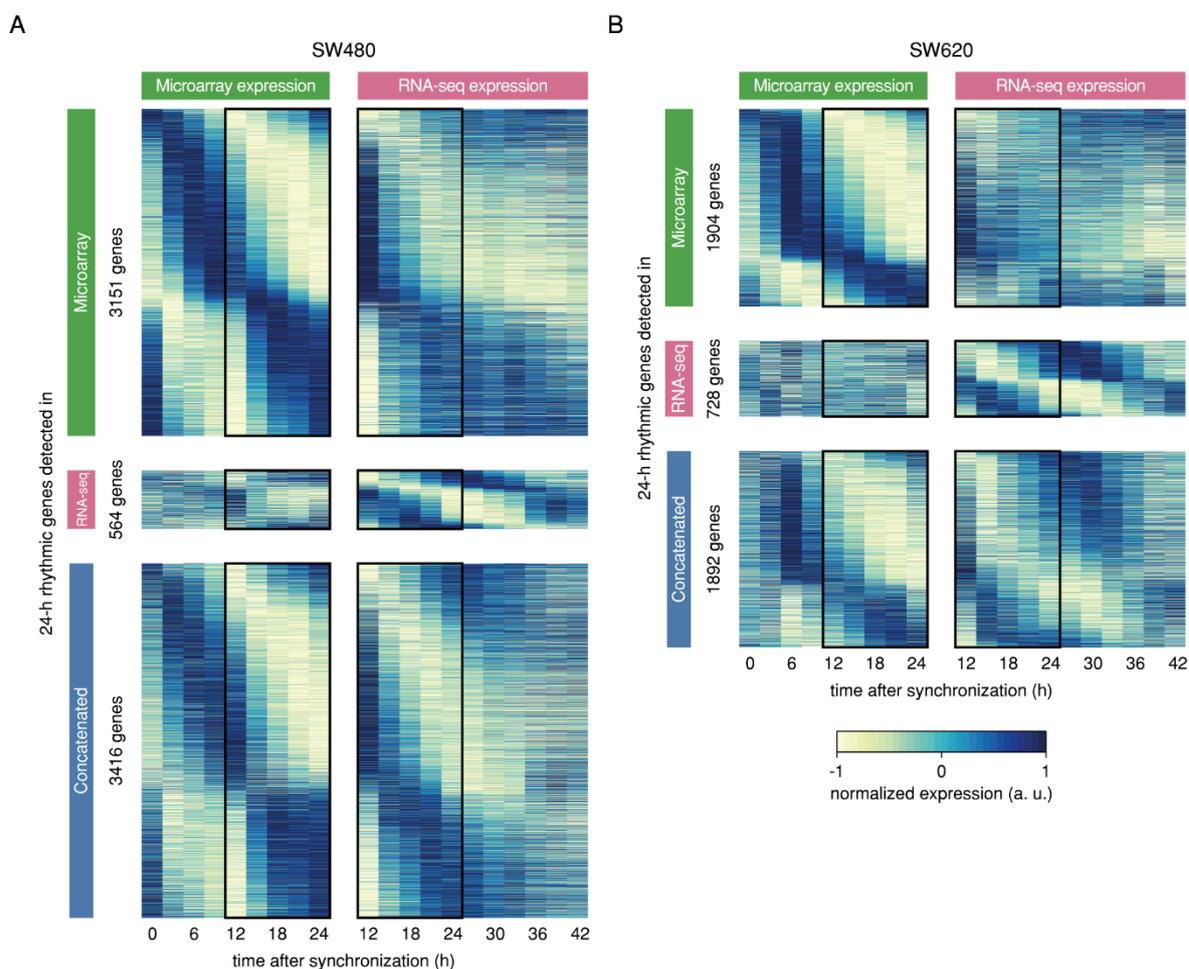


Figure S 12: Expression heatmaps of 24-h rhythmic genes in SW480 and SW620 cells (microarray and RNA-seq data). Range-normalized, phase-ordered microarray (respective left panel) and RNA-seq expression (respective right panel) heatmaps of genes that were identified as 24-h rhythmic ($q < 0.05$ and relative amplitude ≥ 0.1 determined by harmonic regression) in (A) SW480 and (B) SW620 cells in the microarray data (green, top panel), the RNA-seq data (pink, middle panel), and the concatenated data (blue, bottom panel). Each row represents one gene. Phases were estimated based on the dataset in which the 24-h rhythmic genes were identified. Black rectangles mark the shared time points between the microarray and the RNA-seq data.

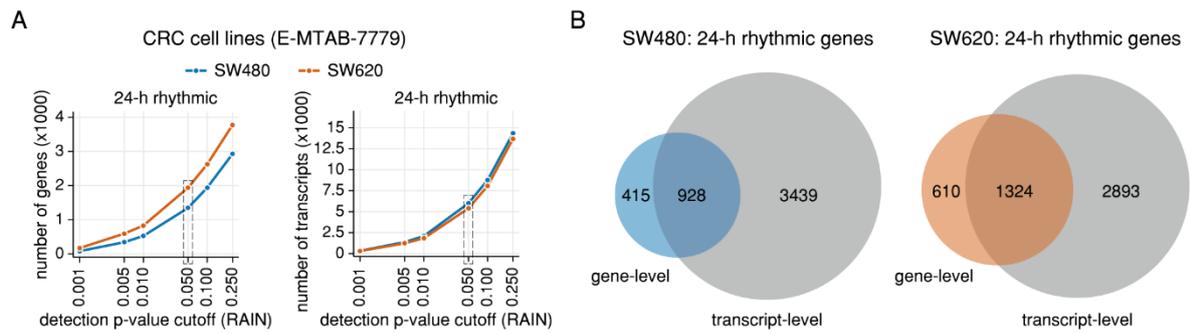


Figure S 13: Rhythmic gene and transcript sets identified in SW480 and SW620 cells (RNA-seq data). (A) 24-h rhythmic genes (left panel) and transcripts (right panel) in SW480 cells (blue) and SW620 cells (orange) for different RAIN p -value cutoffs. Additionally, 24-h rhythmic genes were required to have a relative amplitude ≥ 0.1 . (B) Venn diagrams of gene sets identified as 24-h rhythmic on gene- and on transcript-level for SW480 cells (left panel) and SW620 cells (right panel).

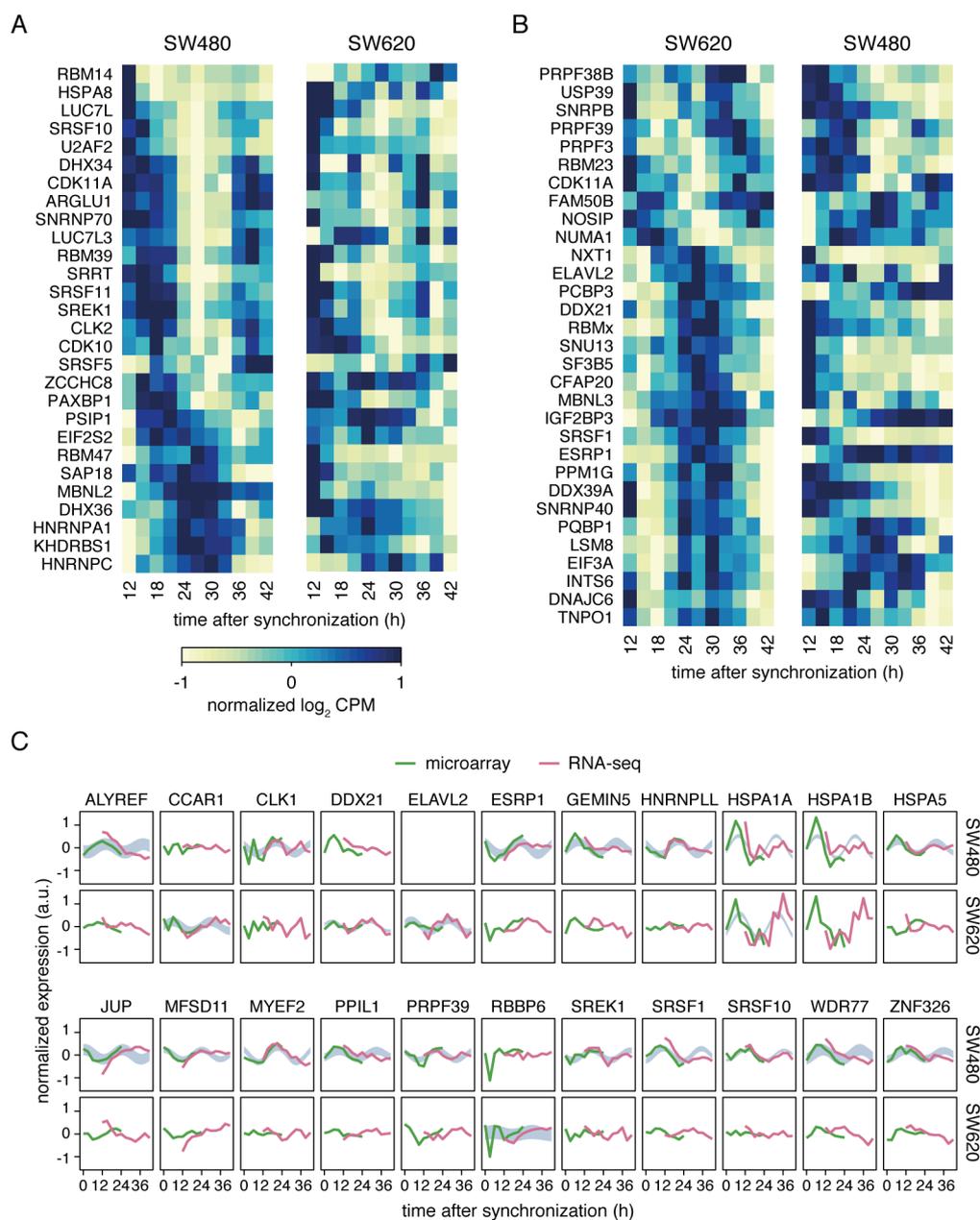


Figure S 14: Expression of 24-h rhythmic splicing-related genes in CRC cell lines (microarray and RNA-seq data). Range-normalized, phase-ordered heatmaps of splicing-related genes that were identified as 24-h rhythmic (RAIN $p < 0.05$ and relative amplitude ≥ 0.1) in (A) SW480 and (B) SW620 cells. The respective right panels show the expression in the other cell line. (C) Mean-normalized time-series expression of splicing-related genes found to be 24-h rhythmic with high amplitudes (harmonic regression $q < 0.05$ and relative amplitude ≥ 0.2) in SW480 cells or in SW620 cells based on the concatenated data. Microarray expression values are represented by green lines and RNA-seq expression values by pink lines. The blue area marks the confidence area of the harmonic regression fitted to the concatenated data for 24-h rhythmic genes.

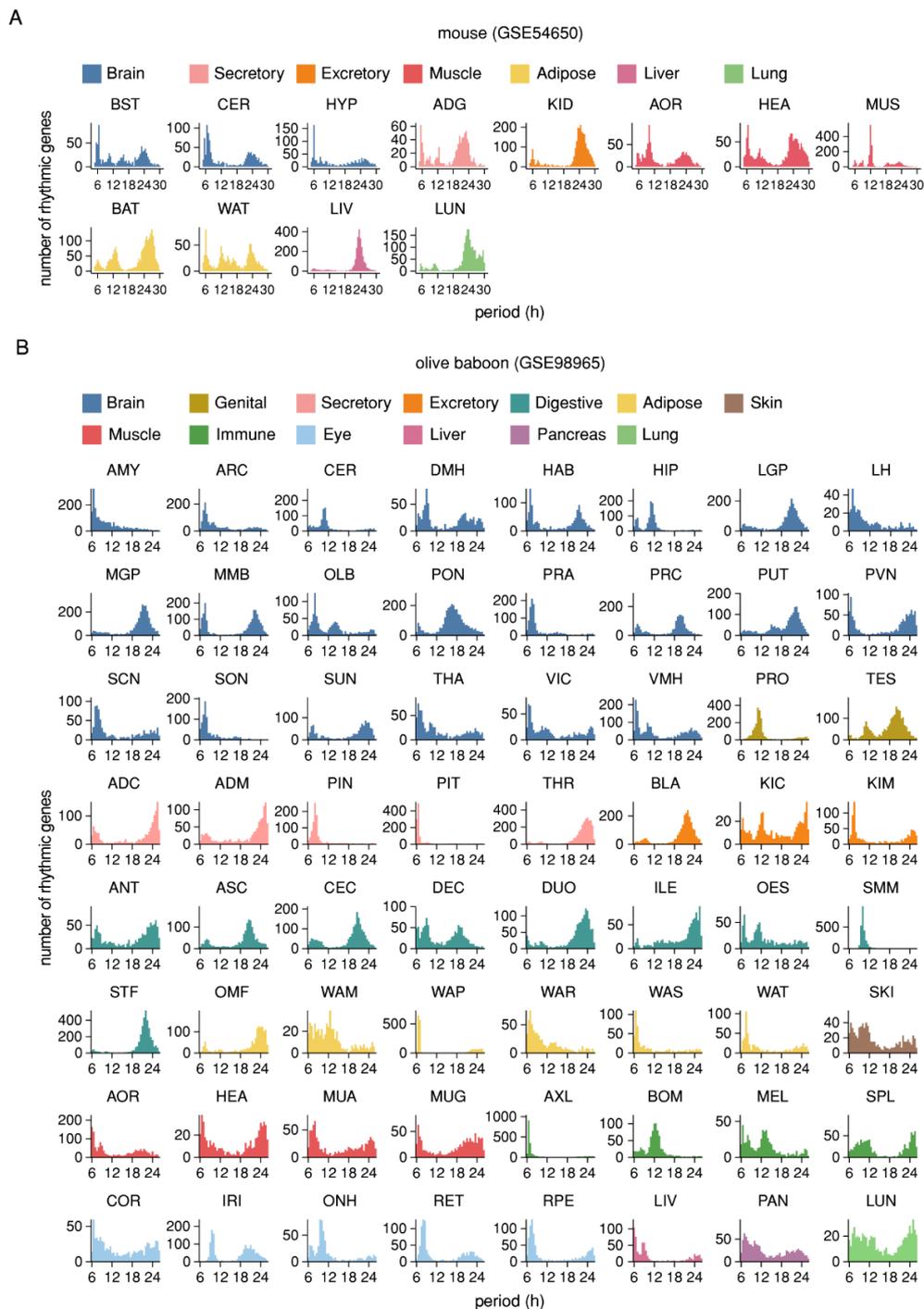


Figure S 15: Period distribution of rhythmic genes in mouse (microarray data, GSE54650) and baboon tissues (RNA-seq data, GSE98965). (A) Histograms of the period distributions of rhythmic genes in twelve murine tissues. Rhythmic genes were determined based on RMA-preprocessed microarray intensity values for a period range of 5–30 h and increments of 0.1 h by fitting a harmonic regression curve ($p < 0.01$). For each rhythmic gene, only the period with the lowest p -value is shown. Periods at the extreme ends of the range (5 h and 30 h) were excluded. (B) Histograms of the period distributions of rhythmic genes in 64 baboon tissues. Rhythmic genes were determined based on CPM values for a period range of 6–26 h and increments of 0.1 h by fitting a harmonic regression curve ($p < 0.01$). For each gene, only the period with the lowest p -value is shown. Periods at the extreme ends of the range (6 h and 26 h) were excluded.

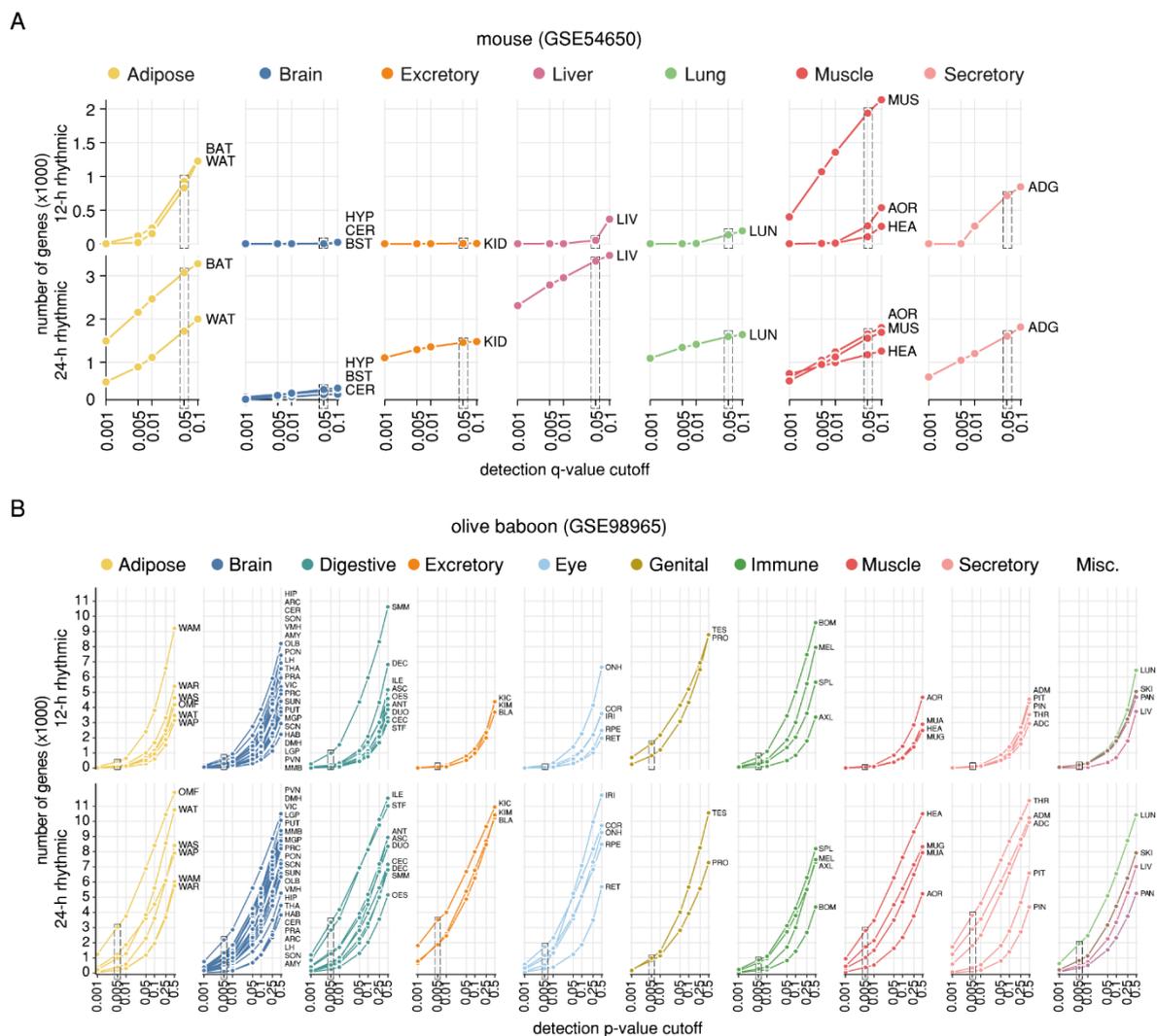
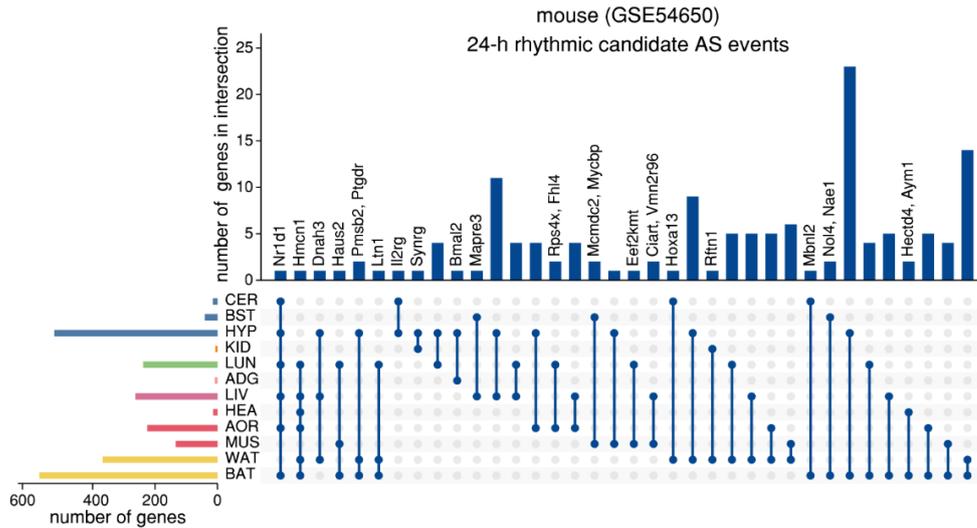
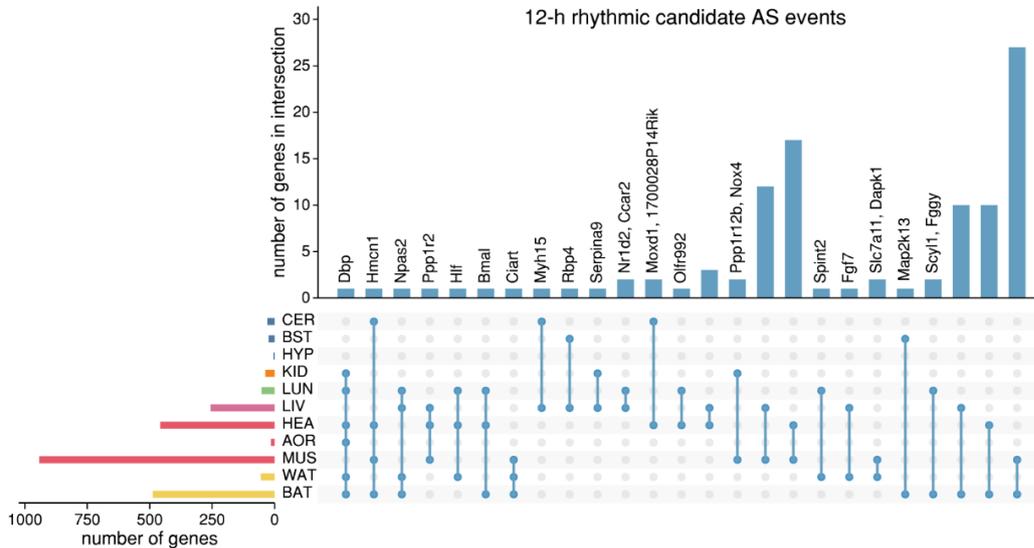


Figure S16: Number of 12-h and 24-h rhythmic genes in mouse (microarray data, GSE54650) and baboon tissues (RNA-seq data, GSE98965) for different cutoffs. (A) Number of 12-h rhythmic genes (upper panel) and 24-h rhythmic genes (upper panel) in twelve murine tissues for different RAIN q -value cutoffs after filtering for genes with a relative amplitude ≥ 0.1 . (B) Number of 12-h rhythmic genes (upper panel) and 24-h rhythmic genes (upper panel) in 64 baboon tissues for different RAIN p -value cutoffs and filtering for genes with a relative amplitude ≥ 0.1 . The gray rectangle marks the cutoffs chosen to identify rhythmic genes for subsequent analyses.

A



B



C

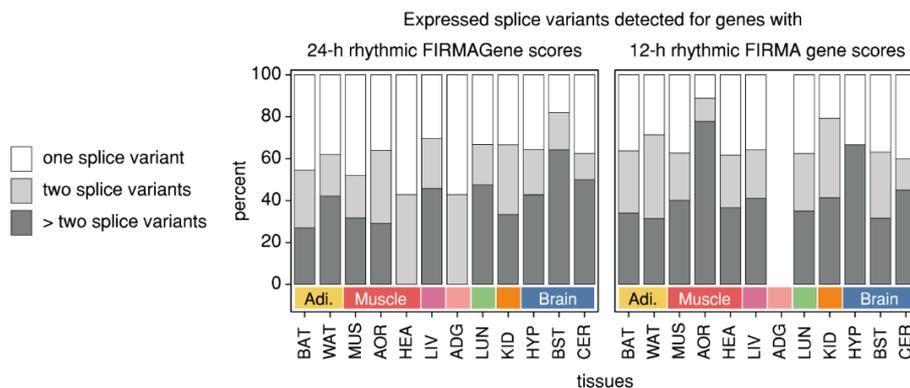


Figure S 17: Murine candidate genes with putative rhythmic AS events (microarray data, GSE54650). UpsetR plots to visualize the intersections between tissues for genes with (A) 24-h rhythmic and (B) 12-h rhythmic FIRMA Gene scores that are detected in at least two tissues. (C) Tissue-wise percentage of the number expressed splice variants (one variant: white, two variants: light gray, three or more variants: dark gray) in the RNA-seq data (GSE54651) for candidate genes with 24-h rhythmic (left panel) and 12-h rhythmic FIRMA Gene scores (right panel).

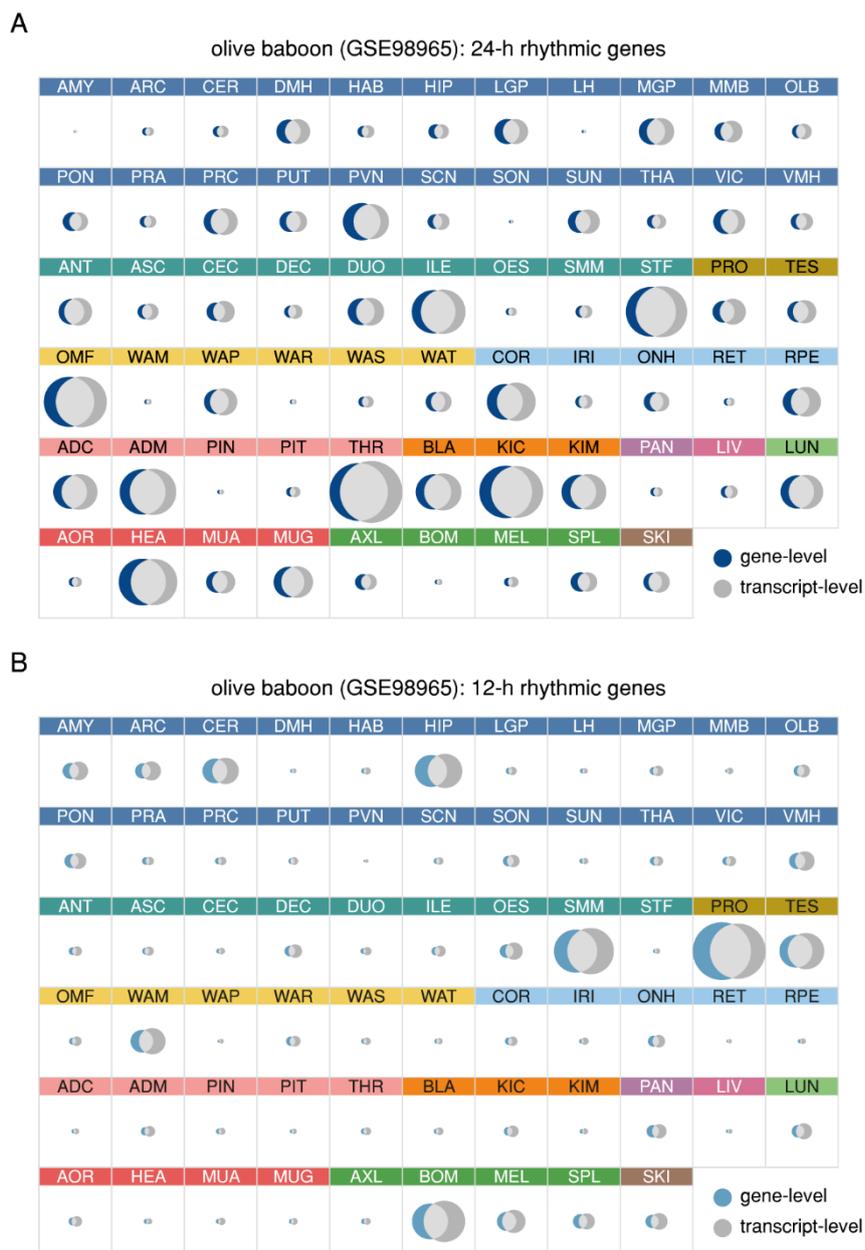


Figure S 18: Intersections between the sets of genes identified as rhythmic on the gene- and on the transcript-level in baboon tissues (RNA-seq data, GSE98965). Tissue-wise area-proportional Venn diagrams of the sets of genes that were identified as (A) 24-h rhythmic and (B) 12-h rhythmic on gene-level (24-h rhythmic: dark blue; 12-h rhythmic: light blue) or on transcript-level (dark gray) for the baboon data. The intersection of the two sets is represented in light gray.

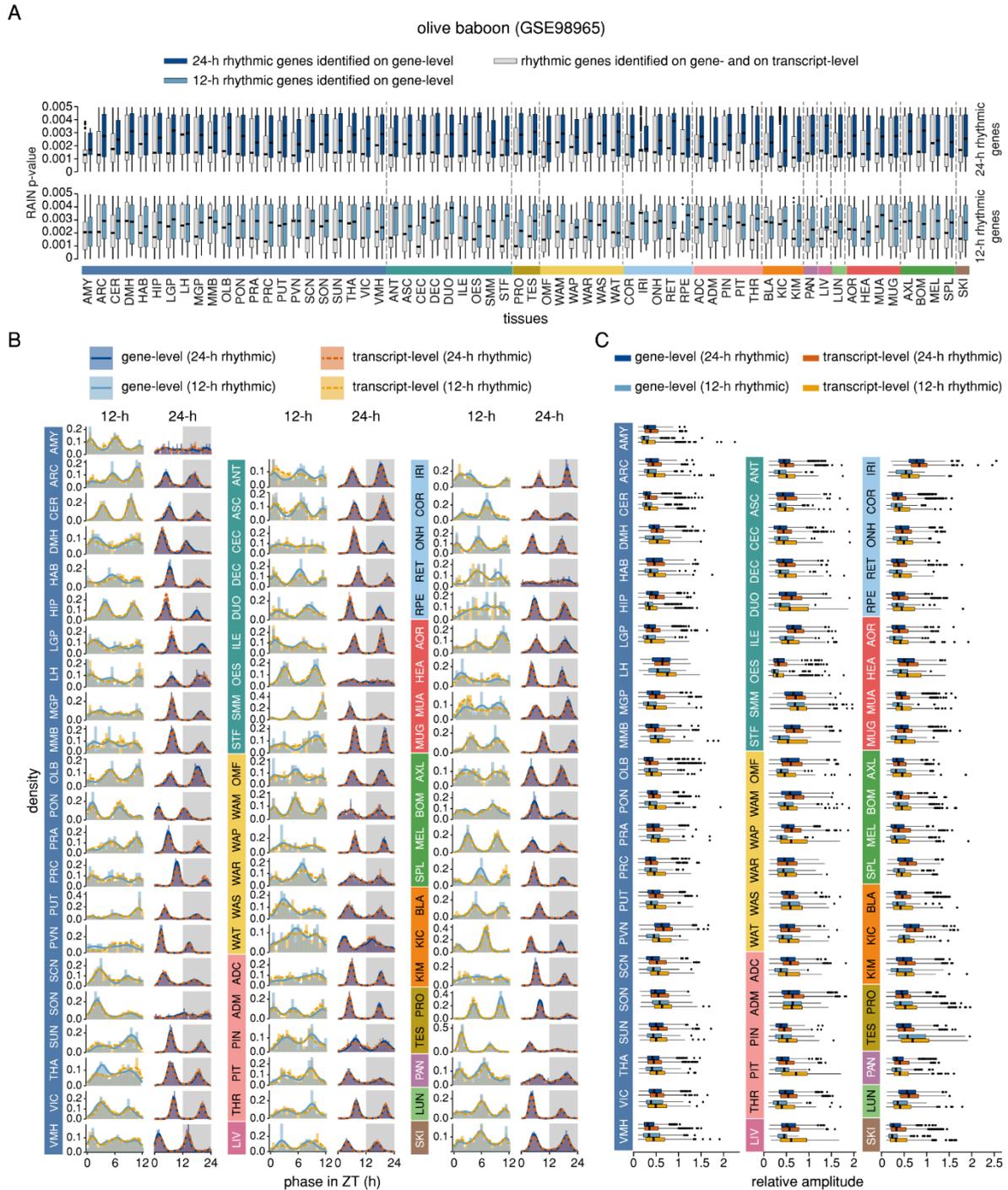


Figure S 19: Comparison of rhythmic parameters between gene- and transcript-level in baboon tissues (RNA-seq data, GSE98965). (A) Boxplots comparing the RAIN p-value distributions of genes from the reverse complement (24-h rhythmic: dark blue; 12-h rhythmic: light blue) and from the intersection (light gray) of 24-h rhythmic genes on gene- and transcript-level (left panel) and the intersection (light gray) of 12-h rhythmic genes on gene- and transcript-level (right panel) identified for each baboon tissue. (B) Distribution of peak phases of 24-h rhythmic genes (dark blue) and transcripts (dark orange) and 12-h rhythmic genes (light blue) and transcripts (light orange) of all baboon tissues. (C) Boxplots of the relative amplitudes of 24-h rhythmic genes (dark blue) and transcripts (dark orange) and 12-h rhythmic genes (light blue) and transcripts (light orange) of all baboon tissues.

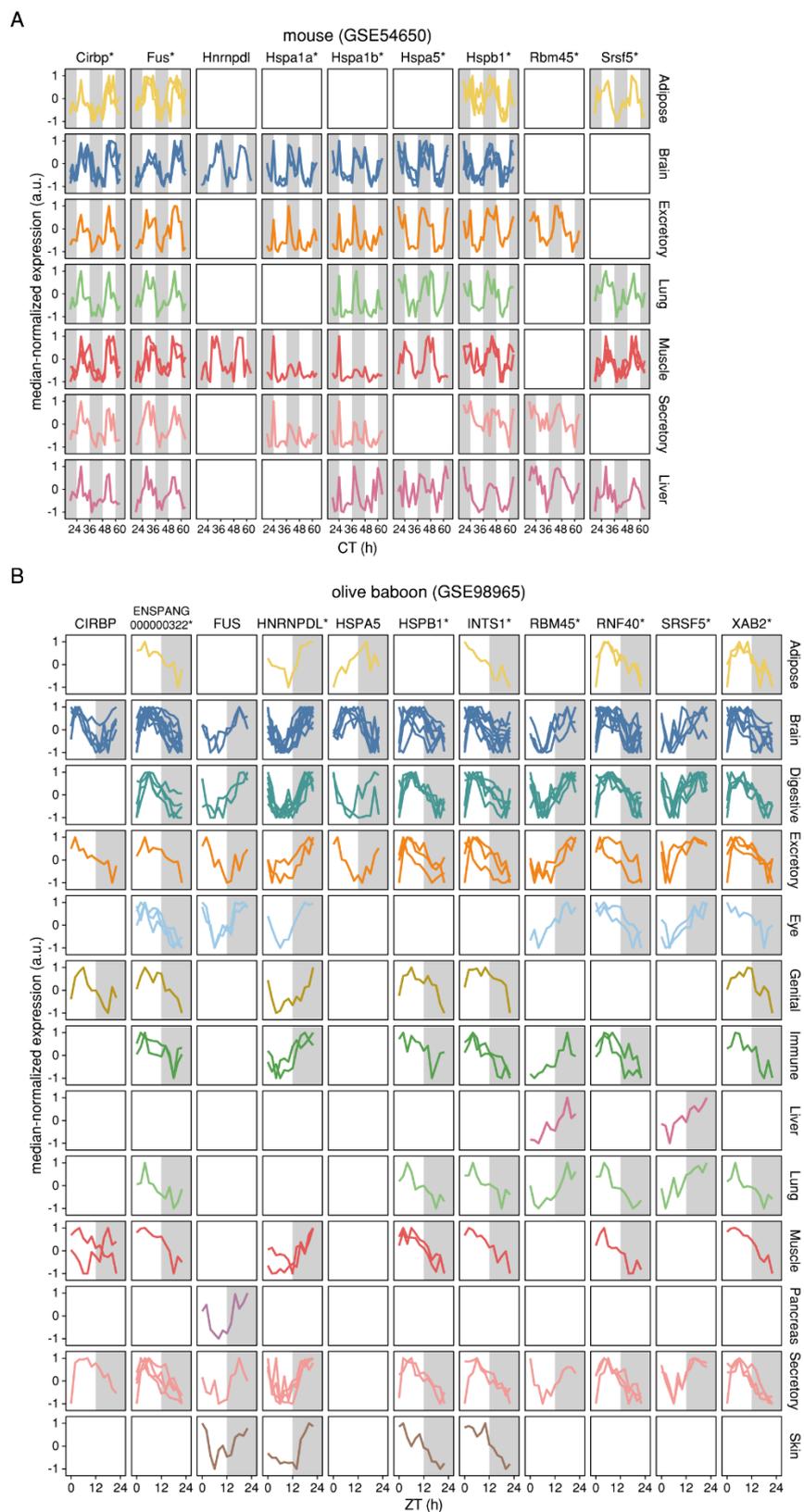


Figure S 20: Time-series expression of 24-h rhythmic splicing-related genes in mouse (GSE54650) and baboon tissues (GSE98965). Median-normalized expression (from -1 to 1) of the topmost five consistently 24-h rhythmic splicing-related genes in (A) murine and (B) baboon tissues. Additionally, splicing-related orthologous genes found to be 24-h rhythmic in both species are shown (marked with an asterisk). Gene expression for tissues from the same tissue type is represented with the same color and in the same panel.

7.1.2 Supplementary Tables

Table S 1: Lists of human splicing-related genes.

Gene	Ensembl ID	Class	List
ACIN1	ENSG00000100813	EJC/mRNP	List 1 & 2
AFF2	ENSG00000155966	linked to splicing and other regulators	List 1 & 2
ALYREF	ENSG00000183684	EJC/mRNP	List 1 & 2
AQR	ENSG00000021776	PRP19 complex and PRP19 related	List 1 & 2
BCAS2	ENSG00000116752	PRP19 complex and PRP19 related	List 1 & 2
BUD13	ENSG00000137656	RES complex	List 1 & 2
BUD31	ENSG00000106245	PRP19 complex and PRP19 related	List 1 & 2
C9orf78	ENSG00000136819	Recruited at C complex	List 1 & 2
CACTIN	ENSG00000105298	Recruited at C complex	List 1 & 2
CASC3	ENSG00000108349	EJC/mRNP	List 1 & 2
CCAR1	ENSG00000060339	Recruited at A complex	List 1 & 2
CCDC12	ENSG00000160799	Recruited at Bact complex	List 1 & 2
CCDC94	ENSG00000105248	Recruited at Bact complex	List 1 & 2
CD2BP2	ENSG00000169217	U5 snRNP	List 1 & 2
CDC40	ENSG00000168438	Second step factors	List 1 & 2
CDC5L	ENSG00000096401	PRP19 complex and PRP19 related	List 1 & 2
CDK10	ENSG00000185324	Recruited at C complex	List 1 & 2
CDK5	ENSG00000164885	linked to splicing and other regulators	List 1 & 2
CELF1	ENSG00000149187	alternative splicing factor	List 1 & 2
CELF2	ENSG00000048740	alternative splicing factor	List 1 & 2
CHERP	ENSG00000085872	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
CLASRP	ENSG00000104859	SR protein and SR related	List 1 & 2
CLK1	ENSG00000013441	SR protein and SR related	List 1 & 2
CLK2	ENSG00000176444	SR protein and SR related	List 1 & 2
CLK3	ENSG00000179335	SR protein and SR related	List 1 & 2
CLK4	ENSG00000113240	linked to splicing and other regulators	List 1 & 2
CRNKL1	ENSG00000101343	PRP19 complex and PRP19 related	List 1 & 2
CTNNBL1	ENSG00000132792	PRP19 complex and PRP19 related	List 1 & 2
CWC15	ENSG00000150316	PRP19 complex and PRP19 related	List 1 & 2
CWC22	ENSG00000163510	Recruited at Bact complex	List 1 & 2
CWC25	ENSG00000273559	Recruited at Bact complex	List 1 & 2
CWC27	ENSG00000153015	Recruited at Bact complex	List 1 & 2
CXorf56	ENSG00000018610	Recruited at C complex	List 1 & 2
DAZAP1	ENSG00000071626	linked to splicing and other regulators	List 1 & 2
DBR1	ENSG00000138231	linked to splicing and other regulators	List 1 & 2
DDX23	ENSG00000174243	U5 snRNP	List 1 & 2
DDX41	ENSG00000183258	Recruited at C complex	List 1 & 2
DDX42	ENSG00000198231	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
DDX46	ENSG00000145833	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2

Gene	Ensembl ID	Class	List
DGCR14	ENSG00000100056	Recruited at C complex	List 1 & 2
DHX15	ENSG00000109606	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
DHX16	ENSG00000204560	Recruited at Bact complex	List 1 & 2
DHX35	ENSG00000101452	Recruited at C complex	List 1 & 2
DHX38	ENSG00000140829	Second step factors	List 1 & 2
DHX8	ENSG00000067596	Second step factors	List 1 & 2
EFTUD2	ENSG00000108883	U5 snRNP	List 1 & 2
EIF4A3	ENSG00000141543	EJC/mRNP	List 1 & 2
ELAVL1	ENSG00000066044	alternative splicing factor	List 1 & 2
ELAVL2	ENSG00000107105	linked to splicing and other regulators	List 1 & 2
ELAVL3	ENSG00000196361	linked to splicing and other regulators	List 1 & 2
ELAVL4	ENSG00000162374	linked to splicing and other regulators	List 1 & 2
ESRP1	ENSG00000104413	alternative splicing factor	List 1 & 2
ESRP2	ENSG00000103067	linked to splicing and other regulators	List 1 & 2
FAM32A	ENSG00000105058	Recruited at C complex	List 1 & 2
FAM50A	ENSG00000071859	Recruited at C complex	List 1 & 2
FMR1	ENSG00000102081	linked to splicing and other regulators	List 1 & 2
FRA10AC1	ENSG00000148690	Recruited at C complex	List 1 & 2
FUBP1	ENSG00000162613	alternative splicing factor	List 1 & 2
FUS	ENSG00000089280	hnRNP	List 1 & 2
FXR1	ENSG00000114416	linked to splicing and other regulators	List 1 & 2
GPATCH1	ENSG00000076650	Recruited at Bact complex	List 1 & 2
GPKOW	ENSG00000068394	Recruited at Bact complex	List 1 & 2
HNRNPA0	ENSG00000177733	hnRNP	List 1 & 2
HNRNPA1	ENSG00000135486	hnRNP	List 1 & 2
HNRNPA2B1	ENSG00000122566	hnRNP	List 1 & 2
HNRNPA3	ENSG00000170144	hnRNP	List 1 & 2
HNRNPAB	ENSG00000197451	hnRNP	List 1 & 2
HNRNPC	ENSG00000092199	hnRNP	List 1 & 2
HNRNPCL1	ENSG00000179172	hnRNP	List 1 & 2
HNRNPD	ENSG00000138668	hnRNP	List 1 & 2
HNRNPD1	ENSG00000152795	hnRNP	List 1 & 2
HNRNPF	ENSG00000169813	hnRNP	List 1 & 2
HNRNPH1	ENSG00000169045	hnRNP	List 1 & 2
HNRNPH2	ENSG00000126945	hnRNP	List 1 & 2
HNRNPH3	ENSG00000096746	hnRNP	List 1 & 2
HNRNPK	ENSG00000165119	hnRNP	List 1 & 2
HNRNPL	ENSG00000104824	hnRNP	List 1 & 2
HNRNPLL	ENSG00000143889	hnRNP	List 1 & 2
HNRNPM	ENSG00000099783	hnRNP	List 1 & 2
HNRNPR	ENSG00000125944	hnRNP	List 1 & 2
HNRNPU	ENSG00000153187	hnRNP	List 1 & 2

Gene	Ensembl ID	Class	List
HNRNPUL1	ENSG00000105323	hnRNP	List 1 & 2
HNRNPUL2	ENSG00000214753	hnRNP	List 1 & 2
HRH1	ENSG00000196639	SR protein and SR related	List 1 & 2
HSPA8	ENSG00000109971	PRP19 complex and PRP19 related	List 1 & 2
HTATSF1	ENSG00000102241	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
IK	ENSG00000113141	Recruited at B complex	List 1 & 2
ISY1	ENSG00000240682	PRP19 complex and PRP19 related	List 1 & 2
KHDRBS1	ENSG00000121774	linked to splicing and other regulators	List 1 & 2
KHDRBS2	ENSG00000112232	linked to splicing and other regulators	List 1 & 2
KHDRBS3	ENSG00000131773	linked to splicing and other regulators	List 1 & 2
KHSRP	ENSG00000088247	alternative splicing factor	List 1 & 2
LENG1	ENSG00000105617	Recruited at C complex	List 1 & 2
LSM1	ENSG00000175324	Sm/LSm	List 1 & 2
LSM2	ENSG00000204392	Sm/LSm	List 1 & 2
LSM3	ENSG00000170860	Sm/LSm	List 1 & 2
LSM4	ENSG00000130520	Sm/LSm	List 1 & 2
LSM5	ENSG00000106355	Sm/LSm	List 1 & 2
LSM6	ENSG00000164167	Sm/LSm	List 1 & 2
LSM7	ENSG00000130332	Sm/LSm	List 1 & 2
LSM8	ENSG00000128534	Sm/LSm	List 1 & 2
LUC7L	ENSG00000007392	U1 snRNP	List 1 & 2
MAGOH	ENSG00000162385	EJC/mRNP	List 1 & 2
MBNL1	ENSG00000152601	alternative splicing factor	List 1 & 2
MBNL2	ENSG00000139793	alternative splicing factor	List 1 & 2
MBNL3	ENSG00000076770	alternative splicing factor	List 1 & 2
MFAP1	ENSG00000140259	Recruited at B complex	List 1 & 2
MIP	ENSG00000135517	hnRNP	List 1 & 2
NONO	ENSG00000147140	linked to splicing and other regulators	List 1 & 2
NOSIP	ENSG00000142546	Recruited at C complex	List 1 & 2
NOVA1	ENSG00000139910	linked to splicing and other regulators	List 1 & 2
NOVA2	ENSG00000104967	linked to splicing and other regulators	List 1 & 2
NUFIP1	ENSG00000083635	linked to splicing and other regulators	List 1 & 2
NXF1	ENSG00000162231	EJC/mRNP	List 1 & 2
NXT1	ENSG00000132661	EJC/mRNP	List 1 & 2
PCBP1	ENSG00000169564	hnRNP	List 1 & 2
PCBP2	ENSG00000197111	hnRNP	List 1 & 2
PHF5A	ENSG00000100410	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
PLRG1	ENSG00000171566	PRP19 complex and PRP19 related	List 1 & 2
PNN	ENSG00000100941	EJC/mRNP	List 1 & 2
PPIE	ENSG00000084072	PRP19 complex and PRP19 related	List 1 & 2
PPIG	ENSG00000138398	Recruited at C complex	List 1 & 2
PPIH	ENSG00000171960	U4/U6 snRNP	List 1 & 2

Gene	Ensembl ID	Class	List
PPIL1	ENSG00000137168	PRP19 complex and PRP19 related	List 1 & 2
PPIL2	ENSG00000100023	Recruited at Bact complex	List 1 & 2
PPIL3	ENSG00000240344	Recruited at C complex	List 1 & 2
PPIL4	ENSG00000131013	linked to splicing and other regulators	List 1 & 2
PPWD1	ENSG00000113593	Recruited at C complex	List 1 & 2
PQBP1	ENSG00000102103	PRP19 complex and PRP19 related	List 1 & 2
PRCC	ENSG00000143294	Recruited at Bact complex	List 1 & 2
PRPF18	ENSG00000165630	Second step factors	List 1 & 2
PRPF19	ENSG00000110107	PRP19 complex and PRP19 related	List 1 & 2
PRPF3	ENSG00000117360	U4/U6 snRNP	List 1 & 2
PRPF31	ENSG00000105618	U4/U6 snRNP	List 1 & 2
PRPF38A	ENSG00000134748	Recruited at B complex	List 1 & 2
PRPF4	ENSG00000136875	U4/U6 snRNP	List 1 & 2
PRPF40A	ENSG00000196504	Recruited at A complex	List 1 & 2
PRPF4B	ENSG00000112739	Recruited at B complex	List 1 & 2
PRPF6	ENSG00000101161	U5 snRNP	List 1 & 2
PRPF8	ENSG00000174231	U5 snRNP	List 1 & 2
PTBP1	ENSG00000011304	alternative splicing factor	List 1 & 2
PTBP2	ENSG00000117569	alternative splicing factor	List 1 & 2
PUF60	ENSG00000179950	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
QKI	ENSG00000112531	alternative splicing factor	List 1 & 2
RALY	ENSG00000125970	hnRNP	List 1 & 2
RALYL	ENSG00000184672	hnRNP	List 1 & 2
RAVER1	ENSG00000161847	alternative splicing factor	List 1 & 2
RAVER2	ENSG00000162437	alternative splicing factor	List 1 & 2
RBFOX1	ENSG00000078328	linked to splicing and other regulators	List 1 & 2
RBFOX2	ENSG00000100320	alternative splicing factor	List 1 & 2
RBM10	ENSG00000182872	Recruited at A complex	List 1 & 2
RBM17	ENSG00000134453	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
RBM22	ENSG00000086589	PRP19 complex and PRP19 related	List 1 & 2
RBM25	ENSG00000119707	Recruited at A complex	List 1 & 2
RBM39	ENSG00000131051	SR protein and SR related	List 1 & 2
RBM4	ENSG00000173933	linked to splicing and other regulators	List 1 & 2
RBM5	ENSG00000003756	Recruited at A complex	List 1 & 2
RBM8A	ENSG00000265241	EJC/mRNP	List 1 & 2
RBMX	ENSG00000147274	hnRNP	List 1 & 2
RBMX2	ENSG00000134597	RES complex	List 1 & 2
RBMXL2	ENSG00000170748	hnRNP	List 1 & 2
RNF113A	ENSG00000125352	Recruited at Bact complex	List 1 & 2
RNPC3	ENSG00000185946	U11/U12 snRNP	List 1 & 2
RNPS1	ENSG00000205937	EJC/mRNP	List 1 & 2
RNU1-1	ENSG00000206652	U1 snRNP	List 1 & 2

Gene	Ensembl ID	Class	List
RNU2-1	ENSG00000274585	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
RNU2-65P	ENSG00000222094	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
RNU4-1	ENSG00000200795	U4/U6 snRNP	List 1 & 2
RNU5A-1	ENSG00000199568	U5 snRNP	List 1 & 2
RNU6-1	ENSG00000206625	U4/U6 snRNP	List 1 & 2
SAP18	ENSG00000150459	EJC/mRNP	List 1 & 2
SART1	ENSG00000175467	tri-snRNP	List 1 & 2
SART3	ENSG00000075856	U4/U6 recycling	List 1 & 2
SDE2	ENSG00000143751	Recruited at C complex	List 1 & 2
SF1	ENSG00000168066	Recruited at A complex	List 1 & 2
SF3A1	ENSG00000099995	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SF3A2	ENSG00000104897	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SF3A3	ENSG00000183431	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SF3B1	ENSG00000115524	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SF3B2	ENSG00000087365	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SF3B3	ENSG00000189091	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SF3B4	ENSG00000143368	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SF3B5	ENSG00000169976	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SF3B6	ENSG00000115128	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SFPQ	ENSG00000116560	linked to splicing and other regulators	List 1 & 2
SFSWAP	ENSG00000061936	SR protein and SR related	List 1 & 2
SLU7	ENSG00000164609	Second step factors	List 1 & 2
SMNDC1	ENSG00000119953	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SMU1	ENSG00000122692	Recruited at B complex	List 1 & 2
SNIP1	ENSG00000163877	RES complex	List 1 & 2
SNRNP200	ENSG00000144028	U5 snRNP	List 1 & 2
SNRNP25	ENSG00000161981	U11/U12 snRNP	List 1 & 2
SNRNP35	ENSG00000184209	U11/U12 snRNP	List 1 & 2
SNRNP40	ENSG00000060688	U5 snRNP	List 1 & 2
SNRNP48	ENSG00000168566	U11/U12 snRNP	List 1 & 2
SNRNP70	ENSG00000104852	U1 snRNP	List 1 & 2
SNRPA	ENSG00000077312	U1 snRNP	List 1 & 2
SNRPA1	ENSG00000131876	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SNRPB	ENSG00000125835	Sm/LSm	List 1 & 2
SNRPB2	ENSG00000125870	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
SNRPC	ENSG00000124562	U1 snRNP	List 1 & 2
SNRPD1	ENSG00000167088	Sm/LSm	List 1 & 2
SNRPD2	ENSG00000125743	Sm/LSm	List 1 & 2
SNRPD3	ENSG00000100028	Sm/LSm	List 1 & 2
SNRPE	ENSG00000182004	Sm/LSm	List 1 & 2
SNRPF	ENSG00000139343	Sm/LSm	List 1 & 2
SNRPG	ENSG00000143977	Sm/LSm	List 1 & 2

Gene	Ensembl ID	Class	List
SNU13	ENSG00000100138	U4/U6 snRNP	List 1 & 2
SNW1	ENSG00000100603	PRP19 complex and PRP19 related	List 1 & 2
SREK1	ENSG00000153914	SR protein and SR related	List 1 & 2
SRPK1	ENSG00000096063	SR protein and SR related	List 1 & 2
SRPK2	ENSG00000135250	SR protein and SR related	List 1 & 2
SRRM1	ENSG00000133226	SR protein and SR related	List 1 & 2
SRRM2	ENSG00000167978	SR protein and SR related	List 1 & 2
SRSF1	ENSG00000136450	SR protein and SR related	List 1 & 2
SRSF10	ENSG00000188529	SR protein and SR related	List 1 & 2
SRSF11	ENSG00000116754	SR protein and SR related	List 1 & 2
SRSF12	ENSG00000154548	SR protein and SR related	List 1 & 2
SRSF2	ENSG00000161547	SR protein and SR related	List 1 & 2
SRSF3	ENSG00000112081	SR protein and SR related	List 1 & 2
SRSF4	ENSG00000116350	SR protein and SR related	List 1 & 2
SRSF5	ENSG00000100650	SR protein and SR related	List 1 & 2
SRSF6	ENSG00000124193	SR protein and SR related	List 1 & 2
SRSF7	ENSG00000115875	SR protein and SR related	List 1 & 2
SRSF8	ENSG00000263465	SR protein and SR related	List 1 & 2
SRSF9	ENSG00000111786	SR protein and SR related	List 1 & 2
SUGP1	ENSG00000105705	Recruited at A complex	List 1 & 2
SYF2	ENSG00000117614	Recruited at C complex	List 1 & 2
SYNCRIP	ENSG00000135316	hnRNP	List 1 & 2
TARDBP	ENSG00000120948	linked to splicing and other regulators	List 1 & 2
TDRD9	ENSG00000156414	linked to splicing and other regulators	List 1 & 2
TFIP11	ENSG00000100109	Recruited at B complex	List 1 & 2
THRAP3	ENSG00000054118	Recruited at A complex	List 1 & 2
TIA1	ENSG00000116001	linked to splicing and other regulators	List 1 & 2
TIAL1	ENSG00000151923	linked to splicing and other regulators	List 1 & 2
TOP1	ENSG00000198900	linked to splicing and other regulators	List 1 & 2
TOPORS	ENSG00000197579	linked to splicing and other regulators	List 1 & 2
TRA2A	ENSG00000164548	SR protein and SR related	List 1 & 2
TRA2B	ENSG00000136527	SR protein and SR related	List 1 & 2
TXNL4A	ENSG00000141759	U5 snRNP	List 1 & 2
U2AF1	ENSG00000160201	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
U2AF2	ENSG00000063244	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
U2SURP	ENSG00000163714	17S U2 snRNP and 17S U2 snRNP associated	List 1 & 2
UPF1	ENSG00000005007	EJC/mRNP	List 1 & 2
USP39	ENSG00000168883	tri-snRNP	List 1 & 2
WBP11	ENSG00000084463	PRP19 complex and PRP19 related	List 1 & 2
WBP4	ENSG00000120688	Recruited at B complex	List 1 & 2
WDR83	ENSG00000123154	Recruited at C complex	List 1 & 2
XAB2	ENSG00000076924	PRP19 complex and PRP19 related	List 1 & 2

Gene	Ensembl ID	Class	List
YBX1	ENSG00000065978	linked to splicing and other regulators	List 1 & 2
ZMAT2	ENSG00000146007	Recruited at B complex	List 1 & 2
ZNF830	ENSG00000198783	Recruited at Bact complex	List 1 & 2
ZRANB2	ENSG00000132485	linked to splicing and other regulators	List 1 & 2
ZRSR2	ENSG00000169249	U11/U12 snRNP	List 1 & 2
AGGF1	ENSG00000164252	NA	List 2
ARGLU1	ENSG00000134884	NA	List 2
BAG2	ENSG00000112208	NA	List 2
BCAS1	ENSG00000064787	NA	List 2
BUB3	ENSG00000154473	NA	List 2
C17orf85	ENSG00000074356	NA	List 2
C19orf43	ENSG00000123144	NA	List 2
C1QBP	ENSG00000108561	NA	List 2
CCDC130	ENSG00000104957	NA	List 2
CCDC75	ENSG00000152133	NA	List 2
CDK11A	ENSG00000008128	NA	List 2
CDK12	ENSG00000167258	NA	List 2
CELF3	ENSG00000159409	NA	List 2
CELF4	ENSG00000101489	NA	List 2
CELF5	ENSG00000161082	NA	List 2
CELF6	ENSG00000140488	NA	List 2
CFAP20	ENSG00000070761	NA	List 2
CIRBP	ENSG00000099622	NA	List 2
CLNS1A	ENSG00000074201	NA	List 2
CPSF6	ENSG00000111605	NA	List 2
CSN3	ENSG00000171209	NA	List 2
DDX1	ENSG00000079785	NA	List 2
DDX17	ENSG00000100201	NA	List 2
DDX18	ENSG00000088205	NA	List 2
DDX19A	ENSG00000168872	NA	List 2
DDX19B	ENSG00000157349	NA	List 2
DDX20	ENSG00000064703	NA	List 2
DDX21	ENSG00000165732	NA	List 2
DDX26B	ENSG00000165359	NA	List 2
DDX27	ENSG00000124228	NA	List 2
DDX39A	ENSG00000123136	NA	List 2
DDX39B	ENSG00000198563	NA	List 2
DDX3X	ENSG00000215301	NA	List 2
DDX3Y	ENSG00000067048	NA	List 2
DDX5	ENSG00000108654	NA	List 2
DDX50	ENSG00000107625	NA	List 2
DDX6	ENSG00000110367	NA	List 2

Gene	Ensembl ID	Class	List
DHX30	ENSG00000132153	NA	List 2
DHX34	ENSG00000134815	NA	List 2
DHX36	ENSG00000174953	NA	List 2
DHX40	ENSG00000108406	NA	List 2
DHX57	ENSG00000163214	NA	List 2
DHX9	ENSG00000135829	NA	List 2
DNAJC6	ENSG00000116675	NA	List 2
DNAJC8	ENSG00000126698	NA	List 2
EEF1A1	ENSG00000156508	NA	List 2
EIF2S2	ENSG00000125977	NA	List 2
EIF3A	ENSG00000107581	NA	List 2
FAM50B	ENSG00000145945	NA	List 2
FAM58A	ENSG00000262919	NA	List 2
FRG1	ENSG00000109536	NA	List 2
FUBP3	ENSG00000107164	NA	List 2
GEMIN2	ENSG00000092208	NA	List 2
GEMIN5	ENSG00000082516	NA	List 2
GNB2L1	ENSG00000204628	NA	List 2
GPATCH3	ENSG00000198746	NA	List 2
GPATCH8	ENSG00000186566	NA	List 2
GRSF1	ENSG00000132463	NA	List 2
HNRNPCL3	ENSG00000277058	NA	List 2
HSPA1A	ENSG00000204389	NA	List 2
HSPA1B	ENSG00000204388	NA	List 2
HSPA5	ENSG00000044574	NA	List 2
HSPB1	ENSG00000106211	NA	List 2
IGF2BP3	ENSG00000136231	NA	List 2
ILF2	ENSG00000143621	NA	List 2
ILF3	ENSG00000129351	NA	List 2
INTS1	ENSG00000164880	NA	List 2
INTS3	ENSG00000143624	NA	List 2
INTS4	ENSG00000149262	NA	List 2
INTS5	ENSG00000185085	NA	List 2
INTS6	ENSG00000102786	NA	List 2
INTS7	ENSG00000143493	NA	List 2
JUP	ENSG00000173801	NA	List 2
KIAA1429	ENSG00000164944	NA	List 2
KIAA1967	ENSG00000158941	NA	List 2
KIN	ENSG00000151657	NA	List 2
LSM10	ENSG00000181817	NA	List 2
LSMD1	ENSG00000183011	NA	List 2
LUC7L2	ENSG00000146963	NA	List 2

Gene	Ensembl ID	Class	List
LUC7L3	ENSG00000108848	NA	List 2
MATR3	ENSG00000015479	NA	List 2
MFSD11	ENSG00000092931	NA	List 2
MOV10	ENSG00000155363	NA	List 2
MSI1	ENSG00000135097	NA	List 2
MSI2	ENSG00000153944	NA	List 2
MYEF2	ENSG00000104177	NA	List 2
NCBP1	ENSG00000136937	NA	List 2
NCBP2	ENSG00000114503	NA	List 2
NELFE	ENSG00000204356	NA	List 2
NKAP	ENSG00000101882	NA	List 2
NRIP2	ENSG00000053702	NA	List 2
NSRP1	ENSG00000126653	NA	List 2
NUDT21	ENSG00000167005	NA	List 2
NUMA1	ENSG00000137497	NA	List 2
PABPC1	ENSG00000070756	NA	List 2
PAXBPI	ENSG00000159086	NA	List 2
PCBP3	ENSG00000183570	NA	List 2
PCBP4	ENSG00000090097	NA	List 2
PDCD7	ENSG00000090470	NA	List 2
PPM1G	ENSG00000115241	NA	List 2
PPP1CA	ENSG00000172531	NA	List 2
PPP1R8	ENSG00000117751	NA	List 2
PRMT5	ENSG00000100462	NA	List 2
PRPF38B	ENSG00000134186	NA	List 2
PRPF39	ENSG00000185246	NA	List 2
PRPF40B	ENSG00000110844	NA	List 2
PSEN1	ENSG00000080815	NA	List 2
PSIP1	ENSG00000164985	NA	List 2
PTBP3	ENSG00000119314	NA	List 2
RBBP6	ENSG00000122257	NA	List 2
RBM14	ENSG00000239306	NA	List 2
RBM15	ENSG00000162775	NA	List 2
RBM15B	ENSG00000259956	NA	List 2
RBM23	ENSG00000100461	NA	List 2
RBM26	ENSG00000139746	NA	List 2
RBM27	ENSG00000091009	NA	List 2
RBM3	ENSG00000102317	NA	List 2
RBM42	ENSG00000126254	NA	List 2
RBM45	ENSG00000155636	NA	List 2
RBM47	ENSG00000163694	NA	List 2
RBM4B	ENSG00000173914	NA	List 2

Gene	Ensembl ID	Class	List
RBM7	ENSG00000076053	NA	List 2
RBMS1	ENSG00000153250	NA	List 2
RBMXL1	ENSG00000213516	NA	List 2
RNF20	ENSG00000155827	NA	List 2
RNF213	ENSG00000173821	NA	List 2
RNF34	ENSG00000170633	NA	List 2
RNF40	ENSG00000103549	NA	List 2
SAP30BP	ENSG00000161526	NA	List 2
SEC31B	ENSG00000075826	NA	List 2
SKIV2L2	ENSG00000039123	NA	List 2
SMN1	ENSG00000172062	NA	List 2
SNRNP27	ENSG00000124380	NA	List 2
SNRPN	ENSG00000128739	NA	List 2
SNURF	ENSG00000273173	NA	List 2
SPEN	ENSG00000065526	NA	List 2
SRPK3	ENSG00000184343	NA	List 2
SRRT	ENSG00000087087	NA	List 2
SSB	ENSG00000138385	NA	List 2
TAF15	ENSG00000270647	NA	List 2
TCERG1	ENSG00000113649	NA	List 2
THOC1	ENSG00000079134	NA	List 2
THOC2	ENSG00000125676	NA	List 2
THOC3	ENSG00000051596	NA	List 2
THOC5	ENSG00000100296	NA	List 2
THOC6	ENSG00000131652	NA	List 2
THOC7	ENSG00000163634	NA	List 2
TNPO1	ENSG00000083312	NA	List 2
TOE1	ENSG00000132773	NA	List 2
TOP1MT	ENSG00000184428	NA	List 2
TRIM24	ENSG00000122779	NA	List 2
TTC14	ENSG00000163728	NA	List 2
U2AF1L4	ENSG00000161265	NA	List 2
UBL5	ENSG00000198258	NA	List 2
WDR77	ENSG00000116455	NA	List 2
WTAP	ENSG00000146457	NA	List 2
YBX3	ENSG00000060138	NA	List 2
ZC3H11A	ENSG00000058673	NA	List 2
ZC3H13	ENSG00000123200	NA	List 2
ZC3H18	ENSG00000158545	NA	List 2
ZC3H4	ENSG00000130749	NA	List 2
ZC3HAV1	ENSG00000105939	NA	List 2
ZCCHC10	ENSG00000155329	NA	List 2

Gene	Ensembl ID	Class	List
ZCCHC8	ENSG00000033030	NA	List 2
ZCRB1	ENSG00000139168	NA	List 2
ZFR	ENSG00000056097	NA	List 2
ZMAT5	ENSG00000100319	NA	List 2
ZMYM3	ENSG00000147130	NA	List 2
ZNF131	ENSG00000172262	NA	List 2
ZNF207	ENSG00000010244	NA	List 2
ZNF326	ENSG00000162664	NA	List 2
ZNF346	ENSG00000113761	NA	List 2
ZRSR2P1	ENSG00000212643	NA	List 2

Table S 2: Candidate exons with differential AS events between SW480 and SW620 cells (microarray data).

Gene	Ensembl gene ID	Ensembl exon ID	mean FIRMA score log ₂ FC (SW480-SW620)
ABHD6	ENSG00000163686	ENSE00001905638	-1.27
ABLIM1	ENSG00000099204	ENSE00001000257	1.17
AC098864.1	ENSG00000177822	ENSE00002042142	1.11
AC244153.1	ENSG00000276170	ENSE00003733044	-1.03
ACD	ENSG00000102977	ENSE00003326886	1.47
ACP6	ENSG00000162836	ENSE00002685668	-1.09
ACSF2	ENSG00000167107	ENSE00002018997	1.47
ACSF2	ENSG00000167107	ENSE00002062313	1.00
ACSS1	ENSG00000154930	ENSE00001018478	1.24
ACSS2	ENSG00000131069	ENSE00003499545	1.16
ACTN4	ENSG00000130402	ENSE00001234718	1.02
ACTN4	ENSG00000130402	ENSE00003479194	1.09
ACTR1B	ENSG00000115073	ENSE00001750306	1.06
ACTR1B	ENSG00000115073	ENSE00002463641	1.09
ADM	ENSG00000148926	ENSE00000988124	1.01
AGPAT3	ENSG00000160216	ENSE00001137610	1.05
AHI1	ENSG00000135541	ENSE00002149823	-1.15
AIG1	ENSG00000146416	ENSE00001760143	-1.13
AK4	ENSG00000162433	ENSE00002341856	-1.32
AKNAD1	ENSG00000162641	ENSE00002035173	1.25
AKR1B1	ENSG00000085662	ENSE00001734693	1.04
AKR1B1	ENSG00000085662	ENSE00001846312	-1.21
AKR1B1	ENSG00000085662	ENSE00003526531	-1.61
AKR1B1	ENSG00000085662	ENSE00003572588	-1.31
AKR1C3	ENSG00000187134	ENSE00003472155	-1.29
AL365277.1	ENSG00000182109	ENSE00001801843	1.02
ALDH18A1	ENSG00000059573	ENSE00000987354	-1.10
ALDH1A3	ENSG00000184254	ENSE00002555275	1.35
AMPD2	ENSG00000116337	ENSE00001889927	1.01
AMPD3	ENSG00000133805	ENSE00002173290	-1.30
ANO7	ENSG00000146205	ENSE00001740295	-1.33
ANXA1	ENSG00000135046	ENSE00001650954	-1.00
ANXA6	ENSG00000197043	ENSE00002118867	-1.14
ANXA6	ENSG00000197043	ENSE00003625652	1.24
AOC1	ENSG00000002726	ENSE00003716442	1.23
AQP3	ENSG00000165272	ENSE00003321062	-1.01
AQP3	ENSG00000165272	ENSE00003467346	-1.12
ARHGAP45	ENSG00000180448	ENSE00003741889	1.33
ARHGAP8	ENSG00000241484	ENSE00001900438	1.26
ARMCX6	ENSG00000198960	ENSE00001941545	-1.13
ASAP1	ENSG00000153317	ENSE00003519417	1.08

Gene	Ensembl gene ID	Ensembl exon ID	mean FIRMA score log ₂ FC (SW480-SW620)
ATP1A3	ENSG00000105409	ENSE00003625885	1.00
ATP5SL	ENSG00000105341	ENSE00003044100	1.16
ATP9B	ENSG00000166377	ENSE00002763592	-1.00
AUTS2	ENSG00000158321	ENSE00001943467	1.11
BCL11A	ENSG00000119866	ENSE00001945539	-1.04
BMP7	ENSG00000101144	ENSE00001607441	1.22
BTBD11	ENSG00000151136	ENSE00002363193	1.35
BTBD6	ENSG00000184887	ENSE00001296252	1.01
C3orf55	ENSG00000174899	ENSE00001907830	1.16
C3orf58	ENSG00000181744	ENSE00003506084	1.06
C9orf152	ENSG00000188959	ENSE00001905392	-1.15
CAB39L	ENSG00000102547	ENSE00001900877	1.10
CADM1	ENSG00000182985	ENSE00002217285	-1.10
CADM1	ENSG00000182985	ENSE00002256764	1.45
CADPS2	ENSG00000081803	ENSE00001776365	-1.03
CALB1	ENSG00000104327	ENSE00000699388	-1.15
CALB1	ENSG00000104327	ENSE00001087859	-1.05
CAMK2D	ENSG00000145349	ENSE00002462786	-1.72
CASP4	ENSG00000196954	ENSE00003732892	-1.10
CAV1	ENSG00000105974	ENSE00001515426	-1.10
CAV1	ENSG00000105974	ENSE00001708326	1.10
CBLB	ENSG00000114423	ENSE00000774844	-1.11
CCND2	ENSG00000118971	ENSE00002239432	1.01
CCNI	ENSG00000118816	ENSE00002086195	1.05
CD22	ENSG00000012124	ENSE00002995594	1.19
CD4	ENSG00000010610	ENSE00002219192	-1.24
CD4	ENSG00000010610	ENSE00002308039	-1.25
CD4	ENSG00000010610	ENSE00003482311	-1.21
CD44	ENSG00000026508	ENSE00002157331	1.44
CD74	ENSG00000019582	ENSE00000841201	1.01
CDH23	ENSG00000107736	ENSE00003738810	1.23
CDHR3	ENSG00000128536	ENSE00001835091	1.02
CHST10	ENSG00000115526	ENSE00001788866	-1.16
CHST11	ENSG00000171310	ENSE00002423376	-1.41
CHTF18	ENSG00000127586	ENSE00001643670	-1.10
CIB2	ENSG00000136425	ENSE00002549216	1.07
CKLF-CMTM1	ENSG00000217555	ENSE00002169715	-1.12
CKMT2	ENSG00000131730	ENSE00002024702	-1.01
CKMT2	ENSG00000131730	ENSE00002047544	1.12
CMTM3	ENSG00000140931	ENSE00003461182	-1.03
COL9A3	ENSG00000092758	ENSE00001952511	-1.40
COL9A3	ENSG00000092758	ENSE00003594174	1.07
CRABP2	ENSG00000143320	ENSE00001446611	1.57

Gene	Ensembl gene ID	Ensembl exon ID	mean FIRMA score log ₂ FC (SW480-SW620)
CSK	ENSG00000103653	ENSE00002624329	1.13
CTNND1	ENSG00000198561	ENSE00001189241	1.10
CUTA	ENSG00000112514	ENSE00002178771	-1.07
CXorf57	ENSG00000147231	ENSE00001458047	-1.36
CYBRD1	ENSG00000071967	ENSE00001895114	1.52
DCLK1	ENSG00000133083	ENSE00001482836	1.14
DDX39B	ENSG00000198563	ENSE00001663180	1.05
DFNB31	ENSG00000095397	ENSE00001141441	-1.10
DGKA	ENSG00000065357	ENSE00002340730	-1.10
DNAH14	ENSG00000185842	ENSE00002268510	-1.26
DPEP1	ENSG00000015413	ENSE00002603136	-1.85
DPP10	ENSG00000175497	ENSE00001881737	1.01
DPYSL3	ENSG00000113657	ENSE00002076325	1.18
EDAR	ENSG00000135960	ENSE00000804459	1.52
EFEMP2	ENSG00000172638	ENSE00002144397	-1.03
EFEMP2	ENSG00000172638	ENSE00002169361	-1.38
EGFR	ENSG00000146648	ENSE00001879915	-1.01
EPAS1	ENSG00000116016	ENSE00001826981	1.19
EPAS1	ENSG00000116016	ENSE00001886817	-1.11
EPB41L4A	ENSG00000129595	ENSE00002084379	-1.11
EPHB4	ENSG00000196411	ENSE00001893555	1.20
EPHB4	ENSG00000196411	ENSE00001925508	-1.20
ERBB3	ENSG00000065361	ENSE00002427203	-1.01
ETV1	ENSG00000006468	ENSE00001824366	-1.14
FABP3	ENSG00000121769	ENSE00003503829	1.21
FAM129A	ENSG00000135842	ENSE00001602351	1.03
FAM73B	ENSG00000148343	ENSE00001685159	-1.04
FAM96B	ENSG00000166595	ENSE00001621441	-1.13
FAM96B	ENSG00000166595	ENSE00002610746	-1.27
FGFBP2	ENSG00000007062	ENSE00002056170	1.28
FGFR2	ENSG00000066468	ENSE00001350377	1.07
FGFR2	ENSG00000066468	ENSE00001643296	-1.13
FILIP1L	ENSG00000168386	ENSE00001822993	1.32
FN1	ENSG00000115414	ENSE00001342442	1.29
FOS	ENSG00000170345	ENSE00002530715	1.23
FOXP1	ENSG00000114861	ENSE00003529965	-1.05
FOXP2	ENSG00000128573	ENSE00001352900	-1.03
FOXP2	ENSG00000128573	ENSE00001515531	1.25
FOXP2	ENSG00000128573	ENSE00001709005	1.20
FXYP3	ENSG00000089356	ENSE00001114491	-1.19
GABARAPL1	ENSG00000139112	ENSE00002283585	-1.09
GALNT6	ENSG00000139629	ENSE00003552051	1.44
GALNT6	ENSG00000139629	ENSE00003565920	1.35

Gene	Ensembl gene ID	Ensembl exon ID	mean FIRMA score log ₂ FC (SW480-SW620)
GALNT7	ENSG00000109586	ENSE00003486021	-1.23
GBP3	ENSG00000117226	ENSE00002587109	1.16
GLI3	ENSG00000106571	ENSE00001748702	-1.12
GPR64	ENSG00000173698	ENSE00001313815	1.37
GRIK2	ENSG00000164418	ENSE00001609672	-1.05
GSTM4	ENSG00000168765	ENSE00001451016	1.20
GULP1	ENSG00000144366	ENSE00001891195	-1.17
GYG2	ENSG00000056998	ENSE00002101522	-1.17
GYLTL1B	ENSG00000165905	ENSE00003547214	1.28
HDX	ENSG00000165259	ENSE00001838765	-1.28
HGD	ENSG00000113924	ENSE00001821570	-1.02
HMGN5	ENSG00000198157	ENSE00001704577	-1.19
HNF1A	ENSG00000135100	ENSE00002538985	-1.13
HNF4A	ENSG00000101076	ENSE00001723444	-1.30
HOXB-AS3	ENSG00000233101	ENSE00001838904	1.48
IGF2BP3	ENSG00000136231	ENSE00001923655	1.02
IGFBP3	ENSG00000146674	ENSE00001487458	1.10
IGSF9B	ENSG00000080854	ENSE00002198977	1.17
IKBKB	ENSG00000104365	ENSE00002127600	1.02
INPP4B	ENSG00000109452	ENSE00002023798	-1.29
IQCH	ENSG00000103599	ENSE00002541845	-1.65
IQGAP2	ENSG00000145703	ENSE00000971762	1.13
KCNAB2	ENSG00000069424	ENSE00001476230	-1.22
KCNAB2	ENSG00000069424	ENSE00001476313	1.04
KCNAB2	ENSG00000069424	ENSE00001738518	1.10
KCNAB2	ENSG00000069424	ENSE00001843042	-1.15
KCNIP3	ENSG00000115041	ENSE00002497162	1.28
KCNN4	ENSG00000104783	ENSE00003017464	-1.02
KHDRBS3	ENSG00000131773	ENSE00002096283	1.05
KHDRBS3	ENSG00000131773	ENSE00002114635	-1.33
KIAA0895L	ENSG00000196123	ENSE00002610565	-1.06
KIAA1217	ENSG00000120549	ENSE00001904791	1.19
KIF21A	ENSG00000139116	ENSE00000936431	-1.11
KLF7	ENSG00000118263	ENSE00001635125	1.29
KLK7	ENSG00000169035	ENSE00001509794	-1.24
LAMB2	ENSG00000172037	ENSE00001906755	-1.03
LCK	ENSG00000182866	ENSE00001292351	-1.04
LGR5	ENSG00000139292	ENSE00000937407	-1.23
LIMCH1	ENSG00000064042	ENSE00002081817	-1.12
LMNA	ENSG00000160789	ENSE00003744694	1.17
LOXL2	ENSG00000134013	ENSE00002106382	1.12
LPCAT2	ENSG00000087253	ENSE00003750395	1.24
LRRTM1	ENSG00000162951	ENSE00001589302	-1.22

Gene	Ensembl gene ID	Ensembl exon ID	mean FIRMA score log ₂ FC (SW480-SW620)
MAD1L1	ENSG00000002822	ENSE00001603752	1.12
MAPKAPK5	ENSG00000089022	ENSE00002340933	-1.03
METTL9	ENSG00000197006	ENSE00003716685	1.07
MFNG	ENSG00000100060	ENSE00001712318	-1.24
MFNG	ENSG00000100060	ENSE00001776020	1.05
MGC32805	ENSG00000250328	ENSE00002060565	1.01
MIB2	ENSG00000197530	ENSE00002081549	-1.05
MKRN1	ENSG00000133606	ENSE00003488289	1.06
MLPH	ENSG00000115648	ENSE00001652650	1.16
MLPH	ENSG00000115648	ENSE00001809540	1.16
MME	ENSG00000196549	ENSE00001019489	-1.13
MT1F	ENSG00000198417	ENSE00002268058	1.18
MT1F	ENSG00000198417	ENSE00002598869	1.01
MTUS1	ENSG00000129422	ENSE00002113270	-1.07
MYB	ENSG00000118513	ENSE00003462536	-1.30
NAE1	ENSG00000159593	ENSE00002602570	-1.14
NCAM1	ENSG00000149294	ENSE00002187646	1.37
NDRG1	ENSG00000104419	ENSE00002097689	1.34
NELL2	ENSG00000184613	ENSE00001357932	-1.24
NOS2	ENSG00000007171	ENSE00003726980	-1.38
NR2F1-AS1	ENSG00000237187	ENSE00003702267	-1.40
NR4A2	ENSG00000153234	ENSE00001782418	-1.06
NRG2	ENSG00000158458	ENSE00002136061	1.13
NRP2	ENSG00000118257	ENSE00001709065	-1.32
NSMF	ENSG00000165802	ENSE00001455299	-1.34
NT5DC4	ENSG00000144130	ENSE00001922295	-1.04
NTM	ENSG00000182667	ENSE00001660288	1.09
OSBP2	ENSG00000184792	ENSE00001671961	-1.09
P2RX4	ENSG00000135124	ENSE00002217721	-1.13
PAX6	ENSG00000007372	ENSE00002193571	-1.07
PCBP1-AS1	ENSG00000179818	ENSE00001741475	-1.02
PCDH19	ENSG00000165194	ENSE00001901871	1.38
PCMT1	ENSG00000120265	ENSE00001909320	-1.19
PDLIM3	ENSG00000154553	ENSE00002067040	1.02
PDS5A	ENSG00000121892	ENSE00002039370	-1.10
PEG10	ENSG00000242265	ENSE00003713547	-1.09
PEG10	ENSG00000242265	ENSE00003729218	1.24
PEG10	ENSG00000242265	ENSE00001882092	1.19
PEG10	ENSG00000242265	ENSE00001904124	-1.20
PFAS	ENSG00000178921	ENSE00001233786	-1.01
PHLDB2	ENSG00000144824	ENSE00003613895	1.10
PIR	ENSG00000087842	ENSE00001833467	1.17
PLAUR	ENSG00000011422	ENSE00003197652	1.10

Gene	Ensembl gene ID	Ensembl exon ID	mean FIRMA score log ₂ FC (SW480-SW620)
PLOD2	ENSG00000152952	ENSE00001008040	-1.27
PMP22	ENSG00000109099	ENSE00002264261	1.22
PPP1R2	ENSG00000184203	ENSE00001662037	1.11
PROM1	ENSG00000007062	ENSE00002075061	1.41
PROM1	ENSG00000007062	ENSE00002077095	-1.04
PSMB9	ENSG00000240065	ENSE00001727868	1.13
PSMD2	ENSG00000145191	ENSE00001902565	1.32
PTCH1	ENSG00000185920	ENSE00002339685	1.07
PTGR1	ENSG00000106853	ENSE00003678205	1.37
PTK2B	ENSG00000120899	ENSE00000818668	1.02
PVT1	ENSG00000249859	ENSE00001565023	1.18
PYCR2	ENSG00000143811	ENSE00003733914	1.17
RAB43	ENSG00000172780	ENSE00001514780	1.12
RAB6B	ENSG00000154917	ENSE00001833196	-1.14
RAB6B	ENSG00000154917	ENSE00001909702	1.05
RAC2	ENSG00000128340	ENSE00001674760	1.08
RAPGEF3	ENSG00000079337	ENSE00002213124	-1.01
RBP1	ENSG00000114115	ENSE00003723860	-1.21
RCC2	ENSG00000179051	ENSE00001881805	1.02
RHCE	ENSG00000188672	ENSE00003720314	-1.14
RHOH	ENSG00000168421	ENSE00002023424	1.03
RIN3	ENSG00000100599	ENSE00003729398	1.11
RNF167	ENSG00000108523	ENSE00002649106	-1.10
RNF8	ENSG00000112130	ENSE00001932563	-1.32
ROBO1	ENSG00000169855	ENSE00001829480	-1.15
S100A6	ENSG00000197956	ENSE00003451566	-1.06
SATB2	ENSG00000119042	ENSE00001831321	1.14
SDCBP2	ENSG00000125775	ENSE00003748885	-1.39
SDK1	ENSG00000146555	ENSE00001594392	1.03
SELPLG	ENSG00000110876	ENSE00001504513	1.55
SEPT4	ENSG00000108387	ENSE00003480007	1.05
SGK2	ENSG00000101049	ENSE00001848773	1.01
SGK2	ENSG00000101049	ENSE00002216908	1.19
SLC12A9	ENSG00000146828	ENSE00001889752	1.07
SLC43A3	ENSG00000134802	ENSE00002156079	-1.10
SLC4A8	ENSG00000050438	ENSE00002359374	-1.47
SLC7A8	ENSG00000092068	ENSE00003745230	-1.03
SLIT1	ENSG00000187122	ENSE00003677425	1.06
SNHG1	ENSG00000255717	ENSE00001438744	1.09
SORBS1	ENSG00000095637	ENSE00001516977	-1.15
SOX5	ENSG00000134532	ENSE00002304701	1.09
SPARC	ENSG00000113140	ENSE00002104832	1.01
SPATS2L	ENSG00000196141	ENSE00001782729	-1.11

Gene	Ensembl gene ID	Ensembl exon ID	mean FIRMA score log ₂ FC (SW480-SW620)
SPON2	ENSG00000159674	ENSE00002048781	1.15
SPON2	ENSG00000159674	ENSE00002062829	1.17
SSBP2	ENSG00000145687	ENSE00002048000	1.01
ST3GAL1	ENSG00000008513	ENSE00002127000	1.12
ST6GALNAC2	ENSG00000070731	ENSE00002763003	1.38
STON2	ENSG00000140022	ENSE00002307945	-1.00
STXBP1	ENSG00000136854	ENSE00001941811	3.49
SWI5	ENSG00000175854	ENSE00001221456	1.03
TAF9B	ENSG00000187325	ENSE00001627287	-1.12
TAF9B	ENSG00000187325	ENSE00001650044	-1.04
TBC1D4	ENSG00000136111	ENSE00001863280	-1.06
TCF7	ENSG00000081059	ENSE00002118128	1.13
TCF7	ENSG00000081059	ENSE00002129213	1.04
TEX261	ENSG00000144043	ENSE00001696853	1.16
THNSL2	ENSG00000144115	ENSE00001915486	-1.17
TIMP2	ENSG00000035862	ENSE00002898727	-1.25
TJP1	ENSG00000104067	ENSE00001493913	1.06
TLE4	ENSG00000106829	ENSE00001912243	1.03
TMC4	ENSG00000167608	ENSE00003735786	1.62
TNFRSF9	ENSG00000049249	ENSE00001831101	-1.13
TP53I11	ENSG00000175274	ENSE00002141764	1.23
TP53I11	ENSG00000175274	ENSE00002186776	-1.38
TRIP6	ENSG00000087077	ENSE00001593746	-1.13
TTC7A	ENSG00000068724	ENSE00001839658	-1.12
TWF2	ENSG00000173366	ENSE00001926724	1.11
UBE2V2	ENSG00000169139	ENSE00001539502	-1.09
UGT1A6	ENSG00000167165	ENSE00003703903	-1.12
USH1C	ENSG00000006611	ENSE00002182900	-1.22
VIPR1	ENSG00000114812	ENSE00003654465	-1.33
VSIG1	ENSG00000101842	ENSE00001339621	-1.06
WDR59	ENSG00000103091	ENSE00002506691	-1.27
WDR59	ENSG00000103091	ENSE00002595213	1.32
WFDC3	ENSG00000124116	ENSE00001911737	1.05
WIPF1	ENSG00000115935	ENSE00001730565	1.19
ZAP70	ENSG00000115085	ENSE00001892068	1.24
ZAP70	ENSG00000115085	ENSE00003503894	1.46
ZBED3	ENSG00000132846	ENSE00002051415	1.10
ZDHHC1	ENSG00000159714	ENSE00002579019	-1.21
ZDHHC20	ENSG00000180776	ENSE00001773330	1.06
ZMYND11	ENSG00000015171	ENSE00003552044	1.18
ZNF385A	ENSG00000161642	ENSE00002344606	1.03
ZNF462	ENSG00000148143	ENSE00001023580	1.41
ZNF706	ENSG00000120963	ENSE00002109851	1.17

Table S3: GO terms (biological processes) enriched for candidate genes with differential AS events between SW480 and SW620 cells (microarray data).

GO term ID	Description	p-value
GO:0007411	axon guidance	5.946E-05
GO:0042127	regulation of cell proliferation	0.0002085
GO:0001666	response to hypoxia	0.000548
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	0.0018831
GO:0042493	response to drug	0.0028185
GO:0007155	cell adhesion	0.0038242
GO:0007165	signal transduction	0.0038418
GO:0014066	regulation of phosphatidylinositol 3-kinase signaling	0.004021
GO:0050900	leukocyte migration	0.006161
GO:0008285	negative regulation of cell proliferation	0.0075429
GO:0051592	response to calcium ion	0.0076838
GO:0042981	regulation of apoptotic process	0.0083246
GO:0048661	positive regulation of smooth muscle cell proliferation	0.0086498
GO:0043066	negative regulation of apoptotic process	0.0088472
GO:0000165	MAPK cascade	0.0092649
GO:0060395	SMAD protein signal transduction	0.009692
GO:0018108	peptidyl-tyrosine phosphorylation	0.017505
GO:0030324	lung development	0.0192872
GO:0006816	calcium ion transport	0.0192872
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	0.0201788
GO:0006805	xenobiotic metabolic process	0.0210066
GO:0042060	wound healing	0.0228173
GO:0034220	ion transmembrane transport	0.023802
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	0.0240901
GO:0008152	metabolic process	0.026306
GO:0007626	locomotory behavior	0.0267163
GO:0046777	protein autophosphorylation	0.0293044
GO:0070374	positive regulation of ERK1 and ERK2 cascade	0.0315391
GO:0001525	angiogenesis	0.0315918
GO:0043524	negative regulation of neuron apoptotic process	0.0330745
GO:0007507	heart development	0.0376352
GO:0046854	phosphatidylinositol phosphorylation	0.0381169
GO:0007420	brain development	0.0438346
GO:0050852	T cell receptor signaling pathway	0.0499184

Table S 4: GO terms (biological processes) enriched for genes with 24-h rhythmic phase-shifted SVPs in SW480 and SW620 cells (RNA-seq data).

Cell line	GO term ID	Description	p-value	p.adjust
SW480	GO:0032799	low-density lipoprotein receptor particle metabolic process	8.14E-04	4.40E-01
SW480	GO:0002695	negative regulation of leukocyte activation	1.24E-03	4.40E-01
SW480	GO:0045727	positive regulation of translation	1.24E-03	4.40E-01
SW480	GO:0010870	positive regulation of receptor biosynthetic process	1.28E-03	4.40E-01
SW480	GO:0010988	regulation of low-density lipoprotein particle clearance	1.28E-03	4.40E-01
SW480	GO:0060765	regulation of androgen receptor signaling pathway	1.38E-03	4.40E-01
SW480	GO:0008380	RNA splicing	1.47E-03	4.40E-01
SW480	GO:0042130	negative regulation of T cell proliferation	1.67E-03	4.40E-01
SW480	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	1.91E-03	4.40E-01
SW480	GO:0000398	mRNA splicing, via spliceosome	1.91E-03	4.40E-01
SW480	GO:0000375	RNA splicing, via transesterification reactions	2.10E-03	4.40E-01
SW480	GO:0006491	N-glycan processing	2.15E-03	4.40E-01
SW480	GO:0033143	regulation of intracellular steroid hormone receptor signaling pathway	2.39E-03	4.40E-01
SW480	GO:2001242	regulation of intrinsic apoptotic signaling pathway	2.41E-03	4.40E-01
SW480	GO:0048012	hepatocyte growth factor receptor signaling pathway	2.70E-03	4.40E-01
SW480	GO:2001224	positive regulation of neuron migration	2.70E-03	4.40E-01
SW480	GO:0051250	negative regulation of lymphocyte activation	2.78E-03	4.40E-01
SW480	GO:0006397	mRNA processing	2.81E-03	4.40E-01
SW480	GO:0043484	regulation of RNA splicing	2.81E-03	4.40E-01
SW480	GO:0050866	negative regulation of cell activation	2.81E-03	4.40E-01
SW480	GO:0032945	negative regulation of mononuclear cell proliferation	3.03E-03	4.40E-01
SW480	GO:0050672	negative regulation of lymphocyte proliferation	3.03E-03	4.40E-01
SW480	GO:0034383	low-density lipoprotein particle clearance	3.22E-03	4.47E-01
SW480	GO:0070664	negative regulation of leukocyte proliferation	3.63E-03	4.55E-01
SW480	GO:0034250	positive regulation of cellular amide metabolic process	3.65E-03	4.55E-01
SW480	GO:0048013	ephrin receptor signaling pathway	3.71E-03	4.55E-01
SW480	GO:0006744	ubiquinone biosynthetic process	4.03E-03	4.57E-01
SW480	GO:1901663	quinone biosynthetic process	4.03E-03	4.57E-01
SW480	GO:0050868	negative regulation of T cell activation	4.25E-03	4.57E-01
SW480	GO:0030522	intracellular receptor signaling pathway	4.30E-03	4.57E-01
SW480	GO:0006743	ubiquinone metabolic process	4.81E-03	4.80E-01
SW480	GO:0010984	regulation of lipoprotein particle clearance	4.81E-03	4.80E-01
SW620	GO:0008380	RNA splicing	2.24E-06	3.75E-03
SW620	GO:0010608	posttranscriptional regulation of gene expression	1.29E-05	7.30E-03
SW620	GO:0034248	regulation of cellular amide metabolic process	1.49E-05	7.30E-03
SW620	GO:0006417	regulation of translation	1.99E-05	7.30E-03
SW620	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	2.79E-05	7.30E-03
SW620	GO:0000398	mRNA splicing, via spliceosome	2.79E-05	7.30E-03
SW620	GO:0000375	RNA splicing, via transesterification reactions	3.05E-05	7.30E-03

Cell line	GO term ID	Description	<i>p</i> -value	<i>p</i> -adjust
SW620	GO:0006397	mRNA processing	4.99E-05	1.04E-02
SW620	GO:0031047	gene silencing by RNA	6.68E-05	1.24E-02
SW620	GO:0006337	nucleosome disassembly	2.93E-04	4.10E-02
SW620	GO:0062033	positive regulation of mitotic sister chromatid segregation	2.93E-04	4.10E-02
SW620	GO:0043484	regulation of RNA splicing	2.94E-04	4.10E-02
SW620	GO:0034250	positive regulation of cellular amide metabolic process	3.69E-04	4.61E-02
SW620	GO:0031498	chromatin disassembly	4.13E-04	4.61E-02
SW620	GO:0032986	protein-DNA complex disassembly	4.13E-04	4.61E-02
SW620	GO:0048024	regulation of mRNA splicing, via spliceosome	4.51E-04	4.71E-02
SW620	GO:0035308	negative regulation of protein dephosphorylation	4.88E-04	4.71E-02
SW620	GO:0035305	negative regulation of dephosphorylation	5.32E-04	4.71E-02
SW620	GO:1903311	regulation of mRNA metabolic process	5.35E-04	4.71E-02

Table S 5: Tissues in the murine and baboon multi-organ circadian transcriptome datasets.

Abbreviation	Tissue	Tissue type	Species	Dataset
ADG	Adrenal gland	Secretory	Mouse	GSE54650
AOR	Aorta	Muscle	Mouse	GSE54650
BAT	Brown fat	Adipose	Mouse	GSE54650
BST	Brainstem	Brain	Mouse	GSE54650
CER	Cerebellum	Brain	Mouse	GSE54650
HEA	Heart	Muscle	Mouse	GSE54650
HYP	Hypothalamus	Brain	Mouse	GSE54650
KID	Kidney	Excretory	Mouse	GSE54650
LIV	Liver	Liver	Mouse	GSE54650
LUN	Lung	Lung	Mouse	GSE54650
MUS	Skeletal muscle	Muscle	Mouse	GSE54650
WAT	White fat	Adipose	Mouse	GSE54650
ADC	Adrenal cortex	Secretory	Olive baboon	GSE98965
ADM	Adrenal medulla	Secretory	Olive baboon	GSE98965
AMY	Amygdala	Brain	Olive baboon	GSE98965
ANT	Antrum	Digestive	Olive baboon	GSE98965
AOR	Aorta	Muscle	Olive baboon	GSE98965
ARC	Arcuate nucleus	Brain	Olive baboon	GSE98965
ASC	Ascending colon	Digestive	Olive baboon	GSE98965
AXL	Axillary lymphonodes	Immune	Olive baboon	GSE98965
BLA	Bladder	Excretory	Olive baboon	GSE98965
BOM	Bone marrow	Immune	Olive baboon	GSE98965
CEC	Cecum	Digestive	Olive baboon	GSE98965
CER	Cerebellum	Brain	Olive baboon	GSE98965
COR	Cornea	Eye	Olive baboon	GSE98965
DEC	Descending colon	Digestive	Olive baboon	GSE98965
DMH	Dorsomedial hypothalamus	Brain	Olive baboon	GSE98965
DUO	Duodenum	Digestive	Olive baboon	GSE98965
HAB	Habenula	Brain	Olive baboon	GSE98965
HEA	Heart	Muscle	Olive baboon	GSE98965
HIP	Hippocampus	Brain	Olive baboon	GSE98965
ILE	Ileum	Digestive	Olive baboon	GSE98965
IRI	Iris	Eye	Olive baboon	GSE98965
KIC	Kidney cortex	Excretory	Olive baboon	GSE98965
KIM	Kidney medulla	Excretory	Olive baboon	GSE98965
LGP	Lateral globus pallidus	Brain	Olive baboon	GSE98965
LH	Lateral hypothalamus	Brain	Olive baboon	GSE98965
LIV	Liver	Liver	Olive baboon	GSE98965
LUN	Lungs	Lung	Olive baboon	GSE98965
MEL	Mesenteric lymphonodes	Immune	Olive baboon	GSE98965
MGP	Medial globus pallidus	Brain	Olive baboon	GSE98965
MMB	Mammillary body	Brain	Olive baboon	GSE98965

Abbreviation	Tissue	Tissue type	Species	Dataset
MUA	Muscle abdominal	Muscle	Olive baboon	GSE98965
MUG	Muscle gastrocnemian	Muscle	Olive baboon	GSE98965
OES	Oesophagus	Digestive	Olive baboon	GSE98965
OLB	Olfactory bulb	Brain	Olive baboon	GSE98965
OMF	Omental fat	Adipose	Olive baboon	GSE98965
ONH	Optic nerve head	Eye	Olive baboon	GSE98965
PAN	Pancreas	Pancreas	Olive baboon	GSE98965
PIN	Pineal gland	Secretory	Olive baboon	GSE98965
PIT	Pituitary	Secretory	Olive baboon	GSE98965
PON	Pons	Brain	Olive baboon	GSE98965
PRA	Preoptic area	Brain	Olive baboon	GSE98965
PRC	Prefrontal cortex	Brain	Olive baboon	GSE98965
PRO	Prostate	Genital	Olive baboon	GSE98965
PUT	Putamen	Brain	Olive baboon	GSE98965
PVN	Paraventricular nuclei	Brain	Olive baboon	GSE98965
RET	Retina	Eye	Olive baboon	GSE98965
RPE	Retinal pigment epithelium	Eye	Olive baboon	GSE98965
SCN	Suprachiasmatic nuclei	Brain	Olive baboon	GSE98965
SKI	Skin	Skin	Olive baboon	GSE98965
SMM	Smooth muscle	Digestive	Olive baboon	GSE98965
SON	Supraoptic nucleus	Brain	Olive baboon	GSE98965
SPL	Spleen	Immune	Olive baboon	GSE98965
STF	Stomach fundus	Digestive	Olive baboon	GSE98965
SUN	Substantia nigra	Brain	Olive baboon	GSE98965
TES	Testicles	Genital	Olive baboon	GSE98965
THA	Thalamus	Brain	Olive baboon	GSE98965
THR	Thyroid	Secretory	Olive baboon	GSE98965
VIC	Visual cortex	Brain	Olive baboon	GSE98965
VMH	Ventromedial hypothalamus	Brain	Olive baboon	GSE98965
WAM	White adipose mesenteric	Adipose	Olive baboon	GSE98965
WAP	White adipose pericardial	Adipose	Olive baboon	GSE98965
WAR	White adipose perirenal	Adipose	Olive baboon	GSE98965
WAS	White adipose subcutaneous	Adipose	Olive baboon	GSE98965
WAT	White adipose retroperitoneal	Adipose	Olive baboon	GSE98965

Table S 6: GO terms (biological processes) enriched for genes with rhythmic FIRMAGene scores and more than one expressed splice variant in murine tissues.

Period	ID	Description	pvalue	p.adjust
24 h	GO:0048588	developmental cell growth	1.01E-08	5.06E-05
24 h	GO:0070997	neuron death	5.11E-07	1.08E-03
24 h	GO:1901214	regulation of neuron death	6.49E-07	1.08E-03
24 h	GO:0007623	circadian rhythm	1.37E-06	1.38E-03
24 h	GO:0022613	ribonucleoprotein complex biogenesis	1.37E-06	1.38E-03
24 h	GO:0007409	Axonogenesis	3.20E-06	2.41E-03
24 h	GO:0042254	ribosome biogenesis	3.57E-06	2.41E-03
24 h	GO:1990138	neuron projection extension	3.85E-06	2.41E-03
24 h	GO:0060560	developmental growth involved in morphogenesis	7.02E-06	3.62E-03
24 h	GO:0008630	intrinsic apoptotic signaling pathway in response to DNA damage	7.23E-06	3.62E-03
24 h	GO:0010498	proteasomal protein catabolic process	8.70E-06	3.83E-03
24 h	GO:1903828	negative regulation of cellular protein localization	9.17E-06	3.83E-03
24 h	GO:0042771	intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator	1.06E-05	4.09E-03
24 h	GO:1902165	regulation of intrinsic apoptotic signaling pathway in response to DNA damage by p53 class mediator	2.20E-05	7.88E-03
24 h	GO:0031647	regulation of protein stability	2.36E-05	7.88E-03
24 h	GO:0048675	axon extension	2.59E-05	8.13E-03
24 h	GO:0032922	circadian regulation of gene expression	2.87E-05	8.14E-03
24 h	GO:0010721	negative regulation of cell development	3.04E-05	8.14E-03
24 h	GO:1902229	regulation of intrinsic apoptotic signaling pathway in response to DNA damage	3.14E-05	8.14E-03
24 h	GO:0048511	rhythmic process	3.25E-05	8.14E-03
24 h	GO:0051402	neuron apoptotic process	3.68E-05	8.78E-03
24 h	GO:0010769	regulation of cell morphogenesis involved in differentiation	4.06E-05	9.26E-03
24 h	GO:0051146	striated muscle cell differentiation	4.58E-05	9.99E-03
12 h	GO:0010498	proteasomal protein catabolic process	2.21E-10	1.13E-06
12 h	GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	5.34E-09	1.36E-05
12 h	GO:0061136	regulation of proteasomal protein catabolic process	9.04E-07	1.40E-03
12 h	GO:1903364	positive regulation of cellular protein catabolic process	1.10E-06	1.40E-03
12 h	GO:1903362	regulation of cellular protein catabolic process	2.60E-06	2.48E-03
12 h	GO:0031647	regulation of protein stability	3.93E-06	2.48E-03
12 h	GO:0032436	positive regulation of proteasomal ubiquitin-dependent protein catabolic process	4.74E-06	2.48E-03
12 h	GO:0045732	positive regulation of protein catabolic process	4.77E-06	2.48E-03
12 h	GO:2000060	positive regulation of ubiquitin-dependent protein catabolic process	5.19E-06	2.48E-03
12 h	GO:1903050	regulation of proteolysis involved in cellular protein catabolic process	5.20E-06	2.48E-03
12 h	GO:1901800	positive regulation of proteasomal protein catabolic process	5.37E-06	2.48E-03

Period	ID	Description	pvalue	p.adjust
12 h	GO:1903052	positive regulation of proteolysis involved in cellular protein catabolic process	6.81E-06	2.86E-03
12 h	GO:0007160	cell-matrix adhesion	7.31E-06	2.86E-03
12 h	GO:0006644	phospholipid metabolic process	1.47E-05	4.54E-03
12 h	GO:0006913	nucleocytoplasmic transport	1.52E-05	4.54E-03
12 h	GO:0051169	nuclear transport	1.52E-05	4.54E-03
12 h	GO:0022613	ribonucleoprotein complex biogenesis	1.52E-05	4.54E-03
12 h	GO:0042176	regulation of protein catabolic process	1.61E-05	4.54E-03
12 h	GO:1903311	regulation of mRNA metabolic process	2.15E-05	5.64E-03
12 h	GO:0002181	cytoplasmic translation	2.22E-05	5.64E-03
12 h	GO:0006397	mRNA processing	2.68E-05	6.49E-03
12 h	GO:0060999	positive regulation of dendritic spine development	2.85E-05	6.59E-03
12 h	GO:0008380	RNA splicing	3.34E-05	6.88E-03
12 h	GO:0035304	regulation of protein dephosphorylation	3.35E-05	6.88E-03
12 h	GO:0033119	negative regulation of RNA splicing	3.38E-05	6.88E-03
12 h	GO:0009896	positive regulation of catabolic process	3.60E-05	7.05E-03
12 h	GO:0031589	cell-substrate adhesion	4.65E-05	8.22E-03
12 h	GO:0034660	ncRNA metabolic process	4.86E-05	8.22E-03
12 h	GO:0006405	RNA export from nucleus	4.92E-05	8.22E-03
12 h	GO:0031570	DNA integrity checkpoint	4.98E-05	8.22E-03
12 h	GO:0032434	regulation of proteasomal ubiquitin-dependent protein catabolic process	5.01E-05	8.22E-03
12 h	GO:0006650	glycerophospholipid metabolic process	5.29E-05	8.31E-03
12 h	GO:0046486	glycerolipid metabolic process	5.47E-05	8.31E-03
12 h	GO:1903008	organelle disassembly	5.63E-05	8.31E-03
12 h	GO:0031331	positive regulation of cellular catabolic process	5.83E-05	8.31E-03
12 h	GO:0060560	developmental growth involved in morphogenesis	5.94E-05	8.31E-03
12 h	GO:0000077	DNA damage checkpoint	6.04E-05	8.31E-03
12 h	GO:0051056	regulation of small GTPase mediated signal transduction	6.47E-05	8.66E-03
12 h	GO:0006914	autophagy	7.07E-05	8.87E-03
12 h	GO:0061919	process utilizing autophagic mechanism	7.07E-05	8.87E-03
12 h	GO:0048025	negative regulation of mRNA splicing, via spliceosome	7.14E-05	8.87E-03

Table S 7: GO terms (biological processes) enriched for genes with rhythmic phase-shifted SVPs in baboon tissues.

Period	ID	Description	p-value	p.adjust
24 h	GO:0008380	RNA splicing	7.81E-14	3.97E-10
24 h	GO:0000375	RNA splicing, via transesterification reactions	5.54E-13	1.18E-09
24 h	GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	1.14E-12	1.18E-09
24 h	GO:0000398	mRNA splicing, via spliceosome	1.14E-12	1.18E-09
24 h	GO:0006397	mRNA processing	1.16E-12	1.18E-09
24 h	GO:1903311	regulation of mRNA metabolic process	3.86E-12	3.27E-09
24 h	GO:0050684	regulation of mRNA processing	3.71E-11	2.70E-08
24 h	GO:0048024	regulation of mRNA splicing, via spliceosome	1.15E-09	7.29E-07
24 h	GO:0043484	regulation of RNA splicing	1.70E-09	9.60E-07
24 h	GO:0048193	Golgi vesicle transport	2.63E-08	1.34E-05
24 h	GO:0016482	cytosolic transport	4.30E-08	1.99E-05
24 h	GO:0000381	regulation of alternative mRNA splicing, via spliceosome	1.07E-06	4.52E-04
24 h	GO:0000380	alternative mRNA splicing, via spliceosome	2.55E-06	9.98E-04
24 h	GO:0010608	posttranscriptional regulation of gene expression	5.26E-06	1.91E-03
24 h	GO:0032469	endoplasmic reticulum calcium ion homeostasis	6.32E-06	2.14E-03
24 h	GO:0006403	RNA localization	9.28E-06	2.95E-03
24 h	GO:0016569	covalent chromatin modification	1.33E-05	3.99E-03
24 h	GO:0016570	histone modification	1.52E-05	4.29E-03
24 h	GO:0034248	regulation of cellular amide metabolic process	1.79E-05	4.79E-03
24 h	GO:0006890	retrograde vesicle-mediated transport, Golgi to ER	2.37E-05	5.75E-03
24 h	GO:0009896	positive regulation of catabolic process	2.38E-05	5.75E-03
24 h	GO:0006417	regulation of translation	3.47E-05	8.01E-03
24 h	GO:2001233	regulation of apoptotic signaling pathway	4.33E-05	9.57E-03
24 h	GO:0006997	nucleus organization	5.31E-05	9.98E-03
24 h	GO:1903362	regulation of cellular protein catabolic process	5.38E-05	9.98E-03
24 h	GO:0050657	nucleic acid transport	5.39E-05	9.98E-03
24 h	GO:0050658	RNA transport	5.39E-05	9.98E-03
24 h	GO:0022029	telencephalon cell migration	5.50E-05	9.98E-03
24 h	GO:0021795	cerebral cortex cell migration	5.81E-05	1.02E-02
24 h	GO:0015931	nucleobase-containing compound transport	7.19E-05	1.20E-02
24 h	GO:0051236	establishment of RNA localization	7.32E-05	1.20E-02
24 h	GO:0042176	regulation of protein catabolic process	7.97E-05	1.24E-02
24 h	GO:0033120	positive regulation of RNA splicing	8.39E-05	1.24E-02
24 h	GO:0006986	response to unfolded protein	8.52E-05	1.24E-02
24 h	GO:0007005	mitochondrion organization	8.53E-05	1.24E-02
24 h	GO:0031331	positive regulation of cellular catabolic process	9.50E-05	1.31E-02
24 h	GO:0043547	positive regulation of GTPase activity	9.80E-05	1.31E-02
24 h	GO:0034620	cellular response to unfolded protein	9.82E-05	1.31E-02
24 h	GO:0021885	forebrain cell migration	1.02E-04	1.32E-02
24 h	GO:0006405	RNA export from nucleus	1.41E-04	1.70E-02
24 h	GO:0034249	negative regulation of cellular amide metabolic process	1.41E-04	1.70E-02

Period	ID	Description	p-value	p.adjust
24 h	GO:0043488	regulation of mRNA stability	1.41E-04	1.70E-02
24 h	GO:0016049	cell growth	1.54E-04	1.80E-02
24 h	GO:0017148	negative regulation of translation	1.55E-04	1.80E-02
24 h	GO:1903313	positive regulation of mRNA metabolic process	1.68E-04	1.88E-02
24 h	GO:0051656	establishment of organelle localization	1.70E-04	1.88E-02
24 h	GO:0031116	positive regulation of microtubule polymerization	1.88E-04	2.03E-02
24 h	GO:0043087	regulation of GTPase activity	2.02E-04	2.11E-02
24 h	GO:0036109	alpha-linolenic acid metabolic process	2.03E-04	2.11E-02
24 h	GO:0031330	negative regulation of cellular catabolic process	2.11E-04	2.12E-02
24 h	GO:0000086	G2/M transition of mitotic cell cycle	2.13E-04	2.12E-02
24 h	GO:0034976	response to endoplasmic reticulum stress	2.20E-04	2.15E-02
24 h	GO:0035966	response to topologically incorrect protein	2.36E-04	2.24E-02
24 h	GO:0061013	regulation of mRNA catabolic process	2.38E-04	2.24E-02
24 h	GO:0035967	cellular response to topologically incorrect protein	2.56E-04	2.36E-02
24 h	GO:0031334	positive regulation of protein complex assembly	2.81E-04	2.55E-02
24 h	GO:0043254	regulation of protein complex assembly	2.96E-04	2.64E-02
24 h	GO:1902905	positive regulation of supramolecular fiber organization	3.12E-04	2.73E-02
24 h	GO:0045732	positive regulation of protein catabolic process	3.25E-04	2.80E-02
24 h	GO:2001252	positive regulation of chromosome organization	3.51E-04	2.98E-02
24 h	GO:0090169	regulation of spindle assembly	3.61E-04	3.01E-02
24 h	GO:0043487	regulation of RNA stability	3.83E-04	3.14E-02
24 h	GO:0010822	positive regulation of mitochondrion organization	4.05E-04	3.27E-02
24 h	GO:0006616	SRP-dependent cotranslational protein targeting to membrane, translocation	4.56E-04	3.51E-02
24 h	GO:0051973	positive regulation of telomerase activity	4.59E-04	3.51E-02
24 h	GO:1903364	positive regulation of cellular protein catabolic process	4.62E-04	3.51E-02
24 h	GO:0006914	autophagy	4.69E-04	3.51E-02
24 h	GO:0061919	process utilizing autophagic mechanism	4.69E-04	3.51E-02
24 h	GO:0040001	establishment of mitotic spindle localization	4.88E-04	3.53E-02
24 h	GO:1901987	regulation of cell cycle phase transition	4.90E-04	3.53E-02
24 h	GO:0044839	cell cycle G2/M phase transition	4.95E-04	3.53E-02
24 h	GO:0090224	regulation of spindle organization	5.07E-04	3.53E-02
24 h	GO:1903146	regulation of autophagy of mitochondrion	5.07E-04	3.53E-02
24 h	GO:0022411	cellular component disassembly	5.45E-04	3.70E-02
24 h	GO:0071826	ribonucleoprotein complex subunit organization	5.47E-04	3.70E-02
24 h	GO:0097711	ciliary basal body-plasma membrane docking	5.68E-04	3.79E-02
24 h	GO:0010389	regulation of G2/M transition of mitotic cell cycle	5.74E-04	3.79E-02
24 h	GO:0042147	retrograde transport, endosome to Golgi	5.89E-04	3.84E-02
24 h	GO:0016197	endosomal transport	6.28E-04	4.04E-02
24 h	GO:0031112	positive regulation of microtubule polymerization or depolymerization	6.47E-04	4.07E-02
24 h	GO:0036498	IRE1-mediated unfolded protein response	6.49E-04	4.07E-02
24 h	GO:0010948	negative regulation of cell cycle process	7.53E-04	4.67E-02

Period	ID	Description	p-value	p.adjust
24 h	GO:0044089	positive regulation of cellular component biogenesis	7.72E-04	4.73E-02
12 h	GO:0030705	cytoskeleton-dependent intracellular transport	1.97E-05	6.26E-02
12 h	GO:0016482	cytosolic transport	3.35E-05	6.26E-02
12 h	GO:0006892	post-Golgi vesicle-mediated transport	4.74E-05	6.26E-02
12 h	GO:0042118	endothelial cell activation	8.72E-05	8.64E-02
12 h	GO:0099518	vesicle cytoskeletal trafficking	1.82E-04	1.44E-01
12 h	GO:0051656	establishment of organelle localization	2.24E-04	1.48E-01
12 h	GO:0048193	Golgi vesicle transport	2.70E-04	1.53E-01
12 h	GO:0010970	transport along microtubule	3.48E-04	1.53E-01
12 h	GO:0099111	microtubule-based transport	3.48E-04	1.53E-01
12 h	GO:0035735	intraciliary transport involved in cilium assembly	4.98E-04	1.69E-01
12 h	GO:0019080	viral gene expression	5.03E-04	1.69E-01
12 h	GO:0048675	axon extension	5.94E-04	1.69E-01
12 h	GO:0051648	vesicle localization	6.09E-04	1.69E-01
12 h	GO:0072384	organelle transport along microtubule	6.43E-04	1.69E-01
12 h	GO:2000651	positive regulation of sodium ion transmembrane transporter activity	6.67E-04	1.69E-01
12 h	GO:0010611	regulation of cardiac muscle hypertrophy	7.02E-04	1.69E-01
12 h	GO:0031503	protein-containing complex localization	7.25E-04	1.69E-01
12 h	GO:0014743	regulation of muscle hypertrophy	8.66E-04	1.91E-01
12 h	GO:0051650	establishment of vesicle localization	9.27E-04	1.94E-01

Table S 8: Overview of 24-h rhythmic phase-shifted SVPs detected in at least six baboon tissues.

Gene	SVPs	Tissues	Circular mean phase \pm SD (CT)	Rayleigh test p-value	Mean phase difference \pm SD (h)
ADD1	ADD1-201	COR, HEA, ILE, KIM, STF, THR, VIC	7.21 \pm 0.42	0.00048	10.18 \pm 2.65
	ADD1-205		20.08 \pm 0.89	0.03497	
CCNL2	CCNL2-202	ADC, ADM, ASC, HEA, ILE, KIC, OMF, SPL, STF	5.52 \pm 0.27	0	10.51 \pm 1.16
	CCNL2-205		18.51 \pm 0.41	0	
CPNE1	CPNE1-203	BLA, CER, COR, ILE, OMF, SKI, THR	19.43 \pm 0.88	0.03391	10.45 \pm 0.59
	CPNE1-204		8.36 \pm 0.92	0.04269	
CSNK2A1	ENSPANT00000043774	HEA, KIC, LUN, OMF, PRO, THR	6.19 \pm 0.17	0.00012	10.86 \pm 0.78
	ENSPANT00000055451		18.53 \pm 0.44	0.00235	
DRG1	DRG1-201	ADM, AXL, DUO, ILE, LUN, STF, WAT	6.06 \pm 0.37	0.00023	10.63 \pm 0.85
	DRG1-202		17.40 \pm 0.41	0.00040	
EIF3H	EIF3H-201	DUO, HEA, SKI, STF, TES, THA, THR	5.31 \pm 0.44	0.00061	10.75 \pm 1.85
	EIF3H-203		17.45 \pm 0.32	8.61E-05	
EXOC3	EXOC3-201	ADC, ADM, ASC, CEC, MGP, MMB, MUA	6.24 \pm 0.43	0.00055	11.30 \pm 0.67
	EXOC3-204		17.78 \pm 0.29	3.04E-05	
HNRNPAB	HNRNPAB-201	ANT, ASC, COR, DEC, KIM, MGP, MUA, OMF, PIT	6.49 \pm 0.45	2.34E-05	10.34 \pm 1.42
	HNRNPAB-202		19.07 \pm 0.37	0	
HSD17B4	HSD17B4-201	BLA, CEC, DMH, MUG, PRO, SUN	8.61 \pm 0.20	0.00017	9.96 \pm 1.94
	HSD17B4-203		19.17 \pm 0.41	0.00186	
MCFD2	MCFD2-201	ADC, ADM, BLA, COR, KIC, LUN, THR	18.52 \pm 0.39	0.00032	9.72 \pm 1.78
	MCFD2-202		5.07 \pm 0.33	0.00011	
MOB4	MOB4-201	ADC, HEA, LUN, OMF, SKI, STF, WAT	5.16 \pm 0.81	0.01936	10.47 \pm 1.83
	MOB4-202		17.47 \pm 0.46	0.00078	
NELFE	NELFE-201	ADC, ADM, ASC, MUA, OMF, PRC, SPL, WAP	7.00 \pm 0.46	0.00019	10.75 \pm 0.82
	NELFE-202		17.82 \pm 0.47	0.00024	
NRDE-2	ENSPANT00000019601	ADM, KIM, MGP, MMB, RPE, SPL, WAP	18.28 \pm 0.42	0.00049	11.66 \pm 0.22
	ENSPANT00000021732		6.44 \pm 0.36	0.00019	
PAF1	PAF1-201	ADC, ADM, ASC, BLA, KIC, WAP	18.07 \pm 0.51	0.00402	11.27 \pm 0.56
	PAF1-203		6.02 \pm 0.45	0.00249	
PCBP2	PCBP2-201	DEC, DMH, MMB, MUG, PRA, PRC	7.15 \pm 0.52	0.00466	10.90 \pm 0.67
	PCBP2-204		20.06 \pm 0.35	0.00098	
PEX6	PEX6-202	ADC, BLA, CEC, DMH, KIM, PON, SUN	15.91 \pm 0.72	0.01000	10.32 \pm 1.25
	PEX6-203		5.49 \pm 0.49	0.00111	
RPS3	RPS3-201	KIM, MMB, MUG, PRC, SKI, THA	7.03 \pm 0.43	0.00207	10.74 \pm 0.59
	RPS3-203		18.93 \pm 0.34	0.00090	
RPS3	RPS3-202		7.29 \pm 0.41	0.00177	10.74 \pm 0.59

Gene	SVPs	Tissues	Circular mean phase \pm SD (CT)	Rayleigh test p-value	Mean phase difference \pm SD (h)
	RPS3-203	KIM, LUN, MUG, OLB, STF, THA	16.86 \pm 0.48	0.00334	
SEPHS1	ENSPANT00000021592	ADM, LUN, MUG, PRO, SPL, STF, THR	6.90 \pm 0.40	0.00034	11.30 \pm 0.39
	ENSPANT00000040491		18.61 \pm 0.44	0.00059	
SERINC3	SERINC3-201	ADC, ANT, HAB, LIV, LUN, THR	6.31 \pm 0.55	0.00600	11.18 \pm 0.80
	SERINC3-202		18.45 \pm 0.49	0.00360	
TAF15	TAF15-201	ADM, KIC, LIV, LUN, PRO, THR	5.62 \pm 0.43	0.00211	11.16 \pm 0.60
	TAF15-202		17.75 \pm 0.36	0.00117	
ZNF207	ZNF207-205	ADM, COR, KIM, LUN, SKI, THR	18.12 \pm 0.35	0.00100	11.56 \pm 0.51
	ZNF207-208		6.22 \pm 0.31	0.00067	

7.1.3 External Data Files

External data file 1: Phase-shifted 24-h rhythmic SVPs in SW480 cells and SW620 cells.

External data file 2: Genes with candidate 24-h and 12-h rhythmic AS events in murine tissues.

External data file 3: Phase-shifted 24-h and 12-h rhythmic SVPs in olive baboon tissues.

7.2 Zusammenfassung

Im Körper zahlreicher Organismen regelt eine innere Uhr den Ablauf physiologischer Prozesse im Einklang mit dem Tagesrhythmus der Umwelt. Auf zellulärer Ebene entsteht die sogenannte circadiane Rhythmik durch ein Zusammenspiel von Uhrgenen und -proteinen, welche durch negative Rückkopplungsschleifen miteinander interagieren und Oszillationen mit einer 24-Stunden-Periodik in der Expression zahlreicher Zielgene auslösen. Die hieraus resultierenden Rhythmen in der Verfügbarkeit von Proteinen und anderen biologischen Molekülen sind wiederum für die zeitliche Organisation einer Vielzahl biologischer Prozesse verantwortlich, darunter möglicherweise alternatives Spleißen. Alternatives Spleißen beschreibt einen Mechanismus der Genregulation, bei dem aus einem einzigen Gen durch variable Kombination von RNA-Teilabschnitten mehrere verschiedene Proteinvarianten mit zum Teil unterschiedlichen Eigenschaften erstellt werden können. Sowohl Störungen der circadianen Uhr als auch anomales Spleißen werden mit der Entstehung und Weiterentwicklung von Krebserkrankungen in Verbindung gebracht.

Die vorliegende Dissertation geht der Frage nach, inwieweit eine circadiane Regulation von alternativem Spleißen in Säugerzellen vorliegt und ob diese sich für Krebszellen unterschiedlicher Tumorstadien unterscheiden. Im Besonderen wird die Hypothese untersucht, ob Veränderungen von circadian regulierten Spleiß-Ereignissen zur Produktion von Proteinvarianten führen könnten, die zur bösartigen Entwicklung von Krebszellen beitragen. Die der Arbeit zugrundeliegenden Daten stammen von zwei menschlichen Darmkrebszelllinien, SW480 und SW620, welche ursprünglich von einem Primärtumor und einer Metastase desselben Patienten etabliert wurden und als *in vitro*-Modell der kolorektalen Tumorprogression dienen. Basierend auf Zeitreihendaten der Genexpression in den beiden Zelllinien wurde eine computergestützte Analyse durchgeführt, bei der 24-Stunden-rhythmische Gene und alternative Spleiß-Ereignisse auf Transkriptomebene identifiziert wurden. Als Vergleichsbasis wurden bereits publizierte Zeitreihendaten zahlreicher gesunder Gewebesorten analysiert, die von Maus- und Pavianorganen stammen.

Die Analyse offenbarte Unterschiede im circadianen Phänotyp zwischen den beiden Zelllinien, und eine stärker deregulierte circadiane Rhythmik der Metastasen-Zelllinie. Es konnte zudem gezeigt werden, dass am Spleißvorgang beteiligte Gene sowie mutmaßliche Spleißereignisse 24-Stunden-Rhythmen aufweisen, die sich ebenfalls zwischen den Krebsstadien unterscheiden. Sowohl in gesunden als auch in Krebszellen waren viele jener Gene rhythmisch gespleißt, die selbst am Spleißvorgang beteiligt sind, was auf eine Autoregulierung des Prozesses schließen lässt. Mehrere der gespleißten Kandidatengene codieren zudem für Proteinvarianten, die an Prozessen beteiligt sind, welche eine Weiterentwicklung des Tumors begünstigen, darunter Zellmigration und Angiogenese. Insgesamt legen die Beobachtungen eine circadiane Regulation von alternativem Spleißen in Säugerzellen nahe, welche einen Einfluss auf die Krebsentwicklung hat.

7.3 Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe. Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik und Resultaten werden von mir verantwortet.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Berlin, Dezember 2019

Rukeia El-Athman