Exploration of Circadian Dysregulation in Cancer

Laura Usselmann, MBiolSci, MSc

A thesis submitted for the degree of
Doctor of Philosophy in Interdisciplinary Biomedical Research
To be awarded by
University of Warwick, Warwick Medical School

July 2021

Supervisors: Dr Robert Dallmann and Dr Sascha Ott

Table of Contents

Acknowledgments

Firstly, I wish to acknowledge and thank Dr Robert Dallmann and Dr Sascha Ott for their excellent supervision, invaluable support and for sharing their extensive expertise throughout my doctoral studies. Both my research and I benefited from their collective enthusiasm and cross-disciplinary experience. I would also like to express my thanks to all members of the Dallmann lab, past and present, for their friendship, advice and humour, through the ups and downs of laboratory research. This thesis is a cross-disciplinary research effort, and as such I am also indebted to Professor David Rand, Dr Denise Vlachou and Dr Maria Veretennikova for their mathematics, statistics and machine learning expertise in relation to the development of TimeTeller.

My PhD research was funded by the Medical Research Council, and thus my gratitude extends to the MRC for their financial support. In particular I would like to thank the leadership team of the University of Warwick's MRC Interdisciplinary Biomedical Research Doctoral Training Partnership. The DTP has provided me with an inter-disciplinary research network, fantastic training opportunities and importantly friendship, improving the quality of both my research and experience.

Last, but by no means least, I would especially like to thank my family Phil, Mum, Dad, Anna, Emily and Martha, and friends Holly and Steph, for their unwavering love and support throughout my marathon student experience, and of course for listening to me natter about science at every opportunity! Thank you.

Declaration

This thesis is submitted to the University of Warwick in support of an application for the degree of Doctor of Philosophy. It has been composed by the author and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author, except in the cases where publicly available transcriptomics datasets were used. Such datasets are cited by publication and GEO accession numbers. Additionally, the work in chapter 4 follows the work of Dr Denise Vlachou and Prof. David Rand (Department of Mathematics, University of Warwick) and expands upon the application of their algorithm "TimeTeller". Some of the ideas in this thesis are also discussed a review of the literature by Stephenson, *Usselmann* et al. (2021). All credit is acknowledged as appropriate in text.

Abstract

The relationship between the circadian molecular clock and cancer is complex and incompletely understood. This thesis attempts to elucidate the circadian behaviour of *in vitro*, *in ovo*, and *in vivo* model systems using a combination of approaches, in order to inform our understanding of the circadian clock in malignancy.

Bioluminescent reporters, qPCR and next generation sequencing methods were employed to demonstrate significant disruption to circadian clock gene activity in MCF10A cells, a non-malignant cell line frequently described by the literature as having healthy clock function. Meanwhile the malignant MCF7 cell line also demonstrated significant circadian clock gene disruption under a multitude of synchronisation/entrainment conditions. These findings highlight the challenges of using 2D *in vitro* models to elucidate circadian behaviour in cancer.

The chick chorioallantoic membrane (CAM) assay was adapted in order to translate 2D circadian cancer models to a 3D environment. The intrinsic molecular clock of the chicken embryo appeared not to be free-running at EDD12-14. The detection of bioluminescent reporters *in ovo*, a relatively new development to the CAM assay, proved challenging for reasons that remain unclear.

The machine learning biological time prediction algorithm TimeTeller, was adapted for use with low-resolution multi-tissue RNA-seq data, to predict biological time and clock dysfunction from single biological samples. The trained TimeTeller model was validated on several independent publicly available datasets. It is hoped that the application of TimeTeller to single biological samples will prove useful in understanding the degree of clock function in tumours and surrounding tissues.

Overall these findings progress our understanding of the state of the molecular circadian oscillator across a range of model systems. It is hoped that the application of this research will enable novel insights into the circadian clock in malignancy.

Notes to Reader

This thesis is an interdisciplinary undertaking, as per the requirements of Warwick Medical School's Medical Research Councils' Doctoral Training Partnership, which funded and supported this research. Therefore, elements of mathematics and statistics are used in order to understand complex biological datasets. It is assumed that the reader has a good understanding of molecular biological systems and transcriptomic datasets, with a basic understanding of mathematics.

Each chapter stands somewhat independent of each previous chapter; therefore, each chapter contains its own introduction and review of the relevant literature. An overall introduction introduces the reader to the field of circadian rhythms, and its relevance to the field of cancer research.

List of Abbreviations

Abbreviation	Unabbreviated
ABC	ATP-binding cassette
AML	acute myeloid leukaemia
AKT	AKT serine/threonine kinase 1
ARNTL	aryl hydrocarbon receptor nuclear translocator like protein
ATAC-seq	assay for transposase-accessible chromatin using sequencing
ATCC	American type culture collection
ASPA	Animals (Scientific Procedures) Act 1986
BAX	BCL2 associated X protein
BCL2L1	BCL2-like 1
B-H	Bonferroni-Hochberg
BHLHE(40/41)	basic helix-loop-helix family member E(40/41)
BID	BH3 interacting domain death agonist
bp	(DNA) base pairs
bZIP	basic leucine zipper domain
Ca^{2+}	calcium ion
CAM	chorioallantoic membrane
cAMP	cyclic adenosine monophosphate
CASP3	caspase 3
CCD	clock correlation distance
CCNA1	cyclin A1
CCNB1	cyclin B1
CCND1	cyclin D1
CDK2	cyclin dependent kinase 2
CDH2	cadherin 2
CEF	chicken embryonic fibroblast
ChIP-seq	chromatin immunoprecipitation
CLOCK	circadian locomotor output cycles kaput protein
CML	chronic myeloid leukaemia
CRE	cAMP response element

CREB	CRE binding protein
CREBBP	CREB binding protein
CRY	cryptochrome
CSNK1	casein kinase 1
CT	circadian time
CTNNB1	catenin beta 1
CycLuc1	cyclic alkylaminoluciferin
DBP	D-box binding PAR bZIP transcription factor
DLMO	dim light melatonin onset
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethyl sulfoxide
EDD	embryonic development day
EDTA	ethylenediaminetetraacetic acid
EMT	epithelial-mesenchymal transition
ERBB2	erb-b2 receptor tyrosine kinase 2 (aka HER2)
ESR1	oestrogen receptor
FBXL3	F-box and leucine rich repeat protein 3
FDR	false discovery rate
fRMA	frozen robust multiarray analysis
GEO	Gene Expression Omnibus
GR	glucocorticoid receptor
GRC	Genome Reference Consortium
GRE	glucocorticoid receptor element
GTEx	Genotype-Tissue Expression project
HCC	hepatocellular carcinoma
HIF1A	hypoxia inducible factor 1 subunit alpha
HOS	human osteosarcoma
H3P(16/23)	histone 3 pseudogene (16/23)
IFNG	interferon gamma
IL6	interleukin 6
logCPM	log2 counts per million
logTMM	log2 trimmed mean of M-values
MAE	mean absolute error

MAPK1	mitogen-activated protein kinase 1
MDa	megadalton
MDM2	MDM2 proto-oncogene
MFS	metastasis free survival
MT	Molecular Timetable
MYC	MYC proto-oncogene, bHLH transcription factor
NCBI	National Center for Biotechnology Information
NFIL3	nuclear factor interleukin 3 regulated
NPAS2	neuronal PAS domain protein 2
NR1D(1/2)	nuclear receptor subfamily 1 group D member 1/2
OS	overall survival
PBS	phospho-buffered saline
PCA	principal components analysis
PLSR	partial least squares regression
PDA	pancreatic ductal adenocarcinoma
PER	period
p-f	post-fertilisation
PGR	progesterone receptor
PK(A/C)	protein kinase (A/C)
PMD	penalised matrix decomposition
PMT	photomultiplier tube
PSA	pressure-sensitive adhesive
qPCR	quantitative polymerase chain reaction
RAP6	RAB5-activating protein 6
RNA-seq	RNA sequencing
ROI	region of interest
ROR	retinoic acid receptor -related orphan receptor
RRE	ROR-response element
RTK	receptor tyrosine kinase
SCN	suprachiasmatic nucleus
smFISH	single molecule fluorescence *in situ* hybridisation
SNP	single nucleotide polymorphism
SPC	sparse principal component

SRA	Sequence Read Archive
SVD	singular value decomposition
TCGA	The Cancer Genome Atlas
TLK2	tousled like kinase 2
TNF	tumour necrosis factor
TP53	tumour protein p53
TSS	transcriptional start site
TSSE scores	TSS enrichment scores
WPRE	woodchuck hepatitis virus post-transcriptional regulatory element
WT	wild type
VEGFA	vascular endothelial growth factor A
ZT	Zeitgeber time

All gene and protein abbreviations and formatting follow the standards set by the HUGO Gene Nomenclature Committee and Mouse Genome Informatics. E.g. human period 2 gene symbol: *PER2*; human protein symbol: PER2; mouse gene symbol: *Per2*; mouse protein symbol: PER2. To avoid confusion with human and mouse genes, chicken (*Gallus gallus*) genes are termed *gPER2* and chicken proteins gPER2.

Many circadian genes have multiple names which are used interchangeably throughout the literature:

Gene symbol	Alternative symbols
ARNTL	BMAL, MOP3
BHLHE40	DEC1, STRA13, SHARP2 or BHLHB2
BHLHE41	DEC2 or SHARP1
NR1D1 & NR1D2	REV-ERBα & REV-ERBβ
CSNK1D/CSNK1E	CK1δ/CK1ε
NFIL3	E4BP4

List of Figures

List of Tables

1.Introduction

1.1 The circadian timing system

Life on planet Earth displays a staggering diversity, which we know to be the result of accumulated changes to inherited genetic material. The process of evolution ensures that organisms are often remarkably well adapted to life within their respective habitats. Naturally, species have evolved molecular mechanisms to cope with predictable changes to their environment. Some of the starkest environmental changes on earth are driven by the day/night cycle, created by the rotation of the earth on its axis every 24 hours (hrs). In order to anticipate and respond to daily environmental changes, organisms from all five kingdoms of life are known to have evolved endogenous time keeping systems from at least two independent evolutionary origins (Rosbash, 2009). Biological processes that oscillate in an endogenous, entrainable and temperature-compensated manner approximately every 24 hrs, are described as circadian rhythms (taken from the Latin 'circa', approximately, and 'diem', day) (Vitaterna et al., 2001). In other words, a circadian rhythm is endogenously generated even in the absence of entrainment factors such as sleep, light/dark cycles or temperature cycles; is able to respond to entrainment factors, e.g. shift work; and runs at the same period regardless of whether the external temperature is 10°C or 30°C (Kidd et al., 2015).

In most mammals, diurnal or nocturnal, the key 'Zeitgeber', external 'time-giver' or entrainment factor, is light. The detection of light by the eye transfers a signal to a region of the brain located in the hypothalamus called the suprachiasmatic nuclei (SCN), which modulates the cellular molecular clock in

the SCN (Ginty et al., 1993). As the SCN is the central timekeeper in mammals, it is capable of inducing the secretion of melatonin and glucocorticoids which can coordinate the molecular clocks of the peripheral organs (Cermakian & Boivin, 2009; Klein & Moore, 1979; Oster et al., 2006). The global coordination of molecular clocks is primarily, but not exclusively, the remit of the SCN. For instance, non-photic Zeitgeber signals including exercise and feeding patterns have been demonstrated to be sufficient to alter the sleep/wake behaviour of rats independently of the SCN (Ruis et al., 1989). In fact, many different external and internal signals are involved in the global coordination of the circadian timing system, as displayed in Figure 1.1 (Kinouchi et al., 2018).

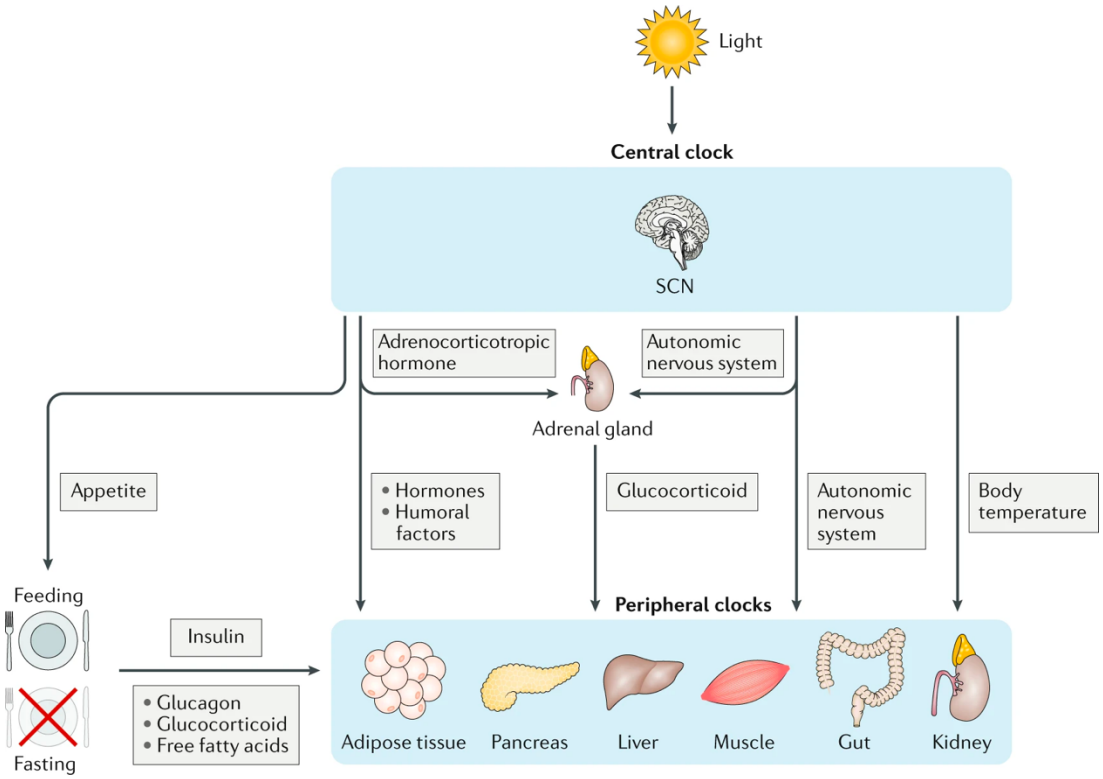


Figure 1.1 Reproduced from Kinouchi and Sassone-Corsi (2020), Figure 1. A representation of the signalling pathways between the central and peripheral oscillators that allow for global coordination of the circadian timing system.

1.2 Molecular architecture of the mammalian circadian oscillator

In mammals, the cellular molecular clock is primarily driven by approximately 15 canonical clock genes which form transcriptional-translational feedback loops that control programmes of oscillatory gene expression over the course of a 24 hr period (Figure 1.2) (K. H. Cox & Takahashi, 2019). Heterodimers of aryl hydrocarbon receptor nuclear translocator like protein (ARNTL) and circadian locomotor output cycles kaput protein (CLOCK) bind to E-box-like enhancer sequences upstream of the period (*PER1*, *2* & *3*) and cryptochrome genes (*CRY1* & *2*) (Gekakis et al., 1998; Hogenesch et al., 1998; Nakahata et al., 2008; Reick et al., 2001; van der Horst et al., 1999). Neuronal PAS domain protein 2 (NPAS2) is understood to be able to substitute for CLOCK in the brain (Reick et al., 2001). As the PERs and CRYs are expressed and begin to accumulate in the cytoplasm, they are subject to extensive regulation and bind with casein kinase 1 (CSNK1), RAB5-activating protein 6 (RAP6) and other proteins to form multimeric complexes approximately 1 megadalton (MDa) in size which are imported into the nucleus (Aryal et al., 2017). Once inside the nucleus these PER/CRY repressor complexes bind further proteins until they reach ~1.9 MDa (Aryal et al., 2017). The PER/CRY repressor complex binds to and degrades ARNTL/CLOCK, removing ARNTL/CLOCK from E-box enhancer sequences. Therefore, high levels of PER and CRY prevents further transcription of *PER* and *CRY* by ARNTL/CLOCK thereby closing the first transcriptional-translational feedback loop (R. Ye et al., 2014).

A second transcriptional-translational feedback loop is formed by the binding of the ARNTL/CLOCK heterodimer to E-box regions upstream of *NR1D1* and *NR1D2*, to promote their expression (Preitner et al., 2002). The concentration

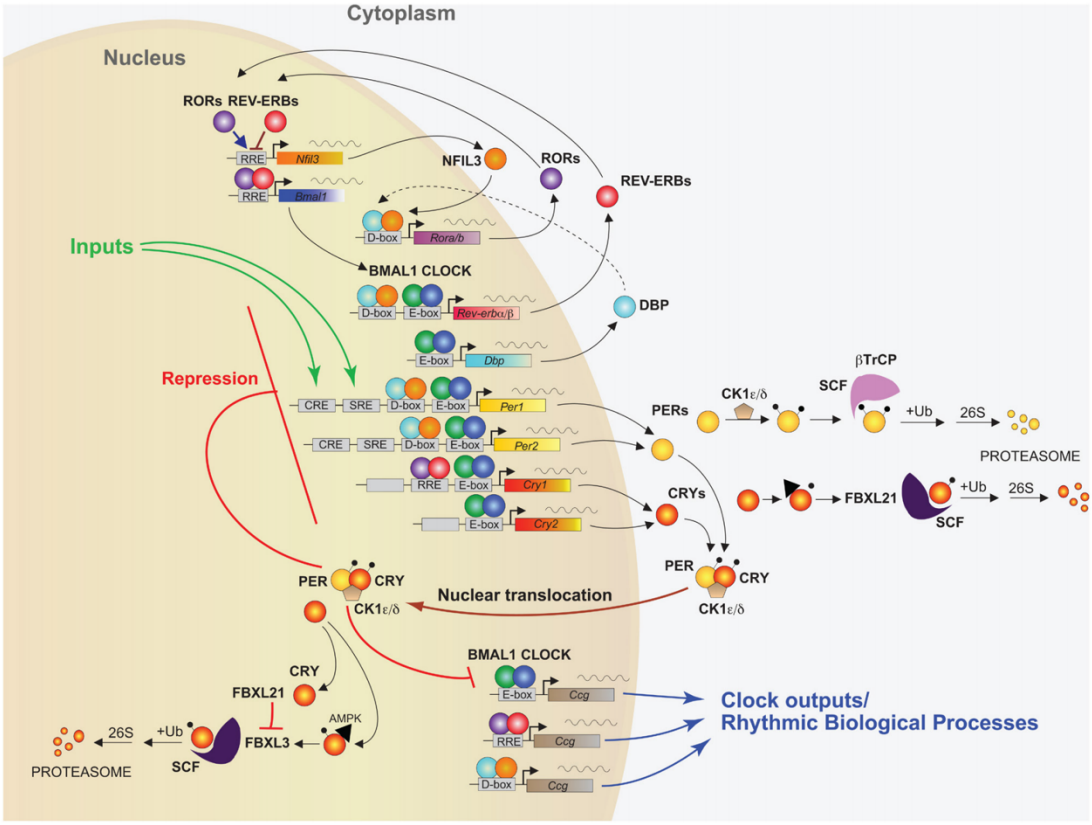


Figure 1.2 The circadian molecular clock. Reproduced from Cox and Takahashi, 2018, Figure 1. E-box binding transcription factors ARNTL (BMAL, green circles) and CLOCK (blue circles), heterodimerise and drive expression of *PER*s (yellow circles), *CRY*s (red/yellow circles), *DBP* (cyan circles), *NR1D*s (*REV-ERB*s, red circles) and other E-box controlled genes. PER and CRY proteins are subject to post-translational modifications in the cytoplasm, such that they dimerise with CSNK1 (CK1, beige hexagon) and are translocated into the cell nucleus. As the PER/CRY complex accumulates in the nucleus it acts to degrade BMAL/CLOCK heterodimers and repress expression of E-box driven genes, thus repressing its own transcription. Further transcriptional-translational control of the molecular oscillator is driven by the competition of NR1D1 and ROR (purple circles) for RRE sites on the *BMAL* and *NFIL3* promoters. This figure also describes a raft of post-translational molecular clock processes which will not be discussed in great detail here, as they are beyond the scope of the work undertaken throughout this thesis.

of the NR1Ds increases until the NR1Ds out-compete RAR-related orphan receptor (RORα & β) binding to ROR-response elements (RREs) upstream of *ARNTL*, acting to repress expression (Guillaumond et al., 2005). Therefore, oscillatory expression of the NR1Ds serves to help regulate the circadian oscillation of ARNTL, PERs and CRYs.

The oscillatory system is further complicated by a series of other molecular interactions with basic leucine zipper domain (bZIP) transcription factors, which lend nuance to the mammalian molecular clock. The ARNTL/CLOCK heterodimer binds to other E-box regions leading to expression of genes such as D-box binding PAR bZIP transcription factor (*DBP*) (Ripperger & Schibler, 2006). Meanwhile expression of other PAR bZIP genes TEF transcription factor (*TEF*) and HLF transcription factor (*HLF*) are known to be regulated both dependently and independently of ARNTL (Hatanaka et al., 2010). The PAR bZIP proteins are thought to bind to D-box regions upstream of the *ROR*s, *NR1D*s, *PER*s and other genes, and are known to promote expression of PER1 and PER2 (Mitsui et al., 2001; Yamaguchi et al., 2000). In contrast, the related bZIP protein nuclear factor interleukin 3 regulated (NFIL3) binds >1,400 D-box sites across the genome that are also known to bind DBP, and is known to compete with DBP to have a repressive effect on PER1 expression (Mitsui et al., 2001; Yamaguchi et al., 2000). NFIL3 has also been demonstrated to heterodimerise with basic helix-loop-helix family member E41 (BHLHE41) and repress expression of ARNTL through binding to the E-box-like element upstream of *ARNTL* (Tanoue et al., 2015).

Since the circadian molecular clock forms a 24 hr oscillator in cells, the phase of expression (commonly taken as the 'acrophase', the maxima of the sinusoidal expression) of each of the genes is consistent relative to the phases of other genes. This is easy to understand, as genes under E boxes promoter control such as *DBP* and *PER2* will be expressed with similar phases but will have different phases to E box binding proteins such as *ARNTL* and *CLOCK*. Phase relationships of clock genes do appear to vary in consistency between different tissues and individuals by several hours – perhaps more so in humans than in mice (Hughey & Butte, 2016; R. Zhang et al., 2014). The best understanding of phase relationship differences across human tissues comes from CYCLOPS,

a model which ordered non-time-stamped samples taken from human organs post-mortem (Anafi et al., 2017). This data set remains to be verified by time-stamped data and will be discussed further in chapter 4.1.2.9. A data set produced from Olive baboons under light/dark entrainment currently provides the best understanding yet of clock gene phase differences across different tissues in a diurnal species closely related to humans, and indicates that there can be quite a large range in circadian gene expression between tissues (Figure 1.3A and B) (Mure et al., 2018). Meanwhile the most comprehensive circadian transcriptome in multiple murine tissues was published by R. Zhang et al. 2014 (Figure 1.3C and D).

Whilst the molecular clockwork is finely tuned, it is not inflexible, as indicated by small differences in the phase of expression of clock genes between different tissues. This is likely to be because the circadian clock interacts with, and is regulated by, other key cellular players. MYC proto-oncogene (MYC), for instance, binds to E-boxes in direct competition with ARNTL/CLOCK, and therefore is capable of 'short-circuiting' the molecular oscillator in over-expression models (Altman et al., 2015). The relevance of this interaction at physiological MYC remains under-explored at present, although MYC is frequently over-expressed in cancer. A further example is that of tumour necrosis factor (TNF), which is implicated in diurnal joint pain in rheumatoid arthritis. TNF has been demonstrated to act in a calcium-dependent manner to increase RORα and decrease NR1D1 binding to the *ARNTL* RRE, and thus is capable of driving ARNTL expression (Yoshida et al., 2018).

The downstream effects of the molecular clock are so pervasive that in mice the core clock genes are known to influence the timing of expression of up to 43% of the genome, in at least one tissue of twelve tissues surveyed (R. Zhang et al., 2014). In fact it is important to understand that although the molecular

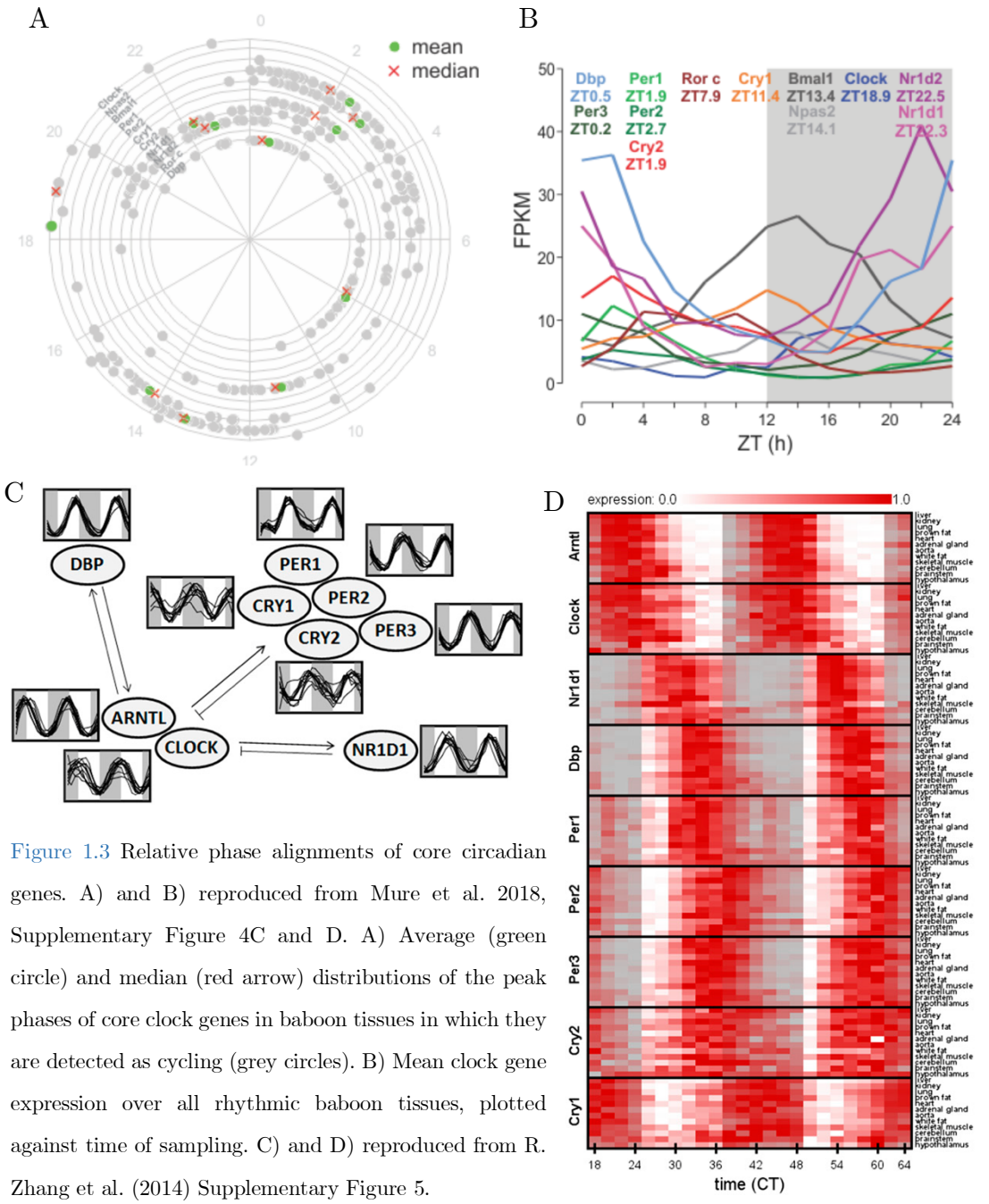


Figure 1.3 Relative phase alignments of core circadian genes. A) and B) reproduced from Mure et al. 2018, Supplementary Figure 4C and D. A) Average (green circle) and median (red arrow) distributions of the peak phases of core clock genes in baboon tissues in which they are detected as cycling (grey circles). B) Mean clock gene expression over all rhythmic baboon tissues, plotted against time of sampling. C) and D) reproduced from R. Zhang et al. (2014) Supplementary Figure 5.
C) Simplified model of the circadian clock network with circadian gene oscillation in 8 mouse organs super-imposed. D) A heatmap representation of the same core clock genes displayed in C).

clock is driven by a relatively small number genes oscillating in a series of transcriptional-translational feedback loops, the result is global circadian changes in the proteome, metabolome, phosphorylome, cistrome, epigenome and microRNAome (Dallmann et al., 2012; Kochan et al., 2015; Koike et al., 2012; Mauvoisin et al., 2014; Robles et al., 2017).

Therefore, it is unsurprising that circadian gene dysregulation in humans is associated with multiple disease states including, sleep-wake disorders, cardiovascular disease, diabetes and cancer (reviewed by Auger et al., 2015; Crnko et al., 2019; Stenvers et al., 2019; Sulli et al., 2019).

1.3 An introduction to cancer

Cancer is the disease caused by uncontrolled division of abnormal cells. It is a major cause of death, second only to heart disease globally (Ferlay et al., 2019). Abnormal cancer cell division is supported by a series of abnormal biological processes known as the 'hallmarks of cancer' (Figure 1.4) (Hanahan & Weinberg, 2011). Broadly speaking, these 'hallmarks' describe the manner in which cells with accumulated mutations exhibit uncontrolled cell growth and division, promote re-structuring of the local environment and manipulate/evade the immune response (Hanahan & Weinberg, 2011). Some of the components of the molecular circadian clock are known to directly interact with some of the key components of processes known to contribute to the hallmarks of cancer, including apoptosis (MYC, p53), cell division (WEE1, Cyclin B1, p53) and mitochondrial metabolism (Farshadi et al., 2020; Schmitt et al., 2018; Stephenson et al., 2021). It is therefore unsurprising that the molecular circadian clock has been implicated in the development of oncogenesis, as will be discussed in detail in the chapter sections that follow.

1.4 The molecular clock and oncogenesis

The relationship between cancer and the circadian clock has long been a topic of research, with epidemiological research suggesting a link between light

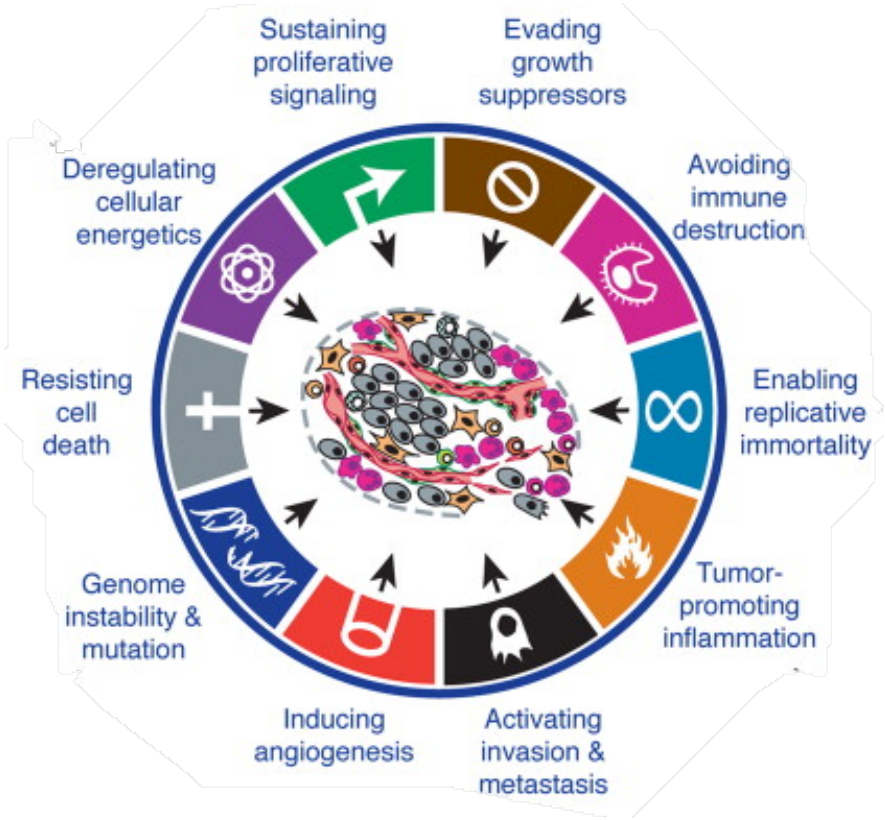


Figure 1.4 The hallmarks of cancer. Adapted from Hanahan and Weinberg, 2011, Figure 6.

exposure at night, and breast cancer incidence as early as 1987, although opinion remains divided regarding the strength of this association (Fagundo-Rivera et al., 2020; Jones et al., 2019; Schernhammer et al., 2001; Stevens, 1987). Additionally, an association between melatonin secretion from the pineal gland and cancer incidence has been a focus of research since the 1970s (Cohen et al., 1978).

A great deal of *in vitro* and *in vivo* research has attempted to elucidate the contribution of each of the core clock genes in the development and clinical outcome of cancer. Whilst single gene over-expression/knock-down models can be undeniably useful to researchers, such a gene-wise approach to circadian cancer research is a somewhat blunt tool because the circadian clock is a finely tuned oscillatory mechanism. Therefore, any change in expression of a single clock gene has a knock-on effect on the expression of other genes in the

circadian network. Additionally, many studies that use over-expression/knock-down models do not investigate whether expression of the gene of interest remains rhythmic upon over-expression/knock-down. Therefore, it is often difficult to predict whether an observed effect is dependent on the absolute expression level of a core clock gene, or degree of oscillatory expression of the core clock gene manipulated, or indeed upon the subsequently altered expression of another core clock gene. Nonetheless, a gene-wise approach to circadian cancer research has yielded some interesting findings which are discussed below with reference to the core clock genes *ARNTL*, *CLOCK*, *PER1*, *2* & *3*, *CRY1* &*2* and *NR1D1* & *2*.

1.4.1 ARNTL

In colorectal cancer and metastatic melanoma patient populations, higher than mean tumour expression of ARNTL has been associated with significantly longer overall survival relative to patients with lower than mean tumour ARNTL expression (de Assis et al., 2018; Zeng et al., 2014). Since the measure of absolute ARNTL expression relative to the tumours of other patients does not give any indication of the degree of functionality of the circadian clocks in these tumours, nor do we know at what time of day the tumour samples were taken from patients – and therefore whether ARNTL in the tumour was higher or lower than to be expected – the implications of the *in vivo* data regarding the role of ARNTL in cancer should be interpreted with caution. These are common problems for circadian research on human tumour samples, and will be addressed further in chapter 4.1.

In vivo data from murine models largely support a tumour-suppressive role for *ARNTL*, as implied by the patient data. For instance, knock-down of *ARNTL*

in murine colon carcinoma C26 cells, has been demonstrated to increase tumour growth in mice (Zeng et al., 2010). Similarly, in the pancreatic cell line BxPC-3, *ARNTL* knock-down was shown to cause accelerated tumour growth in a xenotopic murine model of pancreatic cancer. Moreover, further experiments demonstrated that *ARNTL* over-expression in the pancreatic cell line AsPC-1, resulted in tumour suppression *in vivo* (Jiang et al., 2016). The xenotopic murine pancreatic carcinoma models demonstrated that under/over-expression of *ARNTL* resulted in a corresponding down/up-regulation of phospho-tumour protein p53 (TP53), whilst ARNTL was observed to bind to the promoter of *TP53*. The link between *ARNTL* and *TP53* has been investigated further in a hepatocellular carcinoma cell line where ectopic over-expression of *ARNTL* was shown to increase TP53 expression and reduce tumour load in mice (Fekry et al., 2018). Lastly, whilst *ARNTL* knock-out in a Cre recombinase-induced murine lung adenocarcinoma model caused an increase in tumour burden, the tumour burden in $p53^{-/-}$ background mice could not be increased further by *ARNTL* knock-down in the tumours (Papagiannakopoulos et al., 2016). These findings all support the idea that the tumour-suppressive effect of *ARNTL* expression may be TP53-dependent. In many ways this would make sense as in healthy cells TP53 is capable of halting the cell cycle, and the circadian clock and the cell cycle are also known to interact with one another in healthy cells (Farshadi et al., 2020). It has been hypothesised that one of the ways in which changes to the circadian clock might drive cancer is through changes to the cell cycle, which would drive changes in cell proliferation.

In contrast to the finding that ARNTL can be tumour-suppressive, murine models of acute myeloid leukaemia (AML) have comprehensively demonstrated that when ARNTL is knocked down, leukaemic cells begin to differentiate to a myeloid-lineage monocytic phenotype and cease replication (Puram et al., 2016). Their healthy myeloid progenitor counterparts do not require ARNTL

to maintain replication, so it would seem that something about the leukaemic cell required a functional circadian clock to drive the cell cycle. Unfortunately, TP53 was not investigated by Puram et al. (2016), but as the circadian clock drives different transcriptional programmes in different tissue types, and different cancer types have very different genetic backgrounds, it is perhaps to be expected that different cancer types may react to changes to the circadian clock in different ways (Y. Ye et al., 2018; R. Zhang et al., 2014).

As in *in vivo* data, *in vitro* data on the role of *ARNTL* in tumourigenesis is divided. Experiments in pancreatic adenocarcinoma, colorectal carcinoma, colon adenocarcinoma, breast cancer, glioblastoma cell lines and patient-derived glioblastoma cells, all demonstrate that *ARNTL* has a tumour-suppressive role (Dong et al., 2019; Gwon et al., 2020; Jiang et al., 2016; Ramos et al., 2020; J. Wang et al., 2019; Zeng et al., 2014; Yuan Zhang et al., 2020). However, experiments in mesothelioma and AML cell lines demonstrate a tumour-proliferative role for *ARNTL* (Elshazley et al., 2012; Puram et al., 2016). It is worth noting, that as is often the case in the *in vivo* experiments discussed, *in vitro* experiments are often carried out on populations of cells with an unknown circadian clock background and an unknown degree of circadian synchronicity between cells in the population, both pre- and post-manipulation of *ARNTL* expression.

1.4.2 CLOCK

Data concerning the role of ARNTL's binding partner CLOCK, largely supports a tumour promoting role for CLOCK. For instance, Puram et al. (2016) extended their *in vivo* murine work in AML cells to demonstrate that *CLOCK* knock-down halted cell cycle progression, just as *BMAL* had (Puram

et al., 2016). Other significant *in vivo* murine research into the role of *CLOCK* in cancer used the SW620 and SW480 colon carcinoma cell lines to demonstrate that knock-down of *CLOCK* slowed the metastasis of subcutaneous flank tumours, meanwhile over-expression of *CLOCK* resulted in increased tumour growth (Yaping Wang et al., 2015, 2017). The authors developed their argument that over-expression of *CLOCK* could induce tumourigenesis, by demonstrating *in vitro* that increased *CLOCK* expression caused an increase in epithelial-mesenchymal transition (EMT) markers, an increase in the angiogenesis-inducing factors hypoxia inducible factor 1 subunit alpha (HIF1A), ARNT and vascular endothelial growth factor A (VEGFA), a decrease in the pro-apoptotic factors BCL2 associated X protein (BAX) and BH3 interacting domain death agonist (BID), and an increase in AKT serine/threonine kinase 1 (AKT) and phospho-AKT (Yaping Wang et al., 2015, 2017). Phosphorylation of AKT is known to be involved in the prevention of apoptosis, and also in enabling cell cycle progression, a finding which correlates with the observation of cell cycle arrest upon *CLOCK* knock-down in AML cells from Puram et al. (2016).

Further studies *in vitro* in other cell line models, support the idea that the effect of *CLOCK* on tumourigenesis would appear to be mediated through angiogenesis, cell cycle progression and apoptosis. For example, knock-down of *CLOCK* in the human glioblastoma cell line U87MG was demonstrated to result in increased apoptosis, decreased MYC proto-oncogene, bHLH transcription factor (MYC) and cyclin B1 (CCNB1) (F. Wang et al., 2016). Meanwhile knock-down of *CLOCK* in patient-derived glioblastoma cells reduced their survival, and increased caspase 3 (CASP3) cleavage (Dong et al., 2019).

It is likely that the adverse effects of CLOCK in cancer are at least partially mediated by oestrogen signalling. Histological examination of breast tumours showed that oestrogen receptor (ESR1) expression correlated with *CLOCK* expression, while addition of oestrogen to ESR1+ MCF7 breast cancer cells resulted in an increase in expression of *CLOCK* (Xiao et al., 2014). Earlier work in breast cancer patients identified that particular single nucleotide polymorphism (SNP) variants of *CLOCK* can be associated with either increased or decreased risk of breast cancer diagnosis. Importantly the ESR1/progesterone receptor (PGR) status of the patients' breast cancers also affected the risk associated with each of the *CLOCK* variants, implying an association between the circadian clock and cancer that is dependent upon cancer hormone status (Hoffman et al., 2010).

Whilst the majority of evidence would imply a tumour-promoting role for *CLOCK*, Cadenas et al. (2014) demonstrated that higher than mean *CLOCK* tumour expression correlated with significant improvement to metastasis free survival (MFS) in ESR–/ erb-b2 receptor tyrosine kinase 2 (ERBB2)– breast cancer patients. However, there was no such association in ESR+/ERBB2– patients, which further supports the idea that the association between the circadian clock and cancer is dependent upon cancer hormone status (Cadenas et al., 2014).

1.4.3 PER1, PER2 and PER3

The *PER* genes function as transcriptional repressors, as they remove the ARNTL/CLOCK complex from E box promoter motifs. Evidence *in vitro* and in *in vivo* murine models is largely supportive of a tumour-suppressive role for *PER2*. Of the *PER* genes, *PER2* is the best researched in the literature, and

the same is true in relation to cancer research. In a series of early papers, Hua et al., (2006, 2007) demonstrated that *in vitro* over-expression of *Per2* in mice and in lung and mammary carcinoma cell lines reduced cell proliferation, resulted in the downregulation of *c-Myc*, anti-apoptotic BCL2-like 1 (*Bcl2l1*) and *Bcl2*, and the upregulation of *p53* and pro-apoptotic *Bax* (Hua et al., 2006). Furthermore, *in vivo* ectopic delivery of *Per2* to murine lung carcinomas reduced tumour burden (Hua et al., 2007). Similar results were later observed in human carcinoma cell lines; over-expression of murine *Per2* in human pancreatic carcinoma cells decreased cell proliferation, increased pro-apoptotic BAX and reduced anti-apoptotic BCL2L1, whilst knock-down of *PER2* in osteosarcoma cells increased proliferation via increased phospho-AKT and BCL2 and reduced histone 3 pseudogene 23 (H3P23), H3P16 and cleaved CASP3 (Oda et al., 2009; Qin et al., 2018). Experiments in mammary MCF10A cells have demonstrated that hypoxia resulted in the rapid degradation of PER2, and direct de-repression at the *OCT1* promoter, which triggered the downstream expression of EMT-promoting genes (Hwang-Verslues et al., 2013). This mechanism appeared to be somewhat similar to the HIF1A and EMT response driven by the over-expression of *CLOCK* in colon carcinoma lines, which would make sense as over-expression of *CLOCK* might also have perturbed the activity of PER2 in the colon carcinoma cell lines (Yaping Wang et al., 2015, 2017).

In vivo experiments have confirmed these findings in murine carcinoma models. Over-expression of *PER2* in nasopharyngeal carcinoma cells was demonstrated to slow cancer cell proliferation in murine models, and also caused a decrease in cell cycle and proliferation drivers including phospho- mitogen-activated protein kinase 1 (phospho-MAPK1) and phospho-MAPK14 (Hou et al., 2020). Meanwhile over-expression of *PER2* in chronic myeloid leukaemia (CML) also slowed cancer cell division in a murine model, and resulted in a decrease in the

cell cycle progression drivers MYC and CCNB1 and an increase of TP53 (Sun et al., 2010). Papagiannakopoulos et al. (2016) later demonstrated that when the circadian clocks of mice were disrupted via global *Per2* knock-out, they were more susceptible to heavier tumour burden in a lung carcinoma model. In slight disagreement with Sun et al. (2010) the effect of *PER* knock-out was demonstrated to be independent of the *p53* status of the animal (Papagiannakopoulos et al., 2016). However, the models were different in that Sun et al. (2010) used mice with a healthy clock background, and manipulated the tumour cell clock, whilst Papagiannakopoulos et al. (2016) used mice with a disrupted clock background and did not manipulate the tumour cell clock any further.

Whilst there is ample evidence *in vitro* and in *in vivo* murine models that higher *PER2* expression induces a multitude of tumour-suppressive pathways, data from humans is more divided, and suggests that the effect of *PER2* in cancer is dependent upon cancer type. In a comparison of patient matched tumour/tumour-margin biopsies from 14 cancer types in The Cancer Genome Atlas (TCGA) *PER2* was under-expressed in tumours relative to tumour margins in a majority of cancer types (Y. Ye et al., 2018). However, in a minority of cancer types, *PER2* expression was associated with driving the pro-tumourigenic RAS/MAPK or receptor tyrosine kinase (RTK) signalling pathways. In keeping with the *in vitro* and *in vivo* experimental data discussed thus far, in breast cancer or pancreatic ductal adenocarcinoma (PDA) patients, lower than mean tumour *PER2* expression was associated with poorer MFS and overall survival (OS) respectively (Cadenas et al., 2014; Relles et al., 2013).

Research into the roles of *PER1* and *PER3* in cancer is relatively scant. *In vitro*, *PER1* over-expression was shown to decrease proliferation of HCT116 colon carcinoma cells, whilst knock-down increased proliferation of HCT116

cells (Gery et al., 2006). However, *PER1* knock-down was shown to decrease proliferation in the pancreatic cell line, PaCa-2, and the hepatocarcinoma cell line, HepG2 (Sato et al., 2009). In patient populations, lower than mean *PER1* and *PER3* expression were both shown to correlate with poorer OS in PDA and head and neck squamous cell carcinoma (Oshima et al., 2011). Lower than mean PER1 expression also correlated with increased risk of metastasis in cases of colorectal cancer (Relles et al., 2013).

1.4.4 CRY1 and CRY2

In recent years it has been demonstrated that CRY2 can recruit MYC, tousled like kinase 2 (TLK2) and other proteins to F-box and leucine rich repeat protein 3 (FBXL3) for ubiquitination and resultant degradation (Huber et al., 2016; Papp Correia et al., 2019). TLK2 has previously been demonstrated to be required for cell recovery upon DNA damage, mediated by recovery from the DNA damage response-induced cell cycle checkpoint (Bruinsma et al., 2016). Slightly counterintuitively, the recovery from DNA damage was more successful in cells which lacked *TP53* – it would seem that as long as TLK2 remained present a *TP53* knock-out background pushed cells towards a repair pathway rather than apoptosis (Bruinsma et al., 2016). This information relating the role of CRY and TLK2 to TP53 and the DNA damage response offers an explanation for the earlier observation that *Cry1*$^{-/-}$*Cry2*$^{-/-}$ double knock-out mice are resistant to the increased rate of tumourigenesis usually observed in *p53*$^{-/-}$ mice. It may be that in *Cry1*$^{-/-}$*Cry2*$^{-/-}$*p53*$^{-/-}$ mice TLK2 is not degraded by CRY2 and FBXL3, and therefore is able to induce a successful DNA damage response, resulting in reduced tumourigenesis. This evidence would suggest that *CRY2* expression has a tumour-enhancer role. In agreement with this, higher *CRY1* expression was observed in colorectal tumour biopsies

relative to adjacent noncancerous tissue, and higher tumour *CRY1* expression correlated with decreased OS for colorectal cancer patients (H. Yu et al., 2013). Similarly, *CRY1* was found to be significantly upregulated in higher stage gastric cancers (M. L. Hu et al., 2014).

Despite the findings discussed thus far, data largely supports the view that *CRY2* is tumour-suppressive, not tumour-enhancive. *CRY2* was under-expressed relative to tumour margins in 10 of the 14 cancer types surveyed by Ye et al. (2018) from TCGA, and higher expression has been associated with prolonged MFS in breast cancer, and with better OS in PDA (Cadenas et al., 2014; Relles et al., 2013; Y. Ye et al., 2018). *In vitro CRY2* knock-down in the human osteosarcoma (HOS) cell line has been demonstrated to increase cell proliferation, cell cycle progression, expression of MYC and cyclin D1 (CCND1) and phosphorylation of MAPK3 and MAPK1 (Y. Yu et al., 2018). Meanwhile, similar observations have been made for *CRY1* knock-down in HOS and U2OS osteosarcoma cells both *in vitro* and in *in vitro*. Specifically, *CRY1* knock-down increased cell proliferation, migration, *in vivo* tumour growth, expression of cyclin A (CCNA1/2), cyclin dependent kinase 2 (CDK2), phosphorylation of AKT and MDM2 proto-oncogene (MDM2), and decreased TP53 (L. Zhou et al., 2018). It seems likely that the different effects of the *CRY* genes on tumourigenesis in different models, could be due to different tumour backgrounds, such as *TP53* mutation status. Perhaps this is able to dictate whether the role of *CRY* in cancer cells is tumour-promoting through TLK2 degradation, or tumour-suppressive through MYC degradation.

1.4.5 NR1D1 and NR1D2

The *NR1D* genes are under the transcriptional regulation of the E-box–binding ARNTL/CLOCK complex, and so are expressed in a similar circadian phase as the E-box driven *PER* genes. NR1D proteins act as transcriptional repressors by binding to RRE motifs in gene promoters and recruiting the NCOR1-HDAC3 complex via a heme ligand (Yin & Lazar, 2005; Zamir et al., 1996). HDAC3 deacetylates histones, which allows chromatin to condense and therefore represses transcription locally (Pazin & Kadonaga, 1997). The reason for the significance of this mechanism is that small molecules have been developed that replace the role of heme and therefore act as pharmacological NR1D agonists or antagonists (Grant et al., 2010; Kojetin et al., 2011). Work with NR1D agonists in cancer cells has universally supported an anti-tumourigenic role for the NR1D proteins. The agonist SR9009 was demonstrated to be cytotoxic *in vitro* to T98 glioblastoma cells and HepG2 cells, and also reduced proliferation of glioblastoma stem cells (Dong et al., 2019; Wagner et al., 2019). Meanwhile the agonist SR9011 was demonstrated to reduce proliferation of the same glioblastoma stem cells, and also reduced cell viability in an *NR1D2*-dependent manner in the breast cancer cell lines MCF7, MDA-MB-231, MDA-MB-361 and BT474 (Yongjun Wang et al., 2015). Recent work *in vivo* has confirmed the *NR1D*-dependent anti-tumourigenic effect of SR9009 in a glioblastoma murine model (Sulli et al., 2018). The authors demonstrated that the anti-tumourigenic effect of SR9009 was mediated by an *NR1D*-dependent inhibition of autophagy which resulted in increased p53-independent apoptosis specific to cancer cells. Independent work has demonstrated that SR9009 induces NR1D1 to bind to the promoter regions and repress expression of many core autophagy genes, which was earlier theorised by Sulli et al. (2018) computationally (Shen et al., 2020). Shen et al. (2020) also demonstrated that SR9009 reduced small-cell lung carcinoma

growth in a murine subcutaneous tumour model in an *NR1D*-dependent manner.

The research that has not used *NR1D* agonists is also largely supportive of an anti-tumourigenic role for the *NR1D* genes. *In vitro NR1D2* knock-down in BT474 breast cancer cells was demonstrated to increase cell proliferation, whilst *NR1D1* knock-down in HCT116 colorectal carcinoma cells also cell increased proliferation (Basti et al., 2020; Kourtidis et al., 2010). In contrast to this, Tong et al. (2020) suggested that since high expression of *NR1D2* in patient HCC tumours correlated with poorer OS, *NR1D2* may have a tumour-promoting role in some cancer types. They went on to demonstrate that *NR1D2* knock-down reduced proliferation of HCC cell lines Huh7 and HCCLM3. The mechanism appeared to be dependent on NR1D2-induced over-expression of catenin beta 1 (CTNNB1), and a resultant increase in EMT markers such as cadherin 2 (CDH2) (Tong et al., 2020). This finding suggests that the role of the *NR1D* genes in cancer may vary depending upon the cancer type, as with the other clock genes discussed.

1.5 Challenges to experimental modelling of circadian tumour behaviour

It is challenging to model circadian tumour behaviour experimentally for several reasons. Firstly, it is clear from the research discussed thus far that the expression and role of circadian genes within tumours is likely to be heterogenous between tumour types, which is to be expected given that downstream circadian gene expression varies widely between tissue types even prior to the development of cancer (R. Zhang et al., 2014). Secondly the degree to which the circadian behaviour of an *in situ* tumour is affected by, and

affects, its microenvironment and host, remains mostly unexplored (Hadadi & Acloque, 2021). For instance, the circadian clock is known perturb key immuno-modulators including, but not limited to, TNF, interleukin 6 (IL6) and interferon gamma (IFNG), therefore the circadian state of a tumour is hypothesised to impact the immune microenvironment (Cao et al., 2017; Scheiermann et al., 2018). Recent papers have indicated that chronic circadian disruption in mice, not only promotes tumour growth, but results in tumour microenvironments with immune cell populations that are more immune-suppressive and therefore tumour-supportive in nature (Aiello et al., 2020; Hadadi et al., 2020). This research supports the idea that the clock of the host, the clock of the tumour and the nature of the tumour microenvironment are inter-linked, though presumably crosstalk between tumour cells and the microenvironment for the purpose of synchronisation of local cellular clocks may also vary between tumour types. A further dimension of complexity in modelling the circadian clock in cancer is understanding how the expression of the core clock genes change with time over a 24 hr period, and in relation to one another, as discussed in chapter 1.2. In order to explore how the molecular clock might contribute to the development and maintenance of tumourigenesis, experimental models must be designed to reflect these complex factors.

1.5.1 Contributions of this thesis

The work presented in this thesis was undertaken with the literature discussed thus far in mind. The chapters that follow will develop *in vitro*, *in ovo*, and *in vivo* approaches in order to explore the circadian clocks of cancer cells, with the aim of deepening our understanding of the complex interplay between the functionality of the molecular clockwork over the course of the day and

tumourigenesis. Specifically, the key findings of this thesis that will be presented are as follows:

- The "benign" MCF10A and "malignant" MCF7 breast cell lines cultured under a range of conditions both indicate significantly altered molecular clock function, to that previously described by the literature.
- The relatively novel use of bioluminescence reporters in the chick embryo chorioallantoic membrane tumour graft model should be approached with caution, as tumour bioluminescence may not correlate well with tumour growth and metastasis. The novel use of luciferin-loaded polymer patches is also investigated.
- The semi-supervised time prediction algorithm TimeTeller can be adapted for application to RNA-seq data, and used to detect circadian clock differences in murine models of circadian disruption.

2. Exploring circadian dysregulation in vitro using human breast MCF cell lines

2.1 Introduction

2.1.1 Breast cancer and the circadian clock

Breast cancer is the second most commonly diagnosed cancer worldwide, a significant statistic considering that the majority of the disease burden lies with only half the population (Ferlay et al., 2019). Improving our understanding of breast cancer through experimental research is paramount to being able to understand and treat the disease. In the field of circadian cancer research, breast cancer research in particular has a long history due to studies on the effect of shift work on breast cancer risk, which have often surveyed cohorts of shift-working nurses (Schernhammer et al., 2001; Stevens, 1987). Evidence discussed in the previous chapter appears to indicate that in breast cancer the core clock genes *ARNTL*, *CLOCK*, *PER2*, *CRY2* and *NR1D2* have tumour-suppressive roles. However, there are few studies relating to breast cancer that have examined expression of all of the clock genes in relation to one another, in order to try and further our understanding of how the circadian clock functions in breast cancer.

For instance, Y. Ye et al. (2018) comprehensively examined circadian clock gene expression in the breast carcinomas of patients relative to tumour margins, and established that expression of the core clock genes *PER1*, *PER2*, *PER3*, *NR1D1*, *NR1D2*, *NFIL3*, *CRY2*, *RORA* and *RORB* are all decreased

in the tumours of breast carcinoma patients relative to their tumour margins. However, since this data covers a range of different breast cancer types from TCGA, which were presumably sampled at different times of day (but not night), it is impossible to know what the state of the circadian clock in the individual tumours might have been (Y. Ye et al., 2018). To begin to answer the question of how clock gene expression changes over 24 hrs in cancer, we must look to experimental models where it is possible to sample repeatedly over time. *In vitro* research on cancer versus non-cancer cell lines provides such an opportunity. There are very well-characterised cell lines derived from breast tissues, which are commonly used in breast cancer research. In particular this chapter discusses the use of the benign and malignant MCF breast cell lines as a model for investigating the role of the molecular circadian clock in cancer *in vitro*.

2.1.2 The MCF cell lines and their use in circadian research

The MCF10A breast epithelial cell line was derived in the 1980s from breast tissue cells which spontaneously immortalised *in vitro* and is used to represent benign breast tissue (Qu et al., 2015; Herbert D. Soule et al., 1990). Meanwhile, the MCF7 cell line was developed in 1973 from the metastatic breast adenocarcinoma of a 69 year old woman, and is frequently used as a model of ESR1 and PGR positive breast cancer (Comşa et al., 2015; H D Soule et al., 1973).

Both MCF cell lines have been used in previous circadian rhythms research efforts, with MCF10As frequently touted as "benign good clock" cells and MCF7s as "malignant bad clock" cells (Gutiérrez-Monreal et al., 2016). Table 2.1 compares the expression of core clock genes between the two cell lines in

Table 2.1 A literature-wide summary of the detection of core circadian clock genes in A) MCF10A and B) MCF7 cells. 'U = unclear' is used to denote cases where it is difficult to determine from the data presented by the authors whether a gene is rhythmic or not, or cases where gene expression is rhythmic but it is unclear whether the correct phase relationship occurs relative to the other rhythmic clock genes detected.

A MCF10A

Published since the start of this PhD

Author:	Rossetti et al., 2012			Xiang et al., 2012			Cox, 2012 (Doctoral thesis, UCL)			Gutierrez-Monreal et al., 2016			Chacolla-Huaringa et al., 2017			Zhang, Y. et al., 2018			Lin et al., 2019			Lin et al., 2019		
Experiment:	qPCR, normalised to GAPDH			qPCR, normalised to GAPDH			qPCR, normalised to TBP			qPCR, norm to GAPDH and RPLP0			qPCR, normalised to GAPDH			Lentiviral luciferase reporter			qPCR, norm to GAPDH			Lentiviral luciferase reporters		
Entrainment:	Starve, 2hr serum shock, starve			Starve, 2hr serum shock, starve			2 hr 100 nM dexamethasone			Starve, 2hr serum shock, starve			Starve, 2hr serum shock, starve			Starve, 2hr serum shock, starve			Starve, 2hr shock, starve			Starve, 2hr shock, starve		
Clock Gene	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**
ARNTL	Y	23	Y	Y	24	Y	Y	32	Y	Y	27	Y	Y	24.15	U	Y	19	-	Y	27	Y	Y	27	Y
PER1	Y	24	Y	Y	24	Y	Y	29	Y															
PER2	N	-	-	Y	20	Y	Y	30	Y	Y	24	Y	Y	20.4	U				Y	27	Y	Y	27	Y
PER3	Y	24	Y							N.B. corresponding microarray data set exists, but data is only displayed in the Supplement, probe-wise and without sine fits.														
CRY1	U	-	-	U	-	-																		
CRY2	U	-	-	Y	24	Y																		
NR1D1	U	-	-	Y	23	Y																		
CLOCK	N	-	-	U	-	-	U	-	-															

Y	Yes, rhythmic
U	Unclear. Variation in magnitude of gene expression is observed over time, but cannot be confident of circadian rhythmicity
N	No, not rhythmic
	Not assessed

B MCF7

Author:	Rossetti et al., 2012			Xiang et al., 2012			Cox, 2012 (Doctoral thesis, UCL)			Gutierrez-Monreal et al., 2016			Chacolla-Huaringa et al., 2017			Zhang, Y. et al., 2018			Lellupitiyage Don et al., 2019			Lellupitiyage Don et al., 2019			Lellupitiyage Don et al., 2020		
Experiment:	qPCR, normalised to GAPDH			qPCR, normalised to GAPDH			qPCR, normalised to TBP			qPCR, norm to GAPDH and RPLP0			qPCR, normalised to GAPDH			Lentiviral luciferase reporter			qPCR, norm to GAPDH			Lentiviral luciferase reporters			Lentiviral luciferase reporters		
Entrainment:	Starve, 2hr serum shock, starve			Starve, 2hr serum shock, starve			2 hr 100 nM dexamethasone			Starve, 2hr serum shock, starve			Starve, 2hr serum shock, starve			Starve, 2hr serum shock, starve			Starve, 2hr shock, starve			Starve, 2hr shock, starve			Starve, 2hr shock, starve		
Clock Gene	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**	**Rhythmic?**	**Approx period?**	**Correct phase?**
ARNTL	N	-	-	N	-	-	Y	24	Y	N	-	-	N	-	-	N	-	-	U	-	-	Y	22.2-25.2	Y	Y	24	Y
PER1	N	-	-	N	-	-	Y	28	Y																		
PER2	N	-	-	N	-	-	N	-	-	N	-	-	N	-	-				U	-	-	Y	18.7-23.6	Y	Y	24	Y
PER3	N	-	-							N.B. corresponding microarray data set exists, but data is only displayed in the Supplement, probe-wise and without sine fits.																
CRY1	N	-	-	N	-	-																				
CRY2	N	-	-	N	-	-																				
NR1D1	U	-	-	N	-	-																				
CLOCK	N	-	-	N	-	-	N	-	-																	

the published literature (Chacolla-Huaringa et al., 2017; Cox, 2012; Gutiérrez-Monreal et al., 2016; Lellupitiyage Don et al., 2019, 2020; Lin et al., 2019; Rossetti et al., 2012; Y. Zhang et al., 2018). In the case of the benign MCF10A cell line, rhythmic expression of *ARNTL* and *PER* transcripts is detected in almost all of the published research, however, the period of oscillation varies from as low as approximately 19 hrs, to as high as 32 hrs. It is worth noting that while Cox (2012) observed particularly long periods in *ARNTL*, *PER1* and *PER2* expression (>29 hr periods) the cells were also synchronised differently (using dexamethasone rather than serum starve/shock/starve). When the circadian rhythms of cells are assessed *in vitro*, it is common practice to first synchronise the population of cells to the same phase in their circadian rhythms. This is usually achieved by the addition of an exogenous synchroniser to the media in which the cells are cultured. Dexamethasone, forskolin and serum starve/shock/starve are all acceptable methods of synchronising circadian clocks in cells (Balsalobre et al., 1998, 2000; Yagita & Okamura, 2000). Cell synchronisation methods will be discussed in greater detail in chapter 2.1.3.

Much of the research summarised in Table 2.1 did not identify rhythmic behaviour in the core clock genes of MCF7s, in keeping with the conclusion that MCF7s are "bad clock malignant" cells. However, Cox (2012) demonstrated rhythmic expression of *ARNTL* and *PER2* in MCF7s. Whilst Lellupitiyage Don et al. (2019) did not demonstrate distinct MCF7 circadian rhythms in *ARNTL* and *PER2* transcripts by quantitative polymerase chain reaction (qPCR), they did show oscillatory promoter activation for *ARNTL* and *PER2* using transcriptional luciferase reporter lentiviral vectors (Lellupitiyage Don et al., 2019, 2020), which supports the findings of Cox (2012). It is interesting that Zhang et al. (2018) used transcriptional luciferase reporter lentiviral vectors to observe a particularly short 19 hr period of

MCF10A *ARNTL* oscillation. It should be noted that the transcriptional luciferase reporter lentiviral vectors used were built using the promoters of clock genes from mice, while MCFs are a human cell line, therefore it is possible that the luciferase reporters used by Zhang et al (2018) and Lellupitiyage Don et al. (2019, 2020) provided an inaccurate assessment of the circadian rhythms in the MCF10As and the MCF7s (Ramanathan et al., 2012). However, Ramanathan et al. (2012) demonstrate clear circadian oscillation of transcriptional mouse clock promoter luciferase reporters in U2OS human cells, which indicates that the homology between mouse and human clock gene promoters is high enough to report rhythmic behaviour accurately in either mouse or human cell lines.

The most detailed analysis of circadian rhythms in MCF10A and MCF7 cell lines undoubtedly comes from the DNA microarray analysis of gene expression undertaken by Gutiérrez-Monreal et al. (2016). Samples of serum starved/shocked/starved MCF10A and MCF7 cells were obtained from 8 hrs post return to serum starvation conditions, every 4 hrs, for 28 hrs. As detailed by Table 2.1, Gutiérrez-Monreal et al. (2016) observed clear circadian gene oscillation in MCF10A *ARNTL* and *PER2* by qPCR, however their microarray dataset did not appear to support this finding (Figure 2.1). Only one of the sixteen *PER2* probes demonstrated circadian oscillation in MCF10As in a phase that matched the qPCR data, meanwhile the single *ARNTL* probe appeared to oscillate with a 30.8 hr rhythm. Other circadian clock genes in MCF10A cells also did not appear to be functioning canonically. For instance, *NR1D2* displayed a short 19 hr rhythm in 2 out of 3 probes. One of the *DBP* probes appeared to oscillate in opposition to the *ARNTL* probe, but the other did not. Two *RORA* probes appeared to oscillate in phase with *ARNTL*, but the third was antiphasic, and the fourth was not rhythmic at all. Meanwhile, none of the probes for *CLOCK*, *CRY1*, *CRY2*, *NR1D1*, *PER1* or *PER3*

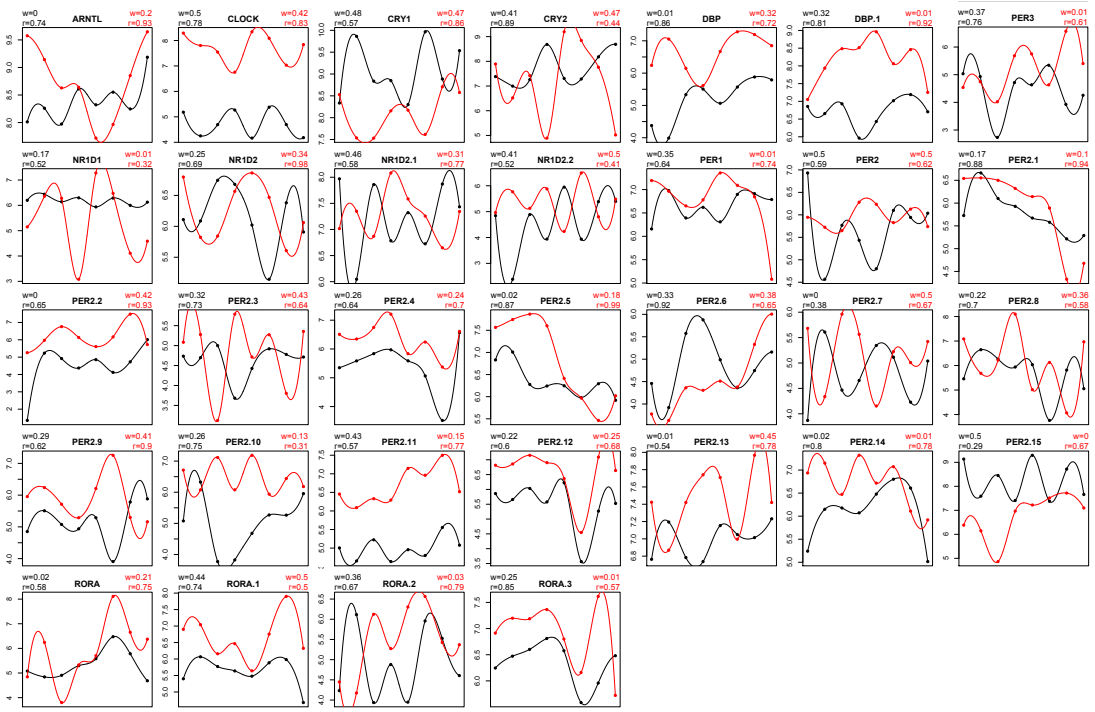


Figure 2.1 Core clock gene expression in MCF7 (black) and MCF10A (red) by microarray. Some genes have more than one probe, additional probes are marked as .1, .2 etc. Period and correlation for each probe is shown at the top of each profile (*"w"*for period and *"r"*for correlation). Y axis denotes normalized log ratioed gene expression. X axis represents time: 0, 4, 8, 12, 16, 20, 24, and 28 hrs delayed 8 hrs post-serum shock. Reproduced from Gutiérrez-Monreal et al. 2016, Supplementary Figure 5.

oscillated in a circadian fashion. The MCF7 cells did not demonstrate appropriate core circadian gene oscillation either, which would support the conclusion that MCF7 is indeed an arrhythmic cell line. It is possible that the unexpected lack of rhythmicity in core clock genes of the MCF10A cells assessed by microarray could be a due to limitations intrinsic to microarray experiments. Microarray experiments can suffer from cross-reactivity between cDNA and off-target probes, washing/imaging issues and normalisation problems (Jaksik et al., 2015). In particular the work presented by Gutiérrez-Monreal et al. (2016) lacks biological duplicates, which makes it hard to determine if the apparent lack of rhythmicity in core clock transcripts in MCF10As is the result of experimental noise, or a true representation of the transcriptome.

2.1.3 Cell circadian clock synchronisation methods

The circadian research performed in MCF7 and MCF10A cells described thus far used either dexamethasone or serum starvation/shock/starvation to synchronise the individual cellular clocks within the population to the same phase. When cells are cultured *in vitro*, synchronisation is necessary as the cells lack the endogenous signals that would have synchronised the cellular clocks within an organism. The use of dexamethasone to synchronise the molecular clocks of a population of cells to the same phase mirrors the afforementioned cooperation of the SCN and adrenal cortex *in vivo*, which results in rhythmic glucocorticoid levels in humans (Ishida et al., 2005; Oster et al., 2006). *In vivo* glucocorticoids bind to glucocorticoid receptors (GRs) and are imported to the nucleus where they bind to GR elements (GREs) allowing for regulation of target gene expression. *PER1* and *PER2* both have GREs in their promoter and intronic regions respectively that are believed to induce *PER* expression, whilst *NR1D1* and *RORα* have GREs that are believed to repress expression (Conway-Campbell et al., 2010; Pineda Torra et al., 2000; So et al., 2009; Stavreva et al., 2009; Surjit et al., 2011). It is also thought that CLOCK and CRYs are both capable of inhibiting the GR complex (Lamia et al., 2011; Nader et al., 2009). These interactions, and perhaps others, are thought to enable dexamethasone to act as a phase resetter of circadian clocks *in vitro* (Dickmeis et al., 2013).

Forskolin is a second commonly used chemical for circadian clock synchronisation *in vitro*. Forskolin is an adenylyl cyclase agonist, therefore it increases cellular cyclic AMP (cAMP) levels (Takeda et al., 1983). cAMP levels have previously been demonstrated to be rhythmic in SCN slices cultured *in*

vitro, and indeed rhythmic cAMP is required for sustenance of circadian clock gene oscillations (O'Neill et al., 2008; Prosser & Gillette, 1989). However, the ability of cAMP to synchronise rhythms in the SCN was found to be independent of the protein kinase A (PKA)-driven phosphorylation of cAMP response element binding protein (CREB) (O'Neill et al., 2008). This is interesting because the *PER1* and *PER2* promoters have cAMP response elements (CREs) that bind phosphorylated CREB in complexes with CREB binding protein (CREBBP), and CREB is phosphorylated in the SCN upon light exposure (Ding et al., 1997; Ginty et al., 1993; Hastings et al., 2007).

Synchronisation by serum deprivation and subsequent replacement is well-known for its use in cell cycle research; cells are forced into G_0 quiescence by serum deprivation, before being released into G_1 upon serum replacement (Davis et al., 2001). Serum deprivation also has a long history as a circadian clock synchroniser of cell populations (Balsalobre et al., 1998). Serum shock protocols are often varied, indeed Balsalobre et al. (1998) did not first deprive cell populations of serum prior to serum shock and subsequent removal, which many protocols do include (Gutiérrez-Monreal et al., 2016; Rossetti et al., 2012; Xiang et al., 2012). Serum-induced circadian clock synchronisation appears to be mediated by a combination of Ca^{2+}-dependent PKC-driven phosphorylation of CLOCK, and CREBBP binding to ARNTL, both of which appear to cause ARNTL/CLOCK dimers to bind to E boxes and drive expression of the *PER* genes (Y. Lee et al., 2010; Shim et al., 2007).

Not all methods of circadian clock synchronisation depend upon chemical manipulation of the cell culture media; for instance, circadian clocks can be entrained by environmental cues such as temperature changes. The link between temperature and circadian rhythms is complex, and not fully understood. Circadian rhythms are by definition 'temperature compensated',

which means that whilst most biochemical processes speed up in response to an increase in temperature, the period of a circadian clock remains approximately 24 hrs regardless of the environmental temperature (Kidd et al., 2015). Despite this, the phase of circadian processes remains temperature sensitive. For example, *in vitro Drosophila melanogaster* organ cultures have been synchronised to 12 hr 25°C / 12 hr 17°C temperature cycles (Glaser & Stanewsky, 2005), whilst *in vitro* organotypic cultures of rat SCN have been demonstrated to entrain to temperature cycles of 12 hr 35.3°C / 12 hr 36.8°C (Herzog & Huckfeldt, 2003). In terms of 2D cell culture, the innate daily peritoneal temperature oscillations of mice have been demonstrated to entrain circadian gene expression in rat fibroblasts (Brown et al., 2002). Meanwhile NIH3T3 fibroblasts have been demonstrated to entrain to temperature cycles of as little as 1°C oscillation (Saini et al., 2012). In fact the core body temperature of humans oscillates by approximately ~1°C every day (e.g. ~36.4.-37.1 (Baschieri et al., 2020), ~36.4-37.4°C (Lericollais et al., 2013)). The oscillation of body temperature in humans is dependent on many variables including sex and age (Baker et al., 2001).

2.1.4 Chapter aims

This chapter is dedicated to trying to better understand the circadian behaviour of MCF10A and MCF7 cell lines, that we might gain a clearer understanding of their use in and to circadian cancer research. A number of different experimental methods will be employed to achieve this.

qPCR is an excellent technique for relative quantification of transcript expression in a population of cells and is a stalwart of molecular biology. Since the transcripts of the core clock genes should oscillate in a circadian manner,

qPCR will be employed in this chapter to assess the functionality of the circadian clock in MCF cell populations. Previous work on the circadian gene expression of MCF cell lines has also employed qPCR as discussed in chapter 2.1.2, e.g. Rossetti et al. (2012).

The use of bioluminescent reporters to assess oscillatory clock promoter activity in MCF cells lines will also be presented. Frequently such reporters are delivered by lentiviral vectors in order to create stably transduced reporter cell lines (Ramanathan et al., 2012). *In vitro* circadian experiments often occur over a period of many days, therefore fluorescent reporter constructs are of limited use due to photo-toxicity. In contrast, bioluminescent reporter constructs allow for population tracking of circadian promoter activity via luciferase-catalysed photon release.

Understanding the circadian system in the wider cell biology context can be aided by the use of ‘‘omics’ technologies. RNA sequencing (RNA-seq) and Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) will be used to comprehensively assess the rhythmic biology of MCF10A and MCF7 cells in greater detail than has been achieved thus far. Whilst RNA-seq profiles the transcriptome of a population of cells, ATAC-seq produces reads in regions of chromatin that are accessible to the cutting action of Tn5 transposase (Buenrostro et al., 2013) (Figure 2.2). ATAC-seq data is somewhat indicative of gene transcription – where DNA is tightly wound around histones, then both Tn5 transposase and the transcriptional machinery are excluded, and genes cannot be transcribed. However, an accessible gene is not *always* transcribed as transcription is also dependent upon other variables, such as the accessibility of distal enhancer/repressor elements, and the presence/absence of transcription factors.

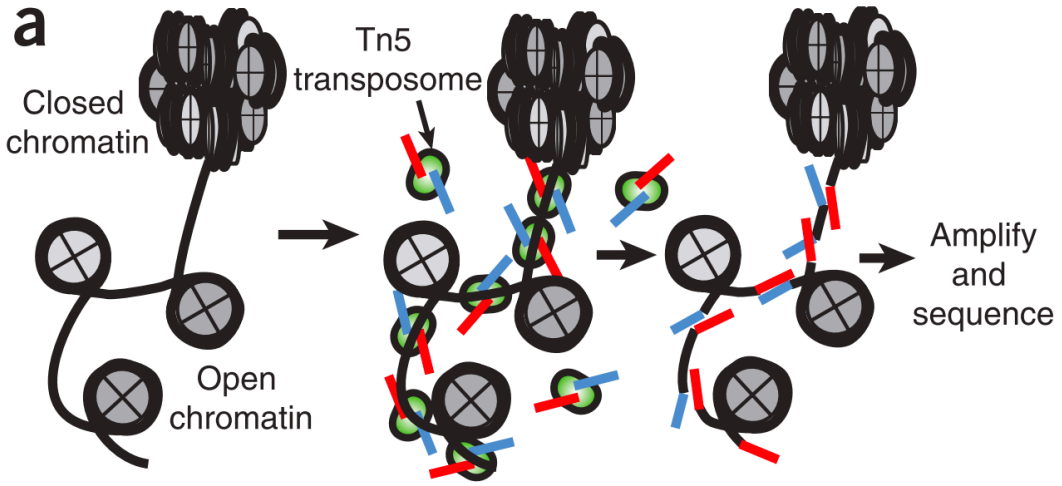


Figure 2.2 A schematic of the ATAC-seq procedure reproduced from Buenrostro et al. (2013). Tn5 transposase is able to fragment DNA whilst tagging it with the adapters that it carries, in a process called tagmentation. Tagmented DNA is amplified by PCR which introduces sample specific Illumina indexes for sequencing.

The circadian clock has long been understood to contribute to chromatin remodelling; the circadian protein CLOCK for instance has been identified as a histone acetyltransferase, in addition to its ARNTL-binding role (Doi et al., 2006). Meanwhile, chromatin immunoprecipitation (ChIP)-seq experiments in mouse liver have identified circadian rhythmicity in histone modifications at transcriptional start sites (TSSs), including changes to H3K4me3, H3K9ac, and H3K27ac marks, which are presumed to result in circadian changes to chromatin accessibility (Koike et al., 2012). Information about rhythmic transcript levels can be provided by RNA-seq, but the circadian activity of the final protein product is regulated by other processes under circadian regulation including splicing, translation and protein post-translational modifications and localisation (Jouffe et al., 2013; McGlincy et al., 2012; Okamoto-Uchida et al., 2019). As changes to chromatin structure are driven by the integration of the aforementioned cellular processes, the same can be expected of circadian changes to chromatin structure (Garcia-Bassets & Wang, 2012). Circadian changes to chromatin structure have even been demonstrated by 4C sequencing

to contribute to tissue type specific expression of downstream rhythmic genes (Yeung et al., 2018). Therefore, ATAC-seq has the potential to enable a deeper understanding of the degree of circadian clock function beyond the transcriptome. ATAC-seq is a relatively young technology, and as of yet it has only been applied in the circadian field to non-mammalian models, or to murine cortex samples in a sleep deprivation context (Hor et al., 2019; Lugena et al., 2019; Weizman et al., 2019).

2.2 Materials and Methods

2.2.1 MCF10A and MCF7 cell culture and media

MCF10A cells were obtained from the American Type Culture Collection (ATCC, CRL-10317) and cultured in humidified incubators at 37°C, 5% CO_2 in complete MCF10A media. Complete MCF10A media was prepared following the standard protocol developed from the original source of the MCF10A cell line Soule et al. (1990) (Arnandis & Godinho, 2015). Complete MCF10A media consisted of the following: Dulbecco's Modified Eagle's Medium/ Hams F-12 50:50 mix (DMEM/F-12) (Corning, 10-092-CV); 5% horse serum (Sigma-Aldrich, H1138); 20 ng/ml human epidermal growth factor (Sigma-Aldrich, E9644); 0.5 µg/ml hydrocortisone; 100 ng/ml cholera toxin (Sigma-Aldrich, C8052); 10 µg/ml insulin (Sigma-Aldrich, 91077C); 20 U/ml penicillin (Sigma, P3032); 20 µg/ml streptomycin (Sigma, S9137) and 300 µM L-glutamine (BDH chemicals, 371077J).

MCF7 cells were obtained from Prof Peter Sadler's lab courtesy of Dr Hannah Bridgewater. Complete MCF7 media consisted of DMEM supplemented with UltraGlutamine (Lonza, LZBE12-604FU1), 10% foetal bovine serum (FBS) (Sigma, F9665-500ML), x20 U/ml penicillin (Sigma, P3032) and 20 µg/ml streptomycin (Sigma, S9137).

All cells were maintained at fewer than 25 passages. For sub-culturing, cells at 90% confluency were detached from tissue culture plates with 0.25% trypsin (VWR, 0458-50) / 1mM EDTA (Fisher Scientific, 10213570) and re-seeded at a 1:4 dilution. Cryopreservation media for storage in liquid nitrogen consisted

of 90% complete media supplemented with 10% dimethyl sulfoxide (DMSO) (Corning, 25-950-CQC).

Imaging media consisted of complete media without phenol red and supplemented with an additional 10 mM HEPES buffer (Gibco, 15630106) in order to provide extra pH buffering capacity whilst cells were incubated for many days without media replacement throughout circadian experiments. 100µM potassium D-luciferin (Promega, E1605) was also added in order to report the activity of D-luciferase reporter constructs.

2.2.2 Generation of stable P(*Per2*)-d*luc* and P(*Arntl*)-d*luc* reporter MCF cell lines

Second-generation lentivirus backbone transfer plasmids pLV7-*Bsd*-P(*Arntl*)-d*luc* and pLV7-*Bsd*-P(*Per2*)-d*luc* were obtained as a kind gift from Dr Andrew Liu's laboratory at the University of Memphis. These lentivirus transfer plasmids are termed 'pLV7' as they were originally generated by modifying pLenti6/R4R2/V5-DEST (available from Invitrogen) to include woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) sequences to promote the expression of the luciferase construct and increase the bioluminescent signal in transduced cells (Ramanathan et al., 2012). The '*Bsd*' refers to the blasticidin resistance gene which acts as a selectable marker in mammalian cells. The 'P(*Per2*)' (or 'P(*Arntl*)') refers to the promoter region of the murine *Per2* gene (or murine *Arntl* gene). Lastly, 'd*luc*' refers to the firefly luciferase gene, which has been modified to include a C-terminal PEST sequence to allow rapid degradation of luciferase, for accurate reporting of clock gene promoter activity (Ramanathan et al., 2012). All plasmid propagation was performed using OneShot Stbl3 *E. coli* (ThermoFisher,

C737303) and the NucleoBond® Xtra Midi plasmid purification kit (Clonentech, 740410.100). The lentiviral packaging plasmid pCMVR8.74 (Addgene, 22036), envelope plasmid pMD2.G (Addgene, 12259) and the clock gene promoter containing backbone transfer plasmid were transfected into HEK293FT cells (gifted by the McAinsh lab, ThermoFisher, R70007) using TransIT-Lenti transfection reagent (Mirus, MIR6600). After 48 hrs the resultant lentivirus-containing media was harvested, sterile filtered and stored at -80°C. Lentiviral transduction of MCF cell lines was achieved by 6 hrs of incubation with 1 ml of lentivirus-containing media and 8 µg/ml polybrene (Sigma, TR-1003-G), per 100 mm diameter tissue culture dish. Blasticidin (Corning, 30-100-RB) was used at 10 µg/ml to select for MCF7 or MCF10A cells which were P(*Arntl*)-luc or P(*Per2*)-luc positive. From this point forwards, the stably transduced cell lines will be referred to in the format "MCF10A *Arntl*-luc".

2.2.3 Synchronisation of MCF cells

2.2.3.1 Dexamethasone synchronisation

MCF cells were seeded with 300,000 cells in 35 mm tissue culture dishes in 3 ml complete media. The following day the cell culture media was replaced with 3 ml complete media supplemented with 100 nM dexamethasone (Sigma, D4902-100MG). Cells were incubated with dexamethasone at 37°C for 20 minutes, after which the cells were washed twice with phospho-buffered saline (PBS) and supplemented with imaging media.

2.2.3.2 Forskolin synchronisation

MCF cells were seeded with 300,000 cells in 35 mm tissue culture dishes in 3 ml complete media. The following day the complete media was replaced with

imaging media supplemented with 10 µM forskolin (APExBIO, B1421) (Balsalobre et al., 2000).

2.2.3.3 Serum synchronisation

MCF cells were seeded with 300,000 cells in 35 mm tissue culture dishes in 3 ml complete media. The following day the complete media was replaced with 3 ml serum-free media. After 22 hrs in serum-deprivation conditions, the medium was exchanged for 3 ml of 50% horse serum / 50% serum-free media. After 2 hrs in serum-rich conditions the media was exchanged for imaging media.

2.2.3.4 Temperature synchronisation

MCF cells were seeded in 35 mm tissue culture dishes in 3 ml complete media. The following day the complete media was exchanged for imaging media, and the cells were entrained with temperature cycles in a Memmert INCO incubator according to Table 2.2 using Memmert's CELSIUS software. Incubator temperature was monitored throughout the experiment. Table 2.2 details the cell seeding densities used.

2.2.4 Bioluminescence assay for clock gene reporter cell lines

Synchronised cells in 35 mm tissue culture dishes were supplemented with imaging media and placed in an ActiMetrics LumiCycle 32 device for continuous monitoring of luciferase reporter activity. Population level bioluminescence (raw counts/sec) was recorded at 10 min intervals for each dish. The resultant raw bioluminescence traces were detrended by calculating the percentage deviation of the raw data at each timepoint from the 24 hr running mean, an example of data normalisation is shown in Figure 2.3. This

Table 2.2 Summary of entrainment protocols for MCF10A and MCF7 clock gene promoter luciferase reporter cell lines

Entrainment	**Protocol**	**Seeding Density** (cells/35 mm dish)
Dexamethasone	100 nM dexamethasone for 20 min, wash out	300,000
Forskolin	10 µM forskolin	300,000
Serum	22 hrs serum-free, 2 hrs 50% serum, return to imaging media for bioluminescence recording	300,000
12 hr 32°C / 12 hr 37°C	1 hr 37°C / 12 hr 32°C / 12 hr 37°C / 12 hr 32°C / constant 37°C	300,000
12 hr 32°C / 12 hr 37°C without HEPES	1 hr 37°C / 12 hr 32°C / 12 hr 37°C / 12 hr 32°C / constant 37°C	300,000
12 hr 36°C / 12 hr 37°C	1 hr 37°C / (12 hr 36°C / 12 hr 37°C)× 4 / 12 hr 36°C / constant 37°C	250,000
12 hr 33°C / 12 hr 38°C	1 hr 38°C / 12 hr 33°C / 12 hr 38°C / 12 hr 33°C / constant 38°C	300,000
10.5 hr 32°C / 10.5 hr 37°C	1 hr 37°C / (10.5 hr 32°C / 10.5 hr 37°C) × 4 / 10.5 hr 32°C / constant 37°C	200,000
13.5 hr 32°C / 13.5 hr 37°C	1 hr 37°C / (13.5 hr 32°C / 13.5 hr 37°C) × 3 / 13.5 hr 32°C / constant 37°C	200,000

accounted for changes in the cell population size with time and for different photomultiplier tube (PMT) sensitivities.

Extra sum-of-squares F tests were implemented using GraphPad Prism v8 in order to identify the best fitting model to each of the detrended bioluminescent traces by iterative hypothesis testing. I.e., sine fits were accepted/rejected in favour of linear fits at a significance threshold of 0.05. Then damped sine fits were accepted/rejected in favour of sine fits at a significance threshold of 0.05. All sine fits were constrained to periods of no less than 18 hrs, and were fitted from 24 hrs post-cessation of synchronisation, as indicated by Figure 2.4. Periods were calculated from GraphPad Prism's best (damped) sine fits. Figure 2.4 demonstrates that amplitudes of sine fits were calculated by inspection as the absolute difference between the maxima and minima of the first oscillation.

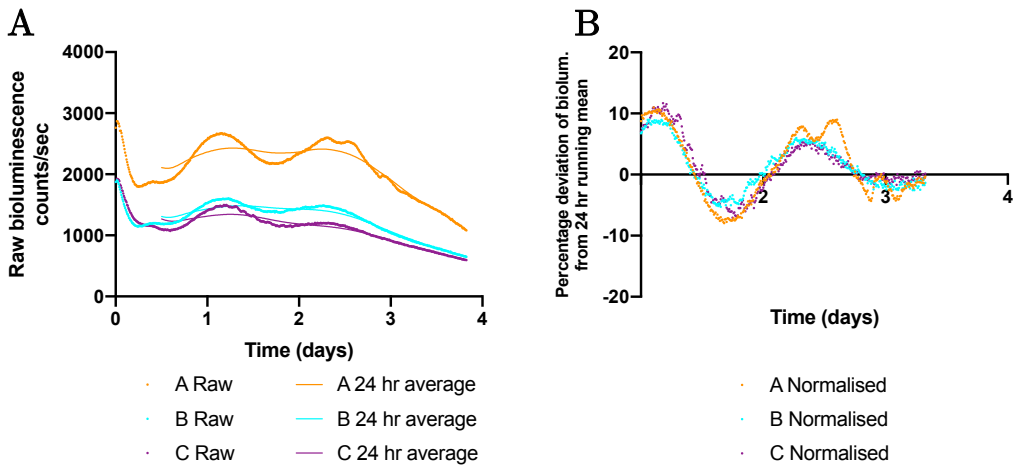


Figure 2.3 Procedure for 'detrending' raw bioluminescence data. A) Raw bioluminescent traces from 3 biological replicate experiments (in this case MCF10A *Per2*-luc cells). 24 hour running means are also plotted. B) Percentage difference between raw data and 24 hr running mean gives 'detrended' data.

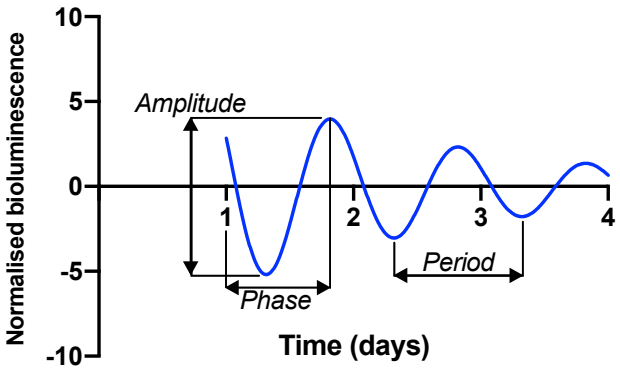


Figure 2.4 Definitions of phase, amplitude and period, as derived from cosine functions fitted to raw data.

Meanwhile, phase measurements were taken as the acrophase of the first peak in the fitted sine function. Welch's unpaired t-tests were applied to compare the periods of fitted cosine functions between conditions. Welch's method does not assume equal standard deviations between the test conditions. No correction for multiple t-testing was applied, therefore each p value of e.g., 0.05 represents an e.g., 5% chance that the null hypothesis (i.e., periods are not significantly different between condition A and condition B) was mistakenly rejected.

2.2.5 RNA extraction from MCF cells

At the time of sampling, dishes of cells were removed from the incubator and the media was removed and replaced with 1 ml of TRIzol (Fisher Scientific, 12034977). The TRIzol-cell mixture was dislodged from the dishes using a cell scraper, transferred to a 1.5 ml Eppendorf and centrifuged for 2 min at 12,000 x g, 4 °C to pellet any cell debris. The supernatant was transferred to a fresh Eppendorf containing 200 µl chloroform (Fisher Chemical, C/4960/PB17) and centrifuged at 12,000 x g, 4 °C for a further 15 min to establish RNA/DNA phase separation. The RNA in the upper aqueous phase was precipitated with 450 µl isopropanol (Fisher Scientific, 11436471) and pelleted by centrifugation at 12,000 x g, 4 °C for 10 min. The supernatant was removed, and the RNA pellet washed twice with 75% ethanol (VWR, 20821.330). The ethanol was removed, and the RNA pellets were allowed to air dry prior to resuspension in RNase/DNase–free water (Invitrogen, 11538646). The Turbo DNA-free kit (Invitrogen, AM1907) was used to remove any trace amounts of DNA from the RNA, and the quality of the RNA was checked by gel electrophoresis (1% w/v agarose gel in Tris/borate/EDTA buffer).

2.2.6 qPCR of MCF cells

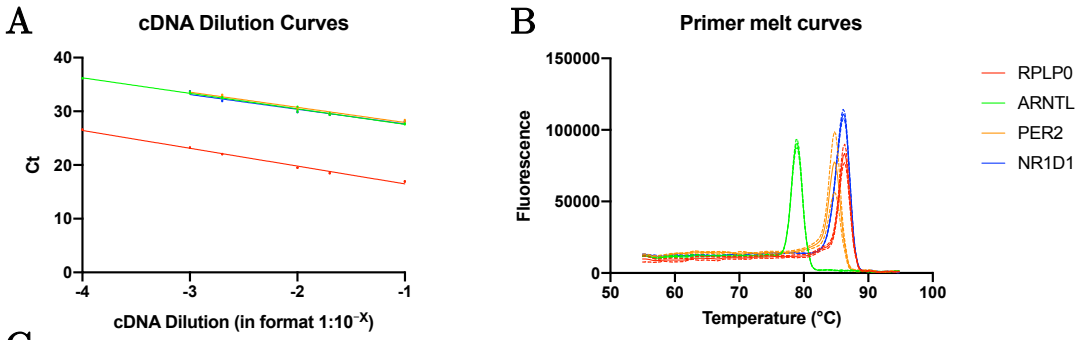
DNase-treated RNA was converted to cDNA using the superscript II reverse transcriptase kit (Invitrogen, 18064022). *RPLP0* was selected as a house-keeping gene as it is known to remain stable over circadian time (Hadadi et al., 2018). The primers used for *RPLP0*, *ARNTL*, *PER2*, and *NR1D1* are listed below in Table 2.3. For each cell line, cells were synchronised to a 12 hr/12 hr

Table 2.3 Forward and reverse primer sequences for human housekeeping and clock genes.

Gene	Forward Primer Sequence	Reverse Primer Sequence
RPLP0	5' AATCCCTGACGCACCGCCGTGATG 3'	5' TGGGTTGTTTTCCAGGTGCCCTCG 3'
ARNTL	5' AAGGATGGCTGTTCAGCACATGA 3'	5' CAAAAATCCATCTGCTGCCCTG 3'
PER2	5' GCAGGTGAAAGCCAATGAAG 3'	5' CACCGCAAACATATCGGCAT 3'
NR1D1	5' AGAGCACCAGCAACATCACCAAGC 3'	5' TTCTTGAAGCGACATTGCTGGCAG 3'

36°C/37°C temperature cycle for 72 hrs before the first samples were taken. Triplicate samples were collected in TRIzol every 4 hrs and snap-frozen in liquid nitrogen, until 6 samples had been obtained under continued temperature entrainment conditions.

RNA was extracted, DNase-treated and converted to cDNA as described in chapters 2.2.5 and 2.2.6. For qPCR experiments samples were assayed in 384 well plate format using an Applied Biosystems QuantStudio 5. Each biological sample was measured in triplicate to account for technical error. Each 10 µl qPCR reaction comprised of 5 µl PowerUp SYBR Green Master Mix (Applied Biosystems, A25741), 1 µl of pre-mixed forward and reverse primer at 0.5 µM each, 1 µl of cDNA (diluted 1:10 from cDNA conversion) and 3 µl DNase/RNase–free water. The temperature profile applied during the experiment was as follows: 2 min 50°C, 2 min 95°C followed by 40 cycles of 15 sec 95°C, 30 sec 55°C, 30 sec 72°C. Primer efficiency was determined by qPCR of serially diluted pooled cDNA to a) confirm that the relationship between the amount of DNA and Ct remained logarithmic, and b) determine the efficiency of amplification (Figure 2.5 A&C). Melt curves of primer pairs (Figure 2.5 B) indicated that even where primer efficiency was >100%, a single PCR product was amplified, which implies that perhaps high primer efficiencies were caused by inhibitors of the qPCR reaction present in higher concentrations of cDNA. The Pfaffl method of relative quantification of DNA accounts for variability in primer efficiencies and was used to calculate the $\log_2$



C

Gene	Slope	Efficiency	E
RPLP0	-3.317	100.21%	2.00
ARNTL	-2.890	121.83%	2.22
PER2	-2.821	126.20%	2.26
NR1D1	-2.737	131.93%	2.32

Figure 2.5 A) cDNA dilution curves for the determination of primer pair amplification efficiency across different concentrations of cDNA. B) Melt curves for each primer pair (line = mean, dotted line = S.D., n=3) C) Tabulation of cDNA dilution curve slopes and corresponding efficiencies of primers.

fold change of each gene at each timepoint relative to the mean of all timepoints (Pfaffl, 2006):

$$\text{Log}_2 \text{ fold change} = \log_2 \left(\frac{{E_{goi}}^{Ct_{b,goi} - Ct_{s,goi}}}{{E_{ref}}^{Ct_{b,ref} - Ct_{s,ref}}} \right)$$

$$where\ goi = gene\ of\ interest;\ ref = reference\ gene;\ s = test\ sample, i.e.\ single\ timepoint;$$
$$b = baseline\ calibrator, i.e.\ mean\ of\ all\ timepoints;\ E = efficiency$$

For each time-course, extra sum-of-squares F tests were implemented using GraphPad Prism v8 in order to determine whether a sine fit constrained to a period of 24 hr was a better fit for the data than a linear fit at a significance threshold of 0.05.

2.2.7 Sample harvesting for RNA-seq and ATAC-seq experiments

Just as for qPCR, MCF cells were entrained to a 12 hr/12 hr 36°C/37°C temperature cycle for 72 hrs before the first samples were taken. For RNA-seq experiments, triplicate samples were collected in TRIzol every 4 hrs and snap-

frozen in liquid nitrogen, until 6 samples had been collected (one sample = 1x 100 mm diameter tissue culture dish). For ATAC-seq experiments cells were washed with PBS and detached from the dish with Trypsin/EDTA. Detached cells were pelleted by centrifugation at 300 x g for 3 min in order to wash the cells with PBS. Centrifugation was repeated in order to remove the PBS, and cell pellets were snap frozen in liquid nitrogen.

2.2.8 RNA-seq experimental procedure

All library preparation for RNA-seq beyond RNA extraction was carried out by the University of Warwick's Genomics Facility. RNA quality was checked with an Agilent Bioanalyzer and concentration was determined by Invitrogen Qubit. Of the triplicate samples harvested, the duplicates with the best RNA integrity numbers (RINs) were selected for RNA-seq; all RINs were >9. Libraries were prepared using the unstranded TruSeq RNA Library Prep Kit (Illumina, RS-122-2001). Samples were sequenced (single–end) using Illumina's NextSeq 500/550 High Output Kit v2.5 (75 Cycles) (Illumina, 20024906) on an Illumina NextSeq 500 platform. Library sizes ranged from 12 to 18 million reads per sample.

2.2.9 ATAC-seq experimental procedure

The protocol described below for nuclear isolation and ATAC library preparation follows the Kaestner laboratory's protocol (accessible at https://www.med.upenn.edu/kaestnerlab/protocols.html, accessed 28/01/21). It was developed from the original ATAC-seq protocols published by Buenrostro et al. and the omni-ATAC-seq protocol developed by Ryan-Corces

et al. (2017) (Buenrostro et al., 2013, 2016; Ryan-Corces et al., 2017). The following buffers were prepared in advance:

Resuspension Buffer:

1 ml	1M Tris-HCl buffer (pH 7.5) (Sigma-Aldrich, T5941)
200 µl	5M NaCl (Fisher Scientific, S671-500)
300 µl	1M $MgCl_2$ (Sigma-Aldrich, M2670)
98.5 ml	RNase/DNase–free water

Lysis Buffer:

4.85 ml	Resuspension buffer
50 µl	10% IGEPAL CA-630 (Sigma-Aldrich, I8896)
50 µl	Tween-20 (Sigma-Aldrich, P9416)
1% v/v	1% Digitonin (final 0.01% v/v, Promega, G9441)

Wash Buffer:

1 ml	10% Tween-20
99 ml	Resuspension Buffer

100,000 cells were washed with 5 ml ice cold PBS and centrifuged at 500 G for 5 min to pellet. Supernatant was discarded and cells were pipetted slowly up/down x3 with 100 µl of cold lysis buffer. Cells were incubated on ice for 3 min. 2 ml of wash buffer was added and tubes were mixed by gentle inversion x3. Nuclei were pelleted by centrifugation at 500 G for 10 min at 4°C. Transposition mix was prepared as follows (scaled as appropriate): 25 µl 2x Tagment DNA Buffer (Illumina, 20034197), 16.5 µl PBS, 0.5 µl 10% Tween-20, 0.5 µl 1% Digitonin, 2.5 µl Tn5 transposase (Tagment DNA Enzyme 1, Illumina, 20034197), 5 µl nuclease-free water. 50 µl was added to each nuclei pellet, and samples were incubated at 37°C for 30 min to allow for transpose

to cut accessible DNA. The DNA was isolated using the MinElute Reaction Cleanup Kit (Qiagen, 28204) and stored at -20°C until the following day.

For library amplification by PCR 10 µl of purified transposed DNA was mixed with 10 µl nuclease-free water, 2.5 µl 25 µM forward primer, 2.5 µl 25 µM reverse primer, and 25 µl NEBNext High-Fidelity 2X PCR Master Mix (New England BioLabs, M0541S). For primers refer to Table 2.4. Library amplification was started using a Hain Q-Cycler 96+ and the following temperature profile: 72°C 5 min, 98°C 30 sec, (98°C 10 sec, 63°C 30 sec, 72°C 1 min) x5 cycles, hold at 4°C. For quantification of library amplification by qPCR, 5 µl of partially amplified library from each sample was mixed with 3.85 µl nuclease-free water, 0.5 µl 25 µM forward primer, 0.5 µl 25 µM reverse primer, 0.15 µl 100X SYBR Green I (Invitrogen, S7563), and 5 µl NEBNext High-Fidelity 2X PCR Master Mix. The qPCR reaction was run on an Agilent

Table 2.4 Illumina primers for ATAC-seq library preparation

Forward Primer Sequence	
AD1_noMX	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Reverse Primer Sequences	
Ad2.1_TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2_CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3_AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4_TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5_GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6_TAGGCATG	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7_CTCTCTAC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8_CAGAGAGG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT
Ad2.9_GCTACGCT	CAAGCAGAAGACGGCATACGAGATAGCGTAGCGTCTCGTGGGCTCGGAGATGT
Ad2.10_CGAGGCTG	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGGAGATGT
Ad2.11_AAGAGGCA	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGGAGATGT
Ad2.12_GTAGAGGA	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGGAGATGT
Ad2.13_GTCGTGAT	CAAGCAGAAGACGGCATACGAGATATCACGACGTCTCGTGGGCTCGGAGATGT
Ad2.14_ACCACTGT	CAAGCAGAAGACGGCATACGAGATACAGTGGTGTCTCGTGGGCTCGGAGATGT
Ad2.15_TGGATCTG	CAAGCAGAAGACGGCATACGAGATCAGATCCAGTCTCGTGGGCTCGGAGATGT
Ad2.16_CCGTTTGT	CAAGCAGAAGACGGCATACGAGATACAAACGGGTCTCGTGGGCTCGGAGATGT
Ad2.17_TGCTGGGT	CAAGCAGAAGACGGCATACGAGATACCCAGCAGTCTCGTGGGCTCGGAGATGT
Ad2.18_GAGGGGTT	CAAGCAGAAGACGGCATACGAGATAACCCCTCGTCTCGTGGGCTCGGAGATGT
Ad2.19_AGGTTGGG	CAAGCAGAAGACGGCATACGAGATCCCAACCTGTCTCGTGGGCTCGGAGATGT
Ad2.20_GTGTGGTG	CAAGCAGAAGACGGCATACGAGATCACCACACGTCTCGTGGGCTCGGAGATGT
Ad2.21_TGGGTTTC	CAAGCAGAAGACGGCATACGAGATGAAACCCAGTCTCGTGGGCTCGGAGATGT
Ad2.22_TGGTCACA	CAAGCAGAAGACGGCATACGAGATTGTGACCAGTCTCGTGGGCTCGGAGATGT
Ad2.23_TTGACCCT	CAAGCAGAAGACGGCATACGAGATAGGGTCAAGTCTCGTGGGCTCGGAGATGT
Ad2.24_CCACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTGGGTCTCGTGGGCTCGGAGATGT

Stratagene Mx3005P machine, using the following temperature profile: 98°C 30 sec, (98°C 10 sec, 63°C 30 sec, 72°C 1 min) x20 cycles. Based on R vs. cycle number plots, a further 6 cycles of PCR were applied to the partially amplified libraries (bringing the total number of PCR cycles required for library amplification to 11). Libraries were stored at -20°C until the following day.

For library purification, Agencourt AMPure XP beads (Beckman Coulter, 10136224) were warmed to room temperature and vortexed. 81 µl of AMPure XP beads were mixed thoroughly with 45 µl of amplified library and incubated at room temperature for 10 min. Samples were placed in a magnetic rack in order to retain the AMPure XP beads whilst the supernatant was removed. The beads were washed with 200 µl of 80% ethanol. Ethanol was removed and the remainder was allowed to evaporate. The AMPure XP beads were re-suspended in 20 µl of nuclease-free water. 17 µl of sample was removed from the beads with the aid of the magnetic rack and stored at -20°C. 1 in 4 dilutions of each sample were prepared in order to run samples on the Agilent High Sensitivity DNA Bioanalyzer and Invitrogen Qubit. Concentrations of libraries were calculated from the Bioanalyzer traces and libraries were pooled and sequenced (paired–end, 35+40) using Illumina's NextSeq 500/550 High Output Kit v2.5 (75 Cycles) (Illumina, 20024906) on an Illumina NextSeq 500 platform. The first sequencing run was over-clustered, therefore two sequencing runs were required to accrue 680 million read pairs over the 24 samples.

2.2.10 RNA-seq data analysis pipeline

All file processing encompassing sequence alignment and transcript counting was achieved using a CentOS Linux 7 kernel on the University of Warwick's CyVerse UK servers (a 'big data' project funded by the Biotechnology and

Biological Sciences Research Council). Illumina .bcl.bgzf files were converted to FASTQ files using Bcl2Fastqv2.20.0.422 (Illumina), Illumina adapter sequences were removed at the same time. Single end FASTQ files were aligned to the human genome (Genome Reference Consortium (GRC), release h38.84) and converted to SAM files using HISAT2 v2.2.0 (Kim et al., 2019). SAM files were compressed to BAM using Samtools v1.10 (Heng Li et al., 2009). Transcript read counts were determined from the BAM files and the human transcriptome (GRCh38.84 .gtf file) using LiBiNorm v2.4, an in-house software package, in HTSeq-count mode with '--stranded=no' (Anders et al., 2015; Dyer et al., 2019). Raw read counts were concatenated for all samples and exported as one text file for all subsequent analysis on a Mac OS.

Raw count normalisation was carried out in R Studio v1.3.1093 (with the underlying R installation v4.0.3). The edgeR package was used to normalise the raw counts to $\log_2$ counts per million (logCPM) or $\log_2$ trimmed mean of M-values (logTMM) (M. D. Robinson et al., 2010).

2.2.11 RNA-seq rhythmicity analysis

Rhythmicity of logCPM normalised transcripts was analysed using the meta2d function of the MetaCycle package in R (Wu et al., 2016). Period length was set to between 23 and 25 hrs. Rhythmicity was evaluated at a Bonferroni-Hochberg (B-H) adjusted p-value of 0.2, which means that ~20% of the positive results are likely to be false positives. A high false discovery rate (FDR) was determined to be necessary by inspection of the data, in order to avoid losing too many true positives. Such an approach is often necessary in the case of relatively low sampling resolution (Hughes et al., 2010), i.e. every 4 hrs as is

the case for the RNA-seq data generated here, and the microarray data of Gutiérrez-Monreal et al. (2016).

2.2.12 ATAC-seq analysis pipeline

As with RNA-seq analysis, Illumina .bcl.bgzf files were converted to FASTQ files using Bcl2Fastq v2.20.0.422 (Illumina), Illumina adapter sequences were removed at the same time. Low quality reads were trimmed using Trimmomatic v0.39, with the TRAILING argument set to '3', and SLIDINGWINDOW to '4:15' (Bolger et al., 2014). FASTQ files were filtered to remove reads that aligned to the GRCh38 mitochondrial genome using Bowtie2 v2.4.1 (Langmead & Salzberg, 2012). These reads were removed as ATAC-seq aims to assess the accessibility of nuclear chromosomes only. Remaining reads were aligned to the GRCh38 chromosomal genome and converted to SAM format. SAM files were compressed to BAM using Samtools v1.10 (Heng Li et al., 2009). BAM files from the same samples but from different sequencing runs were merged using Samtools, resulting in 680 million aligned read pairs across all samples.

The number of reads in samples was down-sampled at random in order to match the sample with the smallest library size. Samples were merged in order to allow MACS2 to detect peak regions (Gaspar, 2018). MACS2 was used in the default mode ($q<0.05$) to identify a preliminary peakset. Those peaks were subsequently filtered further at a generous threshold of $q<10^{-3}$ for finding background regions or a more stringent threshold of $q<10^{-4}$ for finding peak regions. This two-step peak-finding approach ($q=0.05$ followed by a smaller q-value) ensured that peaks were highly significant, but not overly narrow.

Gene-wise views of down-sampled reads in samples were visualised in Integrative Genomics Viewer (J. T. Robinson et al., 2011). Heatmaps of transcriptional start sites (TSSs) were plotted by first using bamCoverage (deepTools) to convert down-sampled BAM files to binned bigWig files (bin width = 50 bp) (Ramírez et al., 2014), and then using the deepTools plotHeatmap function on the open source Galaxy platform (https://usegalaxy.eu). TSS enrichment scores (TSSE scores) were calculated in Linux from down-sampled BAM files using TSS regions for hg38 taken from Ensembl Biomart, according to the formula described by the ENCODE consortium.

Background regions were identified in R according to the strategy described by Blythe & Wieschaus (2016). A series of 25,000 regions that were more than 1000 bp away from any detected peaks (determined at the less stringent threshold of $q<10^{-3}$) were selected at random and any overlapping regions were removed (Blythe & Wieschaus, 2016). Reads in peaks and reads in background regions were calculated for each down-sampled sample using a Linux installation of featureCounts (Liao et al., 2014). For rhythmicity analysis down-sampled read counts in peaks were normalised in edgeR to logCPM with library size set to the number of down-sampled reads in each sample prior to peak calling. This is an important distinction, if counts in peaks are truly higher in one sample relative to another then normalising to total counts in peaks (the default edgeR setting) would eliminate such a true difference. Chapter 2.3.6 explains the normalisation of ATAC-seq data in more detail. Rhythmicity analysis in MetaCycle was implemented as described in chapter 2.2.11.

2.3 Results and Discussion

2.3.1 The effect of circadian clock synchronisation method on apparent clock function in MCF cells

Dishes of MCF10A and MCF7 cells stably transduced with lentiviral clock gene promoter luciferase reporters were subjected to multiple circadian clock synchronisation/entrainment conditions to try and establish if the differences observed in the published literature (see Table 2.1) were the result of synchronisation method.

Dexamethasone induced synchronous oscillatory expression of *Per2*-luc in MCF10A cell populations, albeit with a rather long period of 30.4 hrs (S.D. ± 1.1 hrs) (Figure 2.6). In MCF7s consistency was poorer between replicates, the most rhythmic replicate (C) had a *Per2*-luc period of 24.3 hrs, whilst the least rhythmic (A) did not fit a damped sine wave of period >18 hrs well (although a damped sine wave was still considered to be a better fit than a straight line or standard sine wave by hypothesis testing). Conversely to MCF10As, dexamethasone synchronisation induced better oscillatory expression of *Arntl*-luc than *Per2*-luc in MCF7s. *Arntl*-luc expression oscillated with a period of 23.1 hrs (± 0.9 hrs) in MCF7s, whereas, *Arntl*-luc activity did not oscillate in MCF10As. MCF7 *Arntl* and *Per2*-luc activity oscillated in opposite phase relationships, which is to be expected (Takahashi, 2017).

Figure 2.7 displays the *Arntl* and *Per2*-luc activity in MCF10As and MCF7s resulting from forskolin synchronisation. The findings were broadly similar to the rhythmic expression induced by dexamethasone synchronisation. The MCF10As displayed oscillatory expression of *Per2*-luc with a long period of

29.5 hrs (± 4.3 hrs), whilst MCF7s displayed *Per2*-luc oscillations with a short period of 21.6 hrs (± 1.5 hrs). As was the case after dexamethasone synchronisation, *Arntl*-luc expression was not oscillatory in MCF10As after forskolin synchronisation, meanwhile in MCF7s *Arntl*-luc oscillated with a period of 25.9 hrs (± 4.7 hrs).

Figure 2.8 shows the *Arntl* and *Per2*-luc oscillatory expression in MCF10As and MCF7s in response to serum synchronisation. Serum shock induced rhythmic expression of *Per2*-luc in MCF10As with a long period of 31.4 hrs (± 0.4 hrs), and rhythmic expression of *Arntl*-luc in MCF7s with a period of 24.0 hrs (± 0.8 hrs). *Per2*-luc in MCF7s and *Arntl*-luc in MCF10As remained arrhythmic. Serum synchronisation was similar to forskolin and dexamethasone, in that MCF10A *Per2*-luc was consistently rhythmic with a longer than 24 hr period, whilst MCF7 *Arntl*-luc was consistently rhythmic with ~24 hr period. Whilst the phase relationships between the different cell lines remained the same across the different synchronisation methods presented thus far, the different synchronisation methods all appeared to reset the oscillatory luciferase expression to different phases of the circadian clock (i.e., MCF10A *Per2*-luc phase varies from Figure 2.6 to Figure 2.8). This is unsurprising as dexamethasone, forskolin and serum synchronisation are understood to have differing modes of action.

Figure 2.9 displays the effect of the first of the cycling temperature protocols. Two cycles of 12 hr 32°C / 12 hr 37°C temperature oscillation was sufficient to synchronise the cells such that after return to constant temperature, MCF10A *Per2*-luc expression oscillated with a period of 29.3 hrs (± 1.0 hrs). MCF7 *Arntl*-luc expression also oscillated, although replicates were not particularly consistent, with periods ranging from 18.7 – 41.9 hrs (although, some of the replicates were too variable in amplitude to fit either a damped

sine wave or a standard sine wave particularly well). MCF7 *Per2*-luc expression was also rhythmic with a period of 21.4 hrs (± 1.0 hr), although phases between the replicates varied between 2.0-15.4 hrs. The MCF10A *Arntl*-luc cells did not oscillate, as was the case with dexamethasone, forskolin and serum synchronisation conditions.

The luciferase oscillations during the temperature entrainment protocol are also of interest. Although the temperature protocol was set to 12 hr 32°C / 12 hr 37°C, the incubator heated much faster than it cooled (Figure 2.9C). The abrupt increase in temperature to 37°C caused an abrupt change in luciferase expression in a process known as 'masking' (Spörl et al., 2011), i.e., instead of the free-running circadian gene expression, the abrupt temperature Zeitgeber induces 'forced' circadian gene expression. It has previously been demonstrated in NIH-3T3 fibroblasts, that temperature cycles drive clock gene promoter activity in the correct phase relationships to one another (Saini et al., 2012). Therefore, when temperature increases NIH-3T3 *Dbp*-luc activity decreases, whilst anti-phasic NIH-3T3 *Arntl*-luc activity increases. Similarly, in HaCaT keratinocytes, temperature cycles drive rhythmic transcription of the core clock genes in the canonical phase relationships (Spörl et al., 2011). However it is clear from Figure 2.9 that even though *ARNTL* should be expressed ~11-12 hrs phase advanced of *PER2* in a functional circadian clock (Takahashi, 2017), in MCF10As and MCF7s an increase in temperature drove a decrease in both *Arntl*-luc and *Per2*-luc activity. In MCF10As the trough of *Per2*-luc activity appeared 5-6 hrs prior to the trough of *Arntl*-luc activity, which is indicative of some phase difference, although this was not the case in MCF7s. There are several possible reasons for the breakdown between the phase relationships of *Arntl*-luc and *Per2*-luc activity under temperature entrainment. Firstly, neither the MCF10As nor the MCF7s displayed strong rhythmic *Arntl* and *Per2* promoter activity as a synchronised population in constant temperature

conditions, so it is perhaps unsurprising that clock promoter reporter activity was also disrupted whilst under a cycling temperature entrainment protocol. Secondly, the free-running period of MCF10A *Per2*-luc was particularly long, which could have affected the extent to which MCF10As were capable of entraining to 24 hr temperature cycles. Thirdly, it is unclear whether these phase relationships would have persisted under longer temperature entrainment.

Figure 2.10 displays the effect of 12 hr 32°C / 12 hr 37°C temperature entrainment without HEPES in the media for extra buffering capacity. The findings were broadly similar to that of Figure 2.9 (with HEPES), though the agreement in phase and period between replicates was improved for MCF7 *Arntl*-luc activity. The period of MCF7 *Arntl*-luc activity was longer than that induced by dexamethasone, forskolin or serum shock entrainment at 29.6 hrs (± 1.5 hrs).

The long period in MCF10A *Per2*-luc cells and the lack of oscillation in MCF10A *Arntl*-luc cells demonstrated across all of these synchronisation protocols is contrary to much of the literature discussed in chapter 2.1.2, which suggests that MCF10A cells have a functional circadian clock. With this in mind the temperature protocol was altered to see if a more physiological temperature entrainment protocol of only 1°C temperature change might also induce similar patterns of luciferase activity in cells. The protocol was lengthened to 5 cycles of 36°C/37°C oscillation because the 1°C temperature oscillation was expected to synchronise cells to a lesser degree than 32°C/37°C oscillations (Saini et al., 2012). The cell seeding density was also decreased due to the longer entrainment protocol (see chapter 2.2.4 for details). Figure 2.11 shows the synchronisation effect of 5× 12 hr 36°C / 12 hr 37°C temperature cycles on MCF10A and MCF7 cells. MCF10A *Per2*-luc activity oscillated with

a period of 30.7 hrs (± 1.0 hr, excluding outliers). The phase was quite highly varied between replicates, ranging from 1.0 to 15.7 hrs. The higher variation between replicates was thought to be reflective of the longer entrainment period, which resulted in cells being less healthy by the end of the experiment. As was the case under other entrainment conditions, MCF10A *Arntl*-luc and MCF7 *Per2*-luc activity did not oscillate in a self-sustaining manner with a period of ~24 hrs. MCF7 *Arntl*-luc activity oscillated with a period of 25.5 (± 0.4 hrs, excluding outliers), which was similar to the MCF7 *Arntl*-luc activity observed under other entrainment conditions.

Overall, Figure 2.6-Figure 2.11 demonstrate that MCF10A *Per2*-luc activity robustly oscillated with a mean period of 29.3-31.4 hrs across dexamethasone, forskolin, serum shock and temperature synchronisation protocols. MCF10A *Per2*-luc activity also demonstrated the largest amplitudes of oscillation relative to the other luciferase reporters. In contrast, MCF10A *Arntl*-luc activity was the least rhythmic and did not oscillate in response to any of the synchronisation protocols (Figure 2.6-Figure 2.11). This data is contrary to much of the published literature which found circadian rhythms in MCF10A *ARNTL* and *PER2* expression of ~24 hrs (Chacolla-Huaringa et al., 2017; Gutiérrez-Monreal et al., 2016; Rossetti et al., 2012; Xiang et al., 2012). It agrees most closely with Cox (2012), who demonstrated a period of approx. 30 hrs in *PER2* by qPCR post-dexamethasone synchronisation – although Cox (2012) also demonstrated rhythms of ~32 hrs in *ARNTL* expression, which were not observed here. Rhythmic *Per2*-luc oscillation in MCF10As would imply that the activity of ARNTL/CLOCK complex on the *Per2* promoter is rhythmic. Therefore, it is quite surprising that oscillatory MCF10A *Arntl* promoter activity is not observed in MCF10As in these experiments. It seems unlikely that there was a problem with the *Arntl*-luc reporter construct, as it reports circadian rhythmic activity in MCF7 cells (Figure 2.6).

Figure 2.6-Figure 2.8 demonstrate that MCF7 *Arntl*-luc activity robustly oscillated with a period of 23.1-25.9 hrs across dexamethasone, forskolin and serum shock entrainment. Relatively, temperature synchronisation (Figure 2.9-Figure 2.11) induced less robust oscillation of MCF7 *Arntl*-luc activity, with a broader range of periods, phases and amplitudes. MCF7 *Per2*-luc activity was quite inconsistent and appeared circadian in some entrainment conditions, but not others. For instance, forskolin and 32/37°C temperature entrainment induced periods of 21.6-25.8 hrs, with MCF7 *Per2*-luc activity in the correct phase relationship (~12 hrs phase advanced) relative to MCF7 *Arntl*-luc activity (Figure 2.7, Figure 2.9 & Figure 2.10). However, serum shock did not induce circadian expression of MCF7 *Per2*-luc (Figure 2.8). The amplitude of MCF7 *Arntl*-luc and *Per2*-luc oscillatory activity was lower than the amplitude of MCF10A *Per2*-luc oscillation in all entrainment conditions except forskolin entrainment and dexamethasone entrainment for MCF7 *Per2*-luc only. As discussed in chapter 2.1.2, the literature generally agrees that MCF7 cells do not display circadian expression of core clock genes (Chacolla-Huaringa et al., 2017; Gutiérrez-Monreal et al., 2016; Rossetti et al., 2012; Xiang et al., 2012; Yuan Zhang et al., 2018). However, the data presented here is consistent with Lellupitiyage Don et al. (2019) who demonstrated circadian luciferase reporter activity for *Arntl*-luc and *Per2*-luc reporters in the correct phase relationships. Additionally, Cox (2019) demonstrated circadian oscillation of *ARNTL* and *PER1*, but not *PER2*, in MCF7s by qPCR.

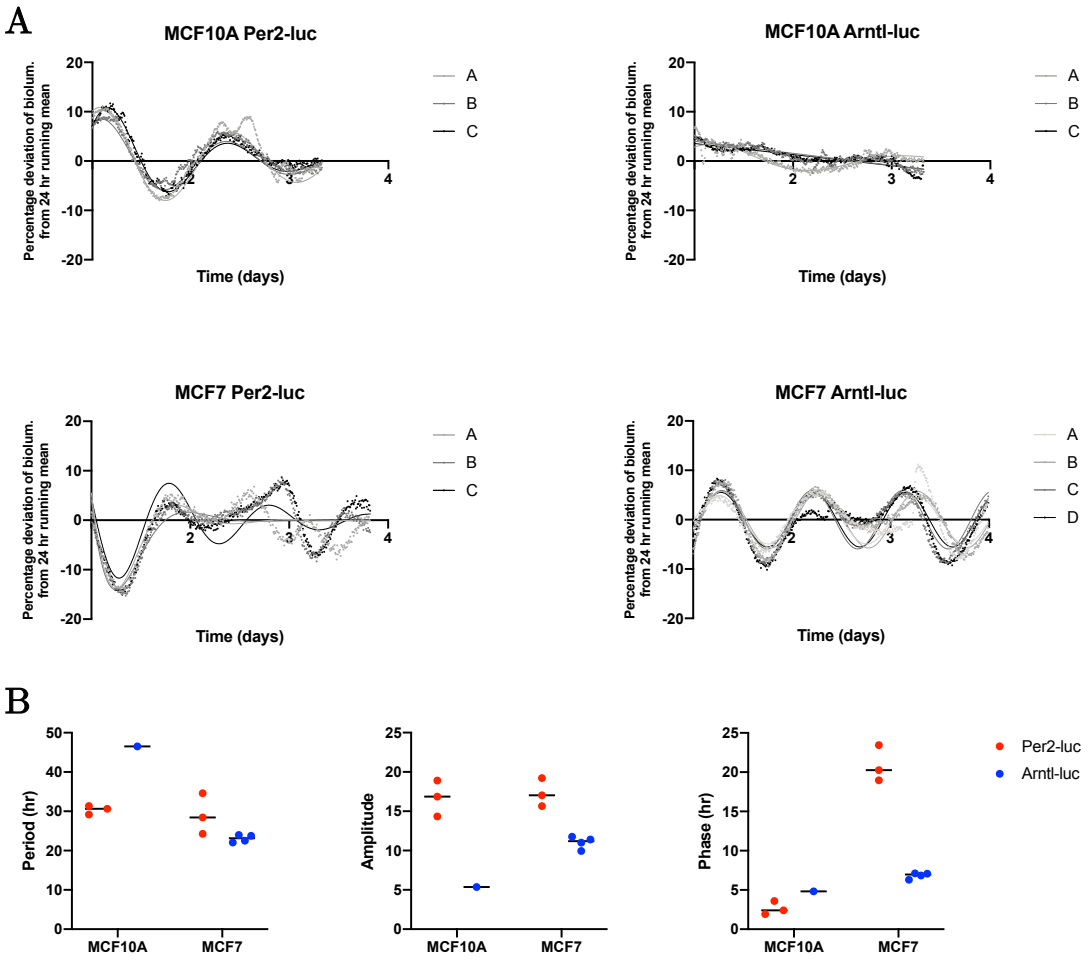


Figure 2.6 A) MCF10A and MCF7 *Per2*-luc and *Arntl*-luc activity 24 hrs post-dexamethasone synchronisation, n=3 and n=4 for MCF7 *Arntl*-luc. B) Phase, amplitude and period graphs for MCF *Per2*-luc and *Arntl*-luc activity.

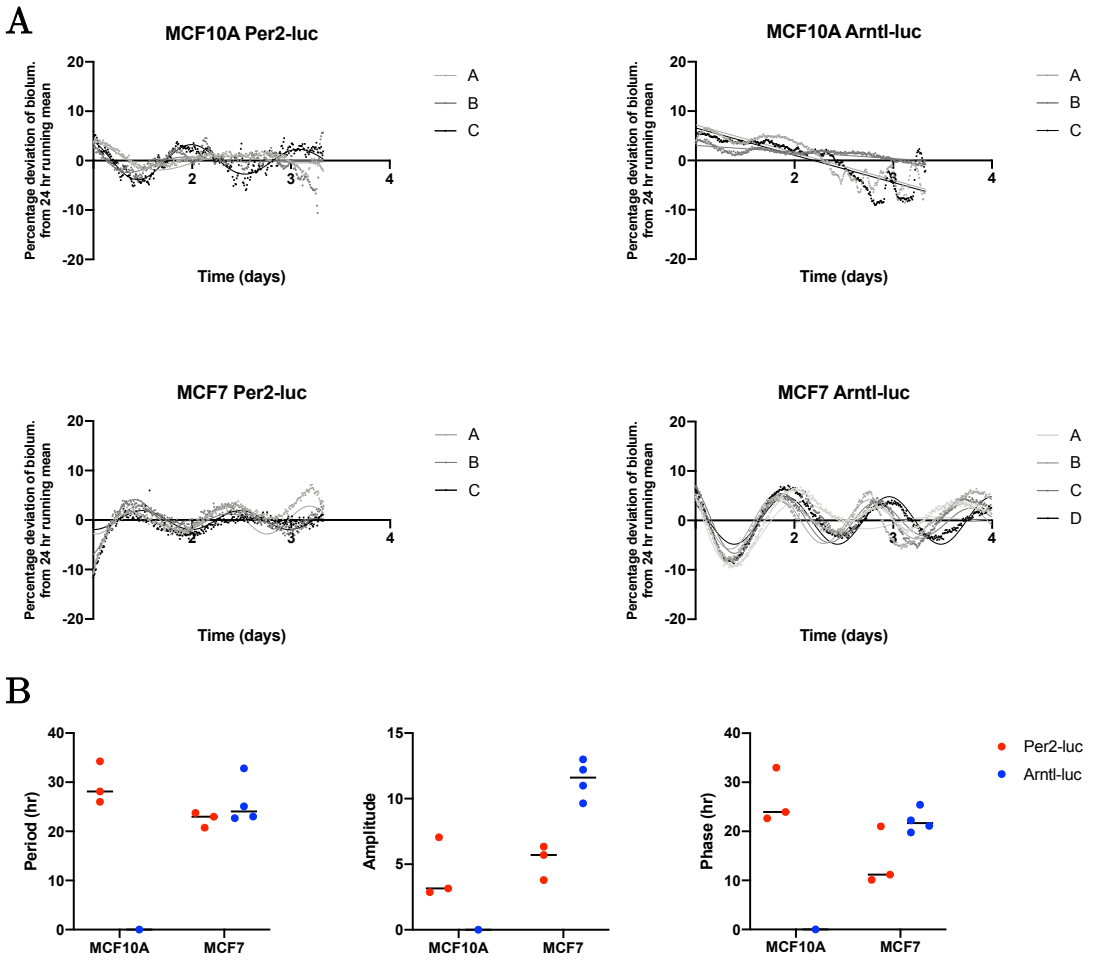


Figure 2.7 A) MCF10A and MCF7 *Per2*-luc and *Arntl*-luc activity 24 hrs post-**forskolin** synchronisation, n=3 and n=4 for MCF7 *Arntl*-luc. B) Phase, amplitude and period graphs for MCF *Per2*-luc and *Arntl*-luc activity.

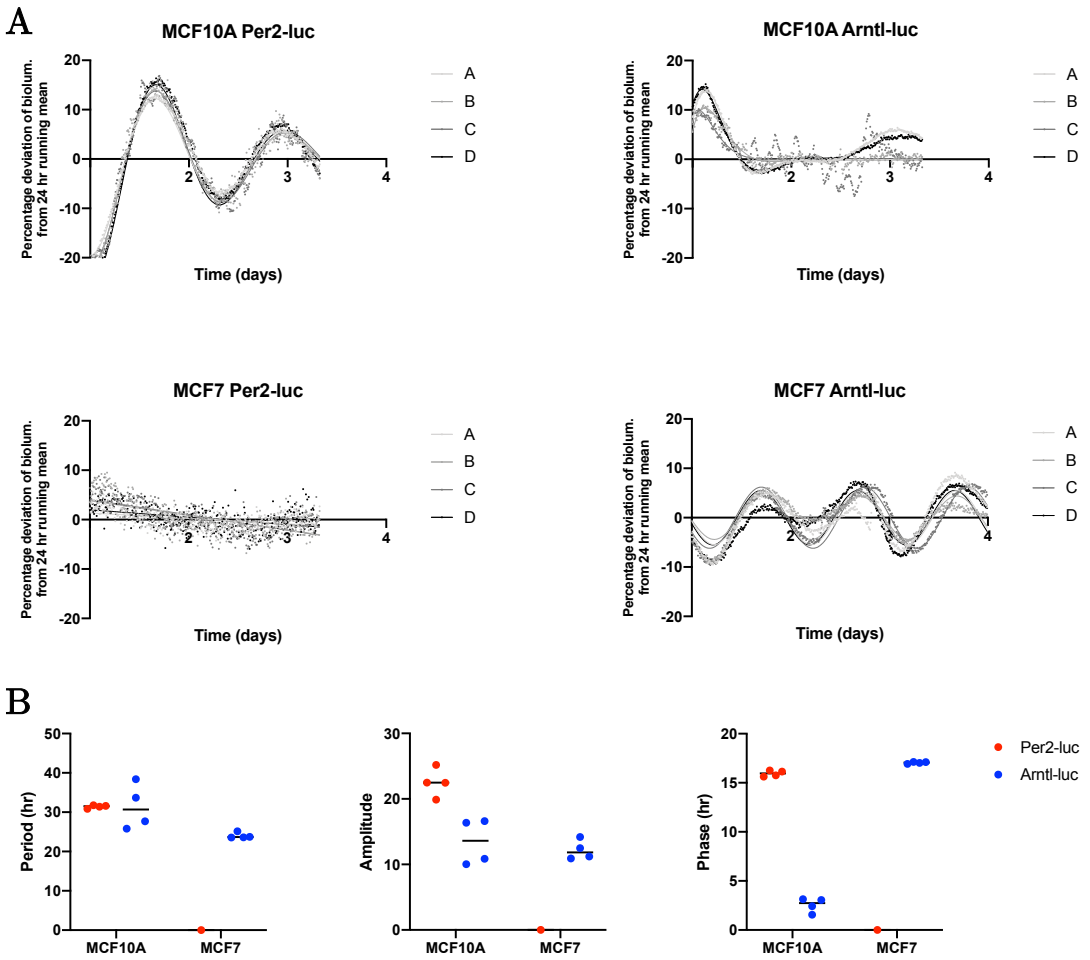


Figure 2.8 A) MCF10A and MCF7 *Per2*-luc and *Arntl*-luc activity 24 hrs post-serum synchronisation, n=4. B) Phase, amplitude and period graphs for MCF *Per2*-luc and *Arntl*-luc activity.

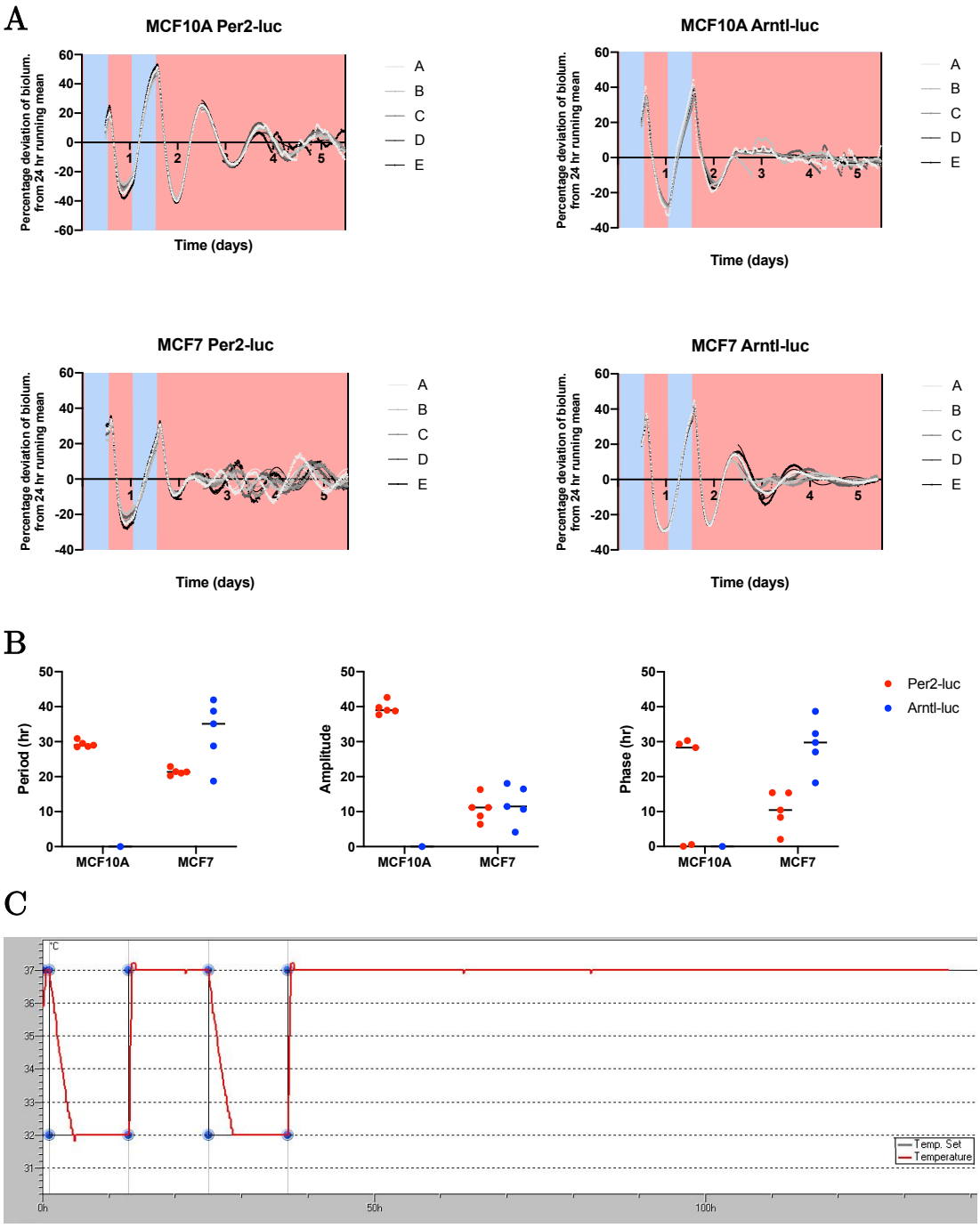


Figure 2.9 A) MCF10A and MCF7 *Per2*-luc and *Arntl*-luc activity induced by **12 hr 32°C / 12 hr 37°C** temperature synchronisation, n=5. B) Phase, amplitude and period graphs for MCF *Per2*-luc and *Arntl*-luc activity. C) The temperature cycling protocol and the resultant temperature profile in the incubator.

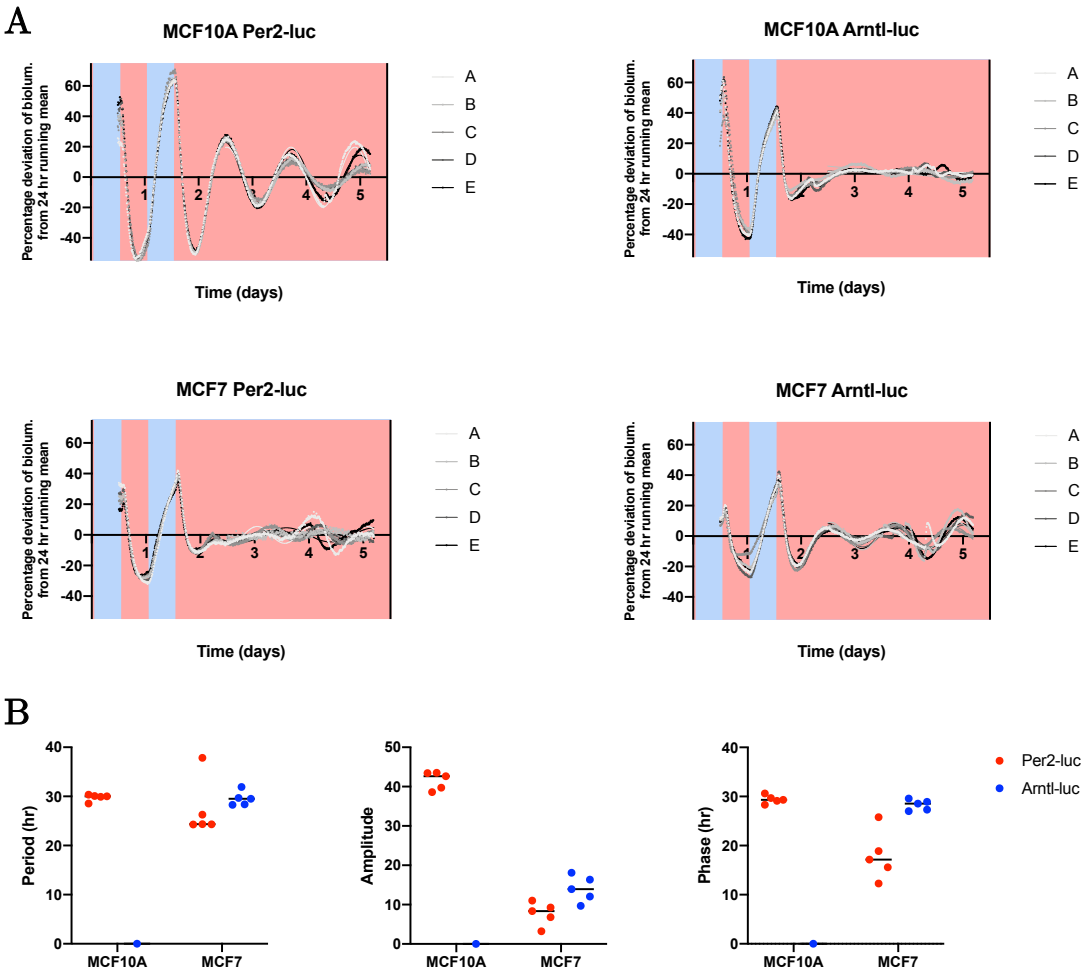


Figure 2.10 A) MCF10A and MCF7 *Per2*-luc and *Arntl*-luc activity induced by **12 hr 32°C / 12 hr 37°C** temperature synchronisation **without addition of HEPES** to the imaging media, n=5. B) Phase, amplitude and period graphs for MCF *Per2*-luc and *Arntl*-luc activity.

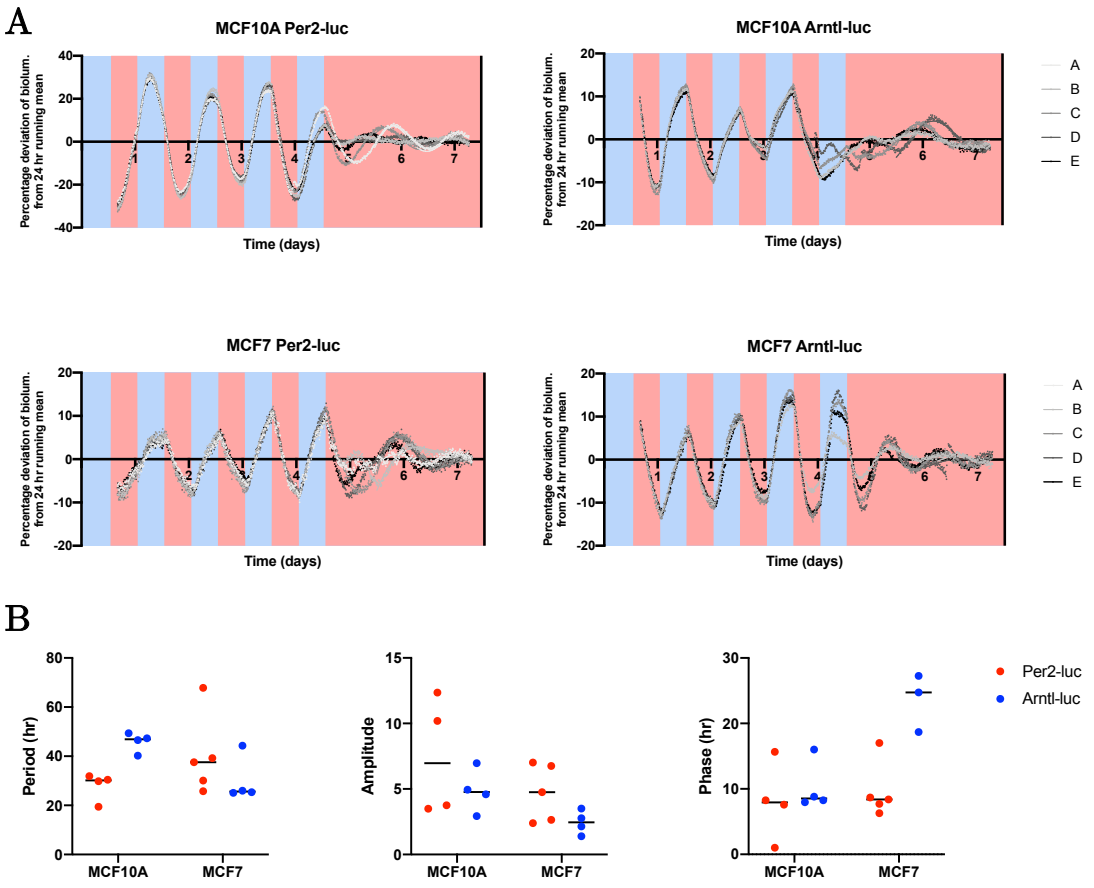


Figure 2.11 A) MCF10A and MCF7 *Per2*-luc and *Arntl*-luc activity induced by 12 hr **36°C / 12 hr 37°C** temperature synchronisation; *Per2*-luc n=5, *Arntl*-luc n=4. B) Phase, amplitude and period graphs for MCF Per2-luc and Arntl-luc activity.

2.3.2 Temperature compensation and period stability of MCF Cells

Circadian rhythms are temperature compensated, which means that the rate of the process, i.e., the period length, remains robust relative to the temperature at which the cells are incubated (Kidd et al., 2015). Therefore, if the long period of the MCF10A *Per2*-luc cells represented a stable temperature compensated circadian clock, then the period would be expected to remain stable at a higher temperature. To test this, MCF10A cells were entrained with 2 hotter cycles of 12 hr 33°C/ 12 hr 38°C before release into constant 38°C (see Figure 2.12). The period of MCF10A *Per2*-luc was subtly but significantly longer when cells were entrained to 33/38°C temperature cycles rather than 32/37°C temperature cycles (Figure 2.15, $p \leq 0.05$). This implies that the circadian timing system in the MCF10As may not be operating in a temperature compensated manner, as it should, though a wider range of temperature entrainment protocols would be required to confirm this. As with dexamethasone, forskolin and serum synchronisation, MCF10A *Per2*-luc activity consistently reported long periods (>24 hrs) across 32/37°C, 33/38°C and 36/37°C temperature entrainment (for which means ranged from 29.3-33.4 hrs, Figure 2.15).

Circadian rhythms should also display a degree of robustness to entrainment cycles of different period lengths (Eelderink-Chen et al., 2021). To investigate this, the MCF cells were exposed to either short T cycles of 21 hrs (10.5 hr 32°C / 10.5 hr 37°C) or long T cycles of 27 hrs (13.5 hr 32°C / 13.5 hr 37°C). Figure 2.13 and Figure 2.14 show that for the most part the free-running construct activity of the MCF cell lines (i.e., activity after release to constant 37°C) was masked by the temperature entrainment protocols. However, there were some instances in which *Arntl*/*Per2* activity was clearly behaving in a manner independent to the short/long T cycles. For instance, MCF7 *Arntl*-luc

activity under 21 hr temperature cycle entrainment appeared just reach a trough, though never quite reach a peak immediately prior to temperature change. This indicates that the oscillator resisted entraining to the short T cycle (Figure 2.13). Conversely, MCF7 *Arntl*-luc activity under 27 hr temperature cycles always peaked, and always troughed, prior to the forced phase adjustment induced by the temperature change, which indicated that the MCF7 *Arntl*-luc activity was shorter than 27 hrs and did not entrain to 27 hr temperature cycles (Figure 2.14). A similar pattern was observed for MCF10A *Per2*-luc activity, which appeared forced to trough and peak immediately after temperature change when under 21 hr temperature cycles, but reached a free-running peak and trough just before the temperature change under 27 hr temperature cycles (Figure 2.13, Figure 2.14). In fact, MCF10A *Per2*-luc activity appeared to entrain quite well to 27 hr temperature cycles, which correlated with the longer oscillations observed for *Per2*-luc in free-running conditions (chapter 2.3.1). The MCF10A *Arntl*-luc and MCF7 *Per2*-luc cell lines did not demonstrate a marked robustness to either 21 or 27 hr temperature entrainment conditions, which correlated with the distinctly less robust oscillatory behaviour observed for these cell lines in free-running conditions.

Figure 2.15 and Figure 2.16 compare the free-running period lengths of each of the MCF reporter cell lines following entrainment to the different temperature entrainment protocols. MCF10A *Per2*-luc oscillation appeared robustly >24 hrs. The shortest period length observed was after 27 hr entrainment, at 28.2 hrs (±1.68 hrs), although this was only significantly shorter (at $p<0.05$) relative to 21 hr and 32/37°C no HEPES entrainment conditions. MCF10A *Arntl*-luc oscillation was observed only after 21 and 27 hr temperature entrainment conditions, and demonstrated a mean free-running period of 26.2 (±1.26 hrs), which was much shorter than for the MCF10A *Per2*-luc free-

running period, suggesting that the molecular clock in MCF10As may not have been functioning canonically even under conditions that entrained both *Per2*-luc and *Arntl*-luc activity. Relatively each MCF7 reporter line showed a greater range of free-running period lengths than each of the MCF10A reporter lines. Although, generally speaking, the conditions with fewer outliers produced period lengths close to 24 hrs (Figure 2.16).

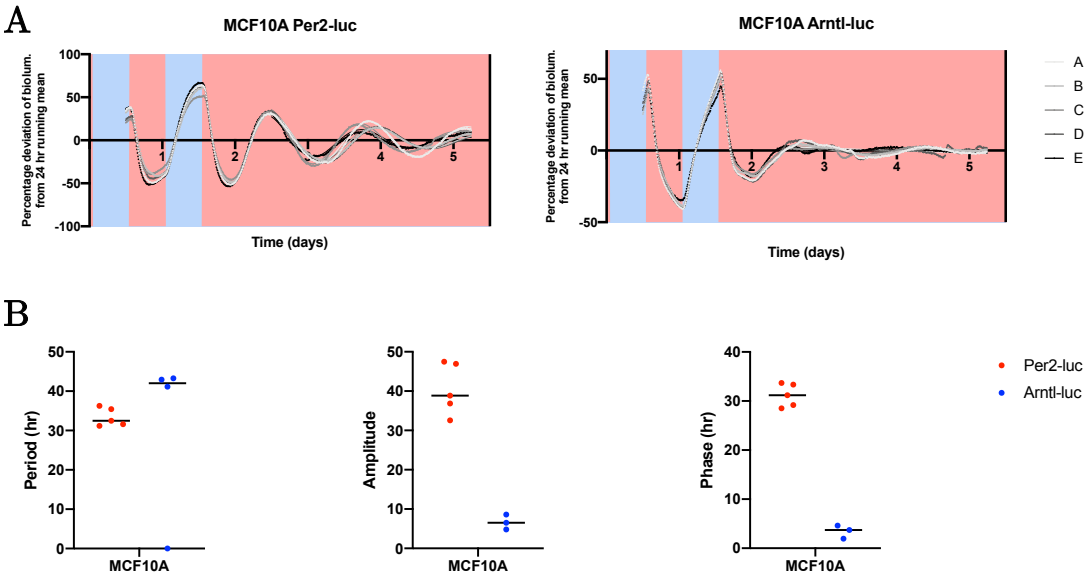


Figure 2.12 A) MCF10A *Per2*-luc and *Arntl*-luc activity induced by **12 hr 33°C / 12 hr 38°C** temperature synchronisation; n=5. B) Phase, amplitude and period graphs for MCF10A *Per2*-luc and *Arntl*-luc activity.

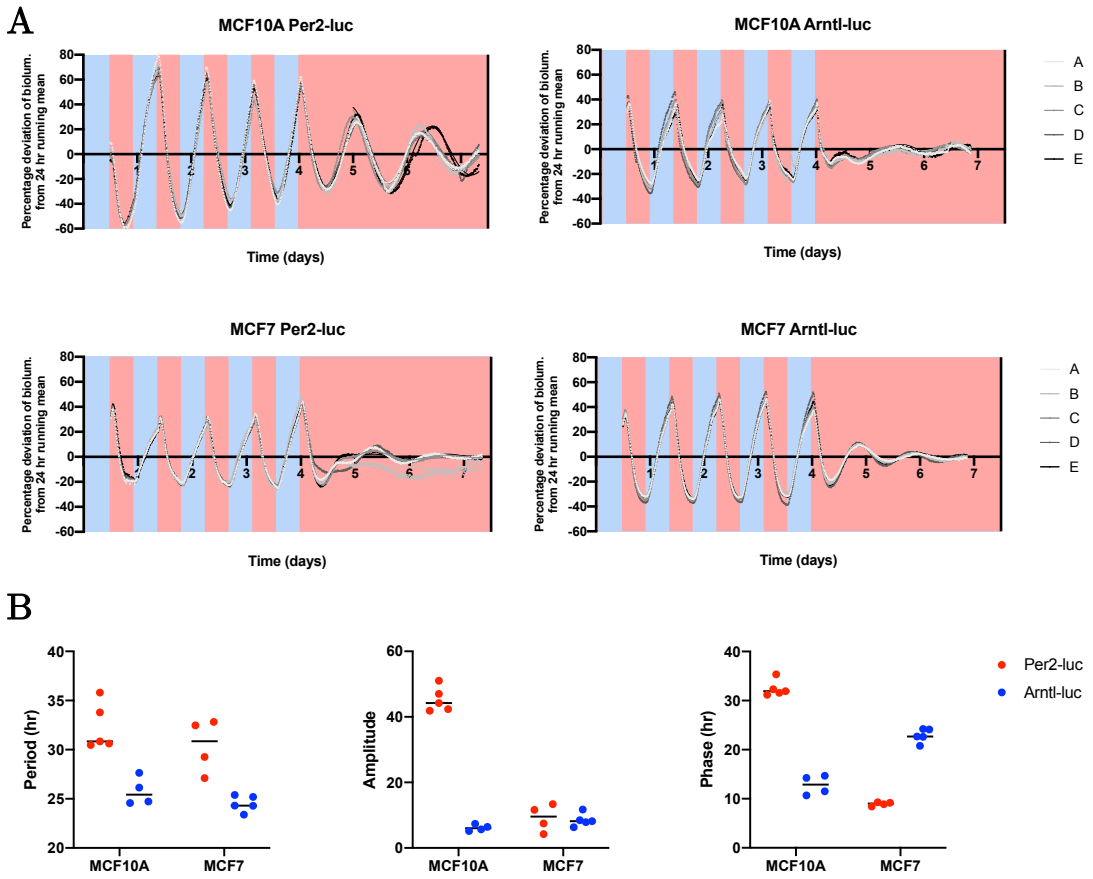


Figure 2.13 A) MCF10A and MCF7 *Per2*-luc and *Arntl*-luc activity induced by **10.5 hr 32°C / 10.5 hr 37°C** temperature synchronisation; n=5. B) Phase, amplitude and period graphs for MCF *Per2*-luc and *Arntl*-luc activity.

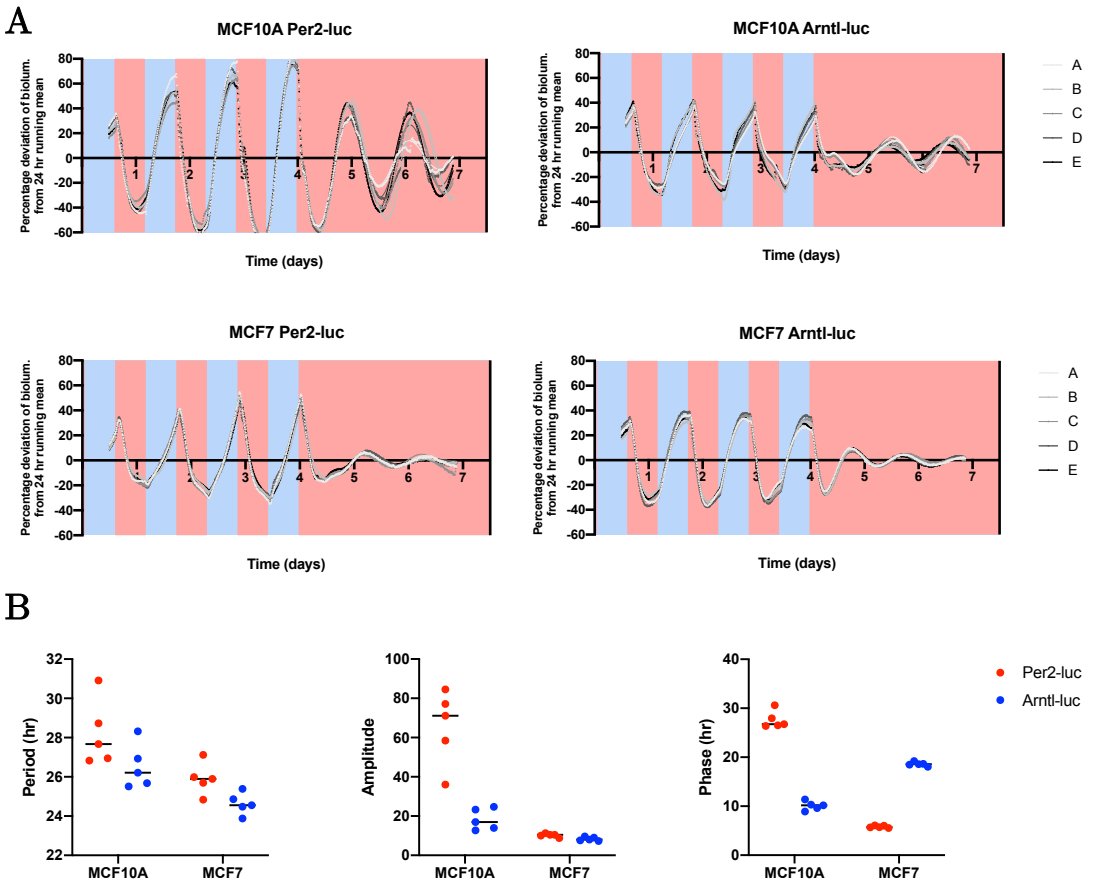
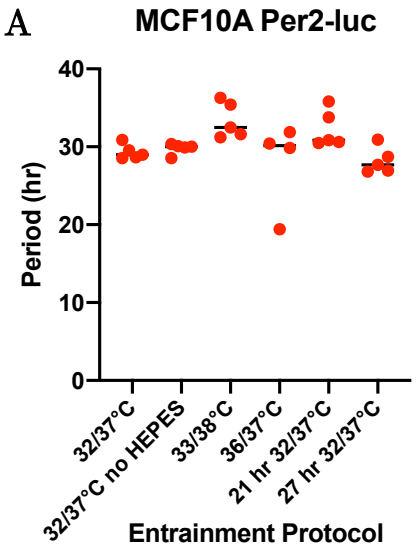


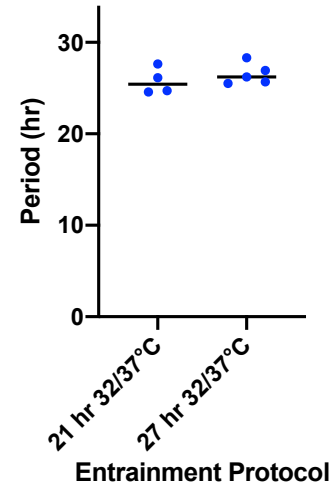
Figure 2.14 A) MCF10A and MCF7 *Per2*-luc and *Arntl*-luc activity induced by **13.5 hr 32°C / 13.5 hr 37°C** temperature synchronisation; n=5. B) Phase, amplitude and period graphs for MCF *Per2*-luc and *Arntl*-luc activity.

A



Welch's t-test p-value	32/37°C	32/37°C no HEPES	33/38°C	36/37°C	21 hr 32/37°C	27 hr 32/37°C
32/37°C		0.4196 n.s.	0.0133 p≤0.05	0.6516 n.s.	0.0464 p≤0.05	0.2465 n.s.
32/37°C no HEPES			0.0220 p≤0.05	0.5562 n.s.	0.0759 n.s.	0.1102 n.s.
33/38°C				0.1481 n.s.	0.4841 n.s.	0.0044 p≤0.01
36/37°C					0.2238 n.s.	0.9166 n.s.
21 hr 32/37°C						0.0160 p≤0.05
27 hr 32/37°C						

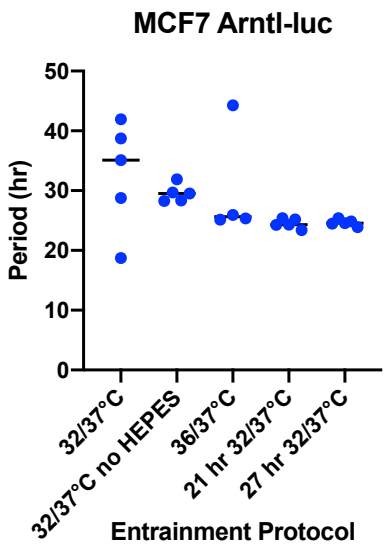
B



Welch's t-test p-value	21 hr 32/37°C	27 hr 32/37°C
21 hr 32/37°C		0.4248 n.s.
27 hr 32/37°C		

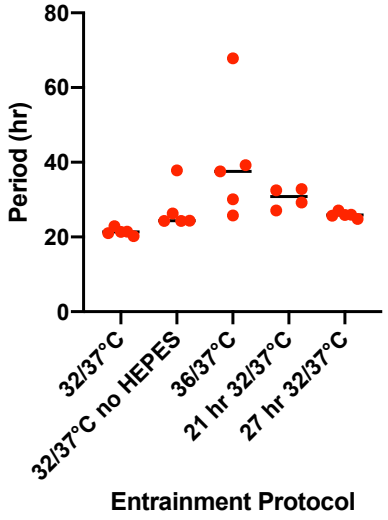
Figure 2.15 A comparison of free-running period length of MCF10A A) *Per2*-luc and B) *Arntl*-luc cells after release from entrainment to different temperature protocols. P-values were determined by Welch's two-tailed unpaired t-tests, n.s. = no significant difference.

A



Welch's t-test p-value	32/37°C	32/37°C no HEPES	36/37°C	21 hr 32/37°C	27 hr 32/37°C
32/37°C		0.4963 n.s.	0.7045 n.s.	0.1192 n.s.	0.1231 n.s.
32/37°C no HEPES			0.9038 n.s.	0.0005 p≤0.001	0.0008 p≤0.001
36/37°C				0.3160 n.s.	0.3241 n.s.
21 hr 32/37°C					0.8009 n.s.
27 hr 32/37°C					

B



Welch's t-test p-value	32/37°C	32/37°C no HEPES	36/37°C	21 hr 32/37°C	27 hr 32/37°C
32/37°C		0.0836 n.s.	0.0635 n.s.	0.0046 p≤0.01	<0.0001 p≤0.001
32/37°C no HEPES			0.1653 n.s.	0.3524 n.s.	0.6001 n.s.
36/37°C				0.2610 n.s.	0.1259 n.s.
21 hr 32/37°C					0.0413 p≤0.05
27 hr 32/37°C					

Figure 2.16 A comparison of free-running period length of MCF7 A) *Per2*-luc and B) *Arntl*-luc cells after release from entrainment to different temperature protocols. P-values were determined by Welch's two-tailed unpaired t-tests, n.s. = no significant difference.

2.3.3 Clock gene transcripts of MCF7 and MCF10A cells under 12 hr 36°C / 12 hr 37°C entrainment by qPCR

In practice, the use of the clock gene promoter bioluminescence assay in combination with temperature entrainment cycles is subject to error. Although firefly luciferase activity is known to be temperature compensated, the bioluminescence output in U2OS cells *in vitro* is not (Feeney et al., 2016). It is thought that a proportion of the change in bioluminescence associated with temperature change is dependent upon a change in the cellular concentration of luciferase's substrate luciferin. Therefore, qPCR was employed as a second method to ascertain clock gene expression in MCF cells under 12 hr 36°C / 12 hr 37°C temperature cycles.

Figure 2.17 displays the results of qPCR for core clock components *PER2*, *ARNTL* and *NR1D1*. For MCF10A cells, straight line fits were not rejected at a significance level of 0.05 in favour of sine fits of period = 24 hrs for any of the three clock genes. However, the error bars between the biological replicates were quite large, so it is not easy to conclude whether MCF10A cells truly lacked rhythmicity in clock gene expression under 36/37°C temperature entrainment or not. For MCF7 cells straight line fits were rejected in favour of sine fits for *PER2* and *NR1D1* expression. The phase of entrainment agreed well between *PER2* and *NR1D1* sine fits – *NR1D1* appeared approx. 1 hr phase advanced of *PER2*. This broadly agrees with *NR1D1* expression being ~4.4 hrs phase advanced of *PER2* expression in baboons (albeit with a large margin of error depending upon tissue) (Chapter 1 Figure X). qPCR of *ARNTL* in MCF7s did not preferentially fit a sine wave of period = 24 hrs over a straight line, though were a sine fit to be forced, the trend of the data indicates that the phase of entrainment would likely be similar to that of *NR1D1* and *PER2*.

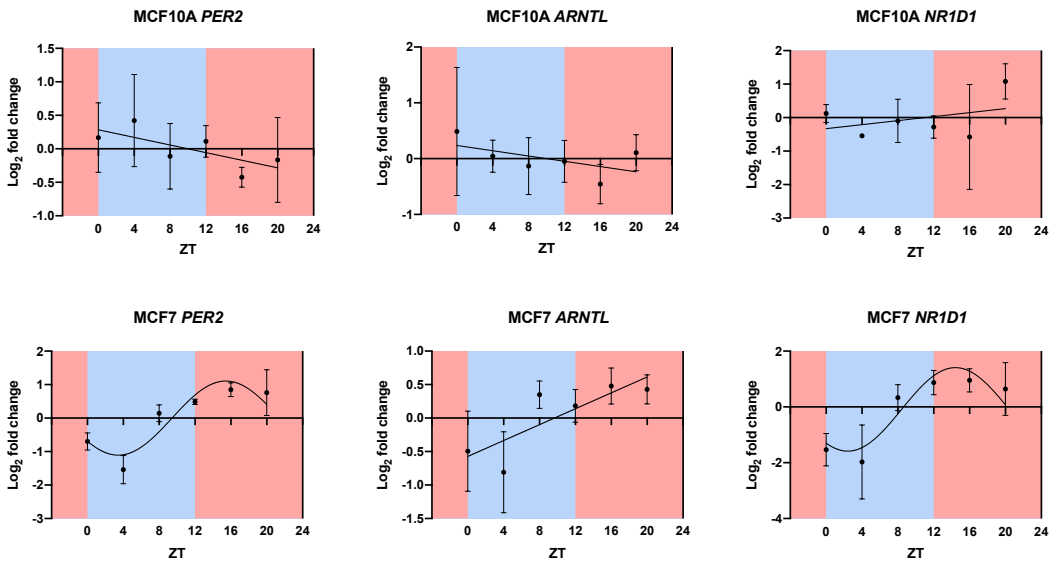


Figure 2.17 Relative abundance of clock transcripts under 12 hr 36°C / 12 hr 37°C temperature entrainment assayed by qPCR. Data is displayed as mean ± S.D., n=3 biological replicates except for MCF10A & MCF7 *NR1D1* ZT4 and MCF7 *ARNTL* ZT12 where n=2 due to disagreement between technical triplicates. Sum of squares F tests determined whether sine fits (period = 24 hrs) surpassed linear fits ($p<0.05$). Sampling commenced after 72 hrs of 36/37°C temperature entrainment.

This was contradictory to the canonical anti-phasic relationships expected between *ARNTL* and *PER2*. MCF7 *Arntl*-luc and *Per2*-luc appeared approximately anti-phasic in free-running conditions after 36/37°C entrainment, however phases appeared aligned whilst under 36/37°C entrainment (Figure 2.11). When the bioluminescence and qPCR data are considered together, the MCF7 cells did not appear to demonstrate canonical circadian clock gene phase alignment under 36/37°C temperature entrainment. This implies the absence of a fully functional temperature-entrainable molecular oscillator in the MCF7 cells, as discussed in chapter 2.3.1.

2.3.4 RNA-sequencing of MCF7 and MCF10A cells under 12 hr 36°C / 12 hr 37°C temperature cycles

2.3.4.1 Data quality

Two methods were considered for normalisation of RNA-seq data to library size, logCPM and logTMM (edgeR R package) (M. D. Robinson et al., 2010). TMM normalisation aims to prevent particularly highly expressed genes from biasing the normalisation and hinges on the presumption that more than half of the genes are not differentially expressed between any two samples (M. D. Robinson et al., 2010). Given that MCF10A and MCF7 cells have different transcriptomes, and were sampled at different circadian timepoints, it seemed possible that this presumption might not have held true. Principal components analysis (PCA) comparison of logCPM and logTMM normalisation methods (Figure 2.18) suggests that in practice logTMM normalisation was not substantially different from logCPM, and so the decision was taken to use the more intuitive logCPM method of normalisation going forwards. Figure 2.19A displays the distributions of logCPM gene expression across all time points and

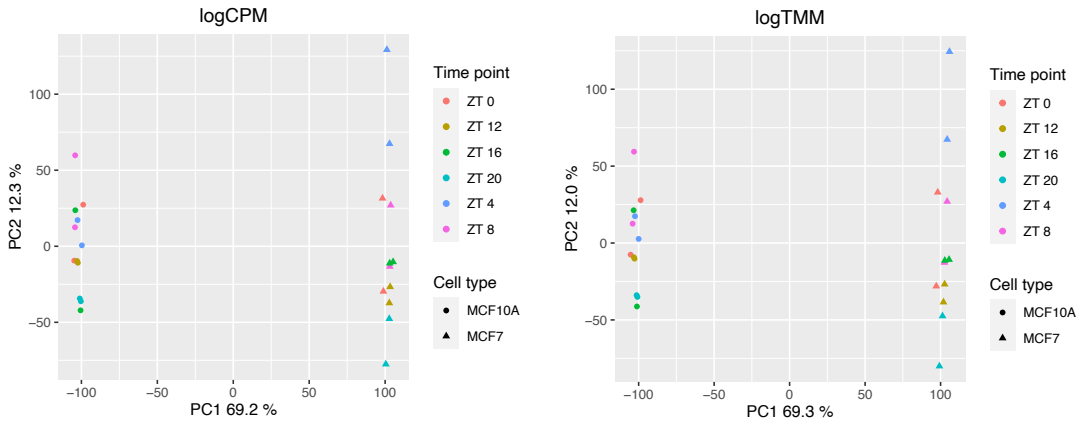


Figure 2.18 Comparison of all RNA-seq samples by PCA. LHS = logCPM normalisation, RHS = logTMM normalisation.

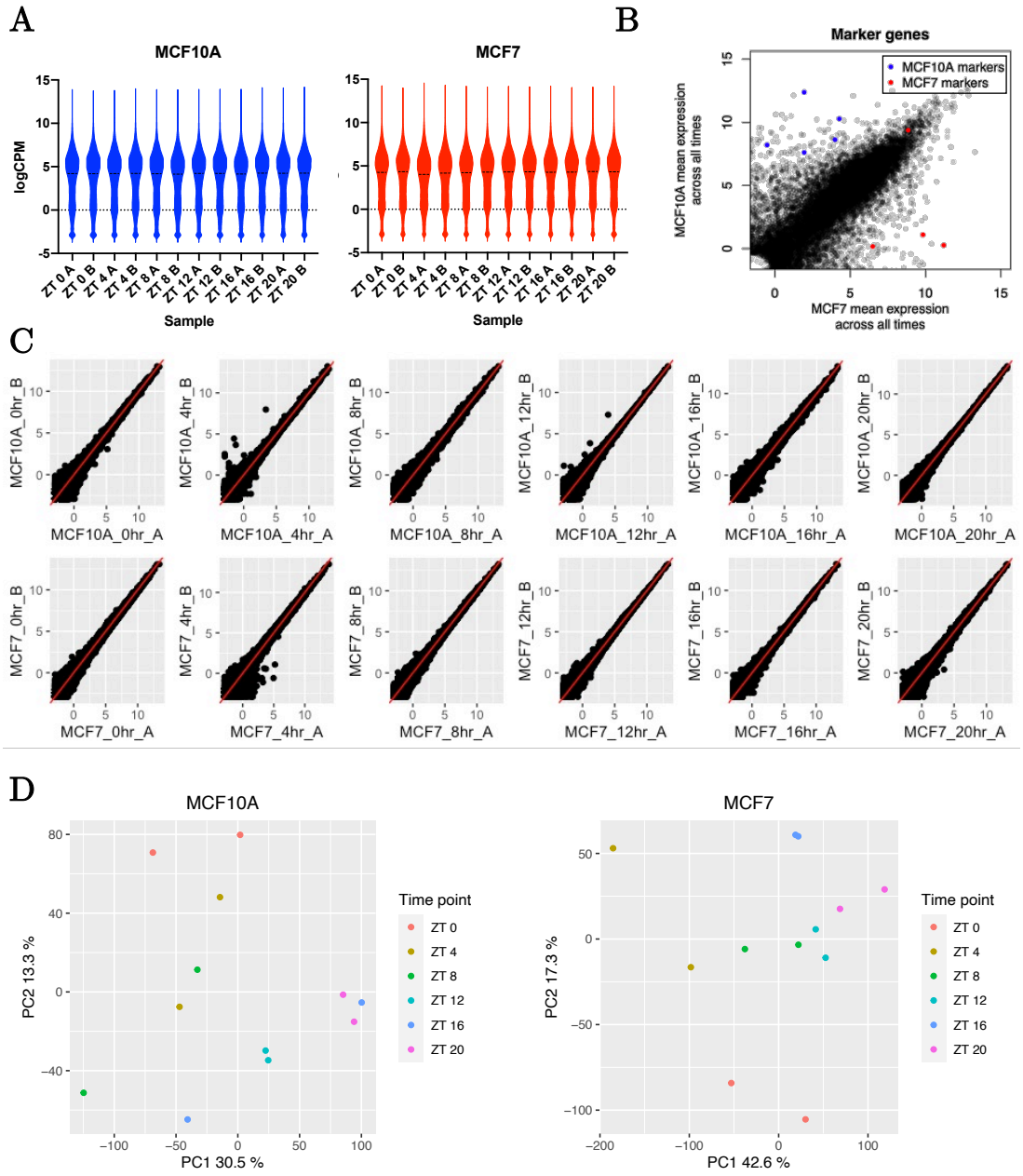


Figure 2.19 Assessment of RNA-seq data quality. A) Violin plots of logCPM normalised data for all genes from all samples in MCF10A (blue) and MCF7 (red) cells. B) Scatter plot of average expression of all genes expressed in MCF7 samples relative to MCF10A samples (logCPM). Marker genes identified by Nagaraja et al. (2006) for distinguishing between MCF10A and MCF7 are highlighted either in red (MCF7) or blue (MCF10A). C) Scatter plots compare expression of all genes between duplicate samples. D) 1st PC vs 2nd PC for MCF10A and MCF7 logCPM normalised RNA-seq transcripts.

both MCF cell types. The distributions were stable from sample to sample and are no cause for concern. It is important to validate that the cells sequenced

were indeed MCF10A and MCF7 cells. To this end, Figure 2.19B displays mean gene expression between all MCF10A samples and all MCF7 samples. Also displayed (in blue and red) are a set of marker genes that are known to be differentially expressed between MCF10A and MCF7 cells (Nagaraja et al., 2006). Based on the general agreement between marker genes and cell line identity, it is reasonable to conclude that Figure 2.19B supports the identity of the MCF10A and MCF7 cell lines. Figure 2.19D compares the variation between and within the timepoints in all of the normalised RNA-seq transcripts for A) MCF10A cells and B) MCF7 cells by PCA. The duplicate time points were clustered reasonably closely in 2D PCA space in MCF7 cells, but appeared to be more variably clustered in MCF10A cells. It is important not to place too much weight on interpreting these PCA graphs, as the first two principal components vary between cell lines in terms of the amount of variability in the dataset that they explain. Additionally, the data presented thus far using promoter-luciferase constructs and qPCR indicate a degree of circadian dysfunction in both cell lines. If the clock in MCF10A and MCF7 cells is not functional then poor separation of timepoints by PCA might be expected. Figure 2.19C demonstrates that duplicate samples are relatively similar to one another in terms of their gene expression with few outliers either side of the red x=y line. Therefore, the loose clustering of duplicate samples in PC1 and PC2 might imply similarity between the timepoints, rather than differences between duplicates.

2.3.4.2 MCF rhythmic gene expression under 12 hr 36°C / 12 hr 37°C temperature cycles

Figure 2.20 displays the genes identified as rhythmic in MCF7 and MCF10A cells using B-H adjusted p-value <0.2. In the heatmaps, genes were ordered by

phase, and colour was scaled by row (i.e., gene-wise). Many more rhythmic genes were identified in the MCF7 cell line vs MCF10As (1503 vs 209 genes). This is contradictory to the literature consensus that MCF10As have a functional molecular oscillator, which presumably governs downstream rhythmic gene expression, whilst MCF7s are arrhythmic. For that reason, Figure 2.19B is key in validating the nature of the cell lines, as discussed in chapter 2.3.4.1. Phase histograms for the MCF10A and MCF7 cells display the

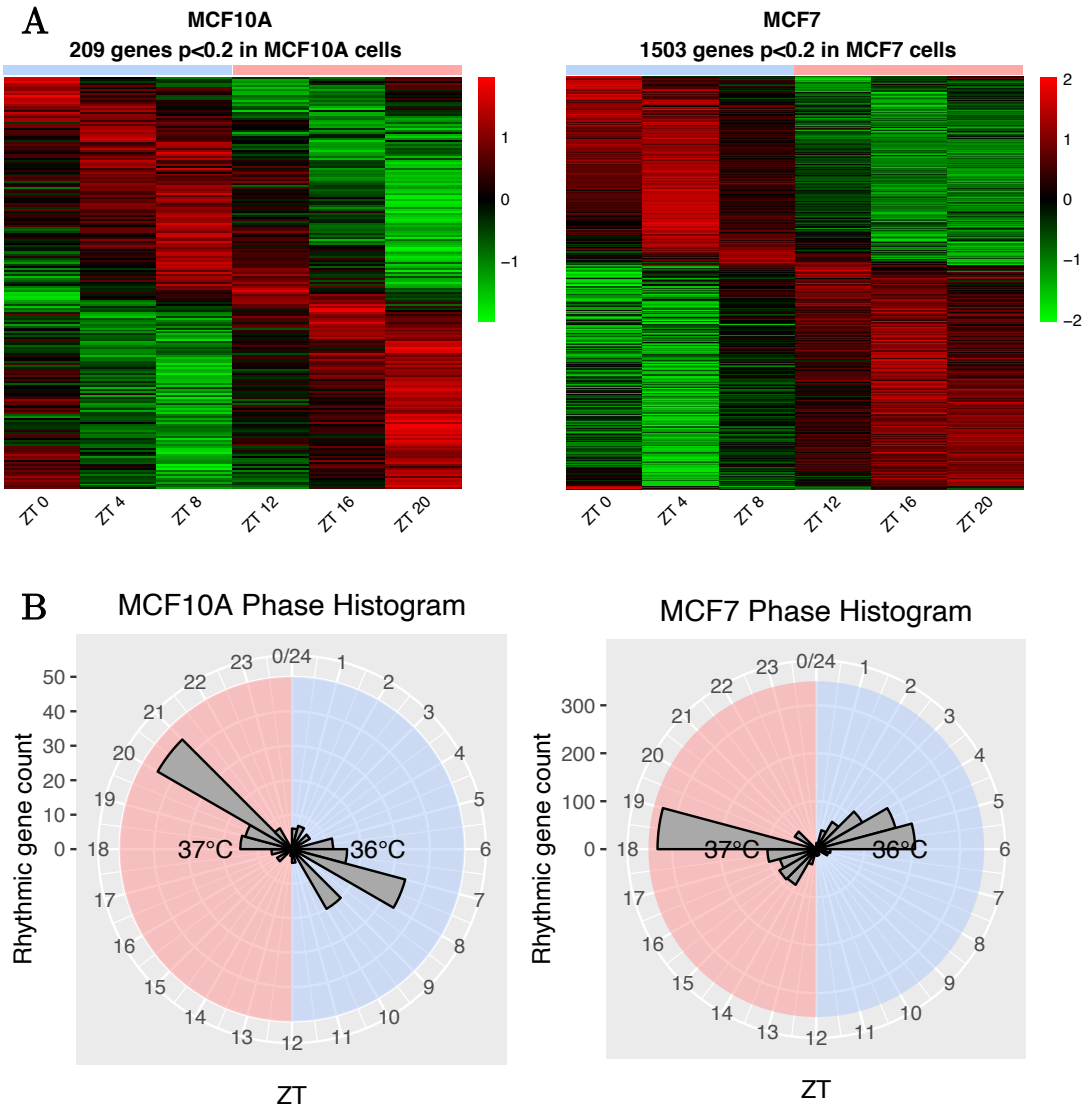


Figure 2.20 A) Heatmaps of rhythmic gene expression under 12 hr 36°C / 12 hr 37°C temperature cycles in MCF10A and MCF7 cells over time. Colour is scaled row-wise. Genes are ordered by phase (B-H adjusted p<0.2; mean of n=2 displayed at each ZT). B) Histograms showing the phase distributions of oscillating genes in MCF10A and MCF7 cells.

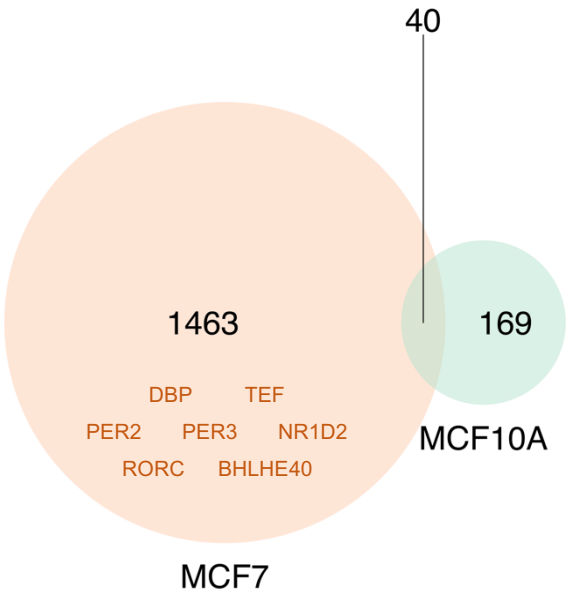


Figure 2.21 Venn diagram of genes identified as rhythmic in MCF7 and MCF10A cells at a B-H adjusted p-value < 0.2. Genes involved in the molecular circadian clock are also denoted.

rhythmic genes by phase of peak expression (Figure 2.20B). The bimodal distribution of rhythmic gene expression is quite normal and has been observed in many tissues (Mure et al., 2018; R. Zhang et al., 2014).

Only 40 genes were identified as rhythmic in both cell lines, implying little overlap in the rhythmic transcriptomes of MCF7 and MCF10A cells (Figure 2.21). This is consistent with the findings of Gutiérrez-Monreal et al. (2016). However, such comparisons should be interpreted with caution. As discussed in chapter 2.2.11, rhythmicity analysis was implemented at a high FDR (0.2), in order to try and limit the number of false negatives (i.e., rhythmic genes incorrectly identified as arrhythmic). However, false negatives will occur, simply as biological and experimental noise cannot be tolerated so well in smaller datasets. Therefore, at an FDR of 0.2 for rhythmic gene detection in both MCF10A and MCF7s, approximately 36% of these 40 doubly rhythmic genes will be false positives. An alternative approach to searching for

differentially rhythmic genes would be to use a hypothesis testing or model selection approach (Pelikan et al., 2020). These approaches compare gene expression in MCF10As and MCF7s directly, thereby reducing the number of differentially rhythmic genes that are falsely identified. However, these methods categorise genes into categories such as 'gain of rhythmicity', 'loss of rhythmicity' and 'change in amplitude of rhythmicity'. Such approaches work well when applied to high-resolution time-courses. However, in the case of low-resolution time-courses, one should be cautious about the potential for over-analysis that such algorithms may encourage. Figure 2.21 also demonstrates that, in addition to the wider transcriptome being more rhythmic in MCF7s than MCF10As, a number of core clock genes were identified as rhythmic in MCF7 cells, whereas none were identified in MCF10A cells.

The expression of MCF10A core clock transcripts are presented in Figure 2.22. Sine fits of period = 24 hrs are displayed for all transcripts. The R^2 values of the sine fits were evaluated in order to assess the goodness of the sine fits. In MCF10A cells, the expression of only two clock transcripts, *NPAS2* and *BHLHE40*, fitted sinusoidal functions with $R^2>0.66$. Peak MCF10A *NPAS2* transcription appeared to be ~6 hrs phase-advanced of MCF10A *BHLHE40*, whereas in multiple mouse tissues *Npas2* and *Bhlhe40* are phase separated by approximately 12 hrs (http://circadb.hogeneschlab.org/mouse provides a searchable online repository of published murine circadian transcriptomes for easy inspection of phase relationships). Transcripts with moderate sine fits ($0.33<R^2<0.66$), also displayed phase desynchrony between clock genes. For instance, in MCF10As *ARNTL* and *PER1* sine fits oscillated with the same phase alignment, which is not observed in the functional circadian clocks of baboons or mice, where *Arntl* and *Per2* oscillate in an anti-phasic manner to one another (Mure et al., 2018; R. Zhang et al., 2014; Chapter 1.2 Figure 1.3). Furthermore, although MCF10A *PER1* and *PER2* were broadly phase-aligned

Figure 2.22 MCF10A expression of core circadian transcripts under 12 hr 36°C / 12 hr 37°C temperature cycles as determined by RNA-seq. Data presented as mean ± S.D.; n=2. Sine fits were set to a period of 24 hr. R^2 values of sine fits are coloured as follows for ease of observation: $R^2<0.33$ coloured red; $0.33<R^2<0.66$ coloured orange; $R^2>0.66$ coloured green.

(acrophases ~3.5 hrs apart) and the PARbZIP transcripts *DBP* and *HLF* (but not *TEF*) were also phase-aligned as expected, the MCF10As *PERs* were phase-advanced of PARbZIP transcripts *DBP* and *HLF* by ~6 hrs, even though

the PARbZIP transcripts should appear ~1-2 hrs phase advanced of the *PER* transcripts (Mure et al., 2018; R. Zhang et al., 2014).

Overall, MCF10A cell populations under 36°C/37°C temperature cycles assayed by RNA-seq, simply did not display clock gene transcription akin to that observed in a functional molecular clock. This conclusion agrees with the MCF10A qPCR data presented in chapter 2.3.3 (Figure 2.17). Although linear fits were not rejected in favour of sine fits for qPCR of MCF10A clock gene expression, the trend of expression appears to mimic that observed by RNA-seq, i.e., *ARNTL* expression decreased with time in RNA-seq and qPCR experiments. *PER2* expression peaked at Zeitgeber time (ZT) 4, and then decreased using both methods. Lastly *NR1D1* expression decreased with time, and then peaked at ZT 20 by RNA-seq and qPCR. This is an excellent confidence builder in the accuracy of the data presented here. The lack of a functional circadian clock in MCF10A cells appears to be reflected in the large reduction of rhythmic genes in the MCF10A transcriptome relative to the MCF7 transcriptome (Figure 2.20).

The expression of MCF7 molecular clock transcripts is presented in Figure 2.23. Overall, sine functions fitted the expression of MCF7 transcripts better than in the MCF10A cells. *PER2*, *PER3*, *CRY2*, *DBP*, *TEF*, *NR1D2*, *RORC* and *BHLHE40* all achieved sine fit R^2 values greater than 0.66. Generally, these genes displayed the canonical phase alignments of a functional molecular clock. That is to say that *DBP*, *TEF* and *NR1D2* were slightly phase advanced of *PER2*, *PER3* and *BHLHE40* as expected (Mure et al., 2018; R. Zhang et al., 2014). The peak of the MCF7 *CRY2* sine fit occurred a few hours phase advanced of where one might have expected, as *CRY2* would normally align with peak *PER2* expression. However, the peak of raw MCF7 *CRY2* expression is delayed by a few hours relative to the fitted sine curve, which is closer to

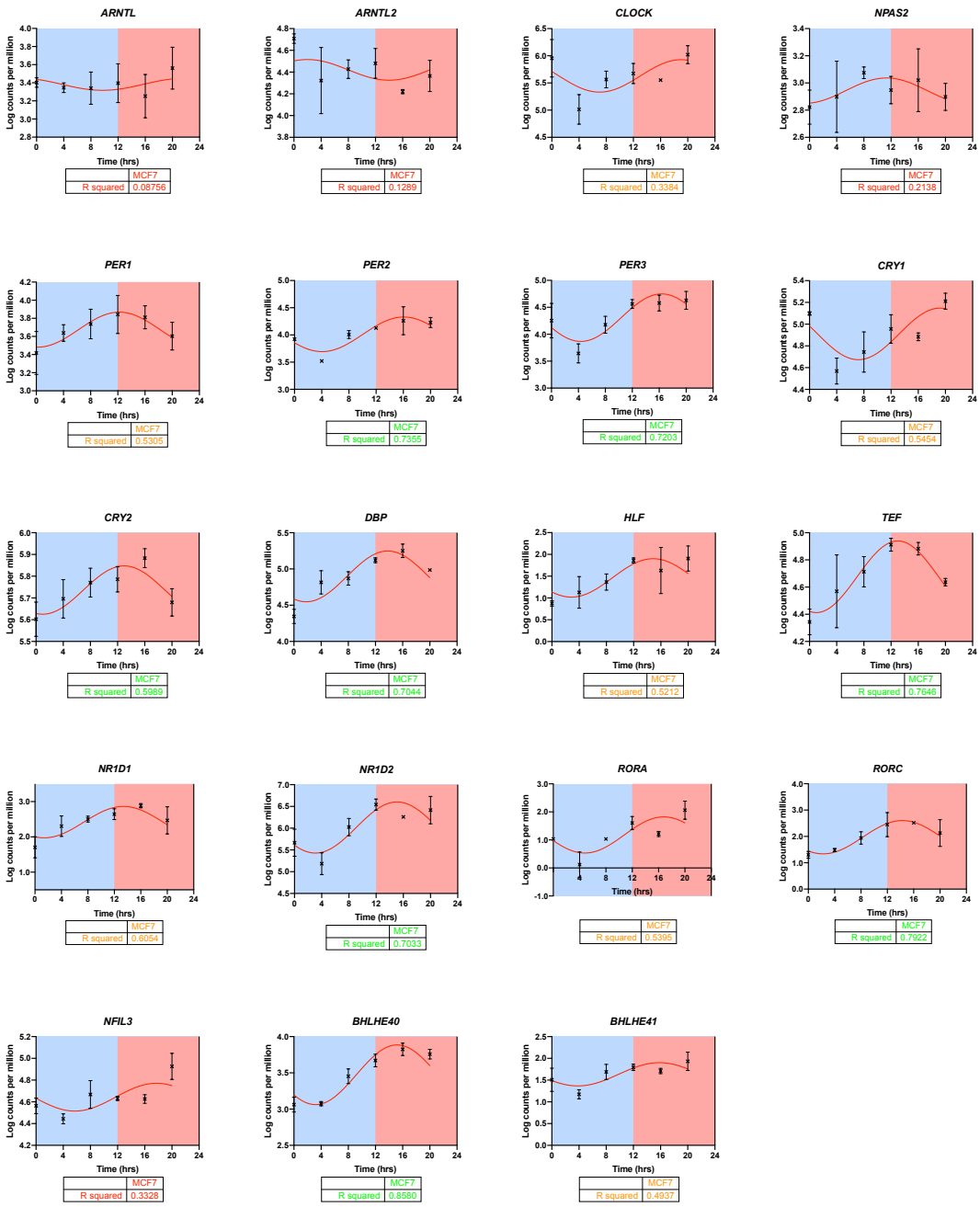


Figure 2.23 MCF7 expression of core circadian transcripts under 12 hr 36°C / 12 hr 37°C temperature cycles as determined by RNA-seq. Data presented as mean ± S.D.; n=2. Sine fits were set to a period of 24 hr. R^2 values of sine fits are coloured as follows for ease of observation: $R^2<0.33$ coloured red; $0.33<R^2<0.66$ coloured orange; $R^2>0.66$ coloured green.

peak MCF7 *PER2*, therefore the phase alignment of MCF7 *CRY2* is not of great cause for concern. More unusually, MCF7 *RORC* was slightly phase advanced of MCF7 *PER2*, whereas in mice and baboons *Rorc* is known to be phase delayed of *Per2* by 4 to 6 hrs.

Figure 2.23 shows that for MCF7 transcripts with an R^2 of >0.33 and <0.66, the phase-alignment of the clock transcripts was more mixed. Unlike in MCF10A cells (Figure 2.22), in MCF7s *HLF* was correctly phase-aligned with the other PAR bZIP proteins *DBP* and *TEF*. Meanwhile whilst *RORA* does not oscillate well in all mouse and baboon tissues, it does oscillate in alignment with *PER2* in the mouse lung, which was also the case in MCF7s (R. Zhang et al., 2014). However relative to MCF7 *PER2* expression, peak MCF7 *PER1* expression was a few hours phase advanced of where it would normally peak in mice or baboons, MCF7 *CRY1* was ~5 hrs phase advanced of where it would normally peak and MCF7 *BHLHE41* was ~2 hrs phase delayed of where it would normally peak. MCF7 *CLOCK* was almost anti-phasic to where it would be expected to peak as it was ~4 hrs phase delayed of MCF7 *PER2*, when in mice *Per2* and *Clock* would normally be expressed anti-phasically. In baboons *Clock* usually has a relatively low amplitude of oscillation, and also exhibits a highly variable phase – although it is never ~4 hrs phase delayed of *Per2,* as was the case in the MCF7 cells.

Strangely, although many MCF7 E-box driven genes (including *PER*s, PAR bZIPs, *NR1D2* and *CRY2*) oscillated in the correct phase alignments, their transcriptional activators *ARNTL*/*CLOCK* or *NPAS2* did not display the rhythmic behaviour that would be associated with a functional circadian clock. As was observed for the MCF10A cells, the MCF7 qPCR data broadly supports the conclusions drawn from the RNA-seq data. For both qPCR and RNA-seq MCF7 *ARNTL* expression was best described by a linear fit, whilst MCF7 *PER2* and *NR1D1* oscillated in phase alignment with one another.

MCF7 cells under 12 hr 36°C / 12 hr 37°C temperature cycles demonstrated better core clock gene oscillation relative to MCF10A cells, and indeed

demonstrated a much greater degree of transcriptome oscillation than MCF10A cells (Figure 2.20). This is contrary to the literature consensus, that 'benign' MCF10A cells generally behave more rhythmically than 'malignant' MCF7s (Table 2.1). Though, even the more rhythmic MCF7 cells did not exhibit canonical clock gene expression.

2.3.5 Comparison to Gutiérrez-Monreal et al. 2016 rhythmic gene expression in MCFs synchronised by serum shock

Gutiérrez-Monreal et al. originally identified 451 probes oscillating with a period of 20-28 hrs in their MCF10A population and 416 probes oscillating in their MCF7 population (with a cosine fit correlation coefficient of >0.9). However, given that their microarray totalled 44,544 probes and no statistical test of rhythmicity was applied, many of their 'rhythmic' probes might have been false positives. Since Gutiérrez-Monreal et al. have made their log ratioed normalised probe counts publicly available (GEO identifier: GSE76370) we re-analysed their data using MetaCycle's meta2d function, as applied to the RNA-seq data presented in chapter 2.3.4.2.

Figure 2.24 demonstrates the expression of MCF10A and MCF7 marker genes in the Gutiérrez-Monreal et al. dataset. Marker genes were previously identified using microarrays and validated by the RNA-seq data presented in chapter 2.3.4.1 (Nagaraja et al., 2006). Of 7 marker probes for MCF10A cells, 3 were expressed at a higher level in MCF10As than MCF7s, whilst 4 were expressed at a low level in both MCF10As and MCF7s. This seems a passable threshold, as not every microarray probe is capable of reporting transcript expression optimally due to processes such as alternative splicing, off-target probe binding and the 3' bias of reverse transcription (Q. Li et al., 2011). However, the

marker probes for MCF7 cells did not appear to be expressed preferentially in MCF7 cells relative to MCF10A cells, which is a cause for concern.

After thresholding for MetaCycle probe rhythmicity at a B-H adjusted p-value <0.2 only 26 probes were identified as rhythmic in MCF10A cells and 55 probes in MCF7 cells (Figure 2.25). It is difficult to compare the re-analysed Gutiérrez-Monreal et al. data directly with the MCF RNA-seq data presented in chapter 2.3.4 for several reasons; the experiments use different technologies and different time course sampling protocols, making direct comparison of rhythmic genes (B-H adjusted p-value <0.2) between the datasets impossible. However, it is clear that MCF10A and MCF7 cells subjected to serum shock by Gutiérrez-Monreal et al. express very few probes that can be confidently identified as rhythmic. This finding is supported by data presented by Gutiérrez-Monreal et al. that the majority of probes for core clock genes were arrhythmic in both MCF10A and MCF7 cells (detailed in chapter 2.1.2). Figure 2.25B demonstrates that the rhythmic probes follow a bimodal distribution, as seen in the MCF RNA-seq data presented in Figure 2.20B.

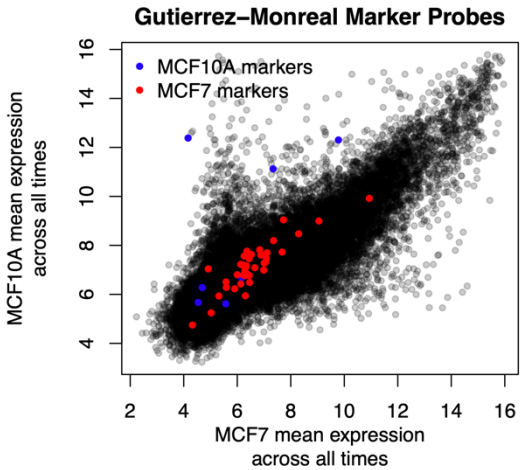


Figure 2.24 Scatter plot of average expression of all probes expressed in MCF7 samples relative to MCF10A samples. Probes for marker genes identified by Nagaraja et al. (2006) for distinguishing between MCF10A and MCF7 cell lines are highlighted either in red (MCF7) or blue (MCF10A).

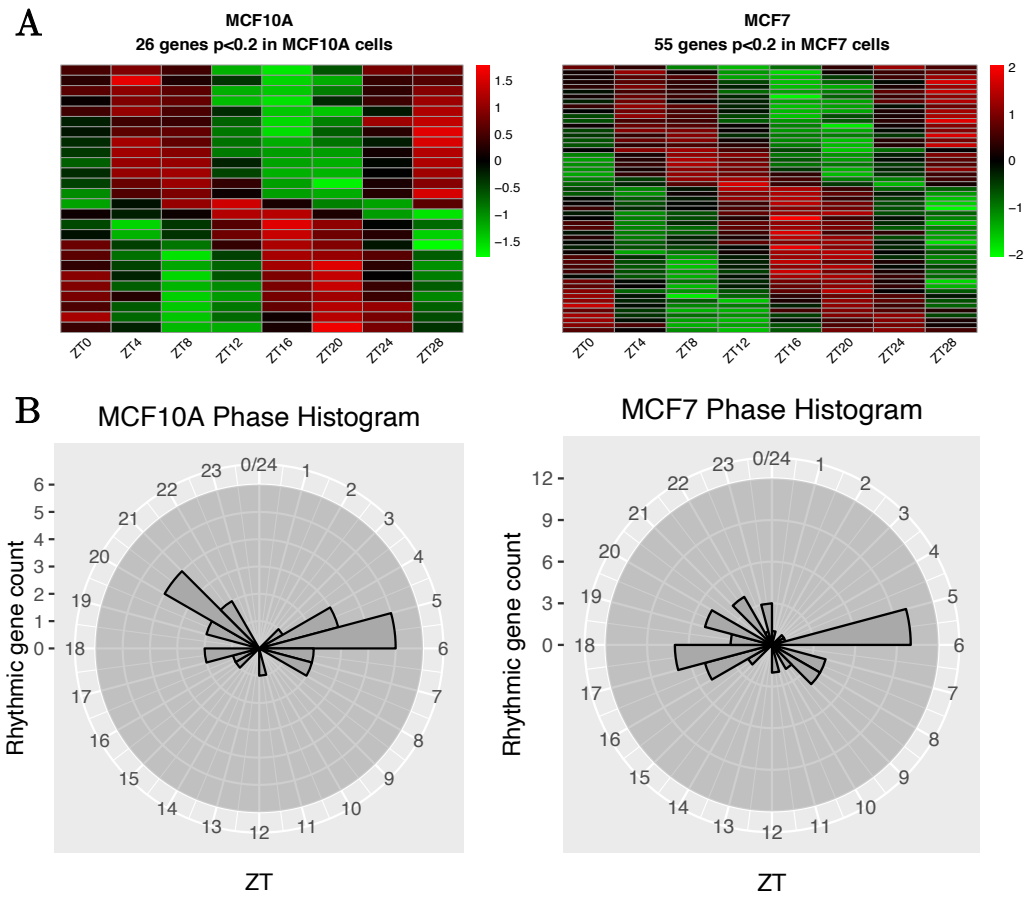


Figure 2.25 Rhythmic probes from the Gutiérrez-Monreal et al. dataset. A) Heatmaps of Gutiérrez-Monreal et al. rhythmic probe expression from 8 hrs post-serum synchronisation in MCF10A and MCF7 cells over time. Colour is scaled row-wise. Probes are ordered by phase (B-H adjusted p<0.2). B) Histograms showing the phase distributions of oscillating probes in MCF10A and MCF7 cells in response to serum synchronisation.

This pattern of rhythmic gene expression has been observed in many tissues (Mure et al., 2018; R. Zhang et al., 2014).

2.3.6 ATAC-seq of 12 hr 36°C / 12 hr 37°C temperature-entrained MCF Cells

MCF10A and MCF7 cells cultured under 12 hr 36°C / 12 hr 37°C temperature cycles were also analysed using ATAC-seq to assess changes in chromatin accessibility over time. Figure 2.26 demonstrates the ATAC-seq library sizes

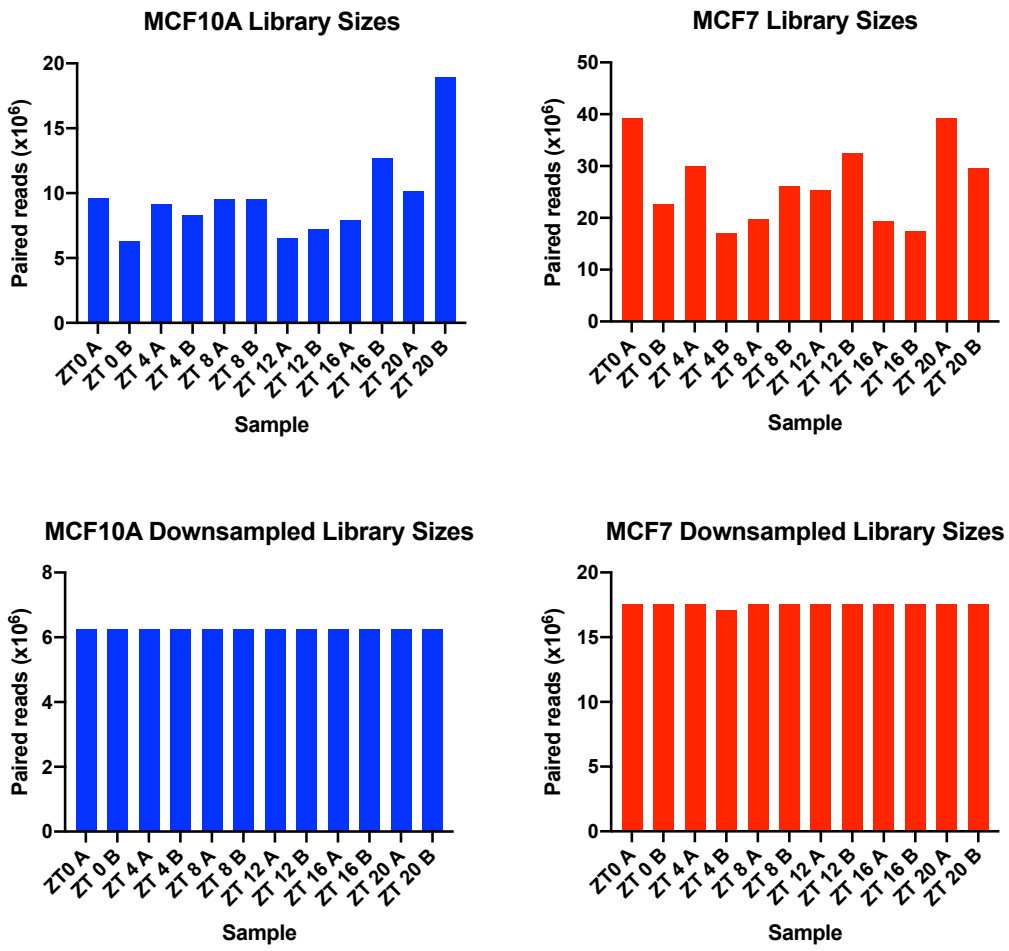


Figure 2.26 MCF10A and MCF7 ATAC-seq raw library sizes, and library sizes post-downsampling to the smallest library size.

of the MCF10A and MCF7 samples. For direct comparison of reads aligned to promoter regions between different time points, the libraries were down-sampled at random to match the size of the smallest library. Figure 2.27 shows the ATAC-seq reads across all time points for the TSSs of a selection of core clock genes. Whilst distinct peaks above the background noise are visible for the MCF10A samples, no such peaks are visible in the MCF7 samples. This was also observed to be the case across all TSSs (Figure 2.28). The lack of MCF7 reads in promoter regions is reflected in the TSSE scores for the MCF7 samples, which fall <5, such that the ENCODE consortium deems the data quality to be 'concerning' (Figure 2.29). Peak calling in the MCF7 samples was confounded by a combination of low signal to noise ratio, and a large number

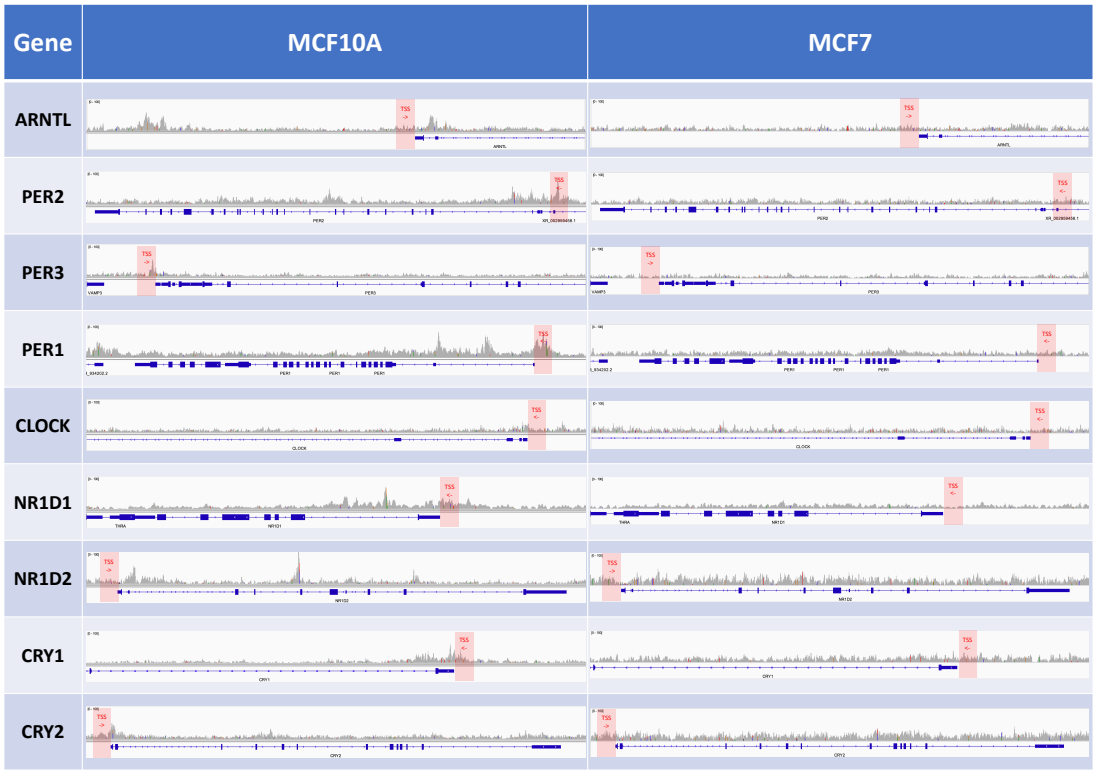


Figure 2.27 Distribution of ATAC-seq reads across core clock genes. TSSs are denoted in red. All track heights were set to the same scale. Reads were summed across timepoints for comparison of MCF7 vs MCF10A.

of reads that fell into the four genomic regions that in MCF7 cells are known to contain clusters of breakpoint mutations resulting in over-expression of genes: 1p13.1-p21.1, 3p14.1-p14.2, 17q22-q24.3, and 20q12-q13.33 (Figure 2.30) (Hampton et al., 2009). Genes in these regions are known to be genuinely over-expressed, therefore blacklisting these regions from peak finding would not have been an ideal solution. The high read count in MCF7 break point regions serves to confirm the identity of the cell line.

It is probable that the poor signal to noise ratio in MCF7 cell promoters resulted from snap-freezing of cells at sample collection times over the course of the experiment. Snap-freezing of cells has previously proven to be successful using the omni-ATAC-seq protocol in mouse embryonic stem cells (Ryan-Corces et al., 2017), and is also the approach that has previously been used

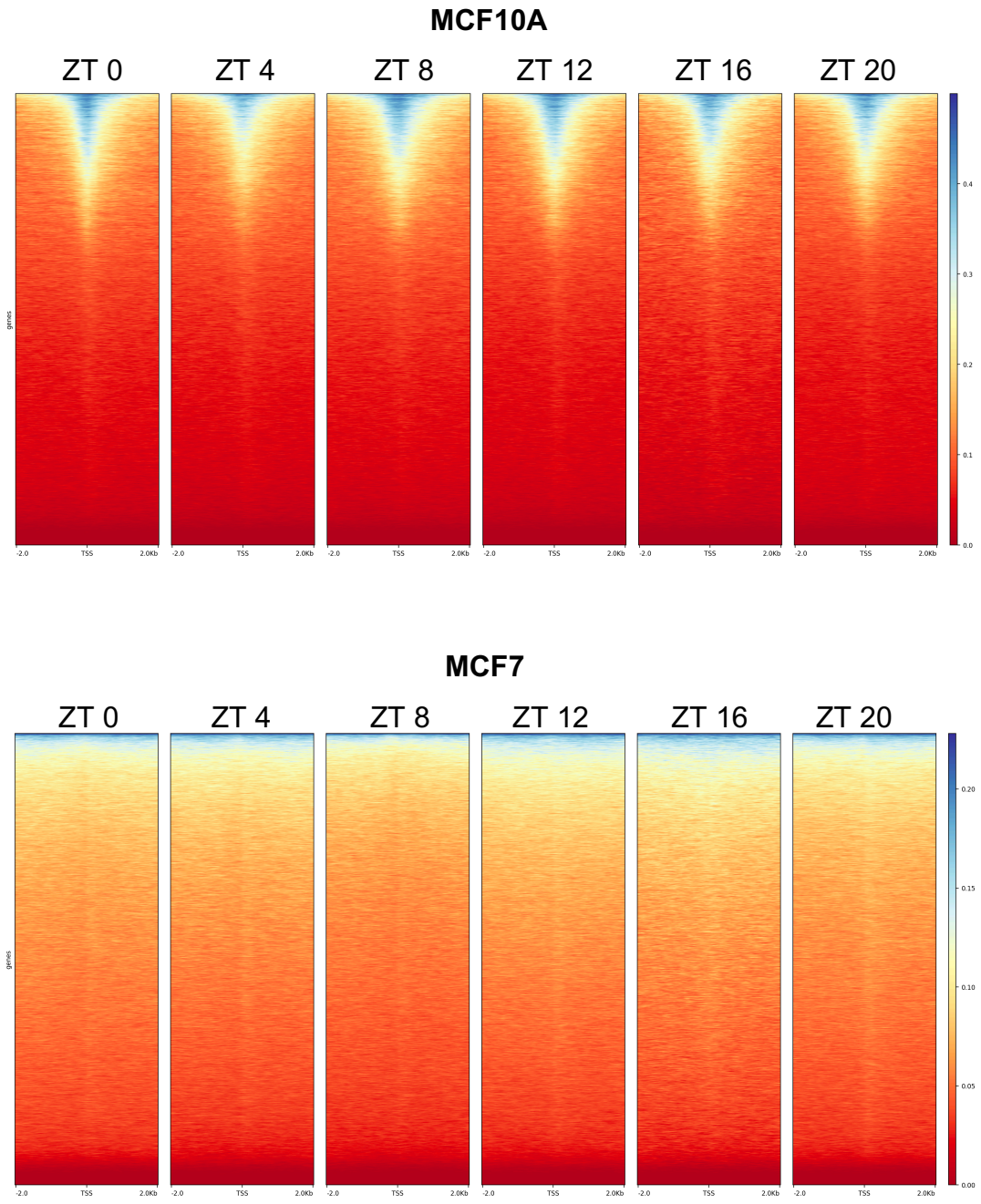


Figure 2.28 Heatmap of ATAC-seq coverage across all TSSs (± 2 kb). Read counts from duplicate samples were summated, prior to binning into genomic regions 50 base pairs in length, color was scaled across all rows and all columns.

with success in the Ott lab. However, other papers have suggested that the decrease in signal:noise resulting from snap-freezing is so great that it becomes untenable (Fujiwara et al., 2019; Milani et al., 2016). It would appear that this applies to the MCF7 data set generated here. It should be noted that the

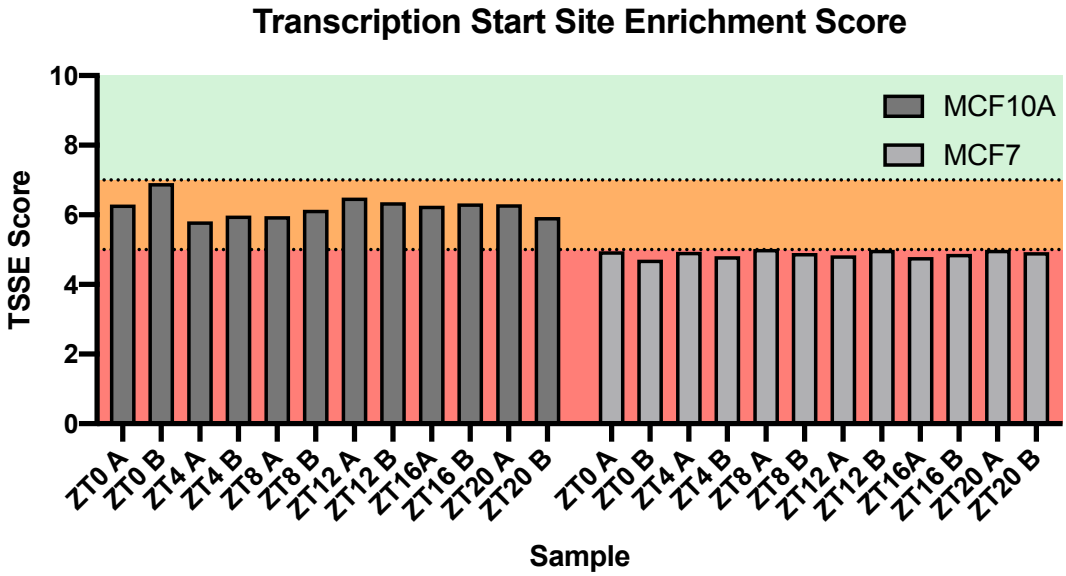


Figure 2.29 TSSE scores for MCF10A vs MCF7 ATAC-seq samples. The ENCODE consortium define the TSSE score as an important measure of data quality. For alignment to the hg38 genome, TSSE scores of >7 are considered 'ideal', 5-7 'acceptable' and <5 'concerning'.

decision to freeze-thaw the MCF7 and MCF10A samples rather than to cryopreserve was not taken lightly. The use of ATAC-seq to study circadian rhythms is essentially the same as assaying time, and freeze-thaw was viewed as being a faster method for halting nuclei at a particular circadian time. Freeze-thaw has been demonstrated to affect the size distributions of ATAC-

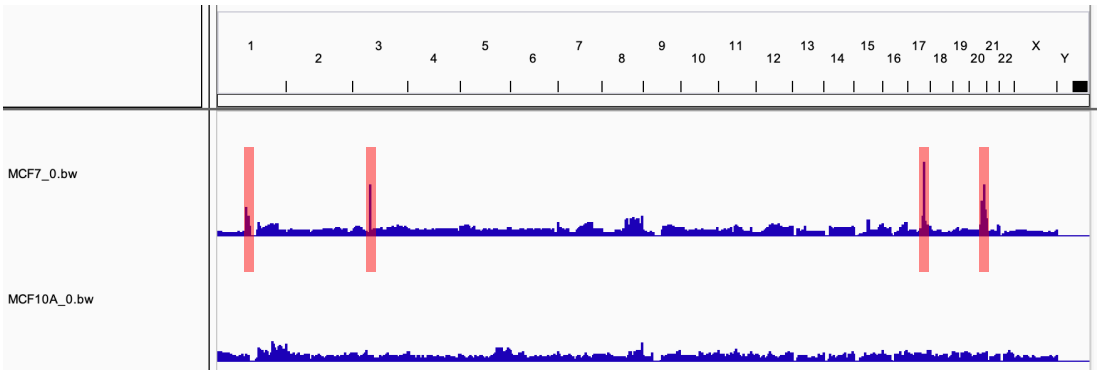


Figure 2.30 Genome-wide overview of the difference in ATAC-seq read density between MCF7 and MCF10A cells. Marked in red are four regions in the MCF7 cells that yield particularly high ATAC read counts. These red regions correspond with four known breakpoints regions that result in increased gene expression in MCF7 cells located at 1p13.1-p21.1, 3p14.1-p14.2, 17q22-q24.3, and 20q12-q13.33 (Hampton et al. 2009). Scale set to the same across the tracks.

seq fragments (Y.-J. Lee et al., 2019). ATAC-seq fragments are usually observed to exhibit periodicity in length of ~10.5 base pairs (bp) and ~200 bp. The 10.5 bp periodicity corresponds with the bias of Tn5 transposase activity resultant from the double-stranded helical structure of the DNA, whilst the 200 bp periodicity is caused by the wrapping of DNA around nucleosomes (Schep et al., 2015). Figure 2.31 displays representative fragment size distribution plots from the MCF10A and MCF7 samples. In the MCF10A samples 10.5 bp periodicity is clear, although 200 bp periodicity is almost completely lost. This pattern is consistent across MCF10A samples and has also previously been observed in the Ott lab from freeze-thawed samples. This pattern of fragment size distribution is not considered to be inhibitory to further data analysis, provided that the fragment size distributions are similar across samples. The MCF7 fragment size distribution plot has a lesser degree of apparent periodicity than the MCF10A plot. This is representative of all MCF7 samples and is in agreement with the hypothesis that the chromatin structure of MCF7 cells was disrupted to a greater degree by freeze-thaw, than the MCF10A cells. It is possible that cryopreservation of cells would have proven to be a less damaging alternative to snap-freezing for preserving the

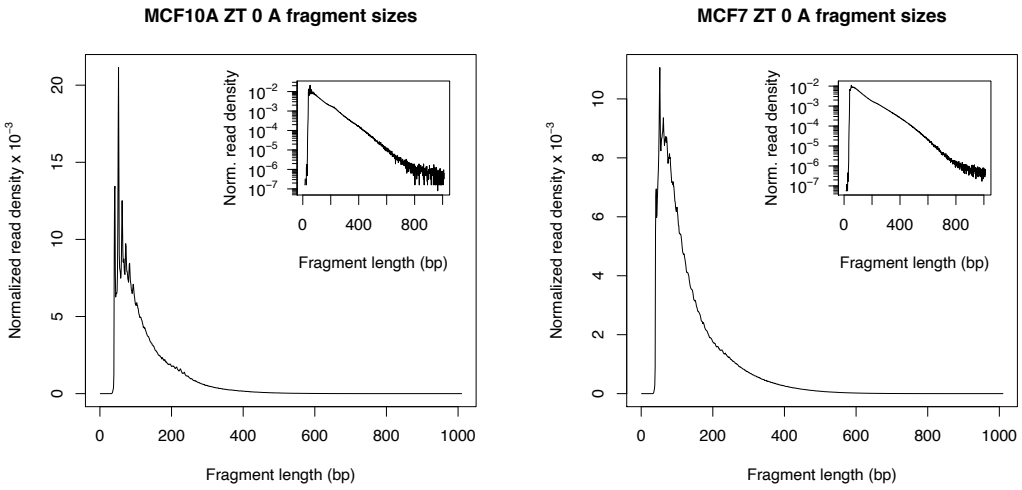


Figure 2.31 Representative fragment size distribution plots of MCF10A and MCF7 ATAC-seq samples.

chromatin structure (Fujiwara et al., 2019). Due to the poorer TSSE scores, fragment size distributions and highly accessible breakpoint regions in the MCF7 samples, only the MCF10A peak calling data will be presented.

The normalisation of ATAC-seq data is a much-discussed topic. Many methods have been proposed including but not limited to: scaling to counts per million either using the total reads in peaks as the library size, or the total number of reads in the sample; quantile normalisation by equilibrating the counts of differentially-accessible regions or of non-differentially-accessible regions; normalising to house-keeping control genes or spike-in control samples; or down-sampling of reads to the same total number per sample (C. Evans et al., 2018; Halstead et al., 2020; Reske et al., 2020). The appropriate normalisation approach depends somewhat on the data. The decision was taken to down-sample the MCF10A samples, as the library sizes were varied between samples and this was a straight-forward approach. The success of down-sampling as a normalisation method hinges on the assumptions that peaks are a relatively rare event and/or all samples have approximately similar degrees of 'open-ness' to their chromatin structure. It was not known if this was the case in MCF10A cells under temperature entrainment – in fact it seemed reasonable to presume that this may not be the case. This question was addressed by peak calling. MACS2 was used to call 67,915 peak regions ($q<10^{-3}$) in the merged, down-sampled MCF10A dataset. Background regions were calculated by selecting a series of regions $>$ 1000 bp away from any detected peaks. Counts in background regions were compared between samples, with the assumption being that if down-sampling was an appropriate method of sample normalisation, background regions would have similar read counts across samples after down-sampling. Figure 2.32A and B demonstrate the read counts in peak and background regions. The similarity in the read counts in background regions across samples after down-sampling supports down-

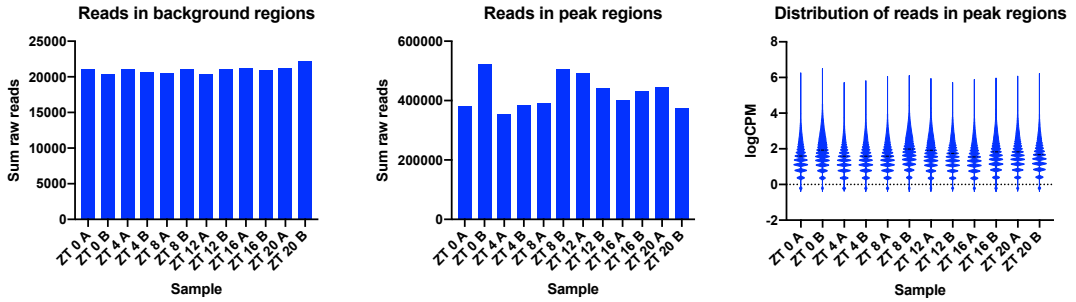


Figure 2.32 MCF10A ATAC-seq samples post-down-sampling normalisation. A) reads in background regions, B) reads in peak regions. C) Normalisation of down-sampled peak reads to logCPM (N.B. library size was taken as the sum of all sample reads, not just peak reads.)

sampling as an appropriate normalisation method for this dataset. Figure 2.32C demonstrates the distributions of reads in peaks after logCPM normalisation.

MACS2 was used to detect 55,183 peaks ($q<10^{-4}$) in the MCF10A dataset. PCA revealed that replicate samples did not cluster particularly closely (Figure 2.34). Indeed, differences between duplicate samples are also clearly visible in Figure 2.32C and Figure 2.33. None of the peaks were found to be rhythmically accessible over a 23-25 hr period by MetaCycle analysis at a B-H adjusted p-value < 0.2. Nor were any peaks rhythmically accessible over a 28-32 hr period (B-H adjusted p-value < 0.2). This may reflect a lack of rhythmicity in chromatin accessibility in MCF10A samples under 12 hr 36°C / 12 hr 37°C temperature cycles. It is certainly the case that the qPCR and RNA-seq data suggests that MCF10A cells do not appear to have a functional molecular clock. However, it is also possible that snap-freezing of the MCF10A cells damaged the chromatin structure to such an extent that the original chromatin structure was lost. Therefore, little information can be gleaned from the ATAC-seq data presented here. It would be interesting to explore the effect of fresh preparation vs cryopreservation vs snap-freezing in liquid nitrogen on signal:noise ratios in a range of different cell lines, as the data presented here,

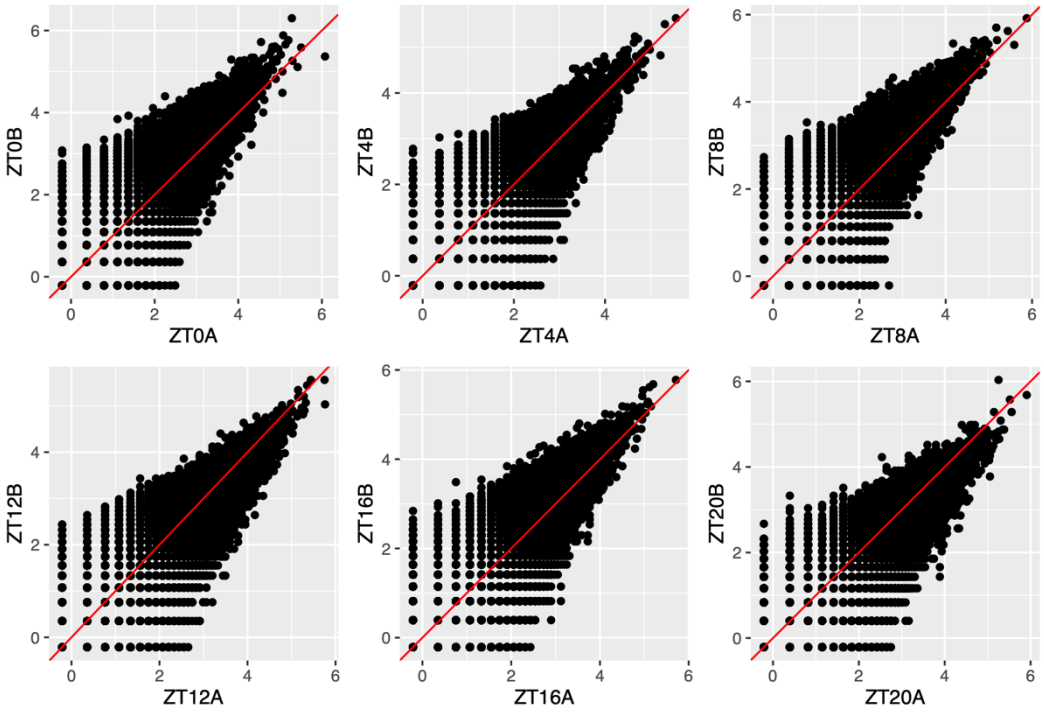


Figure 2.33 Scatter plots of peak counts for duplicate MCF10A time-point samples (logCPM).

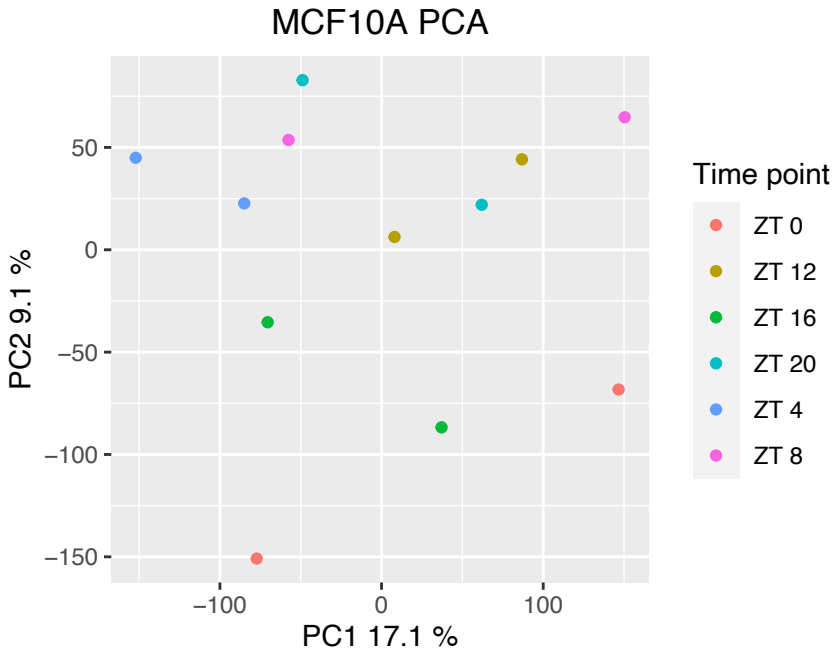


Figure 2.34 PCA of MCF10A samples. 1st PC vs 2nd PC of logCPM normalised ATAC-seq peak region counts.

and that of others, would suggest that some cell lines are more sensitive to snap-freezing than others (Fujiwara et al., 2019; Milani et al., 2016; Ryan-Corces et al., 2017). Though, in all likelihood, snap-freezing is simply best to be avoided.

2.4 Conclusions

The results presented in this chapter demonstrate that neither MCF10A nor MCF7 cells displayed clock gene promoter activity indicative of a functional molecular clock across all of the different synchronisation methods employed. MC10A cell populations generally had long periods (~27-31 hrs) of *Per2*-luc activity and poor oscillation of *Arntl*-luc activity. Entrainment to 13.5 hr 32°C / 13.5 hr 37°C temperature cycles provided the highest free-running amplitude of MCF10A *Arntl*-luc activity, though the period was >26 hrs. Meanwhile the best circadian rhythmic MCF7 promoter activity was achieved by dexamethasone synchronisation and 13.5 hr 32°C / 13.5 hr 37°C temperature oscillations. MCF7s demonstrated periods of oscillation closer to 24 hrs than MCF10As, though rhythms were of lower amplitude than MCF10A *Per2*-luc. These findings broadly support the recent findings of Lin et al. (2019) and Lellupitiyage Don et al. (2020), who also demonstrated longer 27 hr rhythms in MCF10A clock promoter activity and lower-amplitude 24 hr rhythms in MCF7 clock promoters. However, the work of Lin et al. (2019) and Lellupitiyage Don et al. (2020) was achieved in response to serum shock synchronisation protocols, which did not entrain the clocks of MCF10A and MCF7 cells particularly well in the data presented here.

The impact of the synchronisation method on circadian gene promoter activity should provide pause for thought. For instance, from the data presented here, if only the dexamethasone synchronisation data is considered MCF10As could be labelled as 'benign bad clock' cells and MCF7s as 'malignant good clock' cells. Whereas, when considering only the 13.5 hr 32°C / 13.5 hr 37°C temperature cycles it would be easy to label MCF10As as 'benign long clock' cells, and MCF7s as 'malignant damped clock' cells. The implication is that

circadian phenotypes *in vitro* are not stable enough to draw conclusions about the rhythmicity of a cell line from a single synchronisation method.

A further important indicator of a functional circadian oscillator is its ability to entrain to an indicator of external time, i.e., a Zeitgeber. This was assessed by monitoring the expression of circadian genes in MCF10A and MCF7 cells under physiological 12 hr 36°C / 12 hr 37°C temperature entrainment cycles by qPCR and RNA-seq. Both methods demonstrated that MCF7 cell populations appeared to entrain more robustly to 12 hr 36°C / 12 hr 37°C temperature cycles relative to MCF10A cells, although MCF7s also demonstrated a degree of clock dysfunction. This directly contradicts the generally accepted view that MCF10A cells have 'better' molecular clock function relative to the 'worse' clock function of MCF7 cells.

Clearly it is important that the variable clock function of the MCF cell lines – both within this body of research and within the published literature – is not overlooked or over-simplified. One such example would be that of Ye et al. (2018). Ye et al. discuss in excellent detail the differences in clock gene expression between human cancer and non-cancer biopsies. However, they re-analysed the Gutiérrez-Monreal et al. (2016) MCF microarray dataset in a manner that supported both the 2018 literature consensus of 'good clock' MCF10A, 'bad clock' MCF7, and their hypothesis that molecular clocks are disrupted in cancer. Though the original Gutiérrez-Monreal et al. paper noted poor oscillation of core clock probes, Ye et al. presented the same data in a different light by finding 'rhythmic probes' in MCF10A cells ($p<0.05$, no correction applied for multiple statistical tests) and suggesting that as these genes are expressed differently in MCF7 cells, cancer must have disrupted circadian gene expression relative to healthy tissue (Figure 2.35) (Y. Ye et al., 2018). Their supplementary data does provide the converse findings, i.e.,

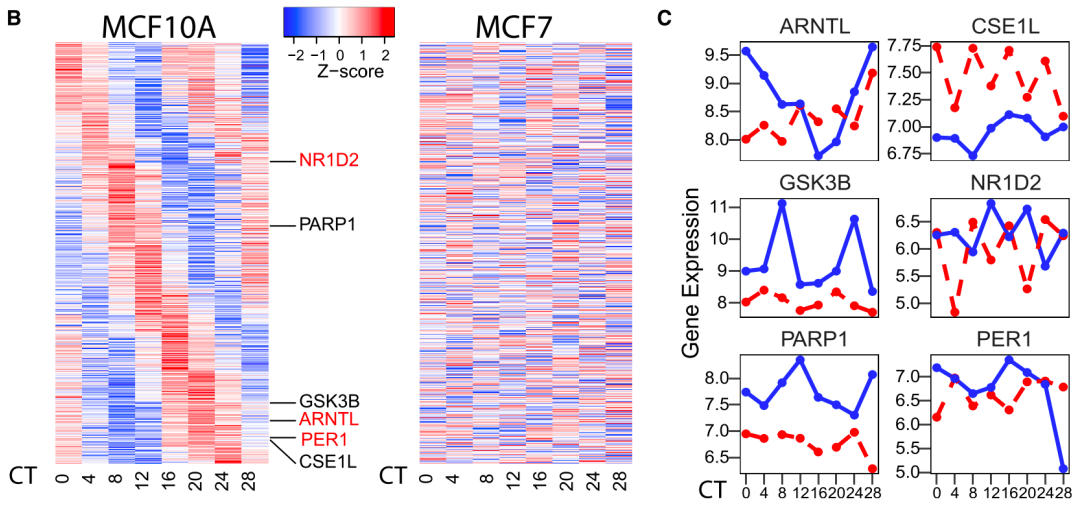


Figure 2.35 Reproduced from Ye et al. (2018), Fig 4. B) This figure presents Ye et al.'s re-analysis of the MCF microarray data set produced by Gutiérrez -Monreal et al. (2016). Circadian oscillating genes were determined by JTK_CYCLE and MetaCycle::meta2d ($p < 0.05$). Red = high expression; blue = low expression. CT refers to time post-serum synchronisation. C) Ye et al. 2018 highlight 6 clock and clock-associated probes that they suggest are more rhythmic in MCF10A cells than MCF7 cells.

rhythmic probes identified in MCF7 cells which appeared to be disrupted in MCF10A cells. However, Ye et al., present and discuss the data in such a way that the reader might be biased towards the presumption that circadian gene expression in the MCF10A cells is representative of canonical clock function in healthy tissue. The data analysis presented in this chapter represents a third interpretation of MCF rhythmicity from the same dataset generated by Gutiérrez-Monreal et al. (2016) and analysed Ye et al. (2018) by employing a more stringent metric of rhythmicity to limit false positives (B-H corrected $p<0.2$). It is important to recognise that the analysis of 'big data' sets is often less than straight-forward, and that therefore there is value in the 'mining' or re-interpretation of published data. As always, for the integrity of the scientific record, scientists should seek to avoid biased data interpretion to support a preconceived hypothesis.

A criticism that could be levelled at the experimental data presented in this chapter is that the biological replicates are not the result of independent

experiments. Whilst the importance of independent experiments is not to be dismissed, circadian experiments are particularly lengthy and are therefore time-consuming to repeat independently. The similarity in the period, phase and amplitude of promoter activity in the MCFs that were synchronised with 12 hr 36°C / 12 hr 37°C cycles in either the presence or absence of HEPES, supports the hypothesis that the data presented here are independently reproducible, given that HEPES is not expected to affect circadian gene expression (Figure 2.9, Figure 2.10).

In recent years many researchers have explored the behaviour of cells in 3D cell culture environments, which are understood to be more physiological than 2D culture (Edmondson et al., 2014). 2D vs 3D research in the breast cell context is primarily focussed on oncology research, and is concerned with how changes in the biology of cells in 2D vs 3D might affect the results of drug screens (Imamura et al., 2015; Riedl et al., 2017). Specifically, MCF10A cells cultured as 3D spheroids are known to express breast cell specific genes such as β-casein (CSN2) and α-lactalbumin, which are not expressed in 2D cultures (Qu et al., 2015). Meanwhile, MCF7 cells grown as 3D spheroids were demonstrated to be less responsive to oestrogen than 2D cultures, but are more structurally differentiated with lumina formation, luminal secretions and apical microvilli (Vantangoli et al., 2015). This is important because 2D and 3D cell culture techniques have also recently been demonstrated to affect the circadian clocks of benign breast cells *in vitro* (Williams et al., 2018). Williams et al. (2018) showed that mammary epithelial cells demonstrated stronger oscillation of PER2 in 3D than in 2D, whilst conversely mammary fibroblasts demonstrated stronger oscillatory PER2 activity in 2D rather than 3D. This finding also held true in the epithelial cells and fibroblasts of lung and skin tissue. Work by the same research group demonstrated that increasing 3D extracellular matrix stiffness, as is found in aging, damped the PER2 oscillation

in mammary epithelial cells (N. Yang et al., 2017). Since both MCF10A and MCF7 cells are of epithelial lineages, it is possible that the 2D culture environment is simply not supportive of circadian gene oscillation in these cell types (Comşa et al., 2015; Qu et al., 2015). This reflects a necessity to move towards more complex and physiological models in order to truly understand how the molecular clock genes interact with other cellular components in benign tissue versus malignant tissue.

3. Exploring circadian dysregulation in cancer *in ovo* using the chick embryo chorioallantoic membrane model

3.1 Introduction

3.1.1 Ethical *in vivo* models for circadian oncology research

The cellular circadian transcriptional-translational clock is not a closed oscillatory system, the molecular clocks of cell populations are entrainable to external stimuli by design. Synchronisation of cellular clocks either in response to entrainment stimuli such as light, or in the absence of entrainment stimuli, is for the most part coordinated by the central SCN oscillator (Cermakian & Boivin, 2009; Ginty et al., 1993; Klein & Moore, 1979; Oster et al., 2006). The cellular clocks of peripheral organ systems such as the liver and the gut are also able to entrain to Zeitgebers such as feeding schedules independently of the SCN (Damiola et al., 2000; Stokkan et al., 2001). Whether the circadian clock in tumours functions in a similar manner to peripheral and central oscillators is the subject of a great deal of research (discussed at length in chapter 1). In order to gain physiological insight into the circadian clocks of cancer cells, it is therefore important to be able to study tumours in an *in vivo* context where central and peripheral molecular clocks are operational. Of course an *in vivo* tumour graft model, also provides a number of non-circadian advantages, including a mechanical 3D environment, supportive cell populations, access to a vascular network for angiogenesis, and a degree of immune function (Holen et al., 2017).

In circadian research, as elsewhere in biology, mammals such as rats and mice are often the *in vivo* mammalian model organism of choice. Their small size, and short generation time makes them relatively easy to work with, and they are relatively well understood due to the large amount of research that has gone before. However, circadian experiments can require lengthy periods of animal isolation in dark/dark conditions (Eckel-Mahan & Sassone-Corsi, 2015). Social isolation is a known stressor for mice, whilst constant darkness has been suggested to have a depressive effect on mice and rats (Gonzalez & Aston-Jones, 2008; Ieraci et al., 2016; Y. Zhou et al., 2018). There is an ethical argument to be made for the replacement, refinement and reduction of animal use in research (Burden et al., 2015). To this end we sought to develop a novel model for *in vivo* circadian cancer research.

A recent example is that of the zebrafish model for tumour grafting (Basti et al., 2020). HCT116 tumours with knock-down of a range of core clock genes were grafted and grown in zebrafish embryos. The experiments were terminated at 7 days post-fertilisation (p-f) – N.B. prior to 5 days p-f zebrafish embryos are not considered to be a protected animal in the UK (Animals (Scientific Procedures) Act 1986, 2014). Knock-down of *ARNTL*, *PER2* or *NR1D1* in HCT116 cells had similar effects *in vitro*. All knock-down cell lines displayed increased cell growth relative to wild type (WT), and the expression of *TP53*, *WEE1* and *MYC* were also all increased. *In vitro* cell migration was reduced in all knock-down cell lines relative to WT. However, in an *in vivo* zebrafish model of tumour grafting, differences between the different knock-down cell lines became apparent. *PER2* knock-down cells formed larger tumours and metastasized faster than the WT cells, which formed larger tumours that metastasised faster than the *NR1D1* knock-down cells. Interestingly these effects were likely to have been independent of the circadian

clock of the zebrafish embryos as the researchers maintained dark/dark conditions throughout the experiment to avoid synchronising the circadian gene expression of the embryos which would normally begin to oscillate in response to light/dark cycles by 5 days p-f (Kaneko & Cahill, 2005). Basti et al. (2020) also demonstrated an absence of 24 hr oscillation of endogenous *per2* expression in the zebrafish embryos in dark/dark conditions at 6 days p-f.

Although zebrafish embryos are cheaper to house and are ethically more acceptable than the use of adult mice, they are limited as a model organism for circadian research and tumour grafting. This is because both tumour grafting and the study of circadian gene expression must be undertaken over a number of days. There are only three days between the earliest opportunity for tumour grafting and the zebrafish beginning to feed and therefore the experiment becoming a regulated procedure under ASPA, which is too short a period of time to allow for tumour engraftment and subsequent monitoring of the circadian system.

In this chapter the chicken chorioallantoic membrane (CAM) is investigated as a model of tumour grafting for a novel circadian research application. The CAM is an extra-embryonic membrane of the developing chicken embryo (Figure 3.1) (Merckx et al., 2020). The CAM is positioned immediately under the eggshell membrane and is highly vascularised to allow for O_2/CO_2 exchange for the developing embryo's respiratory needs. The chicken embryonic CAM has a very long history of use as a reductionist animal model organism in angiogenesis, oncology, virology and tissue engineering research (Komatsu et al., 2019; Marshall et al., 2020; Murphy & Rous, 1911). A chicken embryo takes 21 days to hatch, therefore after two thirds of the incubation period, i.e.,

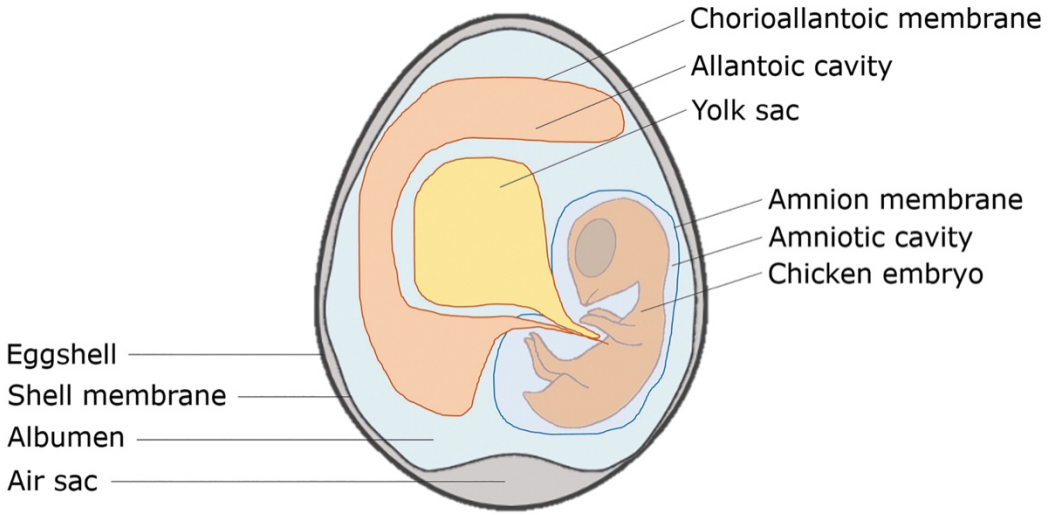


Figure 3.1 A representation of the anatomy of the developing chicken embryo and its associated structures. Reproduced from Merckx et al. (2020).

embryonic development day 14 (EDD14)[1], the developing embryo becomes protected under UK law (Animals (Scientific Procedures) Act 1986, 2014). Grafting of tumour cells onto the CAM is usually performed from EDD8, therefore 6 days are available for monitoring of tumour growth before the chicken embryos are considered to be sentient animals (Merckx et al., 2020).

The use of the early embryonic CAM for tumour grafting, instead of mice, is attractive for a number of reasons aside from its appeal as a more ethical replacement model. The chicken embryo CAM is a low-cost model that is easy to scale up, as an incubator can hold many eggs (tens to hundreds). The imaging of tumour grafts for long term monitoring is theoretically easier in chicken embryos than in mice, as the windowed eggshell provides visible access to a relatively static tumour. Additionally, since the CAM is not innervated,

[1] N.B. The start of the incubation period is denoted as 'EDD0', thus EDD0-0.99 is the 1st day of incubation, EDD13-13.99 is the 14th day of incubation and EDD14 (i.e., EDD14.0) denotes 14 whole days of incubation and marks the end of the first two thirds of the development period.

the risk of distress to the developing early chicken embryo is thought to be low (Ribatti, 2016).

3.1.2 The circadian timing system of the chicken

In humans the circadian clock is synchronised by the SCN, which is in turn entrained by light information received by the eye. However, avian species contain additional photoreceptors in their brains, including the pineal gland, preoptic area, the lateral septum, and the tuberal hypothalamus (Cassone, 2014). In both mammals and chickens the pineal gland is the region of the brain responsible for producing circadian expression of the sleep hormone melatonin, while both mammalian and avian species also produce some melatonin in the retina (Garbarino-Pico et al., 2004; Tosini et al., 2012). In avian species the light-responsive pineal gland functions as a master oscillator, in that it is required for maintenance of activity rhythms in free-running conditions (Menaker, 1968). Meanwhile the avian SCN is thought to consist of two connected structures, the medial SCN and the visual SCN, which together also ensure maintenance of activity rhythms in free-running conditions (Cassone, 2014). Coordination between the avian pineal gland and the avian SCN determines the oscillation of melatonin and clock gene expression in the peripheral tissues (Bell-Pedersen et al., 2005). Birds contain orthologues to all of the human clock genes with the exception of *PER1*, although *PER2* and *PER3* are present (Helfer et al., 2006).

3.1.3 The circadian timing system of the chicken embryo

Circadian patterns of gene expression and behaviour are established during the development of organisms. For instance, human infants display diurnal

oscillations in cortisol expression at 8 weeks of age, oscillatory melatonin at 9 weeks and oscillatory core body temperature at 10 weeks (Joseph et al., 2015). Meanwhile, in the pups of mice and rats, diurnal rhythms are apparent sooner. Oscillatory pineal melatonin has been observed in 5 day old rat pups (Trávníčková & Illnerová, 1997), whilst oscillations of the core clock genes in the SCN have been observed in 2 day old mouse pups (Ansari et al., 2009).

In avian species, circadian rhythmicity has been observed even sooner in development. Rhythmic secretion of melatonin from pineal cells was demonstrated in complete darkness from pineal glands harvested at chicken embryonic development day (EDD) 18 (out of 21 days for development). Pineal cells from EDD13 embryos even demonstrated diurnal secretion of melatonin through entrainment to 12 hr light / 12 hr dark cycles (Akasaka et al., 1995). In fact, it has been demonstrated that exposure to 12 hr light / 12 hr dark cycles on EDD13-15 (with constant light on other days) was sufficient to entrain the body temperature cycle of chicks 7 days post-hatch relative to constant light controls (Hill et al., 2004). Meanwhile 12 hr light / 12 hr dark cycles delivered from EDD7-12 resulted in diurnal variation in vocalisation in chicks 7 days post-hatch relative to constant light controls (Hill et al., 2004). In response to temperature entrainment cycles (36°C/39°C), whole chick embryos have been found to express g*ARNTL*[2], g*PER* and g*CRY* rhythmically as early as EDD4 (Paulose, 2016). This research cumulatively indicates that the chicken embryo CAM may provide a vascularised site for tumour grafting, with an entrainable circadian clock.

[2] 'g' is used throughout this thesis to differentiate chicken (*Gallus gallus*) genes/proteins from those of humans.

3.1.4 Bioluminescence imaging *in vivo* and *in ovo*

As discussed in chapter 2.1.4, bioluminescent reporters are useful tools for circadian research. The combination of bioluminescence reporters and the CAM tumour graft assay is a relatively recent development, dating to 2017 (Hafner et al., 2019; J. Hu et al., 2019; Jefferies et al., 2017; Rovithi et al., 2017). Since chicken eggs remain stationary, many eggs can be placed under one camera, which represents an improvement in throughput relative to mouse experiments. Thus far, none of the researchers that use the CAM assay in combination with bioluminescence imaging have published a protocol for longer-term continuous bioluminescence imaging. This is because for monitoring of tumour growth a 'start' and 'end' image are often sufficient metrics. However, for monitoring of rhythmic gene expression over a number of days, bioluminescence images must be acquired throughout the course of the experiment. Such experiments in mice are difficult, as the substrate of the luciferase reporter, D-luciferin, is rapidly excreted by the kidneys (Berger et al., 2008). Therefore longitudinal bioluminescence imaging in mice over a period of days requires the implantation of an osmotic pump for continual slow-release of D-luciferin (Gross et al., 2007).

It was unknown how long D-luciferin would take to be cleared from the CAM of chicken embryos. As systemic diffusion of D-luciferin is readily observed in mice (Berger et al., 2008), addition of D-luciferin to the chorionic face of the CAM was expected to diffuse quickly into the chicken embryo. The developing chicken embryo has been shown to excrete renal waste into the allantoic sac (which is bounded by the CAM) from at least as early as EDD5 (Bolin & Burggren, 2013; Fiske & Boyden, 1926). The allantoic face of the CAM acts as a barrier impervious to transport of waste urea, uric acid, calcium, potassium and phosphorous from the allantoic sac to the CAM vasculature (Gabrielli &

Accili, 2010). Therefore, D-luciferin may also be compartmentalised to the allantoic fluid by renal filtration in the developing chicken embryo.

3.1.5 Aims and objectives

The following chapter investigates the expression of core clock genes in the chicken embryo using bioluminescent reporters and qPCR. This research lays the groundwork for the subsequent use of the chicken embryo CAM as a tumour xenograft model for a circadian application.

3.2 Materials and Methods

3.2.1 Chicken embryonic fibroblast isolation and culture

Bovans brown fertilised chicken eggs were obtained from Henry Stewart & Co (https://www.medeggs.com/). Eggs were stored for no longer than 4 days between 13-18°C prior to incubation. Eggs were incubated pointed end down and gradually rocked in a Maino MiniPro 108 egg incubator at 38°C, 45% humidity. The protocol for isolation of chicken embryonic fibroblasts is adapted from Hernandez & Brown (2010) (Hernandez & Brown, 2010). On EDD10 sterile scissors were used to open the eggshell and to immediately decapitate the embryo. This is considered to be a humane method of termination for embryonic birds under ASPA, although regulation of experimental procedures under ASPA does not apply to chicken embryos until the latter third of incubation (Animals (Scientific Procedures) Act 1986, 2014). Nonetheless, all procedures were approved by the University's Animal Welfare Ethical Review Body.

Chicken embryos were transferred to a class II biosafety cabinet for the remainder of the procedure. The limb buds and inner organs of the embryo were carefully removed with sterile tweezers and scissors. The remaining tissue was gently washed in sterile PBS. The tissue was macerated using scissors to break down as much of the tissue as possible mechanically. The tissue was transferred to a 50 ml Falcon tube containing 45 ml of 0.25% trypsin (VWR, 0458-50), 1mM EDTA (Fisher Scientific, 10213570) and 2.5 mg/ml amphotericin B (Sigma-Aldrich, A2942). Falcon tubes were incubated at 37°C for 45 min with frequent agitation, after which cell material was pelleted at 500 ×g for 5 min. The supernatant was removed and replaced with fresh

trypsin/EDTA/amphotericin B for a further 15 min incubation period. Tubes were centrifuged at 500 × g for 5 min to pellet the cells, and the supernatant removed. The cells were resuspended in the same media used for MCF7 culture in chapter 2.2.1 (DMEM with UltraGlutamine, 10% FBS, 20 U/ml penicillin and 20 µg/ml streptomycin). Cell suspensions from different embryos were pooled prior to seeding. Two 175 cm^2 flasks were seeded per embryo used, and were incubated at 38°C, 5% CO_2 overnight. The following day, the media in the flasks was replaced to remove cellular debris.

Within one passage fibroblasts growth outstrips that of any other cell type to produce a monoculture of contact inhibited primary chicken embryonic fibroblasts (CEFs). Cells may either be cryopreserved (10% DMSO, 90% FBS), or maintained for 5 or 6 passages in total (maximum 30 days) before they lose their ability to replicate.

3.2.2 Generation of stable *Per2*-luc and *Arntl*-luc reporter CEF cell lines

CEFs were stably transduced with *Per2*-luc and *Arntl*-luc constructs by lentiviral vectors as described in chapter 2.2.2. The protocol differed only slightly in the transfection reagent employed – Fugene HD (Promega, E2311).

3.2.3 Bioluminescence assay for CEF clock promoter reporter cell lines

CEFs were seeded at 50,000 cells per 3 ml of culture media in 35 mm diameter tissue culture dishes. The following day the media was removed from the dishes, and the CEFs were synchronised with 3 ml of fresh media containing

either 10 µM forskolin or 100 nM dexamethasone. Forskolin-containing media also contained 100 µM potassium D-luciferin for bioluminescence monitoring. CEFs synchronised with 100 nM dexamethasone were washed with PBS after 20 min at 37°C in order to remove the dexamethasone. The dishes were replenished with cell culture media containing 100 µM potassium D-luciferin. Bioluminescence was monitored by the ActiMetrics LumiCycle 32 device. Data was detrended as described in chapter 2.2.4. Extra sum-of-squares F tests were implemented to determine whether sinusoidal fits fitted the data significantly better than linear fits ($p<0.05$) as described in chapter 2.2.4.

3.2.4 Chick embryo tissue sampling and subsequent RNA extraction

Chick eggs were incubated in a range of conditions including 38°C constant darkness, 38°C 12 hr light / 12 hr dark, and 12 hr 39°C / 12 hr 36°C constant darkness. For the purposes of ensuring a light-tight environment with oscillatory temperature control, eggs were incubated in a New Brunswick Galaxy 170 S incubator. The CO_2 supply was switched off, and air was pumped in using a simple aquarium pump covered by a light tight cloth to ensure sufficient air flow. A second Galaxy 170 S incubator with an aquarium pump was illuminated every 12 hrs for 12 hrs with a 5 m length strip of 7000 Kelvin LEDs to achieve 12 hr light / 12 hr dark conditions. For constant darkness, constant temperature conditions eggs were incubated in dark-adapted Amerex Instruments hybridisation ovens with a cup of water for increased humidity. The hybridisation ovens were placed in a dark room. Eggs were turned by hand in the dark every 12 hrs for the first 36 hrs to prevent embryos from sticking to the eggshell, which would have reduced the viability.

One hour before onset of EDD12, eggs were removed from the incubators every 4 hrs using a large blackout cloth to protect the incubators from very dim room light upon opening. Embryos were killed in conditions as close to darkness as was possible. Sterile tweezers and scissors were used to remove the embryonic brain and a clean piece of CAM tissue approx. 2 cm in diameter. Tissue was placed into screw cap 2 ml tubes pre-filled with 400 µl of TRIzol and 200 µl of zirconia beads (Thistle Scientific, BSP-11079107zx). The tissue was homogenised in a VelociRuptor V2 set to 7m/s for 20 seconds. Samples were immediately placed on dry ice and stored at -80°C until RNA extraction. As CAM is quite a tough tissue to homogenise, some tubes containing CAM required an additional homogenisation cycle. In this case tubes were allowed to chill on dry ice between cycles to prevent the sample from over-heating during homogenisation.

For RNA extraction all samples were thawed from -80°C and briefly vortexed to ensure a well-mixed sample. 100 µl of sample was transferred to a 1.5 ml tube prefilled with 900 µl of TRIzol and vortexed briefly to mix. TRIzol/chloroform RNA extraction was performed as described for MCF cell material in chapter 2.2.5. The Turbo DNA-free kit was used to remove any trace amounts of DNA from the RNA, and the quality of the RNA was checked by gel electrophoresis (1% w/v agarose gel in Tris/borate/EDTA buffer).

3.2.5 qPCR of chicken embryonic tissue samples

DNase-treated RNA was converted to cDNA using the superscript II reverse transcriptase kit. The primers used for g*ACTB*, g*ARNTL*, g*PER3*, and g*RPLP0* are listed in Table 3.1. All primers were designed in NCBI's Primer Blast tool with melting temperatures of 58.1-60.2°C.

Table 3.1 Forward and reverse primer sequences for chicken housekeeping and clock genes.

Gene	Forward primer sequence	Reverse primer sequence
g*ACTB*	5' ATATTGCTGCGCTCGTTGTT 3'	5' CGACCCACGATAGATGGGAA 3'
g*ARNTL*	5' TACGTGGTGCTACAAACCCA 3'	5' GAAATCCATCTGCTGCCCTG 3'
g*PER3*	5' ATCTGCGGAGCCTTCACTTA 3'	5' GTGCATTCACTAGAAGCGCA 3'
g*RPLP0*	5' GTTTCCTGGAGGGTGTTCGT 3'	5' AAGCACCCGCTTGTAGCC 3'

For qPCR experiments, samples were assayed in 384 well plate format using a BioRad CFX384 Touch Real-Time qPCR System. Each biological sample was measured in triplicate to account for technical error. Each 20 µl qPCR reaction comprised of 10 µl PowerUp SYBR Green Master Mix 2 µl of pre-mixed forward and reverse primer at 0.5 µM each, 2 µl of cDNA (diluted 1:10 from cDNA conversion) and 6 µl DNase/RNase–free water. The temperature profile applied during the experiment was as follows: 2 min 50°C, 2 min 95°C followed by 40 cycles of 15 sec 95°C, 30 sec 60°C, 30 sec 72°C. Primer efficiency was determined by qPCR of serially diluted pooled cDNA to a) confirm that the relationship between the amount of cDNA and Ct remained logarithmic, and b) determine the efficiency of amplification (Figure 3.2). The Pfaffl method of relative quantification of cDNA was used to calculate the $\log_2$ fold change of each gene as described in chapter 2.2.6 (Pfaffl, 2006). For each time-course, extra sum-

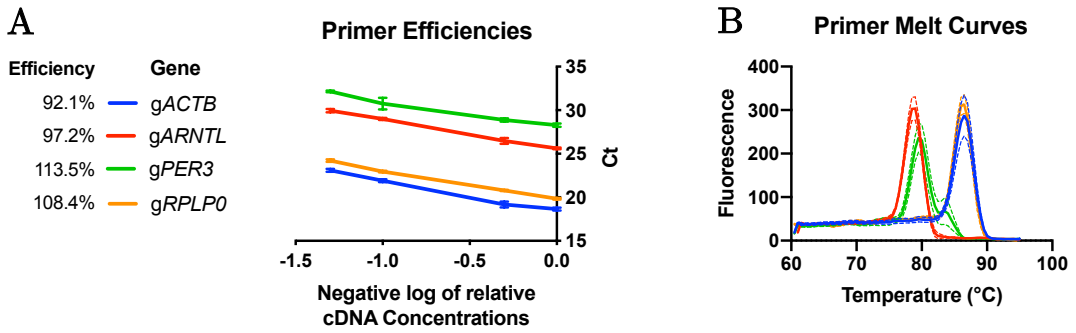


Figure 3.2 qPCR cDNA dilution curves and primer melt curves for chicken genes. A) cDNA dilution curves for the determination of primer pair amplification efficiency across different concentrations of cDNA. B) Melt curves for each primer pair (line = mean, dotted line = S.D., n=3).

of-squares F tests were implemented using GraphPad Prism v8 in order to determine whether a sine fit constrained to a period of 24 hr was a better fit for the data than a linear fit at a significance threshold of 0.05.

3.2.6 Cell culture

MCF10A and MCF7 cells were cultured as described in chapter 2.2.1. MDA-MB-231 cells (human mammary adenocarcinoma cell line), Hepa1-6 cells (murine hepatoma cell line) and C26 cells (murine colon carcinoma cell line) were all cultured in DMEM with Ultraglutamine, 10% FBS, x20 U/ml penicillin and 20 µg/ml streptomycin as described for MCF7 cell culture in chapter 2.2.1. All cell lines were stably transduced with *Per2*-luc or *Arntl*-luc lentiviral constructs as described in chapter 2.2.2.

3.2.7 Windowing of chicken eggs for tumour grafting

Eggs were incubated as described in chapter 3.2.1 in a Maino Mini Pro 108 egg incubator. After 67-77 hrs of incubation (~EDD2.8–EDD3.2) chicken eggs were moved in batches of six to a class II biosafety cabinet for windowing – N.B. a class II biosafety cabinet was used for any procedure which required that the eggs be opened. Eggs were inverted to sit pointed end up upon removal from the incubator to reduce the chances of the embryo sticking to the inner membrane. For the experiment described in chapter 3.3.3, eggs were windowed at the base, top or side or the egg. Windowing was achieved by using sharp sterile tweezers to pierce the eggshell and slowly remove fragments of shell until a window of approx. 2.5 cm diameter had been achieved. The young chick embryo always rotates to the upwards facing side of the egg, which enabled dead embryos or unfertilised eggs to be observed and discarded upon

windowing. The exact size of the window varied slightly from egg-to-egg due to the size and shape differences between eggs. When windows were created at the top or side of the eggs it was necessary to remove 0.5-1.0 ml of albumin from the egg by sliding a 1 ml pipette into the window past the edge of the yolk. This lowered the embryo sufficiently to allow for the window to be sealed with a piece of wide width sticky tape cut to the right size, without risk of injury to the developing embryo. For eggs windowed at the base, removal of albumin was not necessary as the air sac is located directly under the base of the eggshell, creating a void space for windowing. Additionally, for eggs windowed at the base the inner membrane was carefully removed with tweezers in order to expose the embryo, after which the window was sealed with sticky tape. Once eggs were returned to the incubator they were no longer rocked/rolled or agitated in any manner in order to avoid the sticky tape from contacting the embryo.

Cancer cells were usually grafted onto eggs on EDD8, once the CAM had grown sufficiently large to accept and support a graft. Cells were pre-mixed with 20 µl of ice-cold growth factor reduced Matrigel (Corning, 356231) per egg to ensure a defined graft location. Grafts were placed at junctions of large central blood vessels to aid angiogenesis. In later experiments the aim was to graft 4 million cells per egg, though the cell number varied according the number of cells available for grafting. This was not too concerning as the number of cells recommended by the literature for tumour grafting onto the CAM varies widely from 10^3- $>10^6$ depending upon the cell line (Jefferies et al., 2017; Pawlikowska et al., 2020). Table 3.2 details the cell line and graft density used for each of the tumour grafting CAM experiments described in this chapter.

3.2.8 Tumour graft bioluminescence imaging

3.2.8.1 D-luciferin pulse

On the day of imaging, the sticky tape was pierced with a sterile needle in order to deliver potassium D-luciferin dissolved in sterile water directly over the grafted tumour. The concentration of the potassium D-luciferin depended on the experiment (see Table 3.2 for details), whilst the volume varied between 100-300 µl. Eggs were resealed with a small piece of sticky tape and were immediately transferred to the Cairn Alligator bioluminescence imaging system for monitoring of tumour bioluminescence. Fresh potassium D-luciferin was administered at the end of imaging prior to termination of the experiment. See Table 3.2 for details.

3.2.8.2 D-luciferin drip

Eight lengths of tubing (Crystal Clear Tygon R3607, Saint Gobain Performance Plastics, 15113226) were bundled together with black electrical tape to pass through the light-tight port into the Cairn Alligator. Tubing was sterilised by flushing 70% ethanol through and left to soak for 30 minutes. Tubing was subsequently flushed with sterile water (Elga PURELAB system). One end of the tubing was secured just inside the sticky taped window of embryos with tumour grafts. The other was secured to a syringe pre-filled with D-luciferin solution (see Table 3.2 for concentration/volumes). Syringes were placed in an Aladdin 8-channel syringe pump (WPI Instruments, AL-1800) which was covered by a light-tight blackout cloth for the duration of imaging (Figure 3.3).

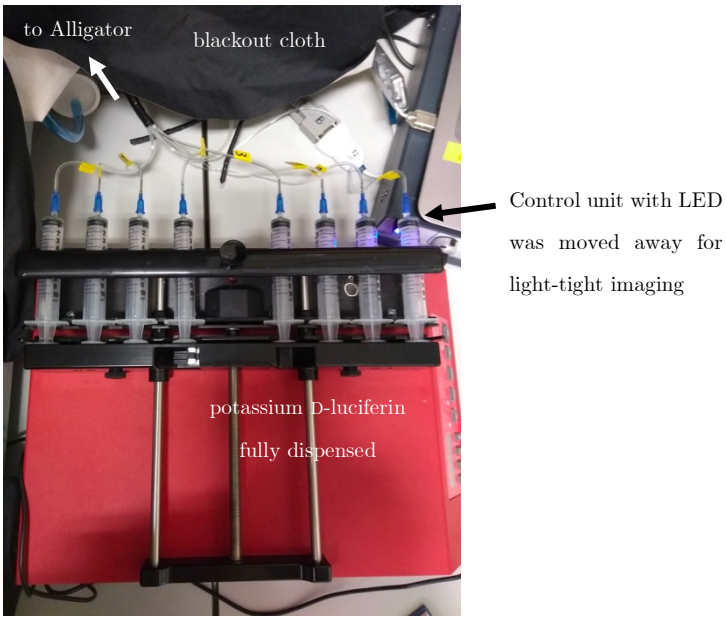


Figure 3.3 Aladdin 8-channel syringe pump (WPI Instruments, AL1800) for continuous delivery of potassium D-luciferin to chicken embryos.

3.2.9 Luciferin loaded polymer patch

Luciferin loaded polymer patches were synthesised in Professor Dave Haddleton's laboratory by Shivani Rughani (Department of Chemistry, University of Warwick) using a protocol developed by Medherant Ltd which was originally intended for the development of polymer patches for transdermal drug delivery. The polymer patch consisted of a poly(ether-urethane)-silicone crosslinked pressure-sensitive adhesive (PSA) mixed with 10% luciferin by weight (Tombs et al., 2018). The PSA was spread onto a thin layer of acetate, and a backing layer was adhered to complete the patch. The luciferin used was either potassium D-luciferin for batch 1 and 2 patches, or cyclic alkylaminoluciferin (CycLuc1) for batch 2 patches (see Figure 3.4). CycLuc1 is a modified luciferin which is >10-fold brighter than D-luciferin *in vivo* (M. S. Evans et al., 2014). However, no difference in CAM graft bioluminescence intensity was observed between patches loaded with potassium D-luciferin compared to patches loaded with CycLuc1 (data not shown), therefore the patches were used interchangeably in experiments. Upon synthesis of the

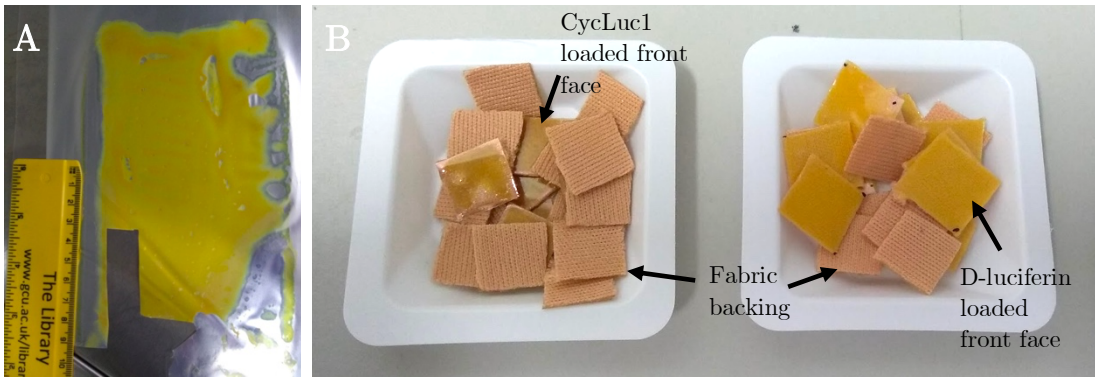


Figure 3.4 Luciferin loaded polymer patches. A) Batch 1, potassium d-luciferin-loaded polymer patch. PSA was coated onto a foiled backing. B) Batch 2, LHS = CycLuc1-loaded polymer patch. RHS = potassium d-luciferin-loaded polymer patch. PSA was coated onto a beige fabric backing.

luciferin-loaded patch the CycLuc1 changed colour to a darker yellow, therefore it is possible that CycLuc1 was modified in the process of the polymer patch production, which might have resulted in similar graft bioluminescence to D-luciferin patches. Since the batch 2 1 cm^2 polymer patches were backed onto non-sterile beige fabric, 50 µl of 2000 U/ml penicillin and 2 mg/ml streptomycin solution was added onto (and was absorbed by) the beige fabric after the placement of the polymer patch onto each chicken CAM to try to limit the risk of infection to the chicken embryo.

3.2.9.1 Image acquisition parameters

Image acquisition parameters varied slightly between experiments. The majority of experiments employed the CAIRN Alligator, whilst two experiments used the Biospace Labs PhotonImager Optima imaging system. All imaging parameters are described in Table 3.2.

3.2.10 Bioluminescence image analysis

All .tif images were exported to FIJI-ImageJ for analysis. Images from the CAIRN Alligator contained a degree of high intensity random background

Table 3.2 Supplementary protocol information for all CAM tumour graft experiments.

EDD0 Date	**01/12/2017**		**20/03/2018**	**21/05/2018**		**29/05/2019**	**20/09/2019**	**04/12/2019**	**04/12/2019**		**07/08/2020**		**09/12/2020**	
Graft day	EDD9		EDD8	EDD8		EDD8	EDD8	EDD8	EDD9		EDD8		EDD9	
Cell type grafted	MCF7 *Arntl*-luc	C26 *Arntl*-luc	MCF10A *Arntl*-luc	Hepa1-6 *Per2*-luc	MDA-MB-231 *Arntl*-luc	Hepa1-6 *Arntl*-luc	MCF10A *Per2*-luc	MCF7 *Arntl*-luc	MDA-MB-231 *Arntl*-luc	MCF7 *Per2*-luc	MCF10A *Per2*-luc	MCF7 *Per2*-luc	MCF10A *Per2*-luc	MCF7 *Per2*-luc
Cell number grafted/egg	152,000	250,000	334,000	6,160,000	6,320,000	6,000,000	4,000,000	4,000,000	4,000,000	4,000,000	4,000,000	4,000,000	4,000,000	4,000,000
Imaging commenced	EDD9		EDD11	EDD11		EDD11	EDD10	EDD10	EDD12		EDD11 A&B, EDD12 C-F		EDD11	
Luciferin delivery method	Pulse		Pulse	Continuous drip		Continuous drip	Polymer patch + pulse	Polymer patch	Polymer patch		Polymer patch		Daily pulse	
Luciferin volume and concentration	100 µl 209 mM potassium D-luciferin + 50 µl 209 mM @ EDD13.9		1 ml 157 mM potassium D-luciferin	13.9 µl/hr 20.7 mM potassium D-luciferin for 72 hr		139 µl/hr 2.07 mM potassium D-luciferin for 72 hr	Batch 1 1 cm^2 patches +100 µl 20 mM potassium D-luciferin @ EDD12 for eggs A,B & C	Batch 2 1cm^2 patches, 10% luciferin	Batch 2 1cm^2 patches, 10% luciferin		Batch 2 1cm^2 patches, 10 % luciferin (where polymer coated patch unevenly, size was increased proportionally)		90 µl 10 mg/ml potassium D-luciferin per day to match Jefferies et al. (2017)	
Camera and image acquisition parameters	Cairn Alligator, exposure = 30 min, bin=2, gain=3, readout mode = 30 MHz		Cairn Alligator, exposure = 30 min, bin=1, gain=3, readout mode = 30 MHz	Cairn Alligator, exposure = 30 min, bin=1, gain=12, readout mode = 1 MHz		Cairn Alligator, exposure = 30 min, bin=1, gain=12, readout mode = 1 MHz	Cairn Alligator, EDD10-11 exposure = 30 min, bin=1, gain=12, readout mode = 1 MHz. EDD12-13, exposure = 60	Cairn Alligator, exposure = 30 min, bin=2, gain=3, readout mode = 1 MHz	PhotonImager Optima, exposure = 20 min		PhotonImager Optima, exposure = 20 min		Cairn Alligator exposure = 15 min, bin = 1, gain = 12, readout mode = 1 MHz	

noise. The FIJI 'remove outliers' function was used to reduce this (threshold: 50, pixels: 2-6). Bioluminescent grafts and a bioluminescence-free (i.e., background) part of the image were selected using the FIJI region of interest (ROI) manager. Mean background intensity was subtracted from mean graft intensity for each graft in each image. The background can be a little variable across the image, for instance a particularly bright tumour graft may slightly increase the noise of neighbouring CAMs under the camera. For most experiments this was negligible and not an issue, however it did present a challenge for quantification of the data presented in Figure 3.14. To account for variable background in Figure 3.14A, the background ROI was set 'locally' i.e., a separate background ROI was created for each tumour graft, as close as possible to each tumour graft. In order to fit a graft ROI and a background ROI into each of the CAM regions, the ROIs were required to be slightly different shapes and sizes to one another (although paired graft and background ROIs always matched in size and shape). Therefore, for the locally background subtracted data presented in Figure 3.14B, mean bioluminescence of graft A cannot be quantitively compared to mean bioluminescence of graft B (because mean bioluminescence is dependent upon the size of the ROI). For analysis of images from the PhotonImager Optima, the 'remove outliers'

function was not necessary. Additionally, the mean bioluminescence was extremely small. Therefore, sum bioluminescence was plotted instead.

3.3 Results and Discussion

3.3.1 Circadian clock function of chicken embryonic fibroblast cells

Primary chicken embryonic fibroblasts (CEFs) extracted from EDD10 embryos were stably transduced with murine luciferase reporters of *Arntl* or *Per2* clock promoter activity. Figure 3.5 demonstrates that both dexamethasone and forskolin were capable of producing oscillatory *Arntl*-luc activity in CEFs with an approx. 24 hr period. The *Arntl*-luc oscillations were quite low in amplitude, which is clear from the degree of noise in the oscillatory signal. Meanwhile,

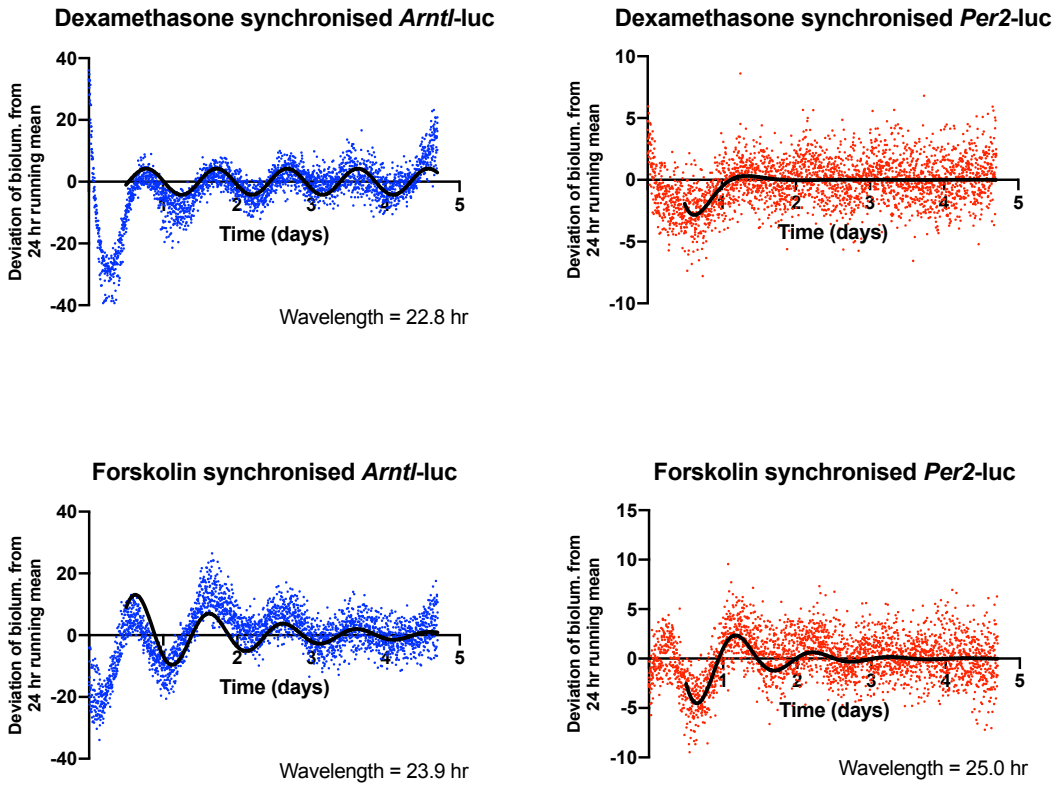


Figure 3.5 Clock promoter activity in chicken embryonic fibroblasts (CEFs). CEFs were synchronised either with dexamethasone or forskolin. Standard or damped sine fits were preferentially fitted to detrended data corresponding to n=4 experiments for each condition in Graphpad Prism. The first 12 hrs of data was omitted from the sine fit to ensure that 24 hrs of data was available for the sliding window used for running average detrending.

oscillatory *Per2*-luc activity was only detectable after forskolin synchronisation, and rhythmicity was lost by 2.5 days post-synchronisation. This could mean that the murine clock promoter constructs were poorly regulated by the chicken transcriptional machinery, or that a majority of cells in the population had a very low amplitude (or lack) of circadian gene activity. The former possibility seems unlikely, as the raw expression of the bioluminescence signal was sufficiently high that activation of clock promoters and detection of luciferase activity did not appear to be a limitation of the assay. The latter scenario that the CEF circadian clock genes simply had a very low amplitude of oscillation seems more likely, as it has already been demonstrated that clock gene expression is not rhythmic in mouse embryonic stem cells, but that rhythmic gene oscillation is established gradually as stem cells differentiate (Yagita et al., 2010). The results presented in Figure 3.5 lend confidence that the cells of the chicken embryo may well express circadian clock genes capable of driving the bioluminescent reporter expression, even if the molecular clock is not yet oscillating at a high amplitude in a free running manner.

3.3.2 Expression of circadian genes in the chick embryo brain and CAM at EDD12-14

qPCR was used to assess the expression of circadian gene transcripts in the developing chick embryo brain and CAM at EDD12-14. The developing chicken embryos were incubated under a variety of conditions described by Figure 3.6. Constant temperature (38°C) and constant darkness was used to ascertain the free-running expression of circadian clock genes. Two Zeitgeber incubation protocols were also explored; firstly 12 hr 39°C / 12 hr 36°C temperature cycles were employed as described by Paulose (2016). The cloacal

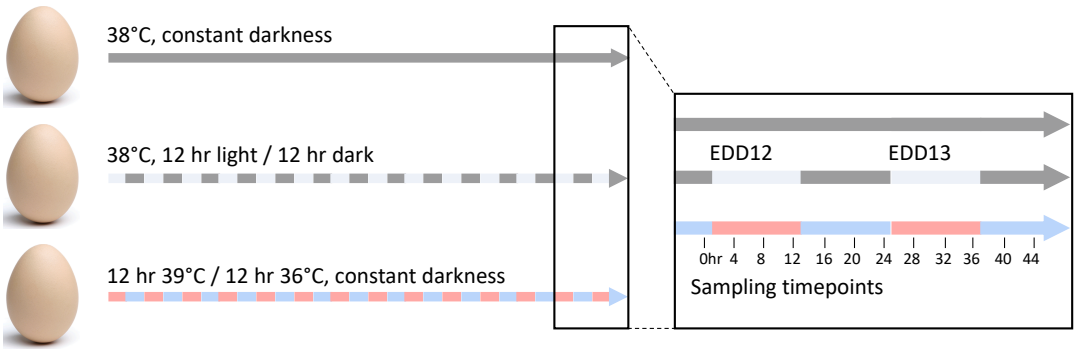


Figure 3.6 Protocol for chicken egg incubation for sample collection for qPCR.

temperature of chickens has been shown to vary by approx. 1°C over the 24 hr period with the peak temperature between midday and mid-afternoon (Aluwong et al., 2017). It is therefore reasonable to presume that between the temperature cycle of the hen and the temperature cycle of the environment, chicken embryos incubated by a hen would experience oscillatory temperature cycles of cooler nights and warmer days. Broody hens will leave a clutch of eggs for no longer than 20 minutes per 24 hr period, therefore the temperature of the eggs should closely track that of the hen. The second entrainment signal investigated by qPCR was a 12 hr light / 12 hr dark cycle. It is likely that very little light would reach developing embryos incubated by broody hens. However, that does not mean that chicken embryos cannot entrain to light/dark cycles, as discussed in chapter 3.1.3. Figure 3.6 describes the protocol followed for incubation of chicken eggs and subsequent tissue sampling for qPCR.

Figure 3.7 shows that g*ARNTL* was uniformly expressed in a non oscillatory manner in the brain and the CAM of the chicken embryos regardless of incubation conditions. g*PER3* expression in the CAM under 12 hr light / 12 hr dark entrainment was the only condition that fitted a 24 hr period sine wave, better than a straight line by sum of squares F test ($p<0.05$). From the mean (black line) it is apparent that g*PER3* appears to be somewhat oscillatory

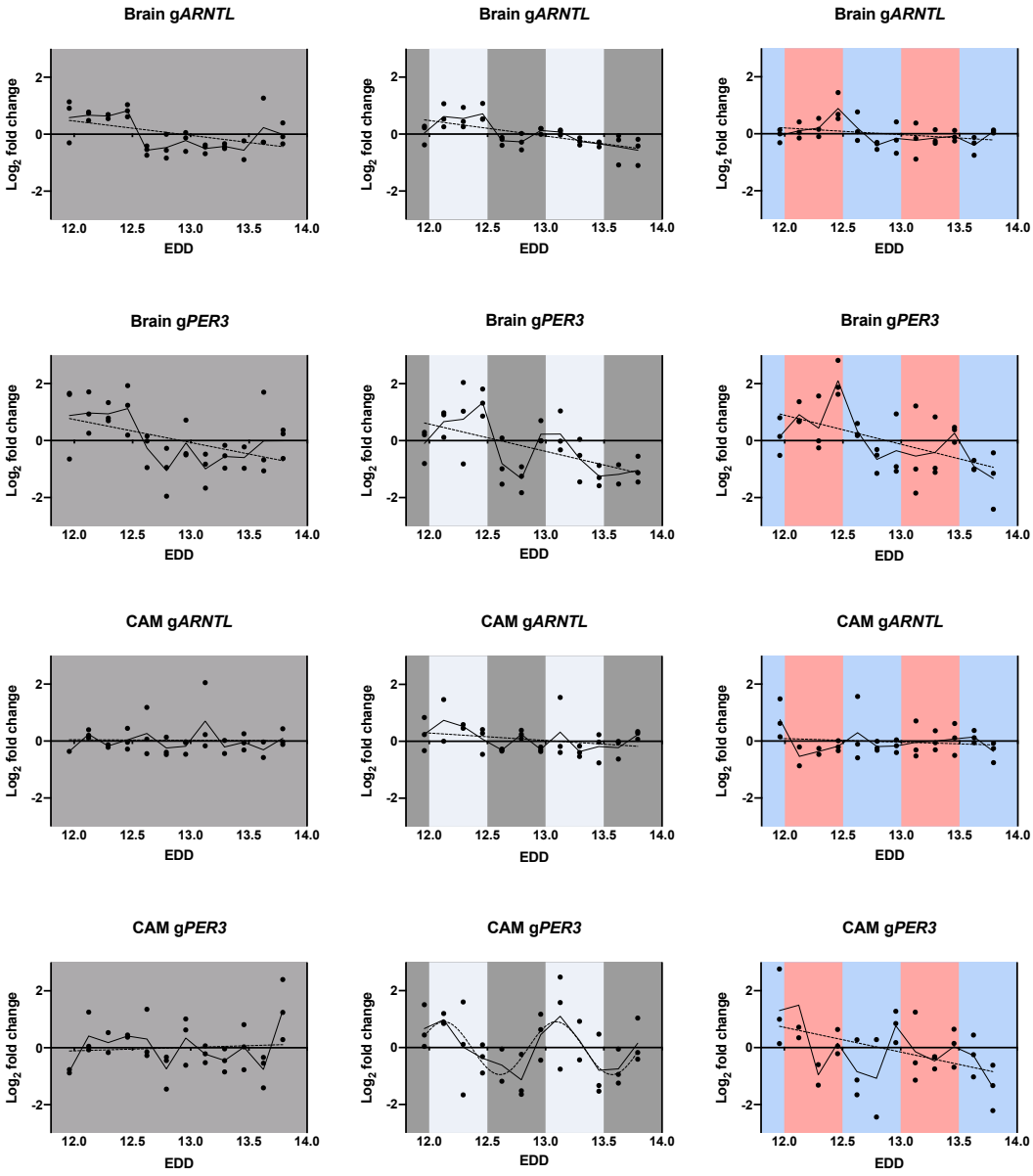


Figure 3.7 qPCR of clock genes g*ARNTL* and g*PER3* in the brains or CAMs of EDD12-14 chicken embryos incubated either under constant temperature, constant darkness (dark grey), 12 hr light / 12 hr dark (light grey / dark grey) or 12 hr 39°C / 12 hr 36°C entrainment (red/blue). ACTB was used as the reference gene. N=3 biological replicates except where technical triplicates did not agree, 19 data points removed out of a total of 432. Solid line represents mean of biological replicates. Sum of squares F tests were used to determine whether sine fits (period = 24 hrs) surpassed linear fits ($p<0.05$), represented by dashed line.

under both light/dark and 39°C/36°C entrainment, with peak expression occuring sooner in the CAM than in the brain.

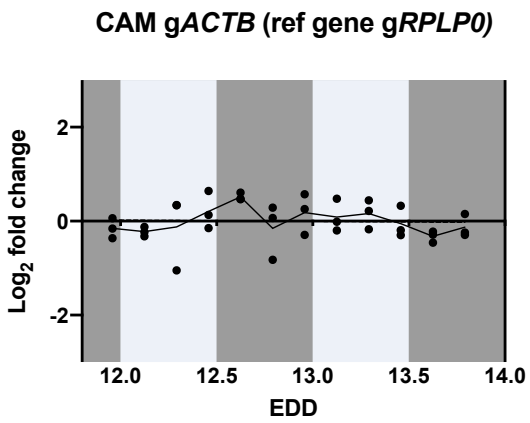


Figure 3.8 RPLP0 expression normalised to ACTB expression in chicken embryos under 12 hr light / 12 hr dark entrainment. N=3 biological replicates, solid line represents mean of biological replicates. No discernible trend, which indicates that g*RPLP0* and g*ACTB* were stably expressed in the chick embryo CAM over time.

g*ACTB* expression normalised to g*RPLP0* expression remained stable over time in the CAM under light/dark entrainment conditions, which supports the use of g*ACTB* as a reference gene for circadian experiments (Figure 3.8).

Whilst research on *PER3* expression during embryonic development is rare, studies on rats have shown that the paralogue *Per2* is expressed at a higher amplitude than *Arntl* in both the foetal SCN and the foetal liver (Houdek & Sumová, 2014; Varcoe et al., 2013). In a similar vein, weak oscillation of g*PER3* in chicken embryos was observed in Figure 3.7 in response to entrainment stimuli, whilst expression of g*ARNTL* was not oscillatory. The implication is that some genes may demonstrate increased oscillatory capability slightly earlier in development than others. Given the absence of oscillatory expression of g*ARNTL*, the weak oscillation of g*PER3* in the presence of Zeitgeber stimuli, does not indicate a functional embryonic molecular clock.

3.3.3 *In ovo* CAM assay development

In order to access the embryonic CAM for tumour grafting the embryos can either be incubated *in ovo* or *ex ovo*. *Ex ovo* culture methods afford better accessibility to the CAM for imaging purposes but yield quite low survival rates (data not shown). *In ovo* incubation requires a window to be created in the eggshell at EDD3, which is resealed in order to provide access to the CAM later in the incubation period for tumour grafting and imaging. Understandably, *in ovo* culture is less disruptive to the embryonic environment, therefore survival rates are generally higher (Naik et al., 2018). Protocols for *in ovo* windowing are abundant in the literature and are varied in many parameters, including the location of the window in the eggshell. However, a search of the literature revealed no studies that compared the effect of the location of the window on embryonic survival.

The viability of a batch of chicken embryos subjected to a combination of different incubation and window location methods is displayed by Figure 3.9. Embryonic survival to EDD13 was found to be very similar between eggs that were either rolled or rocked prior to windowing at the base (62.5% and 66.7% respectively). Nor was there a difference in embryonic survival to EDD13 between eggs incubated either in a purpose built Maino MiniPro 108 humidified egg incubator, or in an Amerex Instruments hybridisation oven with a cup of water for increased humidity (66.7% and 65% respectively)). Windowing at the side of the eggs resulted in a very low EDD13 viability (25%). This came as no surprise, as creating a window in the side of the eggshell was observed to be rather destabilising to the structure of the egg, and easily resulted in the appearance of hairline cracks in the eggshell. Windowing at the top (apex) of the egg yielded the highest EDD13 viability of 81.25%. Control eggs which remained unopened until EDD13 at had a similar endpoint

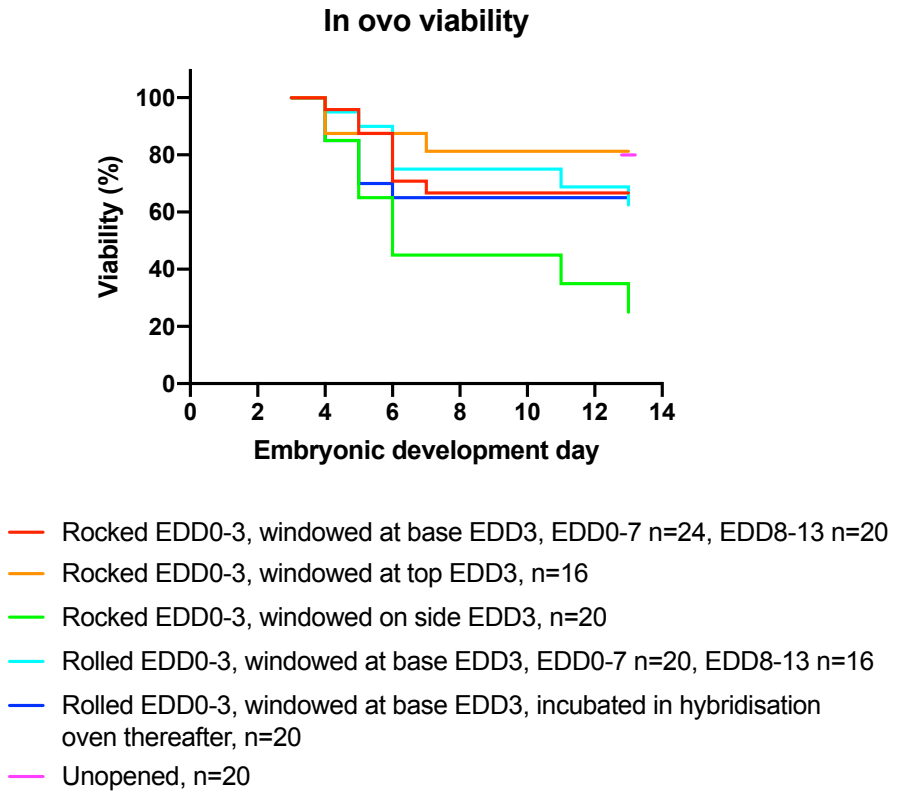


Figure 3.9 The effect of incubation method and window location on chicken embryo viability. From EDD0-3, eggs were either incubated base up and automatically rocked gradually back and forth by 90° every 6 hrs, or eggs were incubated on their sides and allowed to roll as the egg trays automatically rocked back and forth by 90°. The eggs were either windowed at the side, top or base. Eggs were briefly handled each day to check that embryos were viable. Unopened eggs were opened at the termination of the experiment (EDD13) to assess viability.

embryonic viability of 80%, suggesting that windowing at the top of the egg did not adversely affect embryonic viability. When an egg was windowed at the top, the shape of the egg required the size of the window to be smaller (approx. 2 cm diameter) relative to if the egg had been windowed at the base of the egg (2.5-3 cm diameter). Therefore, top windowing may have yielded higher survival relative to base windowing due to the smaller, and presumably less disruptive, window sizes. For some applications a smaller egg window would be sufficient. However, base-windowing was used for the CAM tumour grafting experiments described in this chapter, as a larger window was required

for tumour grafting and potassium D-luciferin delivery for subsequent bioluminescence imaging.

3.3.4 Examples of tumours grafted onto the chicken embryonic CAM

Grafting of tumours onto chicken embryos was easy to confirm by inspection at the end of the experimental procedure at EDD14. Figure 3.10 shows examples of how grafted tumours appeared, white arrows indicate features of biological interest.

3.3.5 Bioluminescence imaging of CAM tumour grafts after a single dose of potassium D-luciferin

Either MCF7 *Arntl*-luc or C26 *Arntl*-luc tumour grafts were introduced to chicken embryos at EDD9. Bioluminescence imaging was commenced at start of EDD10, after the delivery of a 100 µl dose of 209 mM potassium D-luciferin (i.e., 6.6 mg) to each chick embryo. Figure 3.11 demonstrates that the

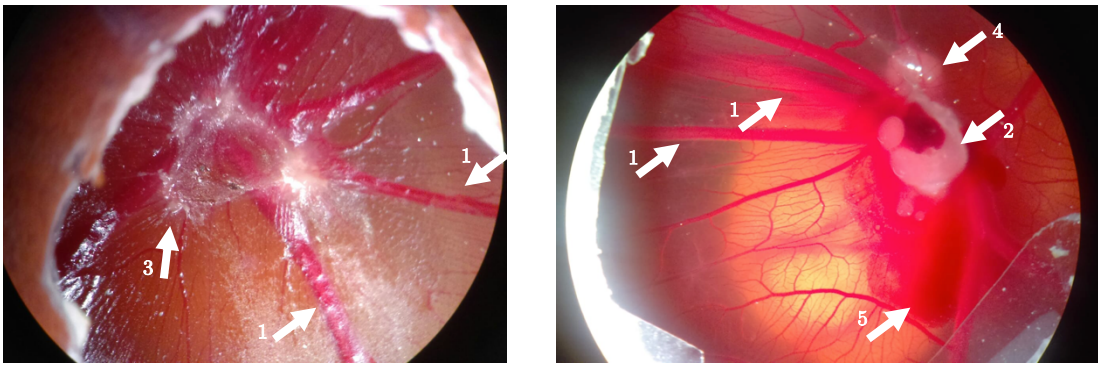


Figure 3.10 A) An MCF7 tumour at EDD14. B) An MDA-MB-231 tumour at EDD14. 1) White arrows indicate examples of increased angiogenesis typical of successful tumour grafting. 2) The original Matrigel graft site. 3) Metastasis of the original tumour graft on the surface of the CAM. 4) Secondary tumour growth beneath the CAM. 5) Blood pooling around the tumour graft, characteristic of leaky tumour-driven angiogenesis.

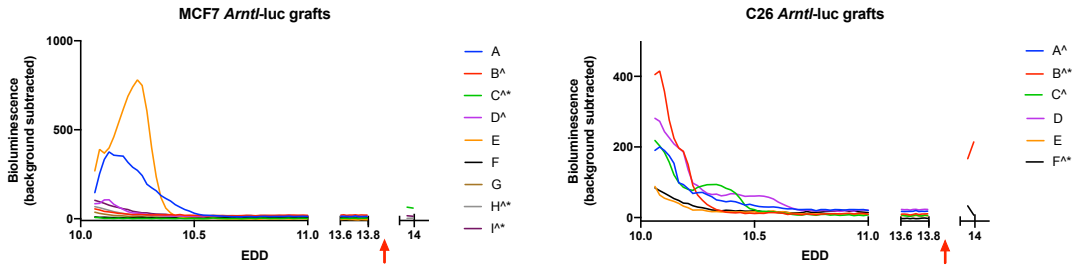


Figure 3.11 Chicken embryos with either A) MCF7 *Arntl*-luc or B) C26 *Arntl*-luc grafts. All embryos received 100 µl 209 mM potassium D-luciferin prior to imaging. ^ denotes embryos which survived to EDD14. * denotes embryos with visible graft bioluminescence at EDD14. Red arrow denotes addition of a further 50 µl 209 mM potassium D-luciferin immediately prior to the end of the experiment.

bioluminescent signal from the tumour grafts was no longer detectable above background within the first 24 hrs of imaging. In some embryos this would be expected as the embryo might have died, or the tumour cells might not have grafted successfully. However, at EDD14 5 MCF7 *Arntl*-luc and 4 C26 *Arntl*-luc grafts were visible on the CAMs of live chick embryos (as denoted by ^ in Figure 3.11). Upon re-dosing with 50 µl 209 mM potassium D-luciferin on EDD14, bioluminescence was detected in 3 of the MCF7 *Arntl*-luc and 2 of the C26 *Arntl*-luc grafts (Figure 3.11). This indicated that either all of the initial D-luciferin dose had been converted completely by luciferase to oxyluciferin within the first 24 hrs of imaging, or that the chicken embryos were capable of eliminating D-luciferin from the CAM within hours of dosing. The latter was perceived to be the more likely scenario, as in *in vitro* experiments 3 ml of 100 µM potassium D-luciferin – only 95 µg of potassium D-luciferin – is sufficient for imaging of >300,000 cells for upwards of 1 week at 37°C (chapter 2.3). In comparison tumour grafts of 150-250,000 cells received more than 5-fold more potassium D-luciferin yet bioluminescence was only detectable for <24 hrs. N.B. the *Arntl*-luciferase construct may be under circadian control in individual tumour cells, however, given that the chicken embryo is unlikely to have a well-developed circadian rhythm it seemed unlikely that *Arntl*-luc (or indeed *Per2*-luc) would be expressed in an oscillatory manner at a tumour

population level. For the purposes of CAM tumour graft assay development, the potential for oscillatory *Arntl*-luc expression was not judged to be a concern.

In an effort to lengthen the period of time for which the bioluminescence signal was detectable, the luciferin dose was increased ~7.5 fold to 1 ml of 157 mM potassium D-luciferin (i.e., 50 mg/ml). This time MCF10A *Arntl*-luc cells were grafted on to the chicken embryos at EDD8 and imaged from EDD11. This allowed embryos to be selected which had survived the graft process and that also had tumour grafts that appeared to be driving a degree of local angiogenesis. Figure 3.12 shows the result of 3 embryos with MCF10A *Arntl*-luc grafts that were dosed with 50 mg of potassium D-luciferin. The bioluminescence signal was undetectable within 12 hours of imaging. After 24 hrs the incubator was opened, and it was confirmed that all three embryos had died. It is possible that the MCF10A *Arntl*-luc tumour grafts killed the chicken embryos, though as the MCF10A cell line is benign, albeit immortalised, it does not metastasise extensively *in ovo*. Therefore, the larger dose of potassium

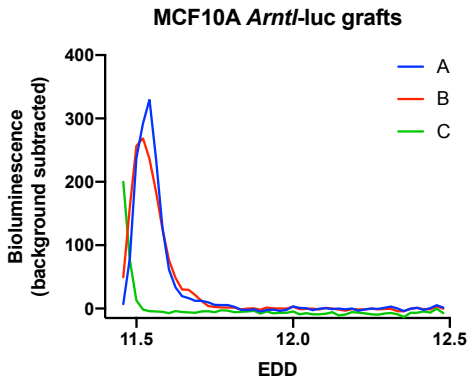


Figure 3.12 Monitoring of bioluminescence from MCF10A *Arntl*-luc tumours grafted onto chicken embryos at EDD8 and imaged from EDD11. MCF10A *Arntl*-luc tumours were dosed with 1 ml of 157 mM potassium D-luciferin. N.B. absolute bioluminescence is not comparable between the MCF10A and MCF7/C26 experiments, as the image acquisition parameters and the number of cells grafted varied between the two experiments, see Table 3.2 for details.

D-luciferin may have been the cause of death. As the potassium D-luciferin was dissolved in sterile H_2O, it is possible that addition of a large volume (1 ml) of water containing potassium ions was osmotically disruptive to the vascularised CAM. It has previously been published that injection of increasing volumes of H_2O up to 300 µl into the chicken egg albumin prior to incubation reduces hatch rates proportionally to as low as 20%, which would support the idea that potassium D-luciferin should be delivered to the egg in a much smaller volume of solvent than 1 ml (Yi lin Wang et al., 2017).

3.3.6 Bioluminescence imaging of CAM tumour grafts with a continuous drip delivery of potassium D-luciferin

In order to try and monitor tumour graft bioluminescence over time, D-luciferin was supplied to the egg continuously via a WPI Aladdin syringe pump for the duration of bioluminescence imaging. MDA-MB-231 *Arntl*-luc and Hepa1-6 *Per2*-luc cells were grafted onto chicken eggs at EDD8. At EDD11 eggs with grafted tumours were selected for bioluminescence imaging. Figure 3.13 shows the result of administration of 20.7 mM potassium D-luciferin solution at a rate of 13.9 µl/hr (i.e., 2.2 mg in 333 µl/day) to eggs with MDA-MB-231 or Hepa1-6 tumour grafts. In some embryos continuous luciferin delivery enabled the monitoring of tumour bioluminescence over >2.5 days, however bioluminescence appeared very uneven over time from embryo to embryo. This was at least partially the result of variance in the delivery rate of luciferin to the CAM, as droplets formed at the end of the luciferin delivery tubing. These droplets took some time to fall and contact the surface of the CAM – presumably at rate dependent upon the distance of the tubing from the eggshell/CAM. The 'drip rate effect' was particularly noticeable in embryo B with a Hepa1-6 *Per2*-luc graft, and embryo C with an MDA-MB-231 *Arntl*-luc

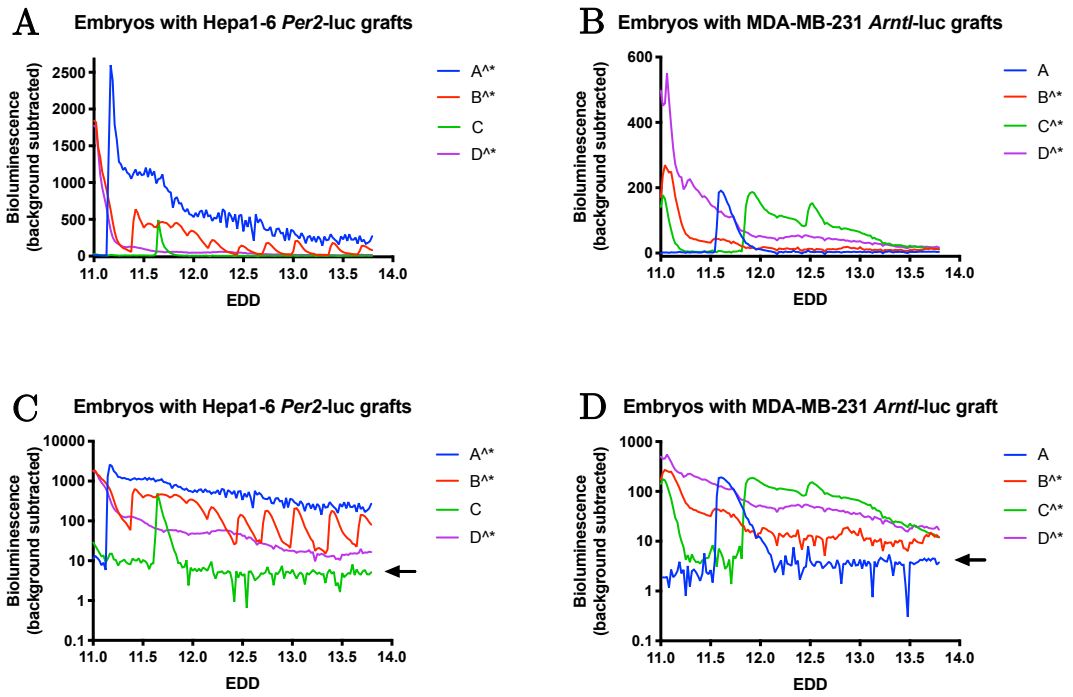


Figure 3.13 Chicken embryo CAMs with A) and C) Hepa1-6 *Per2*-luc grafts and B) and D) MDA-MB-231 *Arntl*-luc tumour grafts. D-luciferin was delivered by continuous drip at 20.7 mM, 13.9 µl/hr. A) and B) y axes are linear. C) and D) y axes are logarithmic. ^ denotes embryos which survived to EDD14. * denotes tumour grafts that bioluminesced at EDD14. Black arrows indicate the baseline of the experiment, i.e. the threshold at which no bioluminescence was visible. As is to be expected the baseline appears quite noisy when data is plotted on logarithmic axes.

graft. This is a difficult issue to address, as there is a limit to how close the tubing can be placed to the CAM – too close and the tubing may contact and damage the CAM whilst the eggs are moved around the Alligator incubator in preparation for imaging. Too far and the 'drip rate effect' becomes more pronounced.

In some cases, the bioluminescence took some time to 'appear', e.g. embryos A and C in the Hepa1-6 *Per2*-luc graft group and embryo A in the MDA-MB-231 *Arntl*-luc graft group. This was likely to have been caused by air bubbles at the end of the luciferin-loaded tubing, as loading a pre-filled syringe into an 8-channel WPI Aladdin syringe pump can sometimes disturb syringes that have already been placed. In contrast, some embryos e.g. B and D in the

Hepa1-6 *Per2*-luc graft group appear to have started the experiment with a burst of luciferin rather than a lack, which was attributed to early luciferin delivery during syringe placement.

In general the bioluminescence decreased with time, although it could be said that embryos B and D in the Hepa1-6 *Per2*-luc graft group, and embryo B in the MDA-MB-231 *Arntl*-luc graft group reached a 'steady-state' bioluminescence signal in the last day of imaging. It is unclear what the cause of the general decrease in signal should be attributed to. Perhaps many of the grafted tumour cells could not be supported by the CAM and died. Alternatively, the decrease in bioluminescence may be related to the storage of D-luciferin within syringes at room temperature – albeit in darkness – for a number of days. D-luciferin solubilised in water with sufficient oxygen exposure is known to degrade to dehydroluciferin, which inhibits luciferase's photon-releasing catalysis of luciferin to oxyluciferin (da Silva & da Silva, 2011; Shi et al., 2020). Therefore, it is possible that the decreasing of bioluminescence with time was the result of increasing inhibitory dehydroluciferin concentration.

In order to address the issue of the 'drip-rate effect', the flow rate of D-luciferin delivery was increased from 13.9 µl/hr (20.7 mM) to 139 µl/hr (2.07 mM). The hypothesis was that drips would fall more frequently and therefore would not limit the rate of bioluminescence. Hepa1-6 *Arntl*-luc cells were grafted at EDD8. At EDD11, eggs with grafted tumours were selected for bioluminescence imaging. Figure 3.14 shows the effect of continuous D-luciferin at a rate of 139 µl/hr. The bioluminescent signals were very weak for the last two days of imaging (Figure 3.14A), therefore background subtraction was applied to each graft individually using a local region for background subtraction to provide a less noisy baseline (Figure 3.14B). Such an approach means that absolute bioluminescence is no longer directly comparable between

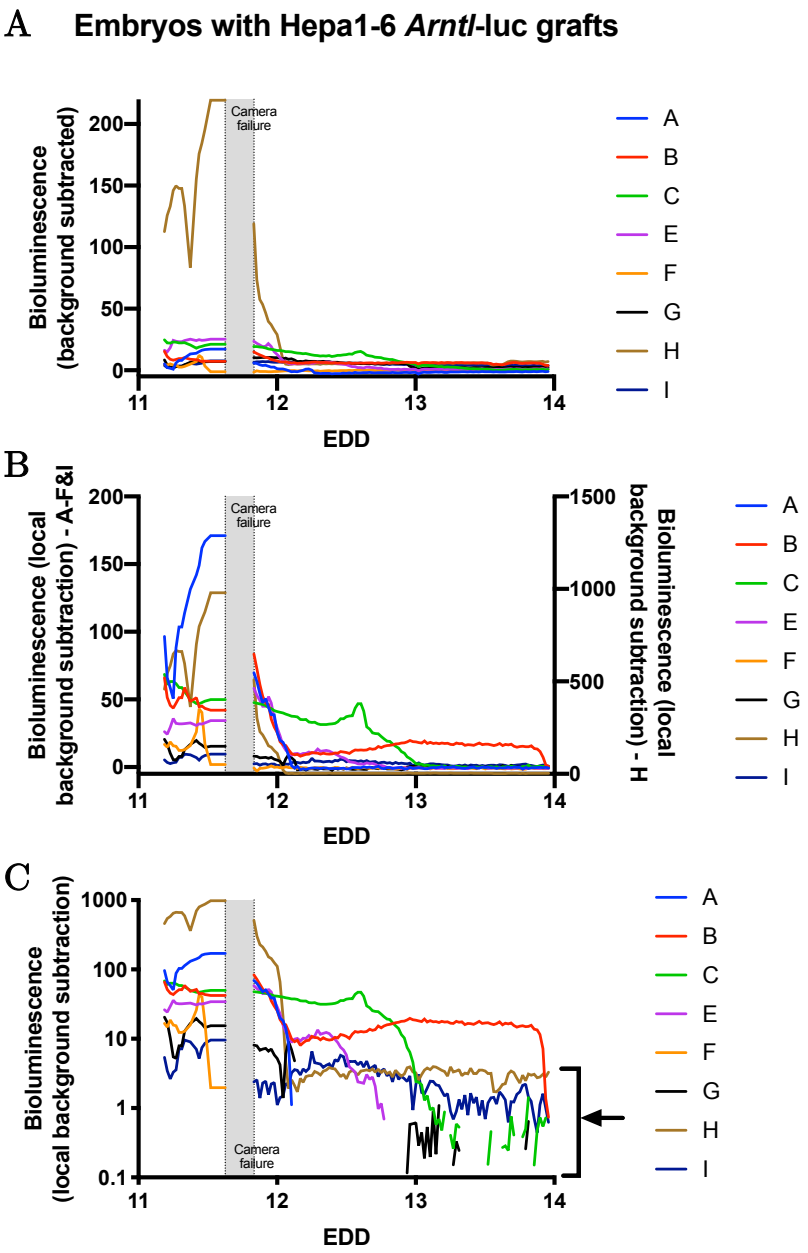


Figure 3.14 A) Chicken embryo CAMs grafted with Hepa1-6 *Arntl*-luc cells. D-luciferin was delivered by continuous drip at 2.07 mM, 139 µl/hr. B) Same embryos plotted using local background subtraction. To allow for an appropriate local region for background subtraction the size of regions of interest are different, therefore absolute mean bioluminescence of embryos A-I cannot be directly compared for this method of background subtraction. C) Same data as B, plotted with a logarithmic y axis. Black arrow indicates the baseline of the experiment, i.e. the threshold at which no bioluminescence was visible. As is to be expected the baseline appeared quite noisy when data was plotted on logarithmic axes. Camera failed for a period of 5 hrs due to the recording computer crashing. N.B. egg incubation conditions were controlled independently and therefore continued unperturbed, despite camera failure. ^ denotes embryos which survived to EDD14, i.e., none.

grafts, see chapter 3.2.10 for more details. The 'drip-rate effect' certainly appeared greatly reduced with a higher drip volume. However, at EDD14 every chicken embryo was found to have died. A disadvantage of bioluminescence imaging is that it is not possible to check every day for host survival without the pausing of image recording. Therefore, it was not possible to ascertain how long the embryos survived. Since all eight drips were confirmed to still be delivering D-luciferin at EDD14, it seems probable that the loss in bioluminescence from five of eight grafts within 24 hrs of imaging, was likely to be due to chick embryo death. The survival in chick embryos grafted with Hepa1-6 *Arntl*-luc cells (0%, n=8), relative to those grafted with Hepa1-6 *Per2*-luc (75%, n=4) or MDA-MB-231 *Arntl*-luc cells (75%, n=8) was stark. This data supports the theory formulated from the results of Figure 3.12, that the delivery of larger volumes of D-luciferin to chick embryos was not conducive to survival.

In order to deliver D-luciferin consistently to the embryo throughout the imaging window it was necessary to seek an alternative method of D-luciferin delivery.

3.3.7 Bioluminescence imaging of CAM tumour grafts with a slow-release luciferin-loaded polymer patch

Advances in polymer chemistry have enabled the development of slow- or controlled-release polymeric formulations of many small molecule drugs (Sung & Kim, 2020). In the same vein, it was proposed that potassium D-luciferin could be encapsulated in a polymer to provide a method of continual luciferin supply to the CAM without the need for a liquid carrier. See chapter 3.2.9 for details of the synthesis of luciferin-loaded polymers.

Figure 3.15 demonstrates the bioluminescence signal resulting from the placement of 1 cm^2 D-luciferin patches (batch 1) on the surface of the CAMs grafted with MCF10A *Per2*-luc tumours at EDD8. The bioluminescent signal that was produced was extremely weak in two of the chicken embryos (D and E), and undetectable in the three other chicken embryos (A, B and C). After 2 days of bioluminescence imaging only one tumour was faintly visible (D). The incubator was opened, and all 5 chicken embryos were found to be alive. 100 µl of 20 mM potassium D-luciferin was added to embryos A-C at EDD12, and the imaging parameters were altered to be more sensitive to bioluminescence (see Table 3.2 for details). Addition of 20 mM potassium D-luciferin appeared to result in large spikes in bioluminescence from embryos A and C, but not B. In fact, the signal from embryo B began to increase at

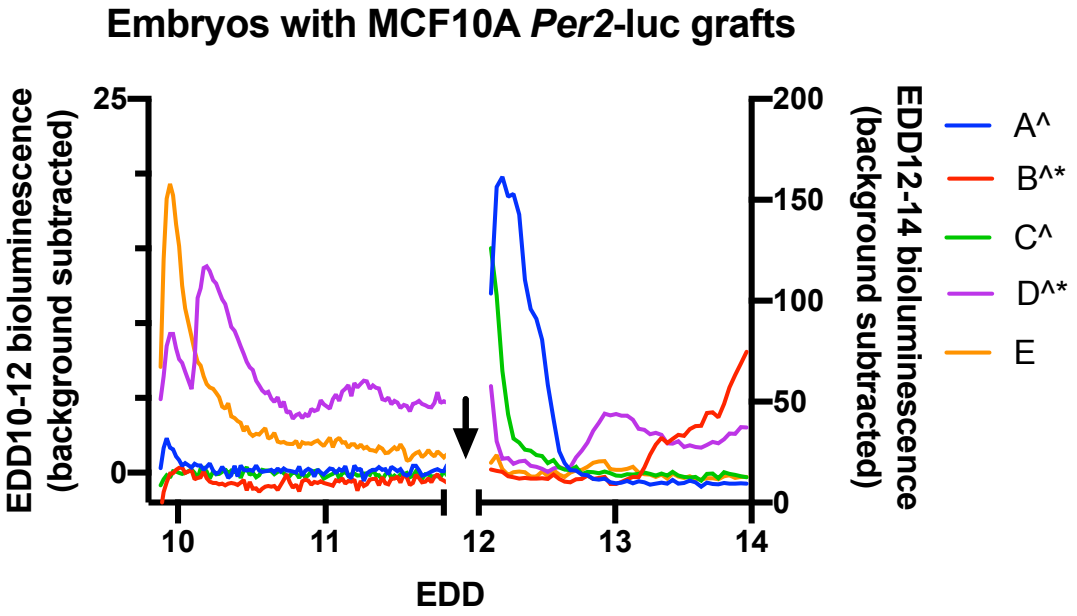


Figure 3.15 Chicken embryo CAMs grafted with MCF10A *Per2*-luc cells at EDD8. 1 cm^2 D-luciferin polymer patches (batch 1) were placed onto the CAM at EDD9.9 to allow for monitoring of graft bioluminescence. The resulting bioluminescence was monitored in the Cairn Alligator. The black arrow indicates addition of 100 µl 20 mM potassium D-luciferin to embryos A-C, and a shift to more sensitive imaging parameters in order to better detect tumour bioluminescence, note the different y axis scales. ^ denotes embryos which survived to EDD14. * denotes tumours that bioluminesced at EDD14.

EDD13, suggesting that a lack of tumour graft growth, rather than a lack of luciferin, may have been the reason for the lack of bioluminescence prior to EDD13. Similarly, embryo D received no luciferin beyond that of the original 1 cm^2 polymer patch, and also remained bioluminescent at EDD14. Therefore, it was concluded that the use of batch 1 D-luciferin polymer patches likely contained enough luciferin for adequate monitoring of MCF10A *Per2*-luc graft bioluminescence over 4 days, but that the rate of graft bioluminescence was highly variable from embryo to embryo.

Figure 3.16 demonstrates the bioluminescence signal resulting from the placement of 1 cm^2 luciferin patches (batch 2) on the surface of CAMs grafted with MCF7 *Arntl*-luc tumours at EDD8. The bioluminescent signal decreased rapidly from all five embryos within the first 24 hrs of imaging. Four chicken embryos survived to EDD14, at which point bioluminescence was detectable only from grafts C and E. However, embryos B and D had visibly grafted tumours (Figure 3.16B). Which would suggest that either luciferin availability or camera sensitivity limited the detection of bioluminescence. Since the luciferin-loaded polymer patch is still early in development, little is known about the release profile of luciferin from the polymer patch onto the CAM. In particular the change in rate of release of luciferin from the patch over time is not well-understood. For instance, if the release of luciferin from the patch follows first order release kinetics rather than zero order, then the bulk of luciferin release from the polymer patch might occur within the first 12 hours of imaging. Therefore, the spike in bioluminescence signal early in the experiment may have been attributable to the rate of luciferin release, rather than the apparent decline of tumour cell growth. It is of interest to note that the signal was of a similar magnitude to that produced by batch 1 luciferin patches from EDD10-12 on embryo CAMs with MCF10A *Per2*-luc grafts (Figure 3.15).

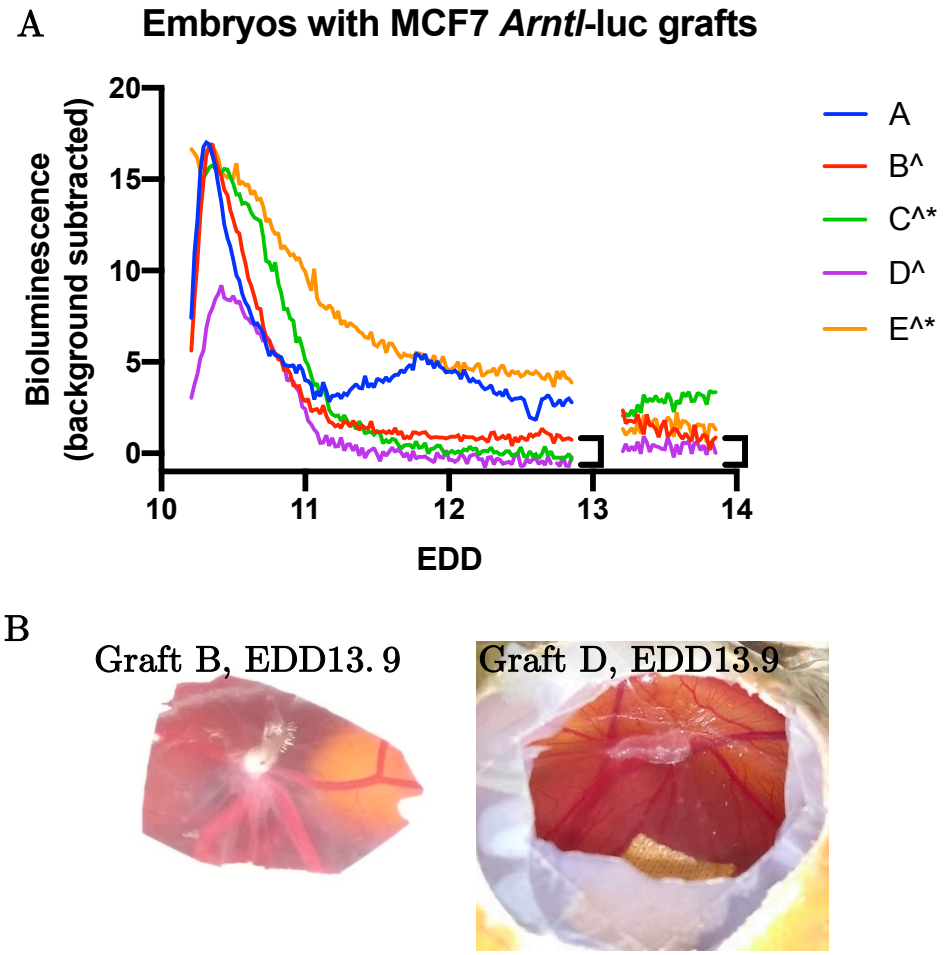


Figure 3.16 A) Chicken embryo CAMs grafted with MCF7 *Arntl*-luc cells at EDD8. 1 cm^2 luciferin polymer patches (batch 2) were placed onto the CAM at EDD10 to allow for monitoring of graft bioluminescence. The resulting bioluminescence was monitored in the Cairn Alligator. ^ denotes embryos which survived to EDD13.9. * denotes tumour grafts with detectable bioluminescence at EDD13.9 B) Embryos B and D both displayed extensive angiogenesis at tumour cell graft sites, indicative of successful grafting, in spite of the absence of a bioluminesce signal at EDD13.9.] denotes the baseline of the experiment, i.e. the threshold at which no bioluminescence was visible.

In order to try to ascertain whether the bioluminescence was present but below the threshold of detection, the luciferin patch technology was next imaged in a PhotonImager Optima imaging system. For this experiment embryos were grafted either with MCF7 *Per2*-luc cells or MDA-MB-231 *Arntl*-luc cells at EDD8. Bioluminescence imaging was commenced at EDD10 after placement

of luciferin polymer patches as shown in Figure 3.17 (batch 2, 1 cm^2). Unfortunately, for the duration of this experiment the temperature control plate would not heat higher than 37°C, therefore the ambient temperature of egg incubation was approx. 5-10°C lower than the target temperature of 38°C. Additionally, the imager and temperature control failed at EDD11.6, causing the temperature to drop further. As a consequence, by EDD12.1 the chicken embryos had reached a 50% mortality rate, accompanied by a rather sharp decrease in graft bioluminescence. The experiment was terminated to avoid further distress to the chicken embryos. Encouragingly, at EDD11.6 bioluminescence remained above the threshold of detection for 10 of 12 grafts. Bioluminescence even increased until EDD11.6 for MCF7 *Per2*-luc grafts A, C and F, which indicated that luciferin release from the batch 2 polymer patches was sufficient to report relative tumour graft growth or decline for >40 hrs. It is unclear why the rapid decrease of bioluminescence observed in Figure 3.15 and Figure 3.16 in the first 12 hrs of imaging was not observed in Figure 3.17. Figure 3.16 and Figure 3.17 embryos both received grafts of $4x10^6$ MCF7 cells at EDD8 (see Table 3.2), so it seems unlikely that this phenomenon was related

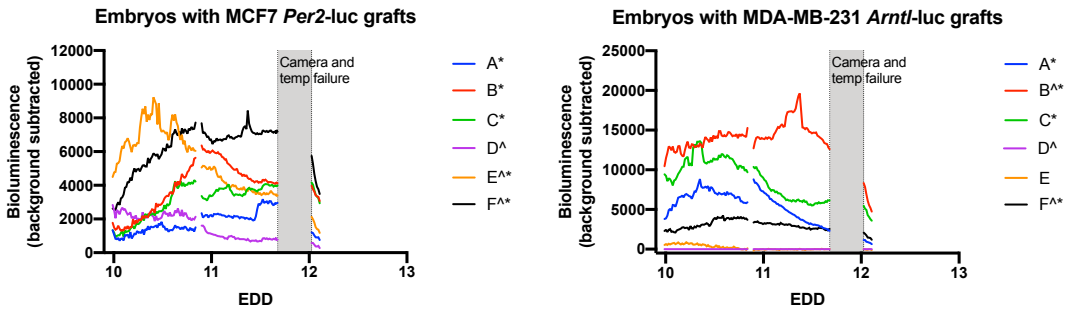


Figure 3.17 Chicken embryo CAMs grafted with MCF7 *Per2*-luc or MDA-MB-231 *Arntl*-luc cells at EDD8. 1 cm^2 luciferin polymer patches (batch 2) were placed onto the CAM at EDD10 to allow for monitoring of graft bioluminescence. The resulting bioluminescence was monitored in the PhotonImager Optima. ^ denotes embryos which survived to EDD12.1. * denotes tumour grafts with detectable bioluminescence at EDD12.1. N.B. ambient egg incubation temperature was low for the duration of this experiment (~27-30°C), and fell further at EDD11.6 at the point of imager failure, therefore the experiment was terminated sooner than EDD14.

to the success of tumour grafting, which should have been similar between the two experiments. It is also unlikely to have been related to any difference in sensitivity of bioluminescence detection between the Cairn Alligator and PhotonImager Optima imaging systems, as it is clear that the difference in bioluminescence signals from Figure 3.15 to Figure 3.17 was not related to the sensitivity of bioluminescence detection, but to the shape of the profile. In other words, none of the bioluminescence signals in Figure 3.17 decreased within the first 12 hrs of imaging, unlike the signals from Figure 3.15 and Figure 3.16.

For subsequent experiments, in order to allow for better temperature control during imaging in the PhotonImager Optima, the chicken eggs were placed inside an adapted polystyrene egg incubator with a clear acrylic lid to allow for imaging. Figure 3.18 demonstrates the use of this system to monitor embryos that were grafted either with MCF7 *Per2*-luc cells or MCF10A *Per2*-luc cells at EDD8. Bioluminescence imaging commenced at EDD11 (grafts A-B) or EDD12 (grafts C-F) after placement of luciferin polymer patches (batch 2, 1 cm^2). The exposure time of the PhotonImager Optima was reduced from 20 min (Figure 3.17) to 5 min (Figure 3.18), since the signal:background ratio was very reasonable in Figure 3.17.

The bioluminescence signals from embryos displayed in Figure 3.18 were remarkably different from those displayed in Figure 3.16. For instance, in Figure 3.17, bioluminescence was still increasing in 3 out of 6 embryos grafted with MCF7 *Per2*-luc cells after 1.6 days of imaging. However, in Figure 3.18 none of the MCF7 or MCF10A cell grafts even displayed detectable bioluminescence after 1.6 days of imaging. In fact, the cell grafts in Figure 3.18 appeared more similar in profile to the MCF7 *Arntl*-luc grafts monitored in Figure 3.16. That is to say that for the most part, bioluminescence signals

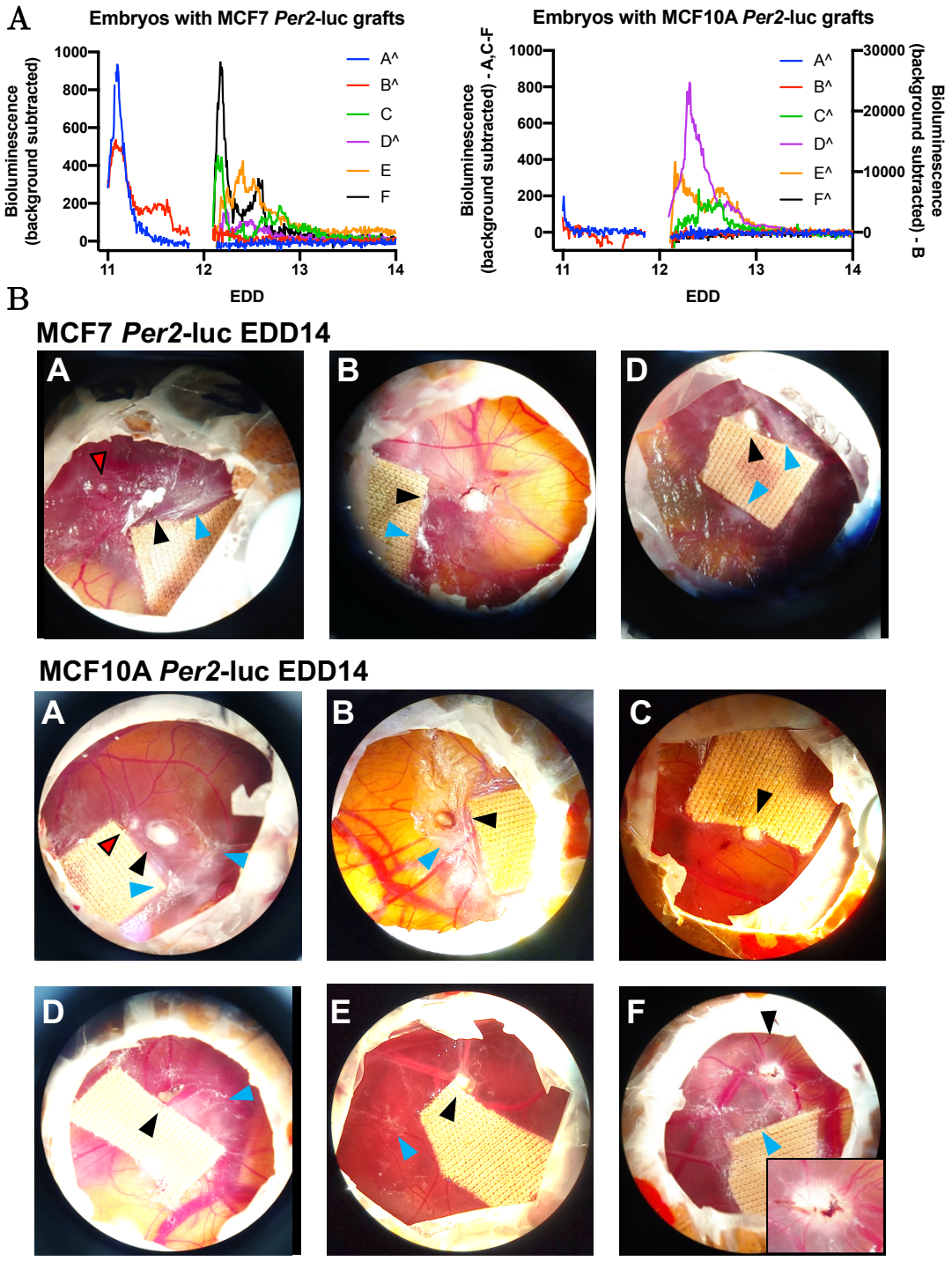


Figure 3.18 A) Chicken embryo CAMs grafted with MCF7 or MCF10A *Per2*-luc cells at EDD8. 1 cm^2 luciferin polymer patches (batch 2) were placed onto the CAM at EDD11 (A&B) or EDD12 (C-F) to allow for monitoring of graft bioluminescence. The resulting bioluminescence was monitored in the PhotonImager Optima. ^ denotes embryos which survived to EDD14. * denotes tumour grafts with detectable bioluminescence at EDD14. B) Photos of embryos alive at EDD14. Black arrowheads indicate the tumour graft site, blue arrowheads denote regions that appear to be infected, likely with a fungal infection. The nature of structures on the CAM can be unclear even to the naked eye, for instance red arrows with black outlines denote regions that may be graft metastasis, or may be the result of fungal infection.

peaked within the first 12 hrs of imaging and then rapidly decayed. In order to try to understand the reason for the rapid loss of bioluminescence, the embryos that were still alive at EDD14 were imaged (see Figure 3.18B). Some grafts, including embryos A and D grafted with MCF10A *Per2*-luc cells, did not appear to be particularly well vascularised at EDD14, which is indicative of graft failure and would explain the lack of bioluminescence within 24 hrs. However, other grafts, including embryo B grafted with MCF7 *Per2*-luc cells and embryos E and F grafted with MCF10A *Per2*-luc cells, were very well vascularised at EDD14, yet also lacked bioluminescent signals within 24 hrs. In fact, embryo F grafted with MCF7 *Per2*-luc cells never displayed any detectable bioluminescence despite excellent tumour graft vascularisation.

A further complication of working with the luciferin polymer patches was that they were not manufactured under sterile conditions. Despite addition of pen/strep to the patch at point of grafting (see chapter 3.2.9), it was not uncommon to observe opportunistic infections on the surface of the CAM in the area surrounding the patch. The infections were likely to have been fungal for two reasons – firstly egg albumin is rich in anti-bacterial lysozyme greatly reducing the chance of bacterial infection (Vilcacundo et al., 2018), and secondly hyphae growth was sometimes observed. Whilst no direct correlation was observed between severity of fungal infection and chicken embryo survival, it is certainly reasonable to speculate that fungal infection is disadvantageous to embryonic development, and that future work with luciferin-loaded polymer patches would benefit from a sterile manufacturing process.

The work with the luciferin-loaded polymer patches is early in its development, and as such many questions remain unanswered. In particular it remains unknown what the release profile of luciferin from the polymer patch into the CAM looks like. This could be answered by direct assay of the CAM for

luciferin by an analysis method such as high-performance liquid chromatography.

3.3.8 Bioluminescence imaging of CAM tumour grafts with daily potassium D-luciferin dosing

Figure 3.18 demonstrated an inconsistency between apparent success of grafting by angiogenesis and detection of graft bioluminescence. In order to confirm that this was not related to the use of the luciferin-loaded polymer patch, MCF10A *Per2*-luc grafts were supplemented with 90 µl 10 mg/ml potassium D-luciferin every day immediately prior to imaging, and were imaged once per day for 4 days. The amount of potassium D-luciferin was selected because it was the same dose that Jefferies et al. (2017) used for bioluminescence imaging of tumour grafts on the CAM. Figure 3.19 demonstrates that daily luciferin dosing and imaging resulted in a rapid decrease of bioluminescent signal (N.B. the logarithmic y axis). Although embryos B, C, D, E and F all demonstrated good angiogenesis indicative of successful tumour engraftment, only graft C demonstrated bioluminescence at the end of the experiment.

Jefferies et al. (2017) demonstrated a similar phenomenon of falling bioluminescence in MG63 and HOS cells grafted onto the CAM in BD Bioscience's Matrigel. However, when MG63 cells were grafted in a gelfoam scaffold (Pfizer), bioluminescence increased for the first two days of imaging prior to falling. Therefore, Jeffries et al. concluded that use of a gelfoam scaffold supported tumour grafting better than Matrigel. For the experiments described in this thesis, growth factor reduced Matrigel was used as the matrix for tumour grafting. However, since good angiogenesis was observed at the graft

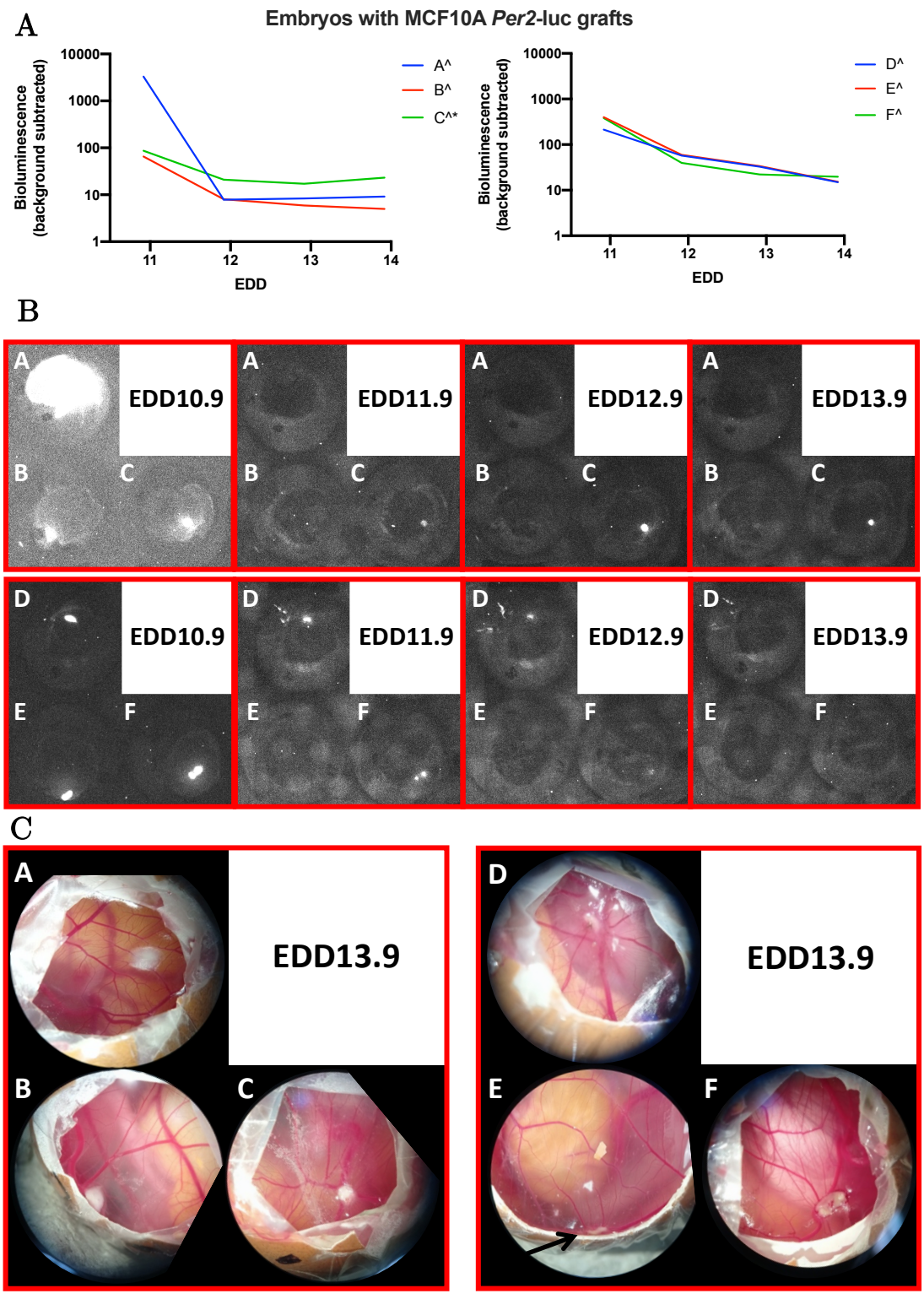


Figure 3.19 A) Chicken embryo CAMs grafted with MCF10A *Per2*-luc cells at EDD9. Imaged for bioluminescence immediately after daily doses of 90 µl 10 mg/ml potassium D-luciferin. ^ denotes embryos which survived to EDD13.9. * denotes tumour grafts with detectable bioluminescence at EDD13.9 B) Bioluminescent images of embryos A-F. Contrast settings are set differently for each image to best allow visualisation of tumour graft bioluminescence. C) Images of embryos at EDD13.9, end of experiment.

site of many tumours, it seems reasonable to conclude that in most cases the growth factor reduced Matrigel was sufficient to support tumour engraftment. Figure 3.20 offers further evidence to support the hypothesis that the Matrigel matrix was sufficient for tumour graft growth and did not cause the bioluminescence to decrease through lack of tumour growth. Two particularly successful MCF7 *Per2*-luc tumour grafts with a large number of metastases at EDD13.9 are depicted in Figure 3.20B. However, even though it can be assumed that the final tumour cell number outstripped the initial number of MCF10A cells grafted, daily dosing with potassium D-luciferin (90 µl, 10 mg/ml) still resulted in bioluminescent signals that decreased rapidly with time (Figure 3.20A).

When Figure 3.20 is taken into account, it is logical to consider that an explanation for the observed decrease in bioluminescence with time, even whilst apparent tumour size increased in the presence of fresh luciferin, might be that expression of the stably transduced lentiviral *Per2*-luc (or *Arntl*-luc) construct was greatly decreased in tumour cells grafted onto the CAM. It should be noted, that cell lines stably transduced with the *Per2*/*Arntl*-luc constructs have been cultured *in vitro* for many passages (>30) in the Dallmann lab without the requirement for reporter maintenance by continual blasticidin selection, and that a significant decrease in cell bioluminescence has never been observed *in vitro* (anecdotal, data not shown). However, it is possible that in a 3D environment gene expression shifted significantly, such that either molecular clock gene activity in the tumour cells was routinely damped, or the lentiviral reporter constructs were transcriptionally silenced. Transcriptional silencing of lentiviral constructs has previously been reported to result largely, though not exclusively, from methylation of DNA, histone deacetylation and resultant condensation of chromatin (Yao et al., 2004; F. Zhang et al., 2007). There is evidence of this occurring both *in vitro* and *in*

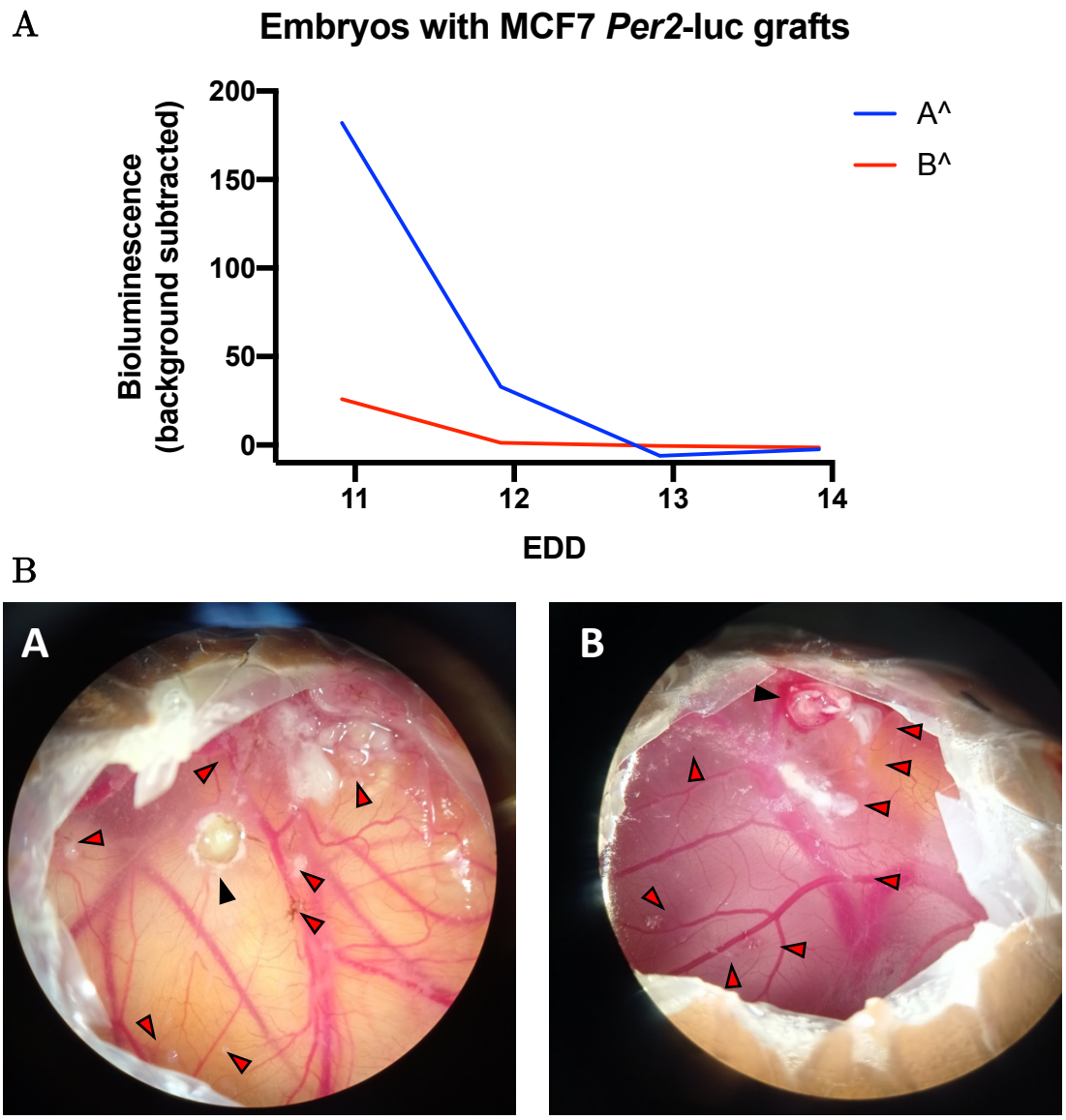


Figure 3.20 A) Chicken embryo CAMs grafted with MCF7 *Per2*-luc cells at EDD9. Imaged for bioluminescence immediately after daily doses of 90 µl 10 mg/ml potassium D-luciferin. ^ denotes embryos which survived to EDD13.9. * denotes tumour grafts with detectable bioluminescence at EDD13.9. B) Images of embryos at EDD13.9, end of experiment. Black arrowheads denote original tumour graft site. Red arrowheards outlined in black indicate visible metastases.

vivo (Palmer et al., 1991; F. Zhang et al., 2007), and transcriptional silencing remains a significant barrier to viral delivery of gene therapy (Tolmachov et al., 2013). Despite this, reports of lentiviral reporter silencing remain uncommon in a tumour graft context. Baklaushev et al. (2017) demonstrated a significant decrease in luciferase expression in 4T1 cells after they were grown

as tumours in mice. However, the authors attributed the decrease to a specific anti-luciferase T-cell response in the immunocompetent mice (Baklaushev et al., 2017). This certainly was not the case in chicken embryos at EDD8-14, as chicken embryos have not yet developed adaptive immune responses. Of course, absence of evidence for lentiviral reporter silencing upon tumour grafting, is not evidence of absence, and is worthy of future investigation. qPCR for luciferase RNA would establish if the decrease in observed bioluminescence with time was attributable, or not, to a concurrent decrease in the luciferase reporter activity.

An alternative possibility worthy of consideration is that with time, tumour gene expression shifted in order to change the import or export of luciferin. The ATP-binding cassette (ABC) family transporter ABCG2 is a known effluxer of D-luciferin, and causes a decrease in tumour graft bioluminescence *in vivo* when upregulated (Yimao Zhang et al., 2007). MCF7 and MDA-MB-231, but not MCF10A cells are known to increase expression ABCG2, which is upregulated further when exposed to cytostatic drugs (Rosenfeldt et al., 2014; Saxena et al., 2011). A decrease in graft bioluminescence with time was observed in both MCF10A cells and MCF7 tumour cell grafts (Figure 3.19 and Figure 3.20), therefore it seems unlikely that ABCG2 specifically is responsible for the phenomenon described in this chapter.

One must also consider the possibility that diminishing oxygen or ATP – the other substrates of the photon-emitting luciferase reaction – progressively limited the rate of the bioluminescence reaction. Decreasing ambient oxygen concentration has been shown to inhibit the rate of the bioluminescence reaction in cells not by a lack of oxygen, but by a resultant lack of intracellular ATP (Moriyama et al., 2008). If intracellular ATP had been limited to such an extent as to prevent the bioluminescence reaction from occurring, then it

seems unlikely that extensive tumour growth and metastasis would have been able to occur as observed in Figure 3.20. Therefore, it seems unlikely that either oxygen or ATP were unavailable to such an extent as to cause the observed decrease in bioluminescence.

3.4 Conclusions

The work presented in this chapter is aimed at developing an alternative model to the mouse, for the investigation of circadian gene expression in a 3D tumour graft model.

Synchronised CEF cultures demonstrated circadian *Per2*-luc and *Arntl*-luc reporter activity, indicating that the molecular clock is oscillating even in embryonic chicken cells. However, the qPCR evidence does not support the presence of a free running circadian clock in the brain or CAM of the 12-14 day old chicken embryo incubated in constant darkness. *gPER3,* but not g*ARNTL*, appeared to be diurnally expressed in response to light/dark and perhaps temperature entrainment cycles. Future work should focus on expanding this research to other circadian genes.[3] A logical conclusion that might be drawn from the data, would be that the individual cells of the developing chicken embryo might have had functional molecular clocks (as observed for the CEFs) which at EDD12-14 remained de-synchronised at a tissue level (as observed by qPCR of CAM/brain under constant darkness), but were partially responsive to environmental entrainment cues (as observed by qPCR of g*PER3* under light/dark entrainment) This conclusion is in agreement with earlier data that suggested rhythmic pineal melatonin secretion could be detected at EDD13 in embryos cultured under light/dark entrainment, but did not occur in the absence of entrainment (Akasaka et al., 1995).

[3] N.B. g*NR1D1* and g*PER2* were also assayed by qPCR, however the data quality was poor, and the arrival of COVID ultimately prevented repeat of these experiments due to the inability to access the site of the 384-well qPCR machine.

Ultimately the chicken embryo remains an interesting model for the study of tumour grafting in a circadian context, even in the absence of a fully functional circadian clock. This is because the chicken embryo could be employed in the same manner that the 3D graft environment of the zebrafish embryo has been used, i.e., to assess how clock gene expression affects tumour growth rate differently in a 3D environment, relative to a 2D cell culture environment (Basti et al., 2020). The chicken embryo CAM model holds at least two advantages over the zebrafish model. Firstly, the chicken embryo allows for 6 days to pass between grafting of tumours and termination of the experiment prior to regulation under ASPA, whereas the zebrafish only allows for 3 days. The longer that cancer cells have to graft and metastasize, the more physiological the tumour environment is likely to become. Secondly, the expression of circadian genes in individual tumours in multiple individual zebrafish cannot not be assayed easily over time, since they would not remain *in situ* under a camera. This is not a concern for the chicken embryo, therefore the CAM provides an excellent opportunity for assaying tumour clock gene expression in a 3D environment over time.

Clearly questions remain concerning how tumours grafted onto the CAM translate to a reproducible, reliable bioluminescent signal. Future work should explore the use of CRISPR/Cas9 to create bioluminescent and fluorescent reporter cell lines at the endogenous clock gene locus. Such cell lines would reveal whether the decrease observed in graft bioluminescence over time was caused by transcriptional silencing of the lentiviral bioluminescence reporter, or perhaps resulted from inhibition of the luciferase reaction via a biochemical mechanism, such as the pharmacokinetics of luciferin availability.

The data presented here also provides proof-of-principle for the use of luciferin-loaded polymer patches to enable bioluminescence imaging for up to 4 days *in*

ovo (Figure 3.15 and Figure 3.16). The development of luciferin-loaded polymer patches is of importance beyond *in ovo* imaging, as they offer a transdermal surgery-free solution to the problem of long-term luciferin delivery to mice.

4. Exploring circadian dysregulation *in vivo* using an algorithmic approach: TimeTeller

4.1 Introduction

If one wishes to understand the degree of circadian dysregulation in cancer and the extent to which it may contribute to tumourigenesis, then the study of the circadian system *in situ* could be highly informative. Clearly, such an approach is met with the same challenge for circadian biologists as for any researcher; namely that sample availability from cancer patients is limited to the *ex vivo* by-products of surgical resection or tumour biopsy. Since the molecular circadian clock is a finely balanced transcriptional-translational feedback system comprising of at least 14 genes which oscillate every 24 hrs (Y. Ye et al., 2018), the inference of clock function from a single biopsy, which is representative of only a single point in time, is not an entirely straight-forward task. Biological time prediction from a single sample has long been a goal of circadian research, and though considerable progress has been made, it remains a work in progress, as will be discussed in this chapter (Ueda et al., 2004).

The combination of multiple variables i.e., expression of clock genes, to predict another variable, i.e., biological time, is a problem well-suited to a machine learning solution. Therefore several machine learning approaches have been developed, which aim to help us to understand more about the circadian state of a single biological sample (Dijk & Duffy, 2020). The primary focus of this thesis chapter is TimeTeller, a semi-supervised machine learning algorithm

which was developed by Professor David Rand and Dr Denise Vlachou at the University of Warwick (Vlachou et al., 2020). TimeTeller was originally designed for biological time prediction of single timepoint microarray data. Here, TimeTeller is extended for use with RNA-seq data sets.

4.1.1 Comparison of the gene expression technologies RNA-seq and microarrays

Although TimeTeller was originally designed for biological time prediction of single timepoint microarray data, working with microarrays has its disadvantages. Like a number of other time-telling algorithms, TimeTeller is trained on publicly available microarray datasets. Therefore, TimeTeller can only predict the time of test samples for which a training dataset comprised of the same microarray probeset exists. This requirement for matching probesets limits the extent to which publicly available microarray datasets can be successfully mined. A second disadvantage to working with microarrays is the relative increase in popularity of RNA-seq methods. PubMed search results per annum for "RNA sequencing" now outnumber those for "microarray", a trend that reflects the experimental advantages of RNA-seq over microarrays. RNA-seq has a larger dynamic range than microarray platforms, does not use probe technology (which risks off-target amplification), and has decreased in cost significantly in recent years (Jaksik et al., 2015; Rao et al., 2019). Given the rising popularity of RNA-seq, this thesis chapter focuses on the expansion of the TimeTeller method to RNA-seq datasets, in order to provide opportunities for novel insights.

The structure of this chapter is as follows:

- An in-depth introduction to the state-of-the-art time-telling algorithms available to date.
- 'Model Methods' describes the structure of the TimeTeller algorithm.
- 'Results and Discussion' describes the novel application of TimeTeller to publicly available RNA-seq datasets.
- 'Conclusions' and future directions

4.1.2 Review of algorithms for time prediction of biological samples

4.1.2.1 Molecular Timetable – Ueda et al., 2004

The original algorithm designed to predict biological time was coined "Molecular Timetable" (MT) (Ueda et al., 2004). MT was trained using the expression of 182 probes expressed rhythmically around the clock in mouse liver and was employed to predict the biological time of mouse liver samples. To train the model, Ueda et al. normalised the 48 hr expression profile of each probe using its mean and standard deviation (i.e., z-score normalised). Cosine functions were subsequently fitted to the normalised probe expression profiles. To estimate the time of a single test sample, probe expression was first normalised using the corresponding probe's profile means and standard deviations of the training data, and then the phase that gave the maximal correlation between the normalised test sample expression of each of the 182 probes and the 182 cosines fitted to the training data was selected. MT demonstrated mean error of <2hrs between predicted biological time and actual time of sampling for 8 independent samples (Ueda et al., 2004).

Whilst MT was an excellent early example of single time-point time-telling, the method had some disadvantages. Firstly, Ueda et al. used many genes to

build the model, which made for lengthy computation times. A reduction in the number of genes has since been demonstrated to greatly reduce the accuracy of time prediction (Hughey et al., 2016). Secondly, the fitting of cosine functions to gene expression profiles presumes that rhythmic gene expression is best described by a cosine function. This biases the model against circadian clock genes that may not be expressed in a cosine fashion. For instance, *DBP* is sometimes detected with sharper peaks and flatter troughs over time, as in Hughes et al.'s mouse liver microarray data (Hughes et al., 2009). Of course, this non-canonical cosine expression pattern may in part reflect the sensitivity of the experimental microarray assay rather than nascent *DBP* expression, but it demonstrates the disadvantage of reliance on cosine fits in time-telling algorithms.

4.1.2.2 ZeitZeiger – Hughey et al., 2016

ZeitZeiger is a supervised dimensionality reduction algorithm which was originally trained on a multi-tissue murine microarray time-course dataset published by Zhang et al. (2014), in order to predict time of other murine microarray time-course datasets. ZeitZeiger is essentially a supervised algorithm as features (i.e., genes) were selected by using dimensionality reduction to identify a small number of genes that optimally predicted time. The ZeitZeiger method can be described as follows:

1) For each of the genes, g, the training data was comprised of observations over time. ZeitZeiger was designed such that the training dataset did not need to consist of sampling points that were evenly distributed through time.
2) Periodic smoothing splines, $f_g(t)$, were fitted to each of the n observations of g genes as a function of time (removing the need for even time-point sampling). The spline was time-discretised to provide a vector of m timepoints τ for each gene g. The mean of the spline and

variance of the observations around the spline were calculated as follows:

$$\bar{f_g} = \frac{1}{m}\sum_{i=1}^{m} f_g(\tau_i)$$

$s_g^2 = \frac{RSS_g}{n}$, RSS=sum of squared residuals.

And each time-discretised spline was normalised to give the $m \times p$ matrix *Z*, where p = number of genes *g*.

$$z_{ig} = \frac{f_g(\tau_i) - f_g}{s_g}$$

The presumption made here was that the variance of gene expression remained constant over time.

3) Penalised matrix decomposition (PMD) was applied to the matrix *Z* in order to find the sparse principal components (SPCs) of the training data. SPCs is a method designed for sparse data, i.e., large and noisy datasets with a low degree of 'significant structure' (D. Yang et al., 2014). This was the case for ZeitZeiger as only a small number of genes in the training data represented time-dependent variation across the Zhang et al. tissues. Each SPC contained information from a linear recombination of a relatively minimal number of features. When the Zhang training data was reduced to 2 dimensions, the first SPC was composed of information from 8 genes, and the second SPC from 6 genes. Only one gene (*Per2*) contributed to both the first and the second SPC.
4) Training data was projected onto the first two SPCs, time-discretised splines were fitted to the projected data and the mean and variance were calculated as was calculated for the original features *g*.
5) Test data was projected onto the first two SPCs calculated from the training data.

6) Projected test data was normalised by the mean and variance of the SPCs, and the likelihood of time, given the projected test data, was estimated for each SPC.
7) The log likelihoods of each SPC were summed for each test sample, and the maximum was identified as the predicted time of sampling.

4.1.2.3 Strengths of ZeitZeiger

ZeitZeiger was demonstrated to perform faster than MT, and to predict time with greater accuracy (Hughey et al., 2016). Additionally, unlike MT, ZeitZeiger does not require that oscillatory genes fit a cosine function. Finally, ZeitZeiger does not require that training data be at discrete time intervals (e.g., ZT0, ZT2, ZT4... etc.), thanks to the splines fitted to features plotted as a function of time. This is a significant advantage as it enables ZeitZeiger to be trained on a wider range of datasets.

4.1.2.4 Weaknesses of ZeitZeiger

A disadvantage of the spline fitting approach is that variance of expression of each of the features is presumed to remain constant over time, which has been observed not to be the case in, for instance, human skin biopsies (Wu et al., 2018). Therefore, the trained ZeitZeiger model is missing information about how the variance of oscillatory gene expression changes over time, and so the likelihood functions used to predict time do not truly contain information that can be related to the confidence of the time prediction (i.e., how close the projected test data truly lies to the projected training data). This means that ZeitZeiger does not provide the confidence of a predicted time. TimeTeller – the approach developed in this thesis chapter for RNA-seq datasets – provides an advantage over ZeitZeiger as it also calculates a confidence metric of predicted time (Vlachou et al., 2020).

4.1.2.5 Applications of ZeitZeiger

ZeitZeiger was originally applied to the multi-organ Zhang et al., (2014) murine microarray dataset and applied to murine test datasets generated from different probeset microarrays. The challenge of working across microarray probesets (introduced in chapter 4.1.1) was addressed by the application of a batch correction method (ComBat) to remove batch effects between training and test datasets generated using different microarray probesets (Hughey & Butte, 2016; Johnson et al., 2007). ComBat was also employed to remove batch effects between different tissue types in the Zhang et al. training dataset, whilst genotype in the test data was provided to ComBat as a covariate. This batch correction approach has several disadvantages. Firstly, batches were not evenly balanced as the training data only contained WT mice, whilst test data batches contained both WT and genetically manipulated knock-out (KO) samples as illustrated by Figure 4.1. The inappropriate application of batch correction

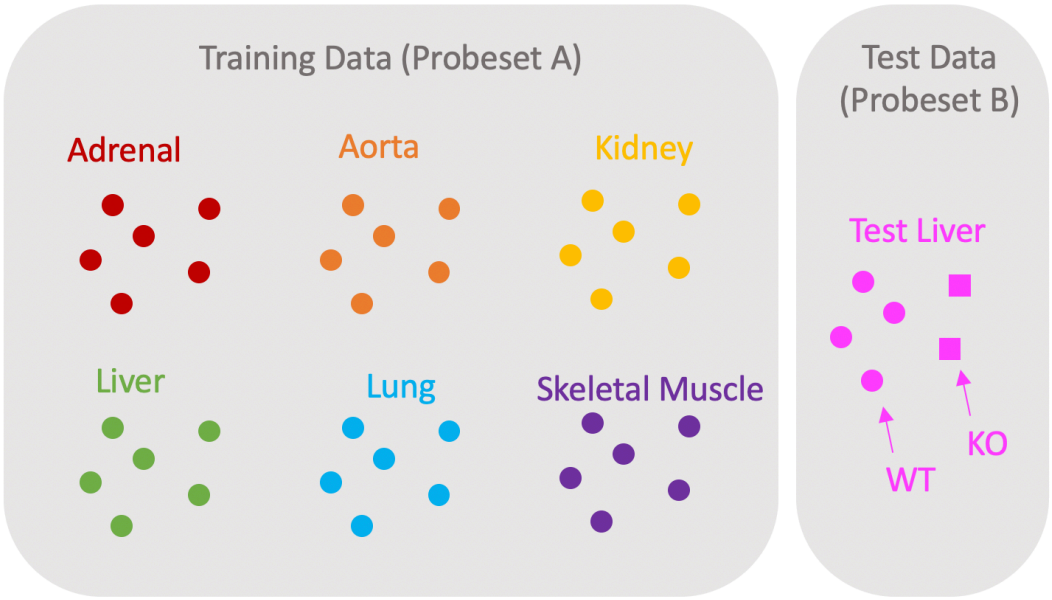


Figure 4.1 An example of the batches that Hughey et al. (2016) identified in order to implement ComBat to enable the comparison of training and test datasets from different microarray probesets. Test data genotype was identified as a covariate by Hughey et al. (2016), however the batches were unbalanced as all of the training data batches from probeset A contained only WT samples.

methods to datasets with unbalanced batches is reviewed elsewhere (Nygaard et al., 2016). A second disadvantage to the use of batch correction is that it necessitates that the model be re-trained for every new training and test dataset combination post- batch correction, therefore it is difficult to directly compare ZeitZeiger's findings from different test datasets. Ultimately, one of the primary aims for the design of models such as ZeitZeiger is to be able to estimate the biological time of human biopsies – both healthy and unhealthy. Batch correction of human data would require *a priori* knowledge of covariates, which may not even be known. In other words, 'real world' data contains many variables which would confound batch correction such as patient age, sex, time of sampling, health status etc. Therefore, a method for time prediction should ideally be applicable without the need for batch correction.

ZeitZeiger has since successfully been applied to human blood microarray datasets, again requiring batch correction prior to time prediction (Hughey, 2017). ZeitZeiger has also been applied to human peripheral blood mononuclear cells assayed by NanoString in order to train a model to predict biological time (relative to dim light melatonin onset (DLMO)) in test data samples (Wittenbrink et al., 2018). The consistency of the NanoString method, with built in housekeeping, negative and positive controls allowed for normalisation to be implemented without the need for batch correction methods. NanoString certainly holds promise for the development of clinical circadian biomarker tools, partly as it is cheaper than either RNA-seq or microarray (Wittenbrink et al., 2018).

4.1.2.6 BIO_CLOCK – Agostinelli et al., (2016)

BIO_CLOCK is described as a 'supervised deep learning algorithm with neural networks' and was trained using a feature-set of 16 oscillatory genes from a mixture of murine datasets (RNA-seq and microarray, multiple organs).

Datasets from various sources were combined and normalised by time-course mean and standard deviation prior to separation into 70% training and 30% test datasets. The model reportedly predicted time of single test samples with a mean absolute error of 1.22 hrs when the following training gene set was used: *Arntl, Per1, Per2, Per3, Cyr1, Cry2, Nr1d1, Nr1d2, Bhlhe40, Bhlhe41, Dbp, Npas2, Tef, Fmo2, Lonrf3* and *Tsc22d3*. However, no figures were provided to support this claim, no rationale was given as to how these genes were selected, and the methods description was quite concise. The amalgamation of multiple datasets prior to separation into 'training' and 'test' sets required normalisation of gene expression to time-course means and standard deviations, which meant that time estimation of single samples was not truly 'single timepoint'. It was reported in 2017 that BIO_CLOCK was not yet available for testing on further datasets since the model was being retrained (Laing et al., 2017).

4.1.2.7 PLSR – Laing et al., 2017

Laing et al., (2017) developed a slightly different approach to single timepoint biological time prediction using partial least squares regression (PLSR) (Laing et al., 2017). Instead of predicting a one dimensional output, i.e., time, PLSR was used to predict 'melatonin phase' a two-dimensional periodic output. The advantage of this approach is that melatonin phase, or more specifically DLMO, is a clinical biomarker that is commonly used as a proxy for SCN phase. Therefore, the predicted time can be related to the patients' own circadian timing. Laing et al., (2017) trained their model on the expression of 100 genes assayed by microarray from round-the-clock blood samples of 26 individuals for whom DLMO data was available. The 100 transcriptomic features that formed the predictor variables were reduced to five dimensions optimised to explain the maximal variance of the melatonin phase response variable (Talamanca & Naef, 2020). The method predicted DLMO with an

error of <2 hrs in 54% of validation samples, although the validation set and training set were formed from different individuals from within the same study, therefore the model requires more robust testing. A point of interest of this method is that DLMO predictions were quite variable in accuracy for individuals that had experienced sleep deprivation (relative to measured DLMO). It is not clear why the method did not perform consistently across the sleep-deprived population of individuals.

4.1.2.8 TimeSignature – Braun et al., 2018

Braun et al., 2018 have described a machine-learning approach that was trained on human blood transcriptomic datasets to predict the biological time of samples. TimeSignature was trained using RNA-seq and transcriptomic data, with the resulting caveat that expression data of each gene must be mean-normalised for each individual. In practice this means that much like BIO_CLOCK, TimeSignature cannot be used for single sample timepoint time prediction. Braun et al. (2018) state that two time-points at opposite phases of the circadian clock (~12 hrs apart to provide an accurate mean estimate) are the minimum requirements for test data samples. This thesis is focused on single timepoint time-telling, therefore this method will not be discussed in further detail.

4.1.2.9 CYCLOPS – Anafi et al., 2017

The methods described thus far require time-annotated training datasets and also, with the exception of ZeitZeiger's application to murine datasets, tissue-matched training datasets. This creates an obvious problem for the clinical application of single timepoint time-telling, namely that time-annotated, tissue-specific training datasets are extremely challenging to obtain. CYCLOPS aims to address this gap by assigning relative circadian time where

time of sampling is unknown (Anafi et al., 2017). CYCLOPS was designed to assign time to samples within large (>250 samples) transcriptomic datasets composed of single samples from different individuals over the course of the day/night period. Though, CYCLOPS is not capable of true single timepoint time-telling, because time is predicted relative to other samples in the dataset, it has the potential to support single timepoint time-telling approaches. Briefly, the CYCLOPS method was designed as follows:

1) The genes in the dataset that were observed to represent the top 2.5% variably expressed genes, and were also known to be rhythmic in mice, were selected to build a feature set.
2) The dimensionality reduction method singular value decomposition (SVD) was applied to the mean-scaled features, such that 85% of the variation over time in the dataset was explained.
3) A neural network subsequently optimally weighted and combined the resultant singular values, dubbed 'eigengenes', using a circular node autoencoder to create an elliptical curve in 2D space, a phenomenon which occurs in oscillatory systems (Alter et al., 2000; Anafi et al., 2017).
4) The authors validated CYCLOPS against 146 pre-frontal cortex biopsies obtained post-mortem with annotated time of death, which as a population had previously been demonstrated to exhibit oscillatory expression of core clock genes (Anafi et al., 2017; Chen et al., 2016). The calculated CYCLOPS phase correlated reasonably well with time of death with a median absolute error of 1.69 hr; some variation was to be expected due to differences in the circadian entrainment of individuals, and the post-mortem nature of the samples (Anafi et al., 2017).

CYCLOPS has also been applied to 249 matched hepatocellular carcinoma (HCC) and HCC tumour margin biopsies (Anafi et al., 2017; Lamb et al.,

2011). As expected, tumour margin samples ordered over time by CYCLOPS demonstrated population-wide oscillations in core clock gene expression. The authors did not take advantage of the paired nature of the data to order tumour samples according to the predicted CYCLOPS phase of tumour margin samples. Instead, the tumour biopsy data was projected onto tumour margin eigenvector space, and then ordered using CYCLOPS. The HCC tumour biopsies demonstrated population-wide oscillations in 8 of 9 core clock genes presented but showed a reduction in amplitude of oscillation and magnitude of expression in some of these genes. It is worth noting that up to approximately 10-20% of 'healthy' tumour margin samples were ordered to a circadian phase which presumably corresponded with night-time. As it seems unlikely that hepatocellular carcinomas were resected overnight, these individuals may have been poorly entrained, which is not wholly unexpected given that up to 30-60% of cancer patients are thought to experience sleep disruption (Fortner et al., 2002; Malone et al., 1994; Savard et al., 2001). A weakness of CYCLOPS is that because the original dataset is often not time-annotated, it is difficult to validate the method. For instance, CYCLOPS has been applied to the Genotype-Tissue Expression (GTEx) dataset of >4000 post-mortem RNA-seq samples from >600 donors in 13 different tissue types in order to apply relative time-ordering to the samples (Ruben et al., 2018). The resultant population-wide gene expression time-courses are available for easy consultation via the publicly available resource: http://circadb.hogeneschlab.org/human. However, validation of the time-ordering method itself is difficult as the GTEx dataset is not time-annotated, sample collection occurred up to 24 hrs post-mortem, and further to this many of the patients were presumably critically ill prior to death. TimeSignature was recently applied to blood samples from critically ill patients that were monitored within 24 hrs of admission to critical care units, to demonstrate that blood samples of these patients displayed dramatically altered circadian gene expression relative to healthy controls (Maas et al.,

2020). CYCLOPS has the significant potential to provide the circadian field with time-ordered training datasets – but it should be remembered that application of CYCLOPS to datasets like GTEx that were never intended to explain the circadian behaviour of healthy individuals, should be approached with caution.

4.1.2.10 Clock Correlation Distance (CCD) – Shilts et al., 2018

The developers of ZeitZeiger have also proposed a method which, like CYCLOPS, but unlike ZeitZeiger, does not rely on the existence of tissue-matched, time-stamped training datasets (Shilts et al., 2018). Shilts et al. (2018) calculated a metric called the 'clock correlation distance' (CCD) to summarise the relative expressions in any given sample of 12 of the core clock genes which were previously used to build ZeitZeiger. It is important to note that CCD is not a time-telling method, instead it is intended to assess the degree of clock gene dysfunction of a population of samples. Briefly, the method can be described as follows:

1) For all samples in each dataset each pair of clock genes was plotted as an ellipse on x-y axes, as when eigengenes are graphed for CYCLOPS.
2) The Spearman correlation between each gene pair was calculated to provide a measure of the shape of the ellipse. 12x12 heatmaps of Spearman correlations were used to visualise the 66 possible relationships between 12 paired clock genes in healthy tissues.
3) The Fisher z-transformed weighted means of Spearman correlations were calculated and normalised to a mouse reference to calculate CCD. Comparison to a mouse reference (comprised of almost 200 samples over 8 tissues from 7 different studies) was possible since the Spearman correlations contain no information about the

different phase of entrainment between humans and mice (Shilts et al., 2018).

Shilts et al. (2018) applied the CCD metric to 12 different cancers types in the TCGA dataset, and also to 8 other publicly available tumour/tumour margin biopsy datasets. The CCDs calculated for each of the 20 tumour datasets were without exception higher than the CCDs of the corresponding non-tumour samples, indicating that clock gene expression of the tumour samples was more different relative to the mouse reference than clock gene expression of non-tumour samples. Interestingly, cancer datasets of the same type appeared different relative to one another in terms of Spearman correlation signatures. For instance, ΔCCD (CCD_{tumour} – $CCD_{non\text{-}tumour}$) differed between the four hepatocarcinoma datasets, and the Spearman correlation signature appeared very different between the lung adenocarcinoma samples from 3 different datasets. This could have been a result of differing times of sampling between different datasets, as although Shilts et al. (2018) claim their method is robust against differing sampling times, their comparisons of day-only or night-only samples in fact yielded different Spearman correlation signatures and different CCDs. Additionally, it is worth noting that the 12 TCGA datasets contain many more tumour samples than non-tumour (normal margin) samples. Therefore, time of sampling may not be consistent between tumour and non-tumour samples and could be reflected as a difference in CCD between the same cancer types. From the work of Shilts et al. (2018) it is possible to conclude that core circadian gene expression is probably disrupted in tumours over a population.

4.1.2.11 TimeTeller – Vlachou et al. 2020

TimeTeller is a semi-supervised algorithm constructed using dimensionality reduction. TimeTeller is unique in that it is designed to output both a prediction of biological time from a single sample, and a metric relating to the

confidence in the time prediction (Vlachou et al., 2020). This thesis will implement the TimeTeller method and extend its application to RNA-seq datasets. The method as applied by Vlachou et al. (2020) will be described briefly as follows, see chapter 4.2.3 for a detailed description of TimeTeller's application to RNA-seq datasets:

1) A transcriptomic training dataset was identified that contained J instances (i.e., number of individuals or tissues) observed over T evenly spaced time-points. Periodic features were selected using a combination of cosine fitting and dimensionality reduction to identify 10-15 probes that behaved similarly over time across J observations.
2) SVD was applied to the training features × observations matrix in order to identify 3 dimensions into which to project the training dataset.
3) Three-dimensional multi-variate Gaussian distributions were fitted to the projected data from each of the time points represented in the training dataset.
4) Discrete Gaussians corresponding to each timepoint were interpolated along a spline through all timepoints in order to extend the model to approximate continuous time.
5) New test samples were projected onto the model, and likelihood functions were generated to describe the proximity of the test data to the training data, throughout model time.
6) The likelihood functions were used to predict the test sample times, and also to infer the degree of confidence in the time prediction termed a 'clock dysfunction metric'.

TimeTeller has been used to detect differences in clock (dys)function between cohorts of healthy and non-healthy samples. Secondly, it has been applied to cancer patients to correlate clock dysfunction with overall survival (Vlachou et al., 2020). Thus far, TimeTeller is the only time-telling method that is both able to predict the biological time or circadian phase of a single biological

sample, and provide an estimate of clock dysfunction. Therefore, TimeTeller's expansion and application to RNA-seq datasets is of great interest and will form the bulk of this chapter.

4.1.3 Summary of time-telling algorithms in relation to circadian cancer research

Comprehensive analysis of publicly available tumour biopsy transcriptomes by Anafi et al. (2017) and Shilts et al. (2018) with the novel mathematical algorithms CYCLOPS and CCD has demonstrated population level dysregulation of rhythmic gene expression in human tumours. However, there consistently appears to be little correlation in the clock gene expression observed across different tumour types and even between cohorts of the same tumour type (Shilts et al., 2018; Y. Ye et al., 2018). It is unclear whether this might result from individual molecular clock difference, tumour difference or time-sampling difference. This demonstrates the importance of well-annotated publicly available datasets. In order to gain the best understanding of how rhythmic gene expression changes between different cancer types and different individuals, publicly available datasets should aim to provide time-stamped, tumour/non-tumour matched biopsies. Ideally data sampling would also be distributed over the 24 hour day, however in a clinical setting this is highly improbable. TimeTeller avoided the requirement for tissue-matched training datasets by normalising each probe in the feature set relative to all other probes in the feature set (Vlachou et al., 2020). Thus, TimeTeller was trained using oral mucosa time course microarray data, and then applied to predict time and clock dysfunction in breast cancer samples. This cross-tissue approach will be investigated with regard to RNA-seq data in this thesis chapter.

For a biological time-telling algorithm to be useful in assessing the molecular clock of a patient tissue sample in a research context, there are a number of key aims that should be considered.

1) The method should predict the biological time of the sample.
2) The method should be applicable to a single patient biopsy – i.e., no requirement for a time-course of samples from the same individual.
3) The method should aim to provide a metric indicating the confidence in the time prediction, i.e., serving as an indicator of clock dysfunction.
4) The method would ideally be applicable across different patient tissues.

After an extensive review of the literature TimeTeller is thus far the only time prediction model that claims to achieve all of these key aims. Each of these aims will be addressed in the application of TimeTeller to RNA-seq datasets.

4.2 TimeTeller Model Method

4.2.1 Publicly available datasets used

The results presented in this thesis chapter are dependent entirely on publicly available transcriptomics datasets that were used to train and test TimeTeller. All datasets used are detailed in Table 4.1.

Table 4.1 Metadata for mined datasets.

Zhang et al. 2014	
Technology	Microarray (Affymetrix MoGene 1.0 ST arrays)
GEO	GSE54652
Tissue(s)	Adrenal, aorta, brown fat, heart, kidney, liver, lung, skeletal muscle, white fat
Entrainment conditions	1 week 12 hr light/ 12 hr dark. Sample collection commenced after 18 hrs constant darkness.
Experimental conditions	Wild type
Timepoints	CT18 – CT64, every 2 hrs.
Number of replicates	1 sample per timepoint consisting of 3 pooled mice
Male or female?	Male
Age	6 weeks
Strain	C57BL/6
Zhang et al. 2014	
Technology	RNA-seq
GEO	GSE54652
Tissue(s)	Adrenal, aorta, brown fat, heart, kidney, liver, lung, skeletal muscle, white fat
Entrainment conditions	1 week 12 hr light/ 12 hr dark. Sample collection commenced after 18 hrs constant darkness.
Experimental conditions	Wild type
Timepoints	CT22 – CT64, every 6 hrs.
Number of replicates	1 sample per timepoint consisting of 3 pooled mice
Male or female?	Male
Age	6 weeks
Strain	C57BL/6
Kinouchi et al., 2018	
Technology	RNA-seq
GEO	GSE107787
Tissue(s)	Liver, skeletal muscle
Entrainment conditions	12 hr light / 12 hr dark
Experimental conditions	*Ad libitum* fed vs 24 hr starved
Timepoints	ZT0 – ZT20, every 4 hrs.
Number of replicates	3 samples per timepoint
Male or female?	Male
Age	8 weeks
Strain	C57BL/6
Weger et al., 2021	
Technology	RNA-seq
GEO	GSE135898
Tissue(s)	Liver
Entrainment conditions	12 hr light / 12 hr dark
Experimental conditions	*Bmal* KO with WT controls and *Cry1/2* double KO with WT controls
Timepoints	ZT0 – ZT20, every 4 hrs.
Number of replicates	2 samples per timepoint
Male or female?	Male
Age	12-16 weeks
Strain	C57BL/6

Yeung et al., 2018	
Technology	RNA-seq
GEO	GSE100457
Tissue(s)	Kidney
Entrainment conditions	12 hr light / 12 hr dark
Experimental conditions	*Bmal* KO with WT controls. Night restricted feeding protocol for all mice.
Timepoints	ZT0 – ZT20, every 4 hrs.
Number of replicates	2 samples per timepoint
Male or female?	Male
Age	8-12 weeks
Strain	C57BL/6

4.2.2 Data analysis pipeline

The Zhang et al. microarray data was normalised using frozen robust multiarray analysis (fRMA) normalisation as described by Vlachou et al. (2020) (McCall et al., 2010).

All RNA-seq data was processed using a CentOS Linux 7 kernel on the University of Warwick's CyVerse UK servers. RNA-seq datasets were downloaded as .sra files from NCBI's SRA database (accessed via the GEO database). SRA files were converted to fastq files using 'fasterq-dump' from NCBI's SRA Toolkit. FASTQ files were aligned to the mouse genome (GRC release m38.84) and converted to SAM files using HISAT2 v2.2.0 (Kim et al., 2019). SAM files were compressed to BAM using Samtools v1.10 (Heng Li et al., 2009). Transcript read counts were determined from the BAM files and the mouse transcriptome (GRCm38.84 .gtf file) using LiBiNorm v2.4, an in-house software package, in HTSeq-count mode (Anders et al., 2015; Dyer et al., 2019). Raw read counts were concatenated for all samples and exported as one text file for all subsequent analysis on a Mac OS. For dataset specific arguments see Table 4.2.

Raw count normalisation to logCPM was carried out in R Studio as described in chapter 2.2.10. Data was also inspected for quality by checking library sizes, replicate plots and PCA plots.

Table 4.2 Dataset specific RNA-sequencing parameters

	Zhang et al., 2014	**Kinouchi et al., 2018**	**Weger et al., 2021**	**Yeung et al., 2018**
Library prep kit	Illumina TruSeq stranded mRNA	Illumina TruSeq stranded mRNA	Illumina TruSeq stranded mRNA	Not stated in the paper, supplement or GSE series matrix txt file.
Reads	Paired end 100 bp	Single end 100 bp	Paired end 100 bp	Single end 100 bp
Strandedness	First read from the opposite/reverse strand	Single read from the opposite/reverse strand	First read from the opposite/reverse strand	Inspection of SAM flags indicated an unstranded protocol
fasterq-dump	Default	Default	Default	Default
hisat2	--rna-strandness RF	--rna-strandness R	--rna-strandness RF	Default
samtools	sort	sort	sort then merge	sort
LiBiNorm	count -z -i gene_id -s reverse	count -z -i gene_id -s reverse	count -z -i gene_id -s reverse	count -z -i gene_id -s no

All RNA-seq data was migrated from R Studio to MATLAB v2020b for use in TimeTeller.

4.2.3 Training dataset tissue and feature selection

The first step to running TimeTeller is to identify a training dataset with J observations over T evenly spaced time-points. For the purposes of explaining the model build in this methods chapter, the Zhang et al., 2014 RNA-seq time-series will be used as the exemplar training dataset (see Figure 4.9 for experimental design). The Zhang et al., dataset contains 12 tissues (i.e., observations) which were sampled repeatedly over time under constant darkness. Three tissues originating from the brain (brainstem, cerebellum, and hypothalamus) were omitted from the model build as rhythmic gene expression

in these brain tissues is of lower amplitude (Vlachou et al., 2020). White fat samples were also omitted from the model build as the white fat tissue sampled at circadian time (CT)40 appeared to have a significantly larger library size relative to the other samples. Subsequent investigation revealed that the CT40 sample may have been sampled from the genital fat pad, rather than the intended subcutaneous fat pad (Figure 4.2).

Periodic features were selected using a combination of cosine fitting and dimensionality reduction to identify genes that behaved rhythmically and synchronously over time across the 8 tissues. Since the shape of the gene expression in each tissue relates directly to clock function, gene expression was first normalised using a shape-preserving method. Expression of each gene in each tissue was z-scored over time, i.e., expression of each gene in each tissue was transformed to a standard normal distribution by subtracting the time-

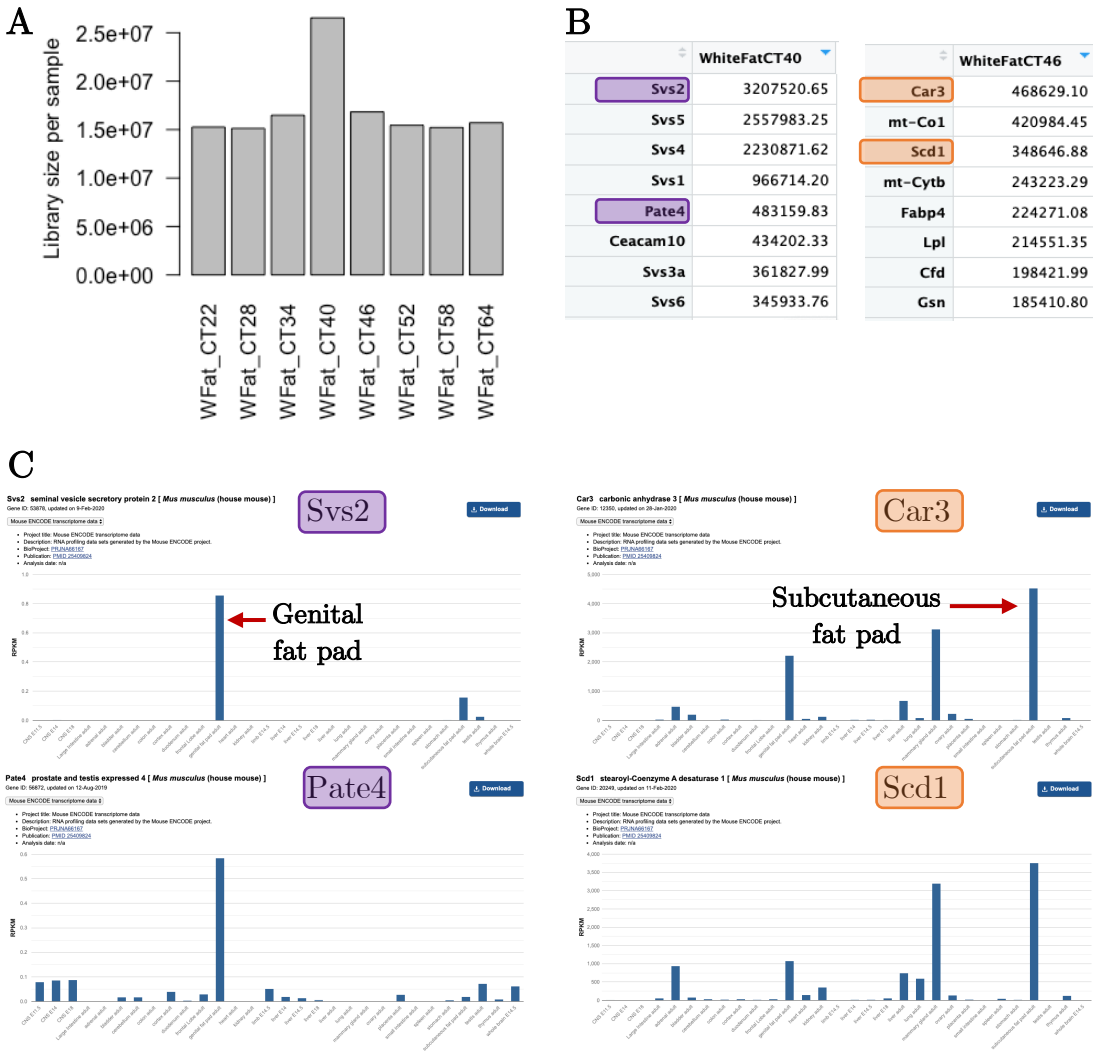


Figure 4.2 Quality analysis of Zhang et al., 2014 white fat time series data. A) Library sizes for the 8 white fat samples (each sampled at a different circadian time) – the library size of the CT40 sample is an outlier. B) The most highly expressed genes in the CT40 are quite different to those in the CT46 sample. C) ENCODE mouse expression atlas data (available at: www.ncbi.nlm.nih.gov/gene) demonstrates that the most highly expressed genes from the CT40 and CT46 samples are differentially expressed in different fat tissues.

course mean and dividing by the time-course standard deviation, as described by the following notation:

For each gene, $g = (1,...,G)$, and each tissue, $j = (1,...,J)$, the gene expression time series was contained within the vector $X_{j,g}$ where the number of samples in the time series was $n = (1,...,N)$ such that:

$$X_{j,g} = \left(x_{j,g_1}, \dots, x_{j,g_N}\right)$$

Each observation, n, of $X_{j,g}$ was normalised to the mean and standard deviation of the time series to calculate Z scores:

$$z_{j,g_n} = \frac{x_{j,g_n} - \bar{X}_{j,g}}{s\left(X_{j,g}\right)}$$

$Z_{j,g}$ was a vector of length N containing normalised time-course information for each gene g and tissue j.

This normalisation method shall henceforth be referred to as "time-course normalisation".

Time-course normalised genes were ranked for goodness of 24 hr period cosine fit using cosinor analysis. Cosinor is a least squares regression approach, originally developed for use with short/sparse datasets (Cornelissen, 2014). The 'cosinor' function written for MATLAB by Casey Cox (2008) was used to determine the genes with the best 24 hr cosine fits. One small change was made to the code, such that the p-value of the zero-amplitude f-test was calculated correctly using the 'fcdf' function rather than the 'fpdf' function. P-values for the zero-amplitude test (i.e., H_0=not rhythmic) were calculated for each gene in each tissue. The mean p-value of all tissues was retained for each gene.

Time-course normalised genes were also ranked for synchronicity using SVD. Matrices (M_g) were constructed for each gene, g, containing J columns of the aforementioned vectors $Z_{j,g}$ (i.e., M_g contained time-course gene expression for

every tissue). The matrices were decomposed such that $M_g = U_g S_g V_g^T$. For the genes that were expressed most similarly across tissues over time, a greater proportion of variation in the matrix was explained by the first principal component, i.e., the first column of U_g, $U_{g,1}$. The proportion of variance explained by the first principal component for each gene was calculated using the singular values σ_g obtained from the diagonal matrix S_g, as follows:

$$\%\ variance = \frac{\sigma_{g,1}^2}{\sum \sigma_g^2} \times 100.$$

Equally weighted ranks were assigned to both rhythmicity and synchronicity scores, and summated, in order to select a feature set for TimeTeller.

4.2.4 Data normalisation approaches

Time-course normalisation has already been described as part of feature selection (chapter 4.2.3). Time-course normalisation is conditional upon knowledge of the mean and standard deviation of feature gene expression in each tissue over time. This information is unknown for independent test data samples which represent a single point in time, n. Therefore, other methods are required for normalisation of independent test data samples.

Vlachou et al. (2020) proposed a data normalisation approach that shall be referred to here as "inter-gene" normalisation. Briefly:

> The data was structured such that for each sample of time $n = (1,...,N)$ and each tissue, $j = (1,...,J)$, expression of the G selected feature genes was contained within the vector $X_{j,n}$:
>
> $$X_{j,n} = \left(x_{j,n_1}, \dots x_{j,n_g}, \dots, x_{j,n_G}\right)$$
>
> Expression of each gth gene in $X_{j,n}$ was normalised to the mean and standard deviation of the expression set to calculate z-scores:

$$z_{j,n_g} = \frac{x_{j,n_g} - \bar{X}_{j,n}}{s(X_{j,n})}$$

$Z_{j,n}$ was a vector of length G containing inter-gene normalised gene expression for each tissue j at a time n ($z_{j,n}$) as its elements.

A model trained using inter-gene normalisation, would also require that test datasets be normalised using inter-gene normalisation.

An alternative approach was used in this thesis to normalise test data when the training data was normalised using time-course normalisation, and is referred to as "quasi-time-course" normalisation. This approach uses the time-dependent mean and standard deviation of gene expression in the corresponding training data tissue to normalise the test data as described by the following notation:

For each independent test data sample originating from a tissue j and a time n, the gene expression for genes $g = (1,...,G)$ was organised in the vector $Y'_{j,n}$ such that:

$$Y'_{j,n} = \left(y'_{j,n_1}, \ldots, y'_{j,n_G}\right)$$

Expression of each gth gene in $Y'_{j,n}$ was normalised using the $\bar{X}_{j,g}$ and $s(X_{j,g})$ from the corresponding tissue and gene in the training data such that:

$$z'_{j,n_g} = \frac{y'_{j,n_g} - \bar{X}_{j,g}}{s(X_{j,g})}$$

$Z'_{j,n}$ was a vector of length G containing quasi-time-course normalised gene expression for a test sample of tissue j at time n (z'_{j,n_g}) as its elements.

4.2.5 TimeTeller model build

4.2.5.1 Dimensionality reduction

After feature selection and either inter-gene or time-course data normalisation, the first step in building TimeTeller is to reduce the dimensionality of the training data. In the case of the Zhang et al. training datasets G=11 features were reduced to dimensions d = (1,...,D) where D=3. Chapter 4.3.1 provides a longer discussion regarding the reasoning for G=11 and D=3.

Normalised training data for each of the timepoints (t_i) contained within the training dataset was organised into a $G \times J$ matrix, M_{t_i}, where the column-wise normalised observations of M_{t_i} were the vectors $Z_{t_i,j}$ of length G. SVD was applied to M_{t_i} in order to extract the eigenvectors for projection of the data onto principal component axes, such that:

$$M_{t_i} = U_{t_i} S_{t_i} {V_{t_i}}^T$$

Where the columns of the $G \times G$ matrix U_{t_i}, $U_{t_{i_d}}$ corresponded to the left singular vectors of M_{t_i} (i.e., the principal components), the columns of the $J \times J$ matrix V_{t_i} corresponded to the right singular vectors of M_{t_i}, and S_{t_i} was a diagonal matrix containing the singular values σ_{t_i} corresponding to each of the column-wise singular vectors.

This was termed a 'local' principal component approach, i.e., a set of principal components was generated for each timepoint t_i in the training dataset. Vlachou et al. (2020) demonstrated that because the projection of training data onto D-dimensional principal component axes may form an inherently asymmetrical shape in D-dimensional space, multiple local projections of the data improved the accuracy of time prediction.

For each observation in the normalised training data set, j, at each time instance, n, the feature gene expression was organised into the vectors $Z_{j,n}$. Each instance of $Z_{j,n}$ were projected into D dimensional space using the first D columns of U_{t_i}, denoted as $U_{t_{i_D}}$:

$$Q_{t_i} = U_{t_{i_D}}{}^T Z_{j,n}$$

The resultant vector $Q_{t_i,j,n}$ of length D contained the projection of a single training data observation of time n and tissue j using the first D principal components of the data associated with each training data timepoint t_i. The $D \times (J \times N)$ matrix Q_{t_i} contained the vectors $Q_{t_i,j,n}$ as its columns. Figure 4.3 displays Q_{t_i} plotted on D=3 principal component axes for each t_i, for the Zhang et al. RNA-seq dataset.

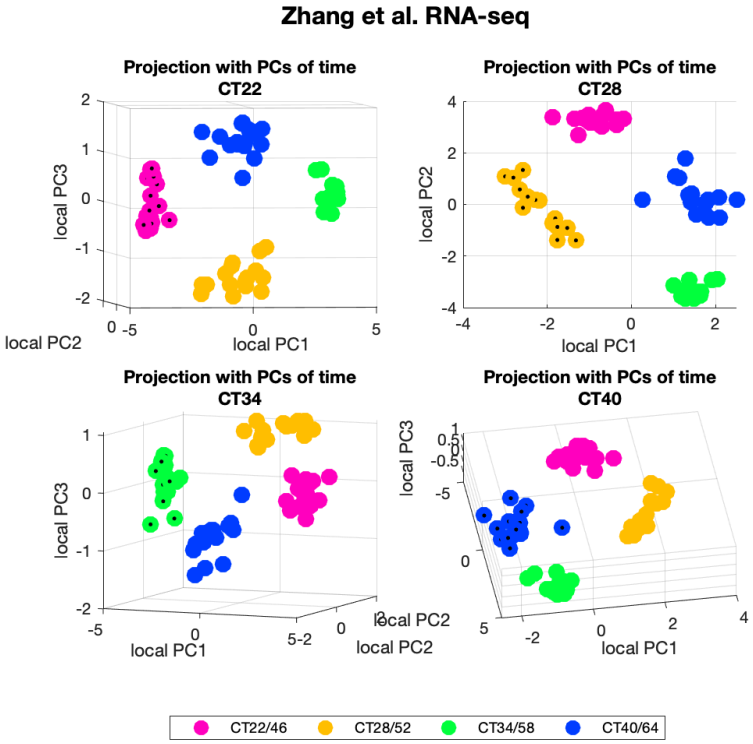


Figure 4.3 Four local projections of the Zhang et al. RNA-seq training data. Data was folded such that samples from CT22 and CT46 were considered to belong to the same timepoint. Each graph displays the training data projected onto the first 3 principal component axes contained within the matrix Q_{t_i}. The four graphs correspond to the four values of t_i. Graphs are oriented to best show the separation of the clusters in 3D space.

4.2.5.2 Gaussian fitting and model interpolation

In order to allow for continuous time prediction, the model was interpolated from discrete time to approximate continuous time. Each local projection data matrix Q_{t_i}, was subdivided into the $D \times J$ matrix $Q_{t_i,n}$ to describe the data for each time n for a fixed t_i projection. $Q_{t_i,n}$ was presumed to be normally distributed. The MATLAB function 'fitgmdist' was used to fit a multivariate Gaussian distribution to each $Q_{t_i,n}$, of mean $\mu_{t_i,n}$ and covariance matrix $\Sigma_{t_i,n}$. The resultant Gaussian distributions are represented in Figure 4.4 by mesh ellipsoids indicating 3 standard deviations of the total probability mass, i.e., encompassing ~97% of the population data within the ellipsoid (plotted using 'plot_gaussian_ellipsoid') (Vallabha, 2021). A cubic spine with periodic end conditions was fitted to interpolate through the means of the Gaussian distributions i.e., $\mu_{t_i,n}$ for all values of n, using the MATLAB function 'cscape'. The spline for each local t_i projection is also displayed in Figure 4.4.

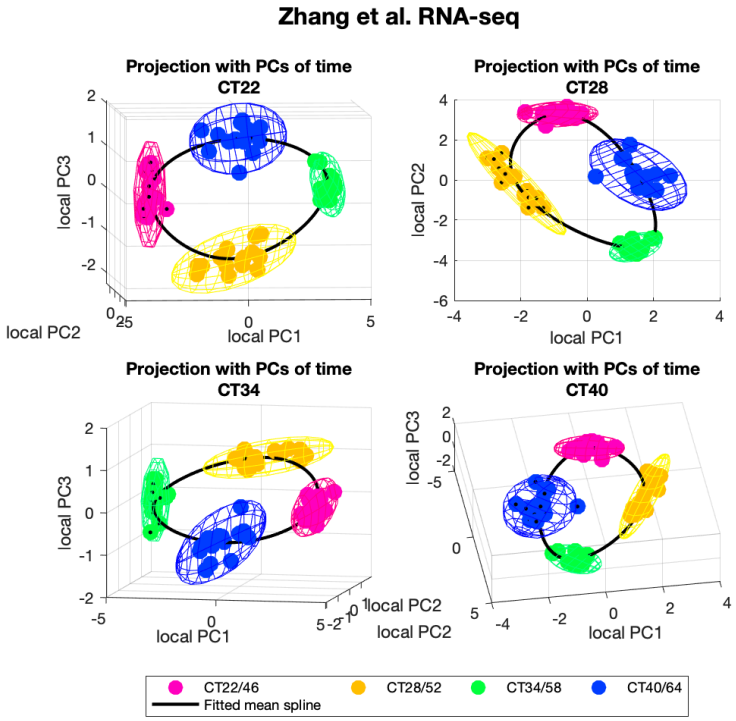


Figure 4.4 Four local projections of the Zhang et al. RNA-seq training data. Training data displayed as for Figure 4.3, but with mesh ellipsoids representing the 97% boundary of the Gaussian distributions fitted to each $Q_{t_i,n}$. The splines represent mean continuous time.

The covariance matrices of the Gaussian distributions, $\Sigma_{t_i,n}$, were interpolated for continuous time along the shape-preserving mean spline using the MATLAB function 'pchip'. The resultant interpolated Gaussian distributions approximated continuous time. In total 384 Gaussians were used to form the model (1 Gaussian per 3.75 min). This was a sufficiently high resolution to predict time accurately. Figure 4.5 demonstrates the result of interpolation.

The Gaussian distributions that were interpolated from the training Gaussian distributions described by $\mu_{t_i,n}$ and $\Sigma_{t_i,n}$, will be referred to as $\mu_{t_i}(t)$ and $\Sigma_{t_i}(t)$, where t = interpolated timepoints (in this instance t=1,...,384).

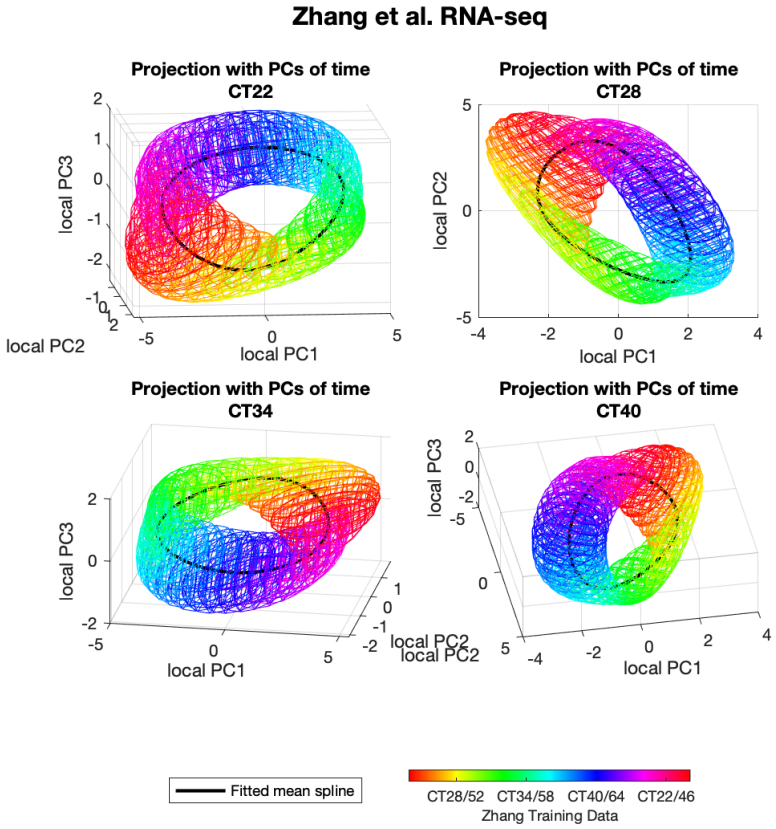


Figure 4.5 Four local projections of the Zhang et al. RNA-seq training data. Splines represent mean continuous time. Mesh ellipsoids represent the 97% boundary of the interpolated Gaussian distributions fitted to each $Q_{t_i,n}$, extending the model to continuous time. N.B. for ease of visualisation only 1 in every 16 of the interpolated Gaussians are plotted.

4.2.5.3 Projection of independent samples onto the model

Independent test observations were projected onto the trained TimeTeller ellipsoids using the same D principal components that were used to transform the training data to principal component space, such that:

$$Q'_{t_i,j,n} = {U_{t_i}}_D^T Z'_{j,n}$$

Where $Z'_{j,n}$ represents an independent test observation sampled at a given time n and tissue j. And $Q'_{t_i,j,n}$ represents a vector of length D generated for each of t_i local principal components.

4.2.5.4 Obtaining time predictions for independent test samples

The likelihood that the projected test data sample belonged to each of the interpolated Gaussian distributions (described by $\mu_{t_i}(t)$ and $\Sigma_{t_i}(t)$) was calculated using the probability density function of each Gaussian (MATLAB function 'mvnpdf') as described by the following equation:

$$L_{Q'_{t_i,j,n}}(t) = \frac{1}{\sqrt{|\Sigma_{t_i}(t)|(2\pi)^3}} e^{\left(-\frac{1}{2}\left(Q'_{t_i,j,n}-\mu_{t_i}(t)\right)\Sigma_{t_i}(t)^{-1}\left(Q'_{t_i,j,n}-\mu_{t_i}(t)\right)^T\right)}$$

Figure 4.6 displays this process pictographically. All resultant likelihood curves, $L_{Q'_{t_i,j,n}}(t)$, were thresholded – such that any likelihoods that were lower than an arbitrarily low threshold (in this case thresh$=e^{-12}$/ logthresh$=-12$) were replaced with the threshold value. The thresholded likelihood curves generated for each local projection, $t_i = 1, \ldots, T$, were combined as follows:

$$L_{Q'_{j,n}}(t) = \frac{1}{T}\sum_{t_i=1}^{T} \log L_{Q'_{t_i,j,n}}(t)$$

Such that the proximity of each independent projected test data point $Q'_{j,n}$ to the trained TimeTeller model was described by a single likelihood curve, $L_{Q'_{j,n}}(t)$ – essentially a geometric mean of all thresholded local likelihood

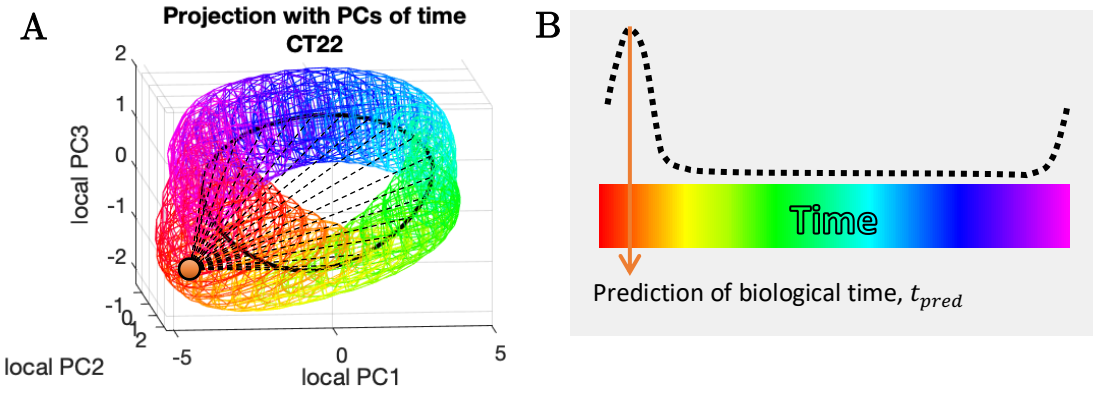


Figure 4.6 A) A representation of an independent test data sample (orange circle) projected onto the trained TimeTeller model of continuous time. Black dashed lines represent the relationship between the projected test sample, $Q'_{t_i,j,n}$, and each of the 3D Gaussian distributions, which is defined by the multivariate normal probability density function, such that $L_{Q'_{t_i,j,n}}(t)$ contains the likelihood of $Q'_{t_i,j,n}$ belonging to time t. B) A representation of the likelihood curve, $L_{Q'_{t_i,j,n}}(t)$, resulting from the scenario depicted in A). The maximum of the likelihood curve corresponds with the model time t_{pred} that best represents the test sample.

curves. The reason for thresholding the likelihood curves was that the minima of the likelihood curves could theoretically have approached 0. Such incredibly low values might have biased the geometric mean of the local likelihood curves, such that true peaks that occurred only in a small number of local projections, but that were nonetheless accurate, would have been smoothed away. Therefore, the threshold helped to prevent extremely low likelihood values from biasing time prediction.

The maximum of the likelihood curve $L_{Q'_{j,n}}(t)$ provided the model time, t_{pred}, that best represented the test sample $Z'_{j,n}$ as indicated by Figure 4.6B.

4.2.5.5 Defining a metric, θ, for confidence in t_{pred}

The shape of the likelihood curve, $L_{Q'_{j,n}}(t)$, generated for each test data sample contained information that relates to the accuracy of the maximum time

prediction, t_{pred}. Three scenarios illustrated by Figure 4.7 display how the raw likelihood curves may vary given different test data samples. The leftmost scenario shows that a test data sample which overlays the training data will have a high maximum likelihood and a distinct peak from which to estimate predicted time. However, the middle scenario shows that when a test data sample is projected towards the centre of the ellipsoidal model, the maximum of the likelihood curve is lower, and the shape is broader. The rightmost side shows that a very distant test data sample will not produce a maximum likelihood above logthresh, and t_{pred} cannot be estimated. Thus far, the absolute value of the maximum likelihood would be sufficient to estimate the confidence in t_{pred}. However, the likelihood curve of each sample, $L_{Q'_{j,n}}(t)$, was

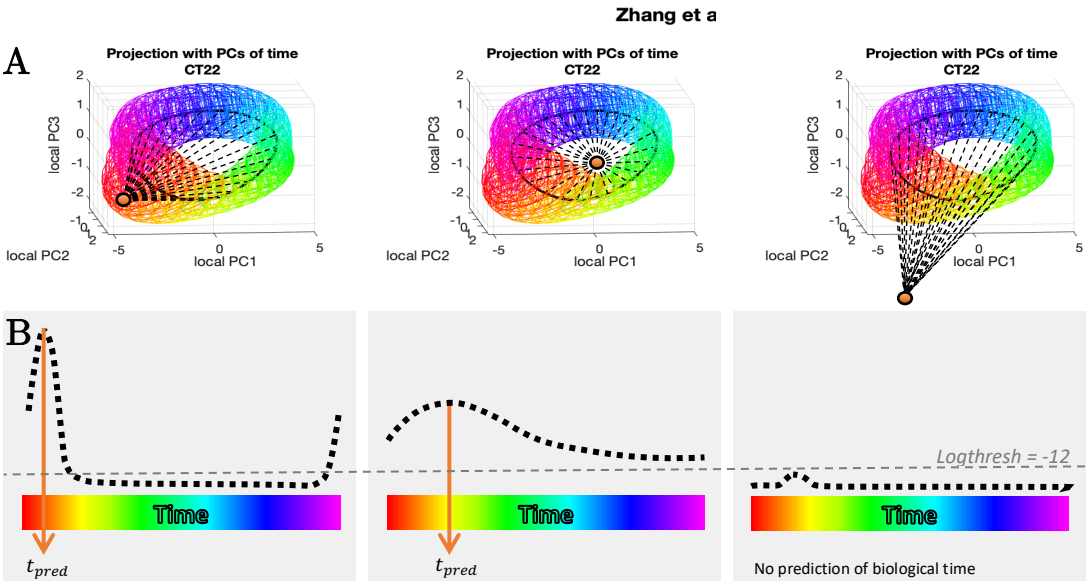


Figure 4.7 A) A representation of different independent test data samples (orange circles) projected onto the trained TimeTeller model of continuous time. Black dashed lines represent the relationship between the projected test sample, $Q'_{t_i,j,n}$, and each of the 3D training Gaussian distributions, which is defined by the multivariate normal probability density function, such that $L_{Q'_{t_i,j,n}}(t)$ contains the likelihood of $Q'_{t_i,j,n}$ belonging to time *t*. B) A representation of the likelihood curve, $L_{Q'_{t_i,j,n}}(t)$, resulting from each of the test data points depicted in A). The maximum of the likelihood curve corresponds with the model time t_{pred} that best represents the test sample. The position of logthresh ensures that relatively low maximum likelihoods do not result in a time prediction.

generated from the geometric mean of T thresholded local likelihood curves, which in turn were generated from asymmetric local projections. Therefore, the resultant likelihood curves frequently contained more than one peak, and a more sophisticated method was required to determine the confidence in t_{pred}.

The further that the primary peak lay from any secondary peaks (to a maximum difference of 12 hrs), the more likely it was that the time prediction from the maximum likelihood estimate, t_{pred}, was inaccurate. Therefore, a cosine function termed a 'clock dysfunction threshold' was implemented to calculate a metric, termed 'theta' (Θ), that progressively penalised secondary peaks which were increasingly distant from the primary peak. To implement the clock dysfunction threshold and calculate Θ, the likelihood curves ${L_{Q'}}_{j,n}(t)$ were first centred on the primary peak and scaled to a maximum of 1 to give ${\Lambda_{Q'}}_{j,n}(t)$, such that for $t = (t_{pred} - 12\ hrs, \dots, t_{pred}, \dots, t_{pred} + 12hrs)$:

$$ {\Lambda_{Q'}}_{j,n}(t) = \frac{{L_{Q'}}_{j,n}(t)}{{L_{Q'}}_{j,n}(t_{pred})} $$

Figure 4.8 provides some examples of how ${\Lambda_{Q'}}_{j,n}(t)$ may appear. The clock dysfunction threshold is also depicted in Figure 4.8, and was defined as follows:

$$ C(t|t_{pred}) = \eta\left(1 + \epsilon + \cos\left(\frac{t - t_{pred}}{24}2\pi\right)\right) $$

where $\epsilon = 0.4$ and $\eta = 0.35$. Chapter 4.2.5.6 discusses the setting of parameters ϵ and η.

The clock dysfunction metric Θ, was defined as the proportion of time t for which the following held true:

$$ {\Lambda_{Q'}}_{j,n}(t) \geq C(t|t_{pred}) $$

Therefore Θ was equal to the proportion of the centred scaled likelihood curve which lay above the clock dysfunction threshold (i.e., in the red region in Figure 4.8).

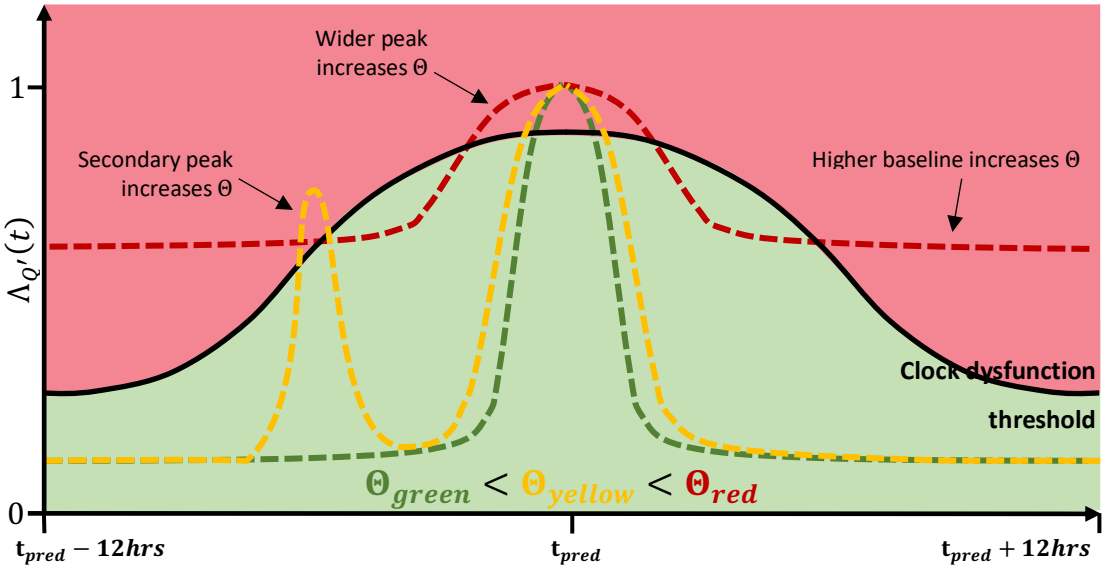


Figure 4.8 Depiction of three exemplar centred scaled likelihood curves, $\Lambda_{Q'}(t)$, (red, yellow and green). The relationship between the shape of the likelihood curve and the clock dysfunction threshold (black) determines the value of Θ.

The Θ metric was designed with secondary peaks in mind, but Figure 4.8 demonstrates that Θ was dependent upon the shape of the likelihood curve as a whole. For instance, the higher the maximum likelihood relative to all other likelihoods, the sharper the primary peak of the likelihood curve. Conversely, a wider flatter primary peak represented a lower confidence in t_{pred}. The relative width of the primary peak contributed to the calculation of Θ. A further factor that contributed to Θ values, was the height of the baseline of the scaled likelihood curve. The height of the baseline is related to both the ratio between the lowest likelihoods and the maximum likelihood, and to the logthresh. In other words, when the maximum likelihood estimate was not that much higher than the minimum likelihoods (or indeed logthresh), the value of Θ was likely to increase.

Higher Θ values served to identify time predictions in which one should have relatively low confidence. Thus, Θ functions as an excellent 'red flag metric' and is termed the 'clock dysfunction metric'. The underlying reason for the low

confidence in a predicted time is not apparent solely from Θ, and usually requires further inspection of the likelihood curves or local projections.

4.2.5.6 Setting ϵ and η to determine the clock dysfunction threshold

The placement of the cosine function $C(t|t_{pred})$ termed the 'clock dysfunction threshold' was somewhat arbitrary. The minima of $C(t|t_{pred})$ should clearly be >0, whilst the maxima should clearly be <1. As the minima of $C(t|t_{pred}) = \eta\epsilon$, and the maxima of $C(t|t_{pred}) = \eta(2+\epsilon)$ it holds that:

$$0 < \eta\epsilon < \eta(2+\epsilon) < 1$$

Here, $\epsilon = 0.4$ and $\eta = 0.35$, such that $\eta\epsilon = 0.14$ and $\eta(2+\epsilon) = 0.84$, satisfying the equation above. These are the same values of ϵ and η that were selected by Vlachou et al. 2020. In practice, the exact values of ϵ and η were not too important, providing that ϵ and η were set such that the values of Θ remained relatively minimal for time predictions of high confidence.

4.3 Results and Discussion

4.3.1 TimeTeller model build applied to microarray vs RNA-seq

There is currently a shortage of time-stamped training datasets in human tissue. Therefore, for the optimisation of time prediction algorithms it is common to use murine datasets (Hughey et al., 2016). For the migration of the TimeTeller method from microarray to RNA-seq datasets, the Zhang et al. (2014) dataset used by Hughey et al. (2016) is an excellent choice because the same samples underwent RNA-seq and microarray analysis. Figure 4.9 describes the sampling protocol that was followed by Zhang et al. (2014). Whilst the microarray dataset contains 24 time points (samples every 2 hrs for 48 hrs), the RNA-seq dataset contains only 8 time points (samples every 6 hrs for 48 hrs). Therefore, the microarray dataset was first used to test whether TimeTeller could perform adequately on such a low-resolution time course.

TimeTeller was trained using an iterative leave-one-tissue-out method, so as to ensure that the test tissue remained independent of the trained model. The

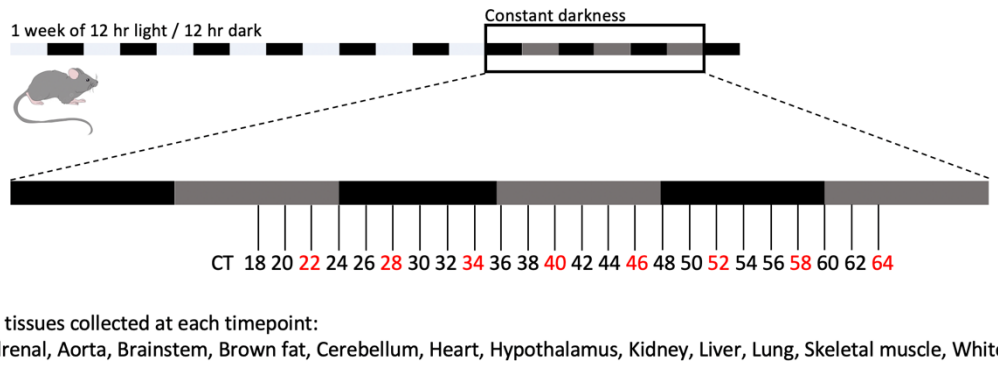


Figure 4.9 Sampling protocol for the mouse circadian transcriptomic datasets published by Zhang et al., 2014. All samples were analysed by microarrays. Red samples were also analysed by RNA-seq. Only 8 tissues from the dataset are used for the model build (i.e., no brain samples, nor white fat, see chapter 4.2.3 for details).

model was built exactly as described by Vlachou et al., (2020), i.e., using inter-gene normalisation (see chapters 4.2.4 & 4.2.5). Figure 4.10A displays the time predictions that TimeTeller produced for each of the samples in the microarray dataset when trained using all 24 time points (CT18 → CT46). Meanwhile, Figure 4.10B shows the time predictions that TimeTeller produced for each of

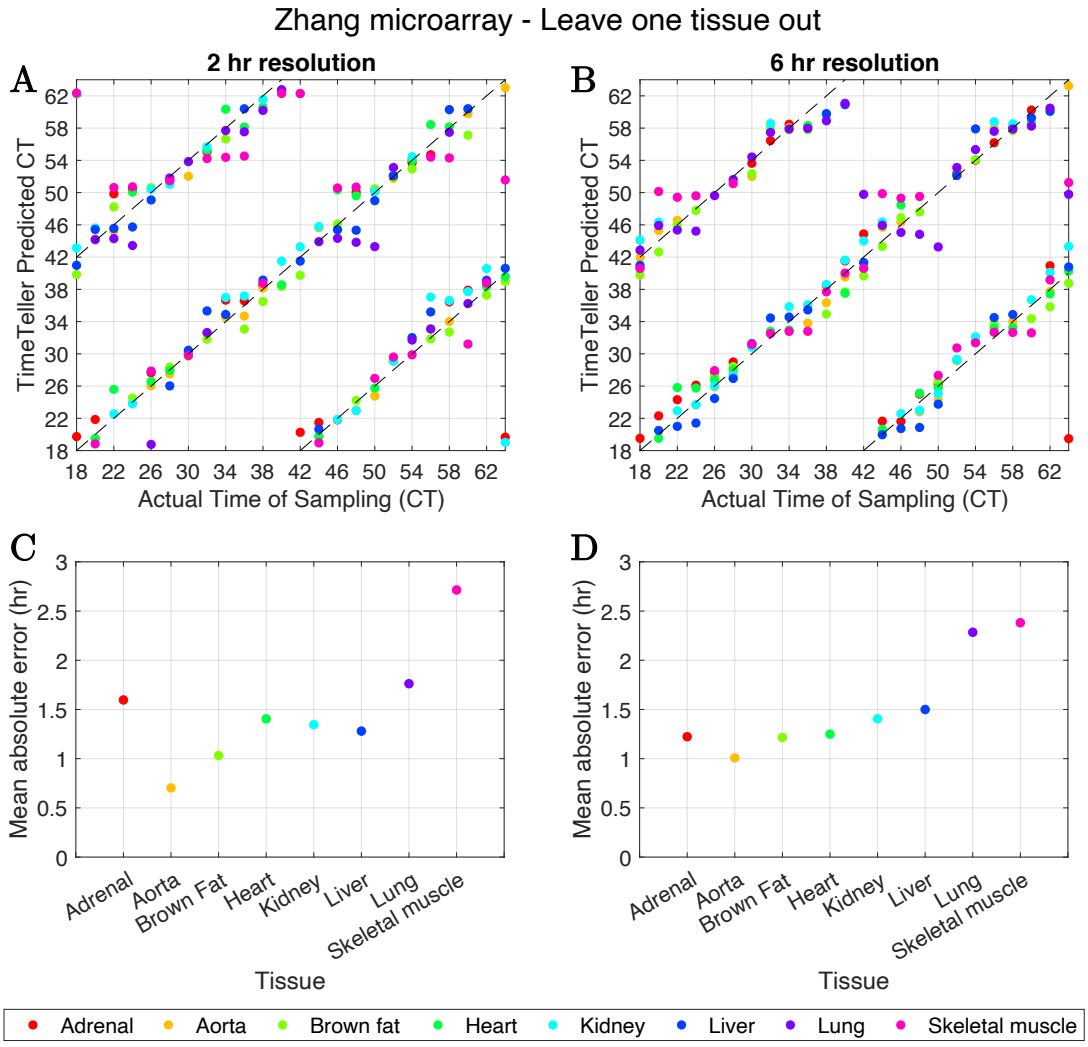


Figure 4.10 Results of TimeTeller trained on Zhang microarray data in a leave-one-tissue-out fashion. A&C) Full 2 hr resolution time course used for training. B&D) Subset 6 hr resolution time course used for training. A&B) TimeTeller's predicted time vs actual time. C&D) Mean of each tissue, (absolute error = absolute difference between predicted time and actual time, accounts for 0hr=24hr). E) Overall MAE for the high- and low-resolution time prediction models.

E

Training resolution	MAE (hr)
2 hr	1.4798
6 hr	1.5334

the samples in the microarray dataset when trained using only the 8 timepoints contained within the RNA-seq dataset (CT22, 28, 34, 40, 46, 52, 58 and 64). Figure 4.10A and B demonstrate that the 2 hr and 6 hr model builds performed very similarly. Figure 4.10E confirms that the mean absolute error (MAE) of all of the test samples was comparable between the 2 and the 6 hr model builds, suggesting that TimeTeller's interpolation of low-resolution training data in principal component space was perfectly adequate for accurate time prediction.

Figure 4.10C&D are of particular interest because they demonstrate that the accuracy of time prediction varied by up to 2 hrs for different tissue types.

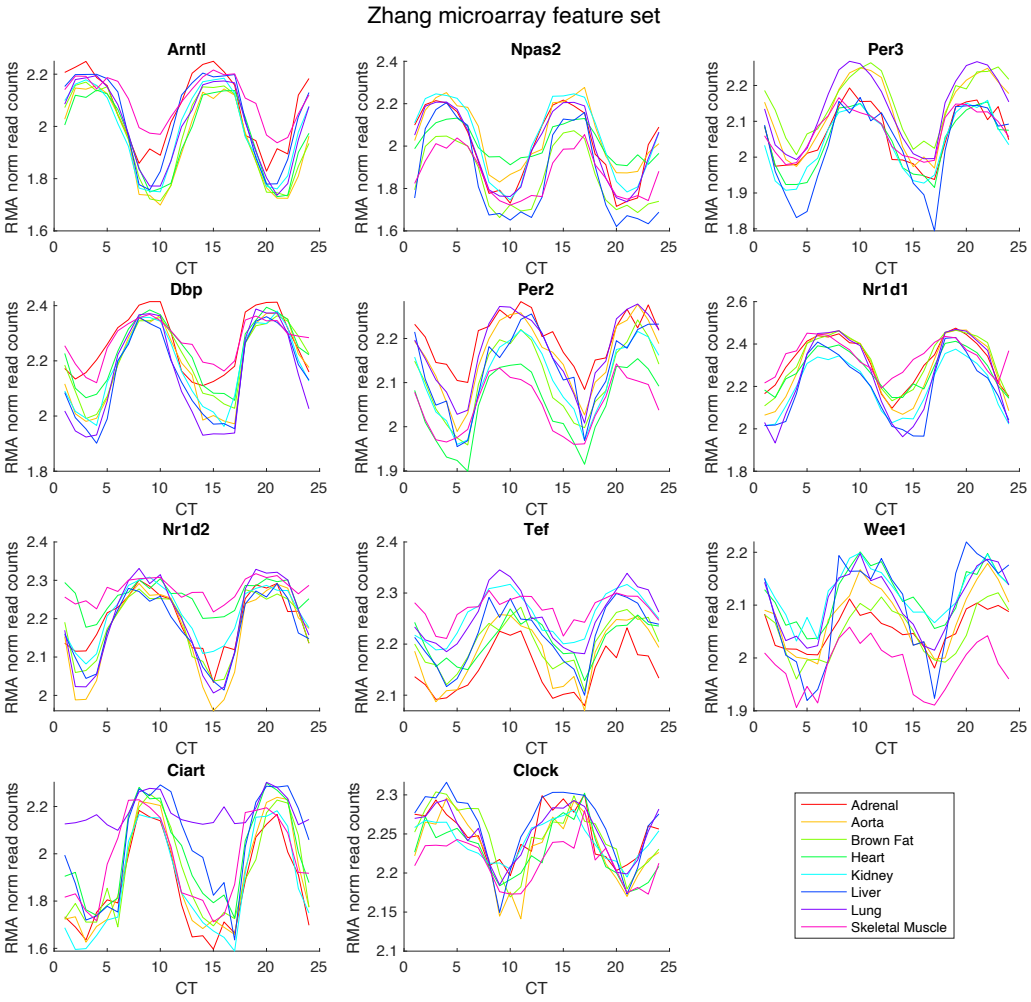


Figure 4.11 Microarray probes used as features to train the TimeTeller model. Different tissues express the oscillatory probes at quite different amplitudes and magnitudes.

Figure 4.11 demonstrates that this bias was likely due to differences between different tissue types in the amplitude and magnitude of the oscillatory probes that were used to train TimeTeller. One of the challenges in building a multi-tissue time predictor is how to deal with tissue-tissue variation in system that is designed to predict time, and ideally would only contain information relating to time-dependent variation between samples. Hughey et al. (2016) addressed tissue-tissue variation by using ComBat to reduce such variation prior to training ZeitZeiger. The disadvantages to using ComBat have been discussed in chapter 4.1.2.5.

Given that the 6 hr resolution Zhang et al. microarray dataset was sufficient to train TimeTeller, the Zhang et al. RNA-seq dataset was screened for rhythmic and synchronous genes as described by Vlachou et al. (2020), and detailed in chapter 4.2.3. Table 4.3 and Figure 4.12 display the results of periodic gene selection for the Zhang et al. RNA-seq data. Eleven genes were selected to train TimeTeller, as 11 probes were selected to train TimeTeller for analysis of the microarray dataset. The highest ranked rhythmic and synchronous genes identified in the Zhang RNA-seq dataset (Figure 4.12) were similar but not identical to the rhythmic and synchronous probes identified in the microarray dataset (Figure 4.11). This is likely to be due in part to the different technologies, and in part to the difference in sampling resolution. The optimal number of features required to train TimeTeller – and indeed other time-telling algorithms – is a question without a precise answer. Vlachou et al. (2020) used 11-16 probes to train TimeTeller. Meanwhile Hughey et al. (2016) trained ZeitZeiger using 10 probes, but also observed remarkable accuracy with as few as 2 probes, and observed no improvement to the MAE of time predictions with more than 13 probes.

Table 4.3 Identification of rhythmic and synchronous genes according to the method developed by Vlachou et al. (2020). Genes were ranked for goodness of cosine fit (indicating rhythmicity) and percentage of variance in the data explained by the first principal component (indicating synchronicity between tissues). Equally weighted ranks were assigned to both scores to select a feature set for TimeTeller.

	Gene Name	ENSEMBL ID	Percentage of variance explained by the 1st PC	Geometric mean of cosinor p-value	Rank of %age of variance explained by the 1st PC	Rank of geometric mean of cosinor p value	Sum rank
1	**Tef**	ENSMUSG00000022389	96.4768	0.00050	49	48	97
2	**Dbp**	ENSMUSG00000059824	95.7964	0.00025	47	50	97
3	**Nr1d2**	ENSMUSG00000021775	97.1282	0.00110	50	46	96
4	**Per3**	ENSMUSG00000028957	94.9269	0.00048	46	49	95
5	**Arntl**	ENSMUSG00000055116	95.9986	0.00120	48	45	93
6	**Hlf**	ENSMUSG00000003949	93.9324	0.00190	43	42	85
7	**Nr1d1**	ENSMUSG00000020889	88.7963	0.00060	35	47	82
8	**Dtx4**	ENSMUSG00000039982	91.3875	0.00200	37	41	78
9	**Npas2**	ENSMUSG00000026077	91.6742	0.00230	38	39	77
10	**Cys1**	ENSMUSG00000062563	88.1739	0.00160	31	44	75
11	**Per2**	ENSMUSG00000055866	94.5974	0.00520	44	23	67
12	Nfil3	ENSMUSG00000056749	86.4695	0.00400	25	32	57
13	Hsp90aa1	ENSMUSG00000021270	83.643	0.00340	18	36	54
14	Bhlhe41	ENSMUSG00000030256	86.186	0.00410	24	29	53
15	Ypel2	ENSMUSG00000018427	85.9972	0.00420	23	28	51
16	Trim65	ENSMUSG00000054517	87.1274	0.00500	27	24	51
17	Clock	ENSMUSG00000029238	84.4957	0.00480	20	25	45
18	Hsp90ab1	ENSMUSG00000023944	80.5717	0.00350	5	35	40
19	Wee1	ENSMUSG00000031016	88.6362	0.00820	33	7	40
20	Tmem57	ENSMUSG00000028826	88.0459	0.00870	30	4	34
21	Naa60	ENSMUSG00000005982	83.3822	0.00650	17	16	33
22	Adamts4	ENSMUSG00000006403	88.214	0.00920	32	1	33
23	Leo1	ENSMUSG00000042487	81.5901	0.00590	9	20	29
24	Rasl11a	ENSMUSG00000029641	80.8371	0.00590	6	19	25
25	Zfp36l1	ENSMUSG00000021127	82.2365	0.00730	11	10	21

TimeTeller was trained in a leave-one-tissue-out fashion on the RNA-seq dataset using the 11 genes selected by rhythmicity and synchronicity analysis, just as for microarray data. The RNA-seq time-course was folded at 24 hrs, such that CT22=CT46. This allowed TimeTeller to predict time on a 0 to 24 hr scale, which are ultimately easier to interpret than 0 to 48 hr time predictions. Figure 4.13 shows TimeTeller's time predictions produced for each of the samples in the RNA-seq dataset. TimeTeller's time predictions of the RNA-seq samples are relatively comparable in accuracy to those of the microarray samples (RNA-seq MAE=1.57 hr vs microarray MAE=1.53 hr). As was observed in the leave-one-out microarray model, Figure 4.13B shows that

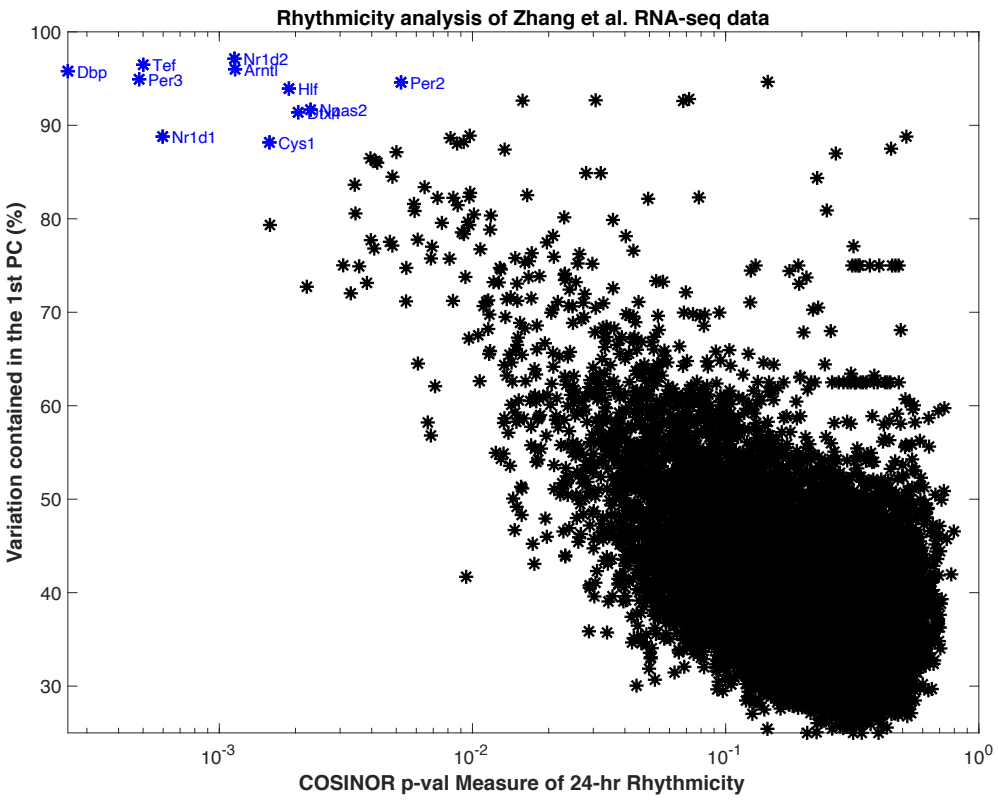


Figure 4.12 Scatter plot of genes expressed in 8 tissues from the Zhang et al. (2014) RNA-seq time-course dataset. Genes are scored for rhythmicity (using the geometric mean of each tissue's cosinor p-value) and percentage of variation explained by the first principal component. The 11 genes marked in blue form the selected feature set (see Table 4.3 for more details).

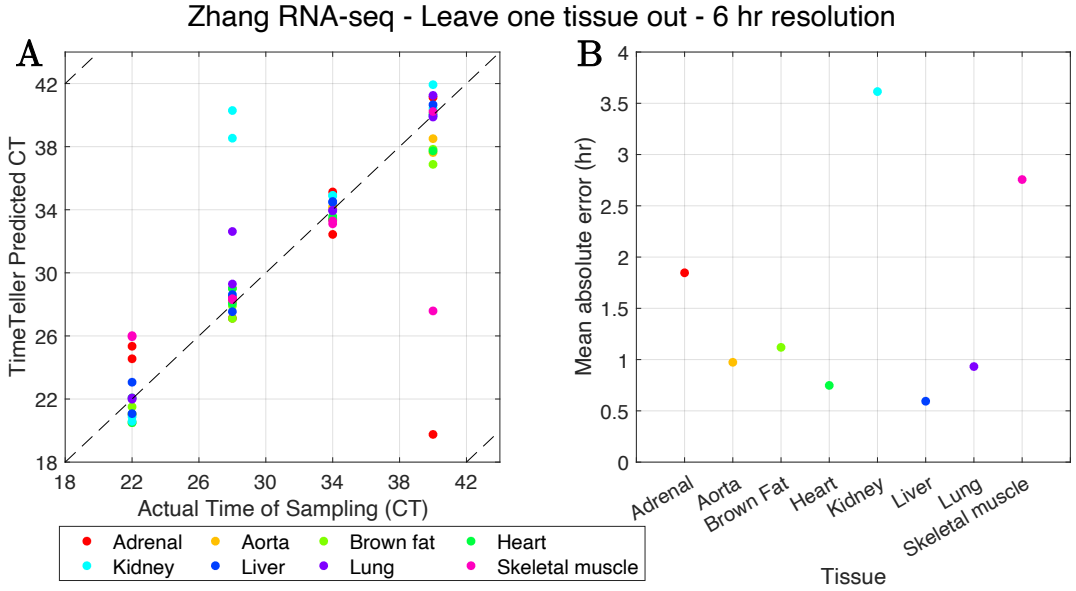


Figure 4.13 Results of TimeTeller trained on Zhang RNA-seq data in a leave-one-tissue-out fashion. A) TimeTeller's predicted time vs actual time of sampling. B) Mean absolute error of all samples from each tissue. Overall MAE = 1.57 hr.

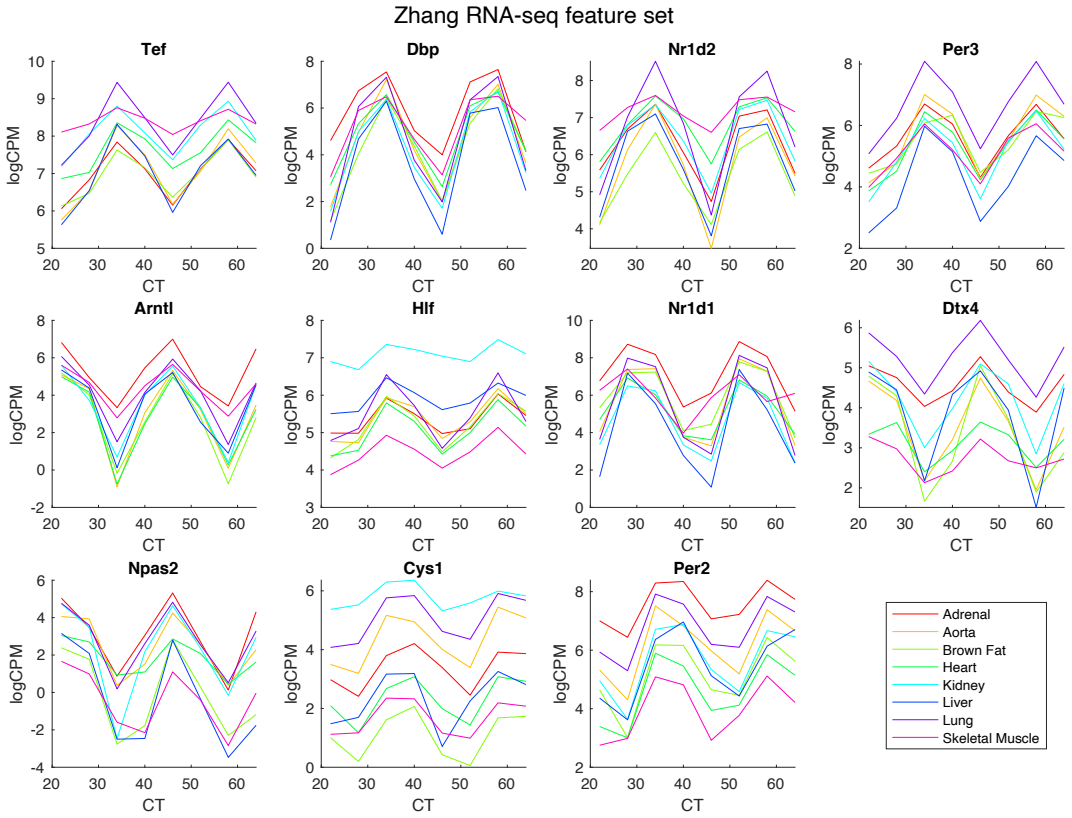


Figure 4.14 Time-dependent expression of the RNA-seq transcripts used as features to train the TimeTeller model in each tissue type used in the model build. Different tissues express the oscillatory genes at quite different amplitudes and magnitudes.

the MAEs of the different tissues in the RNA-seq dataset were highly varied (by ~3 hrs). Figure 4.14 shows that the amplitude/magnitude of expression of the features used to train TimeTeller varied from tissue to tissue, which presumably hindered the accuracy of time prediction in some tissues relative to others. The "inter-gene" data normalisation method implemented by Vlachou et al. (2020) was not sufficient to remove tissue-tissue differences in the data, such that single sample time prediction was not of similar accuracy across tissues.

4.3.2 Improving data normalisation for multi-tissue time prediction

Building a time-telling model from a combination of different tissue types, as described above, is an attractive goal and therefore has been attempted by several research groups (Agostinelli et al., 2016; Hughey et al., 2016; Vlachou et al., 2020). A multi-tissue time predictor could have a broad application to a number of different tissue types, which would make it more useful. Additionally, one of the remaining limitations to the development of biological time prediction algorithms is the availability of tissue-matched training datasets. It therefore makes sense to design a time-telling model that could take advantage of any available datasets, regardless of tissue type. To achieve multi-tissue model builds and/or cross-tissue time prediction, is not an easy goal.

The data presented thus far demonstrates that the normalisation of the oscillatory genes used as TimeTeller's feature set is important in order to remove tissue-tissue variation in oscillatory gene expression. It is clear that the ideal data normalisation method for TimeTeller should be shape-preserving, since the shape of the gene expression in each tissue relates directly to clock function. One such shape-preserving method would be to z-score the training data, i.e., normalise each sample to a standard normal distribution by subtracting the time-course mean and dividing by the time-course standard deviation for each gene in each tissue. This is termed "time-course normalisation" and is described in greater detail in chapter 4.2.3.

Having performed time-course normalisation instead of the aforementioned inter-gene normalisation, TimeTeller was re-run in a leave-one-tissue-out fashion on the Zhang et al. (2014) RNA-seq dataset, in order to test the accuracy of time prediction following time-course normalisation. Figure 4.15

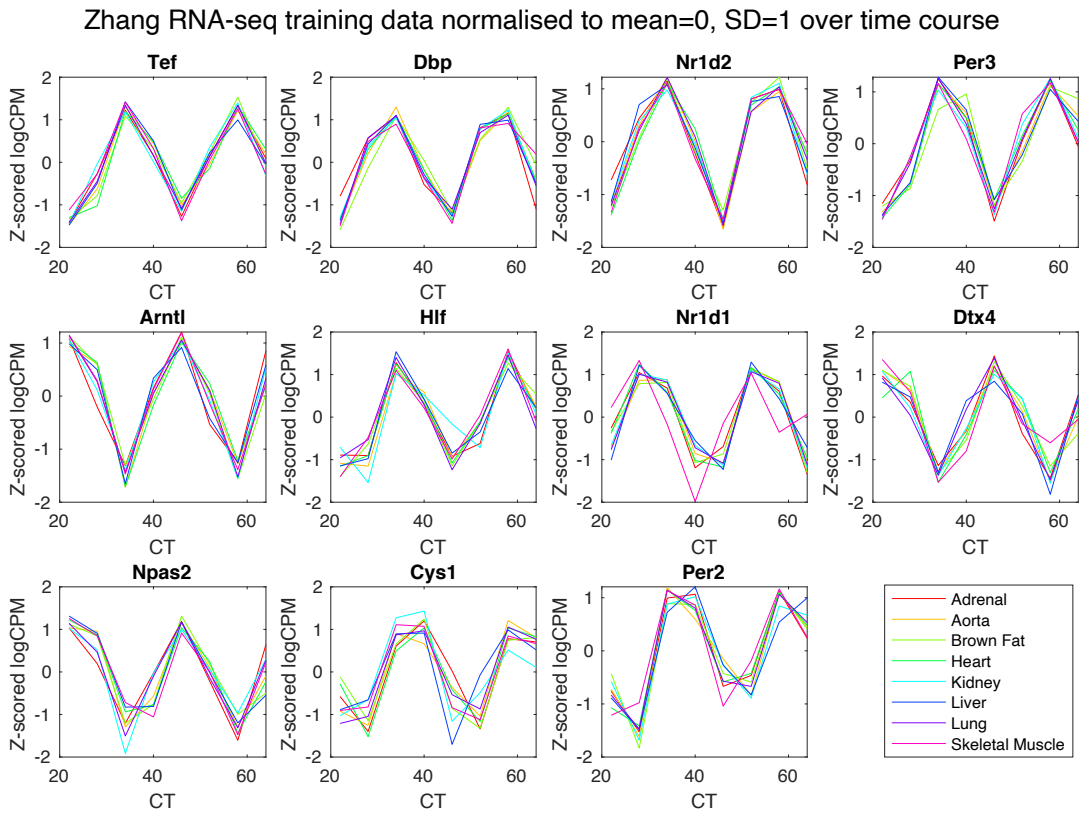


Figure 4.15 Time-course normalised time-dependent expression of the RNA-seq transcripts used as features to train the TimeTeller model in each tissue type used in the model build. Tissue-tissue variation in the dataset was greatly reduced by shape-preserving time-course normalisation.

shows that time-course normalisation removed a great deal of the tissue-tissue variation that was visible in the unnormalised data presented in Figure 4.14. Figure 4.16 demonstrates that time-course normalisation also increased the accuracy of leave-one-tissue-out sample time prediction; the MAE decreased from 1.57 hr using inter-gene normalisation to 0.61 hr when time-course normalisation was applied. Importantly the MAE of different tissues was also more uniform after time-course normalisation (Figure 4.16B). Figure 4.16B shows that the tissue-wise MAE of the adrenal gland and skeletal muscle were slightly higher than that of the other tissues, therefore the raw error of each time prediction was investigated for each tissue. Figure 4.17 demonstrates that relative to the other tissues the skeletal muscle and adrenal gland samples were

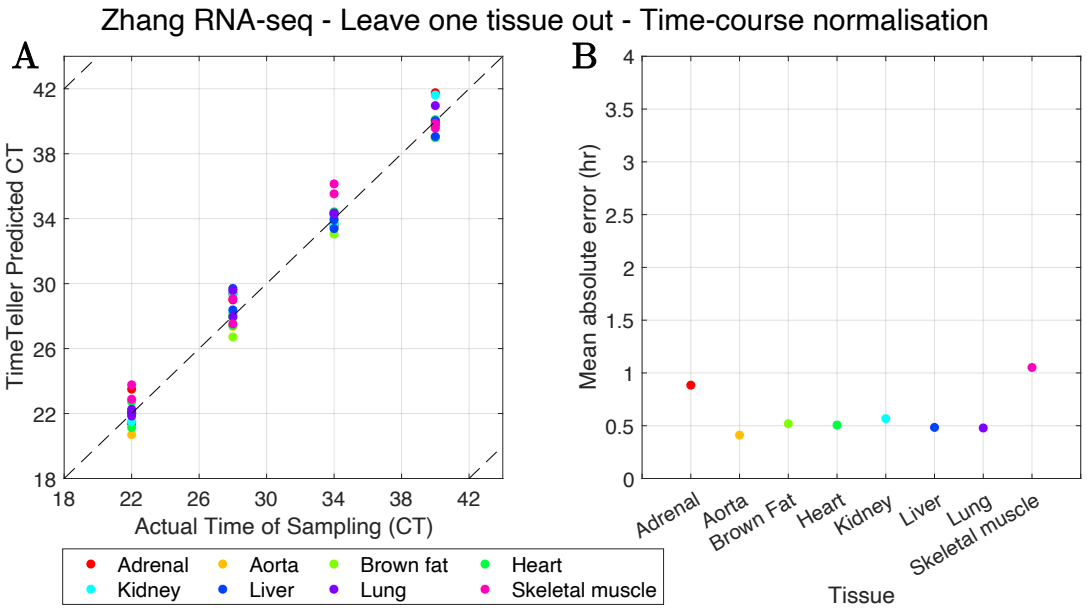


Figure 4.16 Results of TimeTeller trained on Zhang RNA-seq data in a leave-one-tissue-out fashion having applied time-course normalisation. A) TimeTeller's predicted time vs actual time of sampling. B) Mean absolute error of all samples from each tissue. Overall MAE = 0.613 hr.

slightly phase advanced, i.e., TimeTeller assigned a predicted time that was advanced of the known time of sampling. Since the tissue-wise raw error differed from 0 by no more than 1 hr for any tissue, the small phase advance/delays for each tissue were not considered to be a significant cause for concern. It should be noted that TimeTeller is at a disadvantage relative to ZeitZeiger here, as whilst TimeTeller requires training data sampled at regular time intervals around the clock, the ZeitZeiger method accepts training data sampled at any time. Therefore, for ZeitZeiger the relative tissue phase advances/delays could be used to adjust the time of sampling prior to re-training the model. However, thus far it appears that such a tissue-wise phase adjustment has not been implemented in the literature in combination with ZeitZeiger.

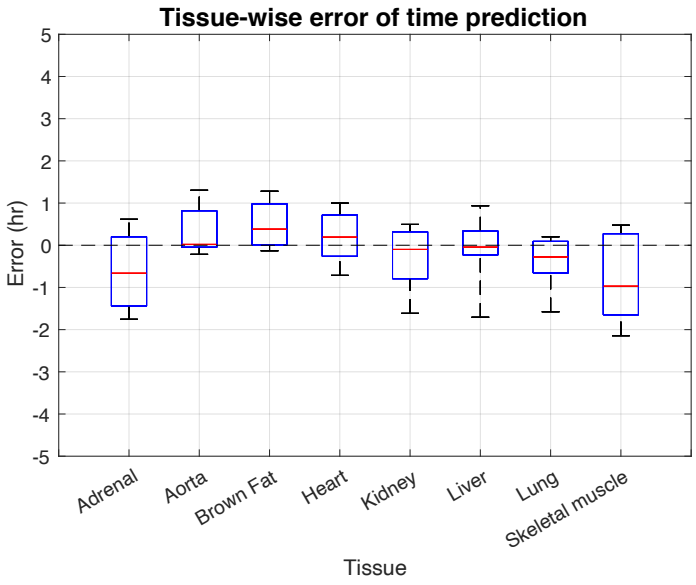


Figure 4.17 The tissue-wise raw error, i.e., TimeTeller's predicted time subtracted from the time of sampling. A negative raw error represents a relatively phase-advanced tissue, whilst a positive raw error represents a tissue that was phase-delayed relative to the other tissues. Each box-and-whisker plot represents the error of 8 samples of the given tissue type. Black whiskers represent the range, blue boxes the interquartile range, and red lines represent the mean.

Having trained the TimeTeller model on the multi-tissue Zhang et al. RNA-seq data, it was important to confirm that the reduction of the feature set to 3 dimensions was appropriate. The idea is to reduce the feature set such that most of the variation in the dataset is described by as few dimensions as possible. Figure 4.18 demonstrates that three dimensions is sufficient to explain $>$90% of the variation in the feature $\times$ observation data matrices that corresponded to each timepoint in the training time series. Vlachou et al. (2020) also found that 3 dimensions was suitable for every application of TimeTeller to microarray data.

The disadvantage of time-course normalisation in this leave-one-tissue-out model is perhaps already apparent. Time-course normalisation is by definition time-course dependent. Therefore, the goal of single time point time prediction

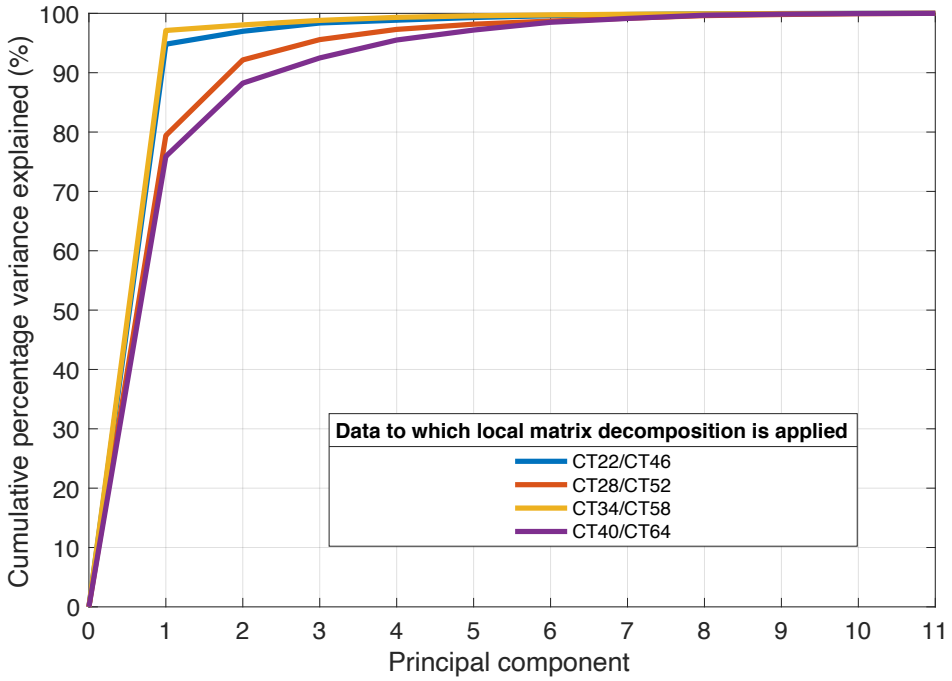


Figure 4.18 Cumulative percentage of variance explained by each of the principal components found by SVD. Data for day 1 and day 2 (e.g., CT22 and CT46) is combined to build a 24 hr model. 3 dimensions are sufficient to explain more than 90% of the variation in the dataset for each local projection.

is not achieved. However, the notion behind time-course normalisation is important. That is that shape-preserving normalisation can be achieved using the mean and standard deviation of gene expression in each tissue over time.

4.3.3 Adapting time-course normalisation for single sample time prediction: testing quasi-time-course normalisation on an independent test dataset

Since time-course normalisation is not an appropriate approach for the normalisation of a single independent test data sample for which no time wise means and standard deviations of gene expression are available, "quasi-time-course normalisation" will be tested as an alternative approach for single-timepoint time-telling. Quasi-time-course normalisation is described in greater

detail in chapter 4.2.4; essentially gene expression in the independent test data sample is normalised to the time series mean and standard deviation of gene expression in the training tissue that corresponds to the test sample. To test the efficacy of "quasi-time-course normalisation" an independent test dataset is required. An appropriate test dataset should be comprised of murine samples in at least one of the following tissues: adrenal, aorta, brown fat, heart, kidney, liver, lung or skeletal muscle. Secondly, the test data samples should be time-stamped, so that TimeTeller's biological time prediction can be compared with the known time of sampling. Lastly, a time series would ideally contain samples collected over a 24 hr period in order to understand how TimeTeller functions over 24 hr time. Possible test datasets were identified by literature search and are detailed in Table 4.1, chapter 4.2.1.

The Kinouchi et al. (2018) dataset contains samples from mouse liver and skeletal muscle taken around the clock in light/dark entrainment conditions (Kinouchi et al., 2018). Sample collection commenced at ZT0 which corresponded to the 'lights on' time. The Zhang et al. (2014) training dataset contains samples taken under dark/dark conditions from CT22, i.e., 22 hrs after the lights would have been switched on, were the mice not in dark/dark conditions. For the purposes of this study, CT28 and CT52 in the Zhang et al. dataset is considered equivalent to ZT4 in the Kinouchi et al. dataset. In truth, there is likely to be a small difference in the oscillatory gene expression between CT28, CT52 and ZT4 as the free-running murine circadian clock is slightly shorter than 24 hrs (23.8 hr in C57BL/6 (Schwartz & Zimmerman, 1990)). However, a recent comparison of the period lengths of core clock genes in light/dark entrained vs short time dark/dark conditioned mice showed no significant difference between the two datasets (Huan Li et al., 2020).

The Kinouchi et al. data was downloaded from NCBI's SRA database as raw SRA files (GEO identifier: GSE107787) and normalised to logCPM as described in chapter 4.2.2. For both skeletal muscle and liver, there were two conditions. Firstly, mice that had been fed *ad libitum*, and secondly mice that had been starved for exactly 24 hrs prior to point of sampling (Figure 4.19). Kinouchi et al. examined the differences in the oscillatory gene expression between the *ad libitum* fed mice and the starved mice and concluded that oscillatory transcription was significantly altered in the liver and skeletal muscle of starved mice. It should be noted that the study design was unusual, as the oscillatory gene expression of the starved mice was analysed by Kinouchi et al. as though the starved samples belonged to a continuous time series, even though each mouse had been starved for exactly 24 hrs regardless of where it appeared in the time-course (Figure 4.19). Here, TimeTeller has an opportunity to provide information about the molecular clock in single independent samples.

Firstly, it was important to examine the raw Zhang et al. and Kinouchi et al. data in order to assess whether quasi-time-course normalisation seemed an appropriate approach. Figure 4.20A displays the raw Zhang et al. data previously displayed in Figure 4.14, however this time the data was plotted folded at 24 hrs to better represent how the data was folded to train

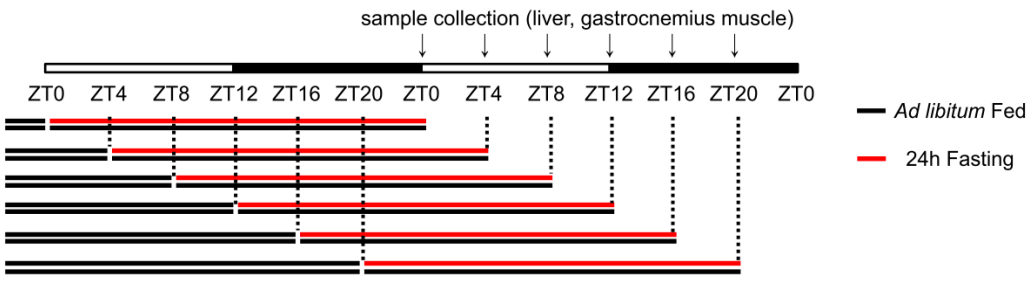


Figure 4.19 Sampling protocol for Kinouchi et al. dataset. Reproduced from Kinouchi et al. (2018) Figure 1.

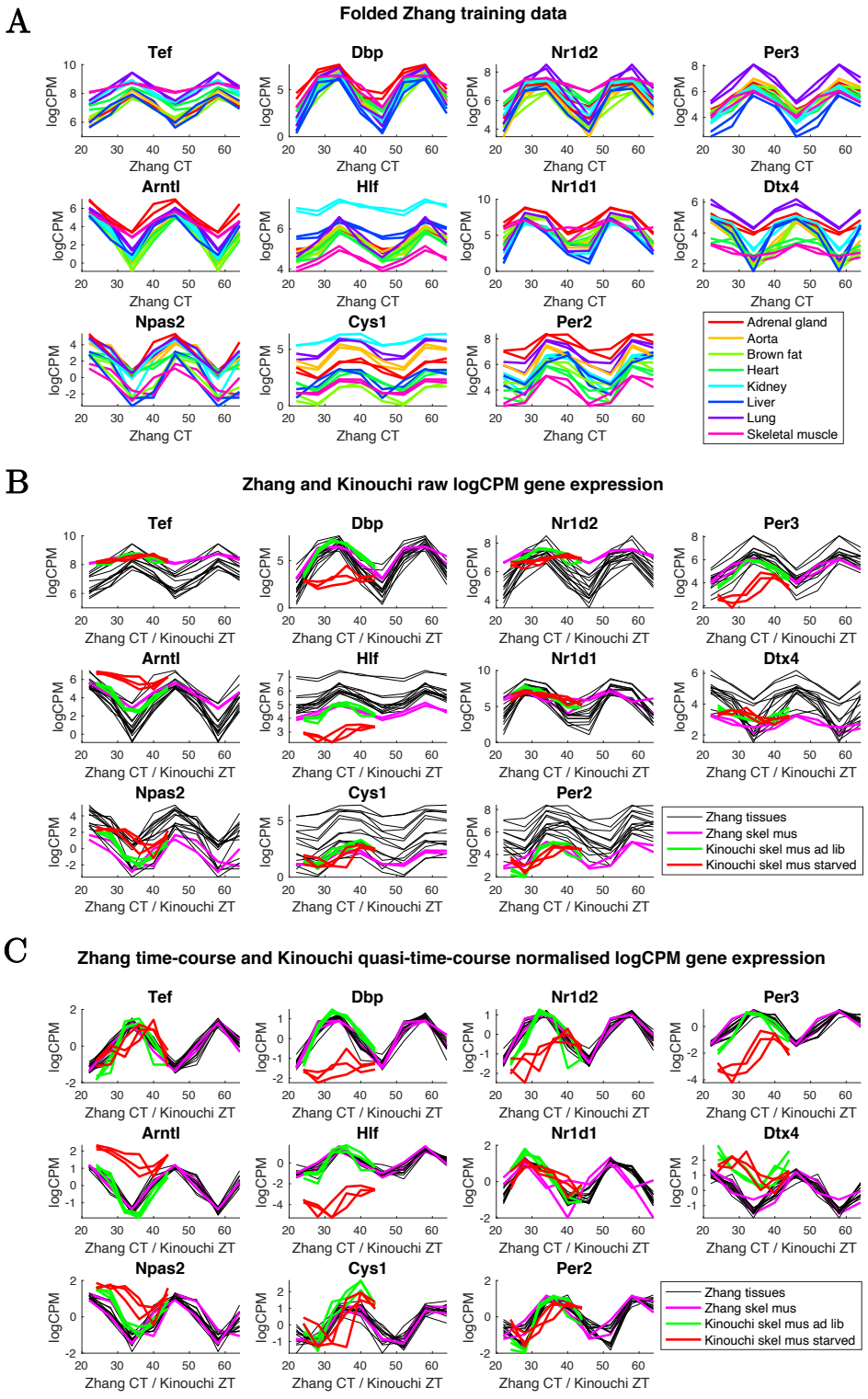


Figure 4.20 logCPM normalised Zhang and Kinouchi data. A) Raw folded, double-plotted Zhang training data coloured by tissue. B) Raw Zhang training data plotted in black, skeletal muscle in pink. Raw Kinouchi test data plotted in green (*ad libitum* fed) and red (starved). C) Time-course normalised Zhang training data plotted in black, skeletal muscle in pink. Quasi-time-course normalised Kinouchi test data plotted in green (*ad libitum* fed) and red (starved). N.B. each sample is from a different mouse to the sample at the following timepoint. Therefore lines are drawn only for ease of data visualisation.

TimeTeller. The skeletal muscle data was considered first. Figure 4.20B demonstrates that expression of the training genes in the *ad libitum* fed skeletal muscle samples from Kinouchi et al. aligned remarkably well to expression of the training genes in the skeletal muscle Zhang samples. This was a promising indicator that quasi-time-course normalisation would be an appropriate approach for independent normalisation of single samples. Rhythmic gene expression in the starved skeletal muscle samples plotted in Figure 4.20B appeared different to that in the *ad libitum* fed samples, as described by Kinouchi et al. (2018). Figure 4.20C shows the result of normalising the Kinouchi et al. skeletal muscle gene expression to the time-course mean and standard deviation of skeletal muscle gene expression in the Zhang et al. data, i.e., the result of implementing the aforementioned quasi-time-course normalisation on the Kinouchi et al. skeletal muscle dataset. The quasi-time-course Kinouchi et al. *ad libitum* skeletal muscle data aligned well to the time-course normalised Zhang et al. data (Figure 4.20C). The most differently expressed gene was *Dtx4* which appeared to be more highly expressed in the Kinouchi skeletal muscle samples than the Zhang skeletal muscle samples. *Dtx4* is involved in the *Notch* signalling pathway which can help to determine cell fate amongst other processes (Chastagner et al., 2017). *Dtx4* is consistently expressed rhythmically in the eight Zhang et al. tissues, though the relationship between *Dtx4* and the core circadian clock genes is unknown (R. Zhang et al., 2014).

TimeTeller was able to predict the sampling time of the Kinouchi et al. *ad libitum* fed skeletal muscle samples with remarkable accuracy (Figure 4.21A, C); overall, the MAE of *ad libitum* fed skeletal muscle samples was under 30 minutes. With a logthresh of -12, the Θ values ranged from 0.02 to 0.08. This gave an indication of the magnitude of Θ associated with samples known to have 'normal' time predictions. On average the time predictions for the starved

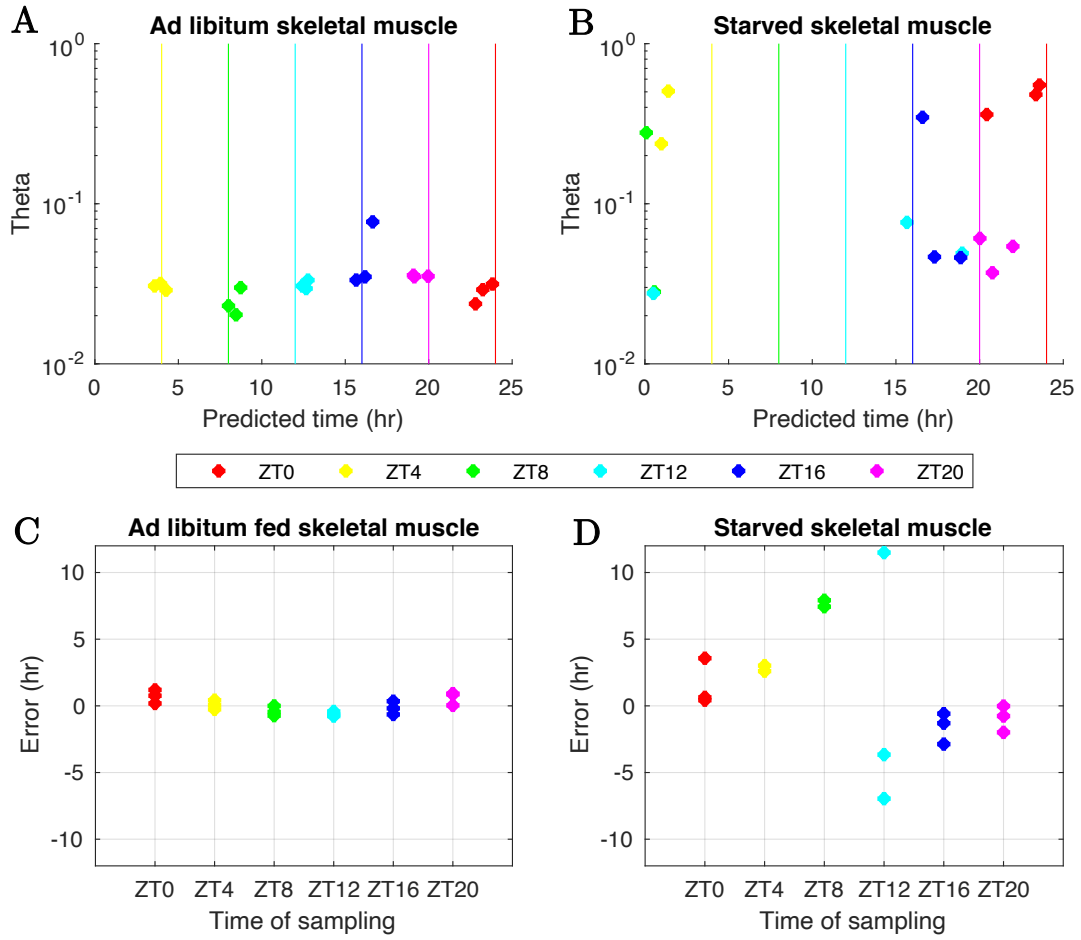


Figure 4.21 A) TimeTeller's predicted times for Kinouchi et al.'s ***ad libitum* fed skeletal muscle** samples, plotted against corresponding Θ values (logthresh=e^{-12}). Vertical coloured lines represent the actual time of sampling of all samples. B) TimeTeller's predicted times for Kinouchi et al.'s **starved skeletal muscle** samples, plotted against corresponding Θ values. C) Error (difference between time of sampling and TimeTeller's predicted time) for ***ad libitum* fed skeletal muscle** samples. MAE= 0.496 hr. D) Error for **starved skeletal muscle** samples. MAE = 3.45 hr excluding one ZT4 and one ZT8 sample for which the maximum likelihood fell below logthresh.

skeletal muscle samples displayed a much higher error (MAE=3.79 hr, Figure 4.21D), which was to be expected, given that Kinouchi et al. (2018) reported a very different rhythmic transcriptional programme in starved mice relative to *ad libitum* fed mice. Additionally, the Θ values were more varied, and generally higher, than for the *ad libitum* fed samples.

The advantage of applying to TimeTeller to the Kinouchi et al. starved samples is that they can be considered independently from one another – as each of the mice was starved for exactly 24 hrs the data should not truly be considered as a continuous time-course. In fact, TimeTeller demonstrates that the ZT20 starved skeletal muscle samples have time predictions and Θ values that are most comparable to the *ad libitum* samples. This suggests that 24 hrs of food restriction prior to ZT20 did not have a very large effect on the expression of the genes used to build TimeTeller, i.e., the genes that are most consistently circadian in mouse tissues. The time predictions for ZT0 and ZT16 starved skeletal muscle samples were more abnormal than the ZT20 samples. Two of the ZT0 samples had excellent time predictions, though on average Θ was higher for all of the ZT0 samples. Meanwhile the ZT16 starved samples had Θ values within the 'normal' *ad libitum* range, but the time predictions were phase delayed relative to time of sampling. The 'daytime' starved skeletal muscle samples (ZT4, 8 and 12) all displayed much poorer time predictions relative to time of sampling, and/or much larger Θ values. In fact, the maximum likelihoods were so low for two samples (one ZT4 and one ZT8) that they fell below logthresh (e^{-12}) and no time prediction was able to be estimated.

The Θ metric contains interesting information about the degree of clock dysfunction, however it is somewhat opaque to interpretation since the magnitude of Θ depends upon several variables (discussed in detail in chapter 4.2.5.5). Therefore, it is helpful to view the centred scaled likelihood curves and the raw data projections for a better understanding of the key drivers behind the apparent increase clock dysfunction in the starved skeletal muscle samples. Figure 4.22 shows that the likelihood curves were generally more disordered for the starved samples relative to the *ad libitum* samples. The higher baselines of the starved sample likelihood curves relative to the *ad*

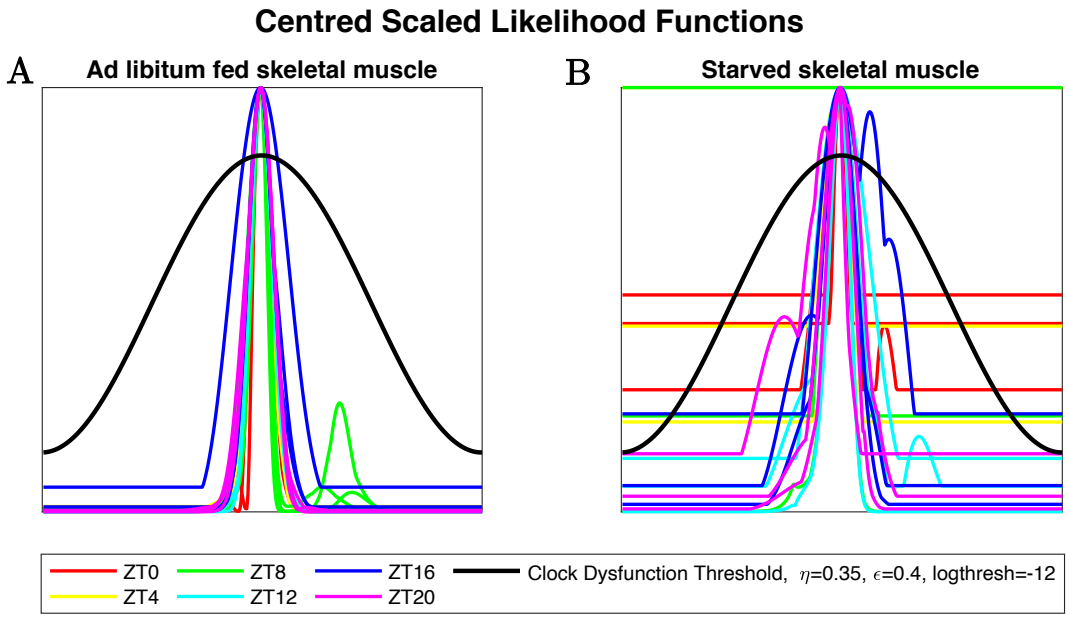


Figure 4.22 Centred scaled likelihood functions of A) Kinouchi et al.'s *ad libitum* fed skeletal muscle samples and B) Kinouchi et al.'s starved skeletal muscle samples. All functions are coloured by time of sampling. Likelihood functions were thresholded such that the minimum likelihood was set to e^{-12} (i.e., a logthresh value > 0). Then all likelihood curves were scaled such that the maximum likelihood was equal to 1. Functions were centred around the primary peak such that the shapes of different functions were comparable to one another. Θ was calculated as the proportion of the centred scaled likelihood function that crosses above the black clock dysfunction threshold.

libitum samples, reveals that the maximum likelihoods were frequently lower for the starved samples. Generally lower maximum likelihoods are reflective of test data that is projected to a low dimensional location that is relatively distant to the location of the projected training data. If the maximum likelihood is so low that the baseline of the likelihood curve is raised above the clock dysfunction threshold (shown in black, Figure 4.22) then the value of Θ will be higher to reflect the fact that TimeTeller's time predictions are of relatively low maximum likelihood, as observed for some of the starved skeletal muscle samples. Figure 4.22 also reveals that a greater proportion of the starved samples' likelihood curves had secondary peaks than the *ad libitum* samples' likelihood curves. Secondary peaks may occur when different local

projections of the test data give different time predictions, which implies that the training genes were expressed in a combination that may not be observed in 'normal clock' samples. The further a secondary peak lies from the primary peak, the more it is likely to reflect an abnormal combination of expressed genes, and the more likely it becomes that the secondary peak will cross the clock dysfunction threshold (black line) and increase the value of Θ. Figure 4.22 shows that some of the secondary peaks in the starved skeletal muscle samples crossed the clock dysfunction threshold and contributed to an increase in the value of Θ.

Sometimes visualising the data in the low dimensional projection space can help to understand high Θ values. Figure 4.23A shows that the *ad libitum* fed mice lie neatly on the trained TimeTeller model (which is visualised using mesh ellipsoids to represent interpolated time – see chapter 4.2.5 for details). The 3-dimensional graphs are rotated such that the relative locations of the training and test data samples can be appreciated as best as possible. From Figure 4.23B it is clear that the starved skeletal muscle samples behave differently to the Zhang et al. training data. Interestingly the starved skeletal muscle samples appear to form an elliptical oscillator that is 'off-centre' from the model such that it oscillates in the vicinity of the CT22/46 training data. This observation should be interpreted with caution, given that the starved samples do not form a conventional time course.

It is interesting to consider that food intake in mice is a rhythmic process, and that the majority of caloric intake occurs during the dark phase (ZT12-ZT24/0) (Kinouchi et al., 2018). Therefore, the starved samples taken at ZT12 had arguably endured 24 hrs starvation which was immediately preceded by 12 hrs of relatively low caloric intake. Conversely, the starved samples taken at ZT0 had endured 24 hrs starvation which was immediately preceded by 12

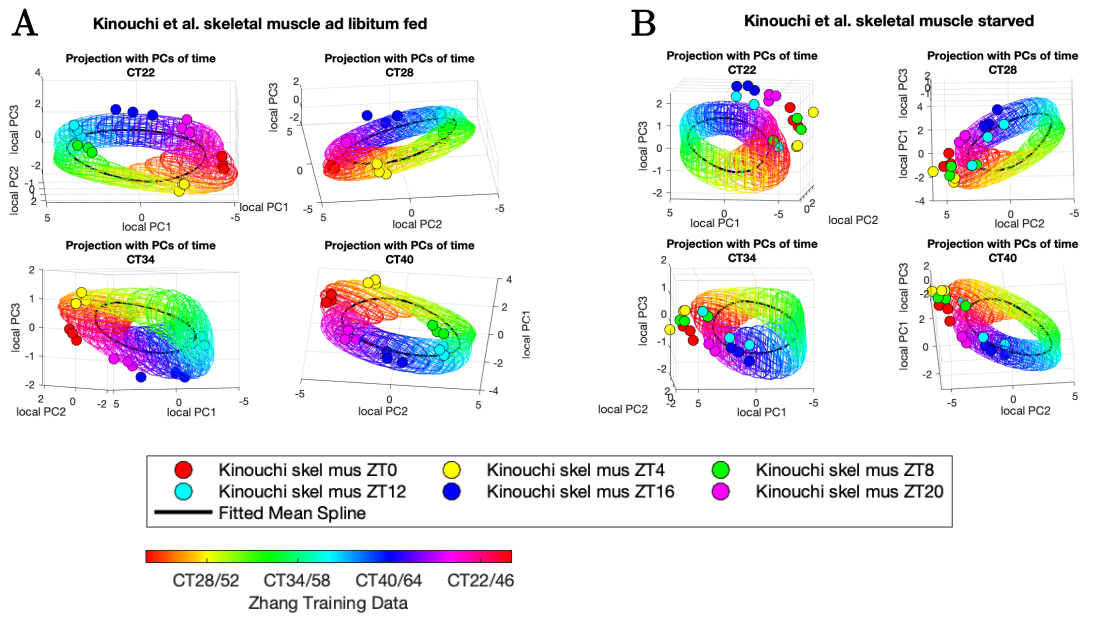


Figure 4.23 Projected Zhang et al. training and Kinouchi et al. skeletal muscle test data onto the first 3 principal components identified from each of the original training data timepoints (t=4). A) Projected *ad libitum* fed samples. B) Projected starved samples. The mesh ellipsoids which represent the training data are colour-matched to the test data, such that ZT0 = CT24 = red. The mesh ellipsoids represent 3 standard deviations from the mean of the interpolated Gaussian ellipsoids. The means of the interpolated Gaussian ellipsoids lie along the spline fitted to the training data distributions (black line).

hrs of relatively high caloric intake. In other words, over the 36 hr prior to sampling, mice sampled at ZT12 were 'more starved' than the ZT0 mice. It is therefore interesting to compare the ZT0 (red) and ZT12 (cyan) samples in Figure 4.23A & B. The predicted times of the samples corresponding to ZT0 behave much more similarly in the *ad libitum* and starved conditions, than the predicted times of the ZT12 samples in the *ad libitum* vs starved conditions. This might indicate that the degree of rhythmic gene disruption was related to the of the relative length of caloric deprivation in the mice. This observation is only possible by considering each sample as an independent entity in TimeTeller, rather than as a part of a time-course, and therefore was not a point of discussion in the Kinouchi et al. paper.

The Kinouchi et al. (2018) paper also contained RNA-seq data corresponding to liver samples from *ad libitum* fed and starved mice. Figure 4.24 demonstrates expression of the training genes in the Zhang et al. training data, and also the liver samples from Kinouchi et al. (2018). Quasi-time-course normalisation of the Kinouchi et al. liver gene expression to the time-course mean and standard deviation of the Zhang et al. liver gene expression also

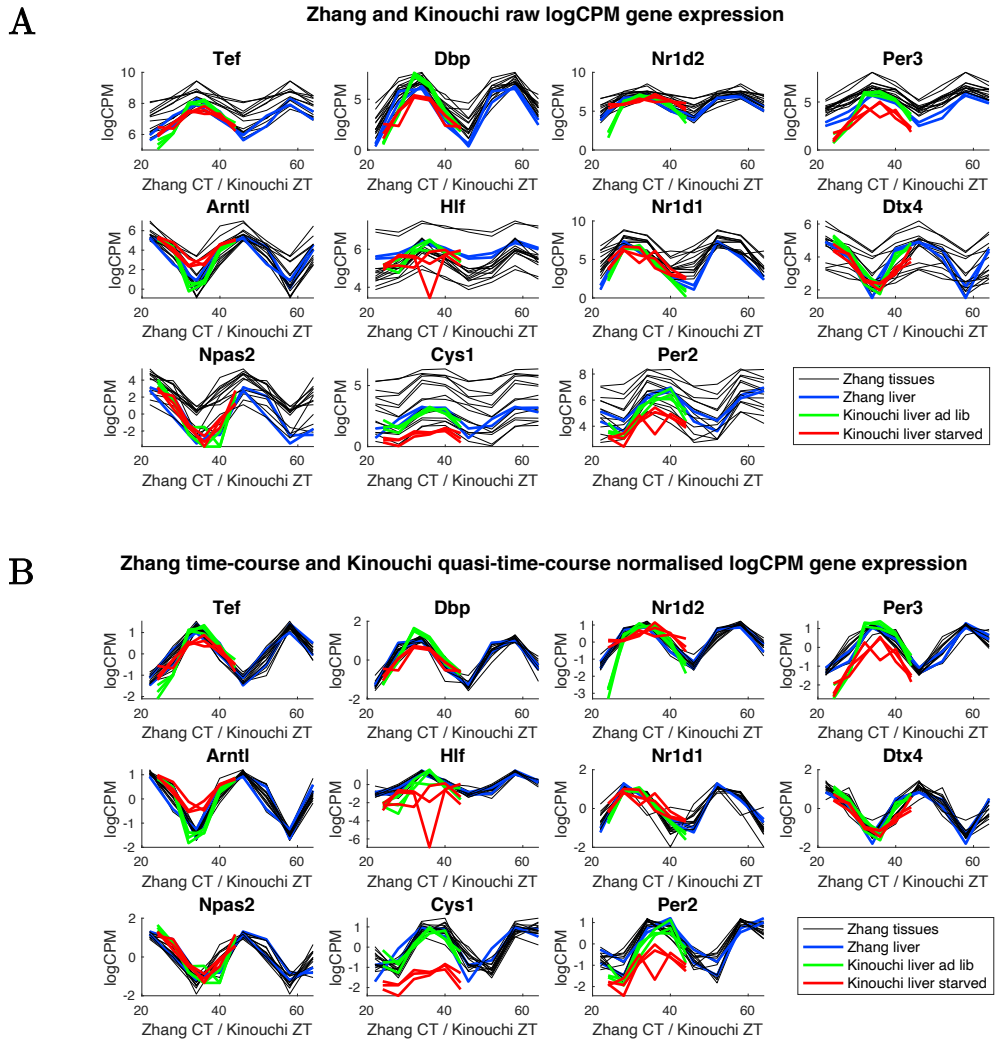


Figure 4.24 A) Raw Zhang training data plotted in black, Zhang liver in blue. Raw Kinouchi liver test data plotted in green (*ad libitum* fed) and red (starved). C) Time-course normalised Zhang training data plotted in black, Zhang liver in blue. Quasi-time-course normalised Kinouchi liver test data plotted in green (*ad libitum* fed) and red (starved). N.B. each sample is from a different mouse to the sample at the following timepoint. Therefore connecting lines are plotted only for ease of data visualisation.

appeared to be an appropriate approach for independent normalisation of single Kinouchi et al. liver samples (Figure 4.24B). Expression of the rhythmic training genes in the starved liver samples appeared different to that in the *ad libitum* fed liver samples, as described by Kinouchi et al. (2018). It is interesting to note that expression of the rhythmic training genes in the starved liver samples was also slightly different to training gene expression in the starved skeletal muscle samples. For instance, in the starved skeletal muscle tissue *Cys1* appeared slightly phase delayed relative to *ad libitum* fed samples. However, in the starved liver tissue *Cys1* appeared to be under-expressed relative to the *ad libitum* fed samples. This suggests that starvation conditions caused a tissue-specific response in circadian gene expression.

TimeTeller was able to predict the sampling time of the Kinouchi et al. *ad libitum* fed liver samples with very good accuracy (Figure 4.25A, C); overall, the MAE of *ad libitum* fed liver samples was 0.87 hr. With a logthresh of -12, the Θ values ranged from 0.01 to 0.05. This gave an indication of the magnitude of Θ associated with samples known to have 'normal' clock function. Unlike the starved skeletal muscle samples, the Θ values of the starved liver samples were consistent with the Θ values of the *ad libitum* fed liver samples (Figure 4.26), which indicated high confidence in TimeTeller's time predictions for the starved liver samples. However, the MAE of TimeTeller's time predictions for the starved liver samples was much higher (3.87 hr, Figure 4.25D), than would be expected of functional clocks – which given that Kinouchi et al. (2018) reported a different rhythmic transcriptional programme in the livers of starved mice relative to *ad libitum* fed mice, was not wholly unexpected. For instance, ZT4 starved liver samples had what appeared to be almost perfectly functional clocks with low Θs and a time prediction error approaching 0, indicating that rhythmic training gene expression in the liver was apparently unperturbed by 24 hrs of starvation. On the other hand, the ZT16 samples

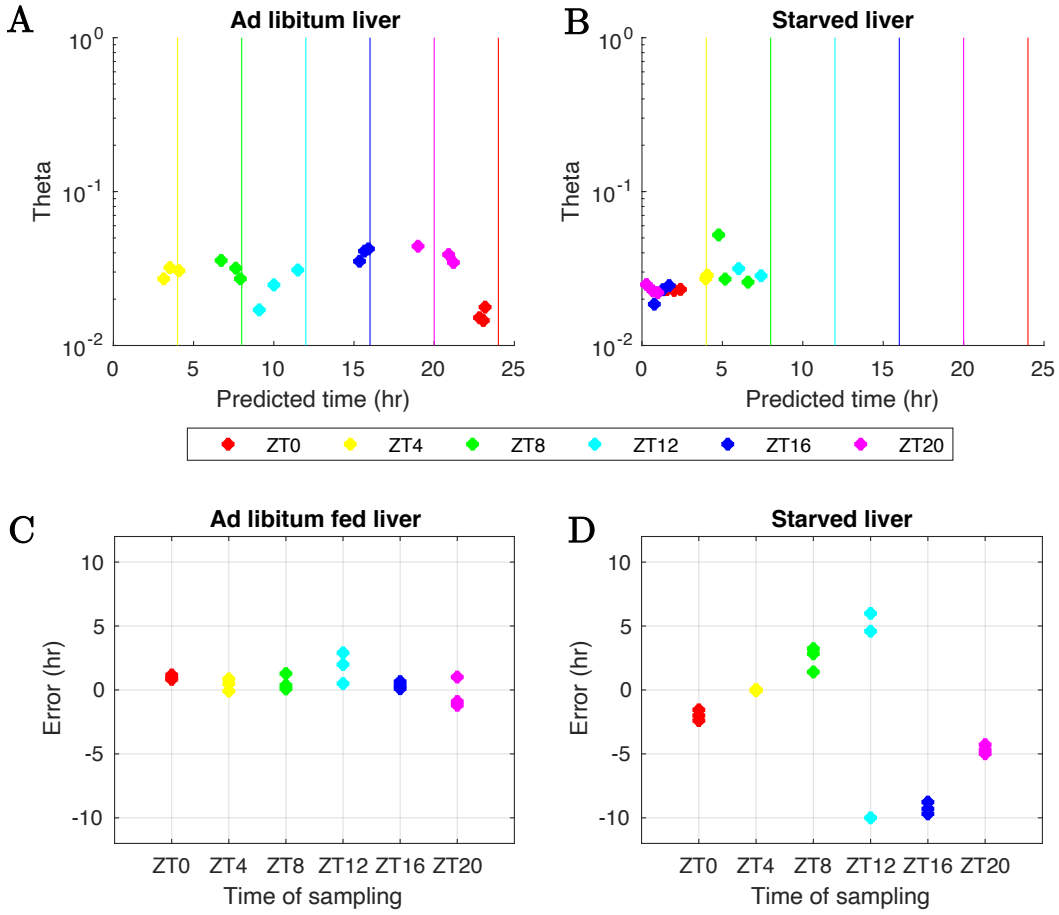


Figure 4.25 A) TimeTeller's predicted times for Kinouchi et al.'s ***ad libitum*** **fed liver** samples, plotted against corresponding Θ values (logthresh=e^{12}). Vertical coloured lines represent the actual time of sampling of all samples. B) TimeTeller's predicted times for Kinouchi et al.'s **starved liver** samples, plotted against corresponding Θ values. C) Error (difference between time of sampling and TimeTeller's predicted time) for ***ad libitum*** **fed liver** samples. MAE= 0.87 hr. D) Error for **starved liver** samples. MAE = 3.87 hr excluding one ZT12 sample for which the maximum likelihood fell below logthresh.

appeared to have extremely inaccurate time predictions, with a very high (magnitude) error of -9.25 hr. It is apparent from examining the data in the low dimensional projection space that the *ad libitum* fed liver samples aligned well with the trained TimeTeller model (Figure 4.27). However, the starved skeletal liver samples appeared to form an elliptical oscillator 'off-centre' from the model, such that it oscillated in the vicinity of the CT28/52 training data

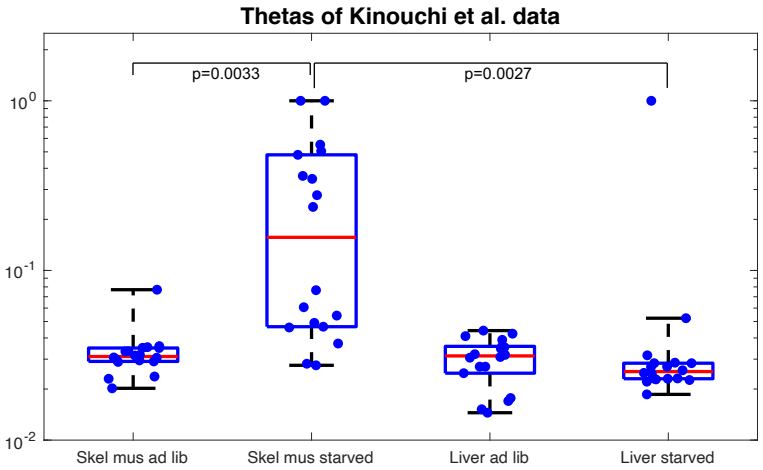


Figure 4.26 Θ values for all samples in Kinouchi et al. (2018) data set. The Θ values indicate significant clock dysfunction in the starved skeletal muscle samples. P values correspond to Welch's two sample t-test and indicate >99% confidence that the marked distributions do not belong to populations of the same mean.

– about 4-6 hrs phase-delayed of the time that the skeletal muscle samples clustered to most closely. This observation should be interpreted with caution, given that the starved samples do not originate from a conventional time course.

The Kinouchi et al. skeletal muscle and liver data discussed above demonstrates for the first time that TimeTeller is capable of predicting both time, and confidence in the predicted time (Θ), in a single time point fashion from RNA-seq data, provided the following holds true:

- The test and training data sets are processed and normalised using the same RNA-seq pipeline.
- The test data matches one of the tissues in the training data set to allow for quasi-time-course normalisation.

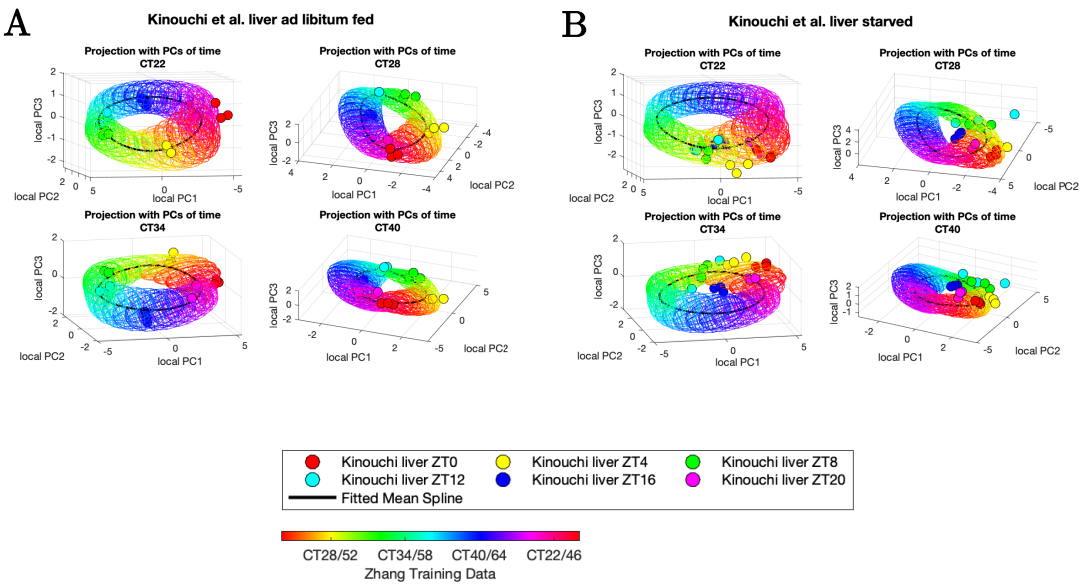


Figure 4.27 Projected Zhang et al. training and Kinouchi et al. **liver** test data onto the first 3 principal components identified from each of the original training data timepoints (t=4). A) Projected *ad libitum* fed samples. B) Projected starved samples. The mesh ellipsoids which represent the training data are colour-matched to the test data, such that ZT0 = CT24 = red. The mesh ellipsoids represent 3 standard deviations from the mean of the interpolated Gaussian ellipsoids. The means of the interpolated Gaussian ellipsoids lie along the spline fitted to the training data distributions (black line).

The Kinouchi et al. data highlights one of the significant hurdles to the implementation of time-telling algorithms to single samples. That is that a single sample can only provide information about the clock gene expression at a single instance in time. For instance, the starved liver samples at ZT4 had predicted times and Θ values consistent with a healthy oscillating clock. However, given that many of the other 24 hr starved liver samples displayed a disrupted oscillator, it is likely that the molecular clock in the ZT4 samples was also not oscillating as normal over time. For this reason, the use of a single sample to draw conclusions about the function of the molecular oscillator in a single individual should be approached with caution.

This conclusion is relevant for analysis of human data. For instance, Wittenbrink et al. (2018) applied ZeitZeiger to NanoString data to build a

diagnostic tool that predicts the biological time (relative to DLMO) of a single individual from a single human blood sample. However, the application of TimeTeller to the Kinouchi et al. data – particularly the starved liver samples – demonstrates that time prediction from a single sample in a single individual can be misleading. Therefore, it is likely that the strength of time-prediction algorithms lies in protocols that rely on repeated sampling, relative sampling and/or population sampling. Blood samples for instance are possible to obtain repeatedly, e.g., 2 samples, sampled 4 hrs apart. Meanwhile, for tissues that can only be biopsied once, such as a tumour, it might be possible to gain greater insight by comparing the clock function in 'healthy' tumour margin relative to malignant tumour stroma. Finally, TimeTeller's time predictions and Θ values can be compared between conditions sampled over a population of individuals, in order to make the method more resilient to the effect of misleading samples.

4.3.4 Testing quasi-time-course normalisation on further independent test datasets

A number of datasets were identified to determine whether the Zhang tissue trained TimeTeller could also assess circadian clock function using quasi-time-course normalisation in datasets with genetically altered clocks. The datasets selected included *Arntl* WT vs KO samples from liver and kidney, and *Cry1/Cry2* double KO samples from liver (Weger et al., 2021; Yeung et al., 2018) (see Table 4.1, chapter 4.2.1 for further details of datasets). When testing TimeTeller on data that had a knockout of a rhythmic training gene, that training gene was excluded from the model build, and replaced with the next most rhythmic and synchronous gene, in order to avoid biasing the time prediction. In practice this meant that the feature set remained the same for

all datasets, except those containing *Arntl* KO samples, in which case *Arntl* was replaced with *Nfil3*. *Nfil3* is a known component of the circadian clock (see chapter 1.2 for further details). All datasets were downloaded from NCBI's SRA database and were normalised to logCPM as described in chapter 4.2.2.

4.3.4.1 Weger et al. 2021 Liver *Arntl* WT vs *Arntl* KO

Figure 4.28A demonstrates that the majority of TimeTeller's time predictions for the *Arntl* WT and *Arntl* KO liver samples were of relatively high confidence (Θ value remained < 0.08). The time predictions for the *Arntl* KO liver samples displayed a much higher error than the *Arntl* WT liver samples (Figure 4.28 A&B) as the time predictions for the *Arntl* KO liver samples all clustered between CT20 and CT24, regardless of the original sampling time. The clustering of *Arntl* KO liver samples to CT20-24 is also noticeable when the data is observed in the low dimensional local projection spaces (Figure 4.28C). This clustering of samples to a particular point in time implies that when *Arntl* was knocked out the oscillatory clock in the liver became 'stuck or 'frozen' at a particular point in time.

Hughey et al. have previously demonstrated using ZeitZeiger that *Arntl* KO microarray samples had a demonstrable bias for CT20-24, but the bias was not so extreme that the molecular oscillatory system appeared frozen (Figure 4.29). There are multiple reasons why TimeTeller's findings might differ from ZeitZeiger's; ZeitZeiger was trained using a slightly different gene set to TimeTeller, ZeitZeiger was tested on different test datasets from different tissues that used different knock out techniques relative to the test datasets used here, and Hughey et al. even applied an unbalanced batch correction method to their data prior to running ZeitZeiger. The *Arntl* WT and KO liver samples analysed by TimeTeller support the conclusion that *Arntl* KO biases the molecular clock to a state that limits the expression of E-box driven genes

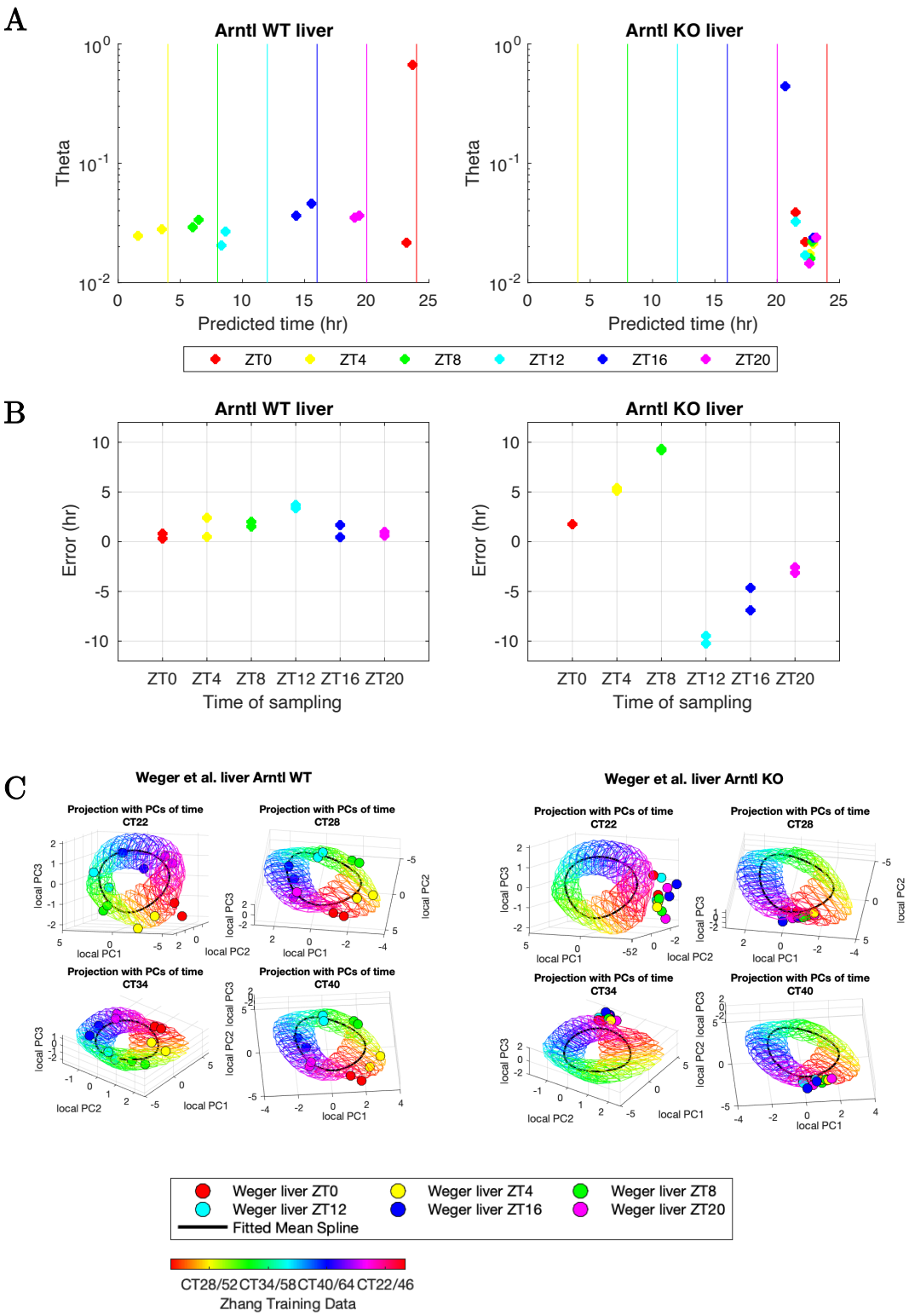


Figure 4.28 A) TimeTeller's predicted times for **Weger et al's Arntl WT and KO liver** samples, plotted against corresponding Θ values (logthresh=e^{12}). Vertical coloured lines represent the actual time of sampling of all samples. B) Error (difference between time of sampling and TimeTeller's predicted time) of samples. C) Projected Zhang et al. training and Weger et al. liver Arntl WT and KO test data onto the first 3 principal components identified from each of the original training data timepoints (t=4).

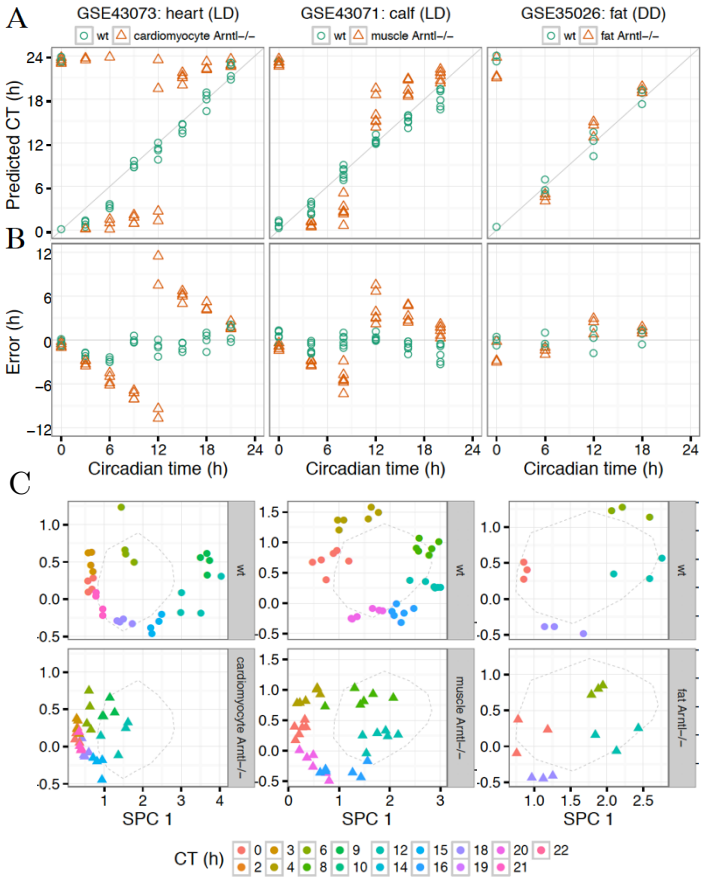


Figure 4.29 Figure reproduced from Hughey et al. 2016, supplementary Figure 11. ZeitZeiger's findings for three different *Arntl* KO test microarray datasets. A) Time of sampling vs ZeitZeiger predicted time. B) Error, N.B. ZeitZeiger error trends appear reversed relative to TimeTeller's error trends because ZeitZeiger calculates error as 'predicted time' – 'sampling time', rather than 'sampling time' – 'predicted time'. C) Low dimensional data projections. Dashed line represents the model built from the Zhang microarray dataset. Coloured circles/triangles indicate projected test data. *Arntl* KO gave a varied phenotype, but tended to result in a damped oscillator that was biased towards ~CT20-CT24.

such as *Per2* and *Dbp*, and that this may occur to such an extent as to limit oscillatory core circadian gene expression altogether. Since E-box driven genes are at their lowest expression at CT20-24, this is the timepoint that best represents the gene expression that results from *Arntl* KO.

4.3.4.2 Yeung et al. 2018 Kidney *Arntl* WT vs *Arntl* KO

As was observed for Weger et al.'s liver samples, Figure 4.30A demonstrates that the majority of TimeTeller's time predictions for the Yeung et al. *Arntl* WT and *Arntl* KO kidney samples were of relatively high confidence (i.e., Θ values remained < 0.08). The time predictions for the *Arntl* KO kidney samples displayed a much higher error than the *Arntl* WT samples (Figure 4.30 A&B), as the time predictions for the *Arntl* KO kidney samples all clustered between CT20 and CT24, regardless of the original sampling time. The clustering of *Arntl* KO kidney samples to CT20-24 was also noticeable when the data was observed in the low dimensional local projection spaces (Figure 4.30C). This clustering of samples to a particular point in time implies that when *Arntl* was knocked out the oscillatory clock in the kidney became 'stuck or 'frozen' at a particular point in time, just as was observed for the Weger et al. liver samples.

It is worth noting that one of the ZT4 *Arntl* KO samples behaved exactly as the ZT 4 WT samples did (Figure 4.30, red arrows). The raw gene expression of this sample was inspected to confirm that this sample was not an accidental duplication of a WT sample upon upload/download to/from the SRA database. The implication is that a WT mouse was mistakenly included in the *Arntl* KO group, though Yeung et al. do not address this issue in their paper. When using TimeTeller, such data quality issues are very apparent to the researcher. Yeung et al.'s dataset confirms TimeTeller's ability to predict biological time of a third tissue – kidney.

4.3.4.3 Weger et al. 2021 liver *Cry1/Cry2* WT versus *Cry1/Cry2* double KO

Figure 4.31A demonstrates that the majority of TimeTeller's time predictions for the *Cry1/2* WT and *Cry1/2* KO liver samples were of relatively high

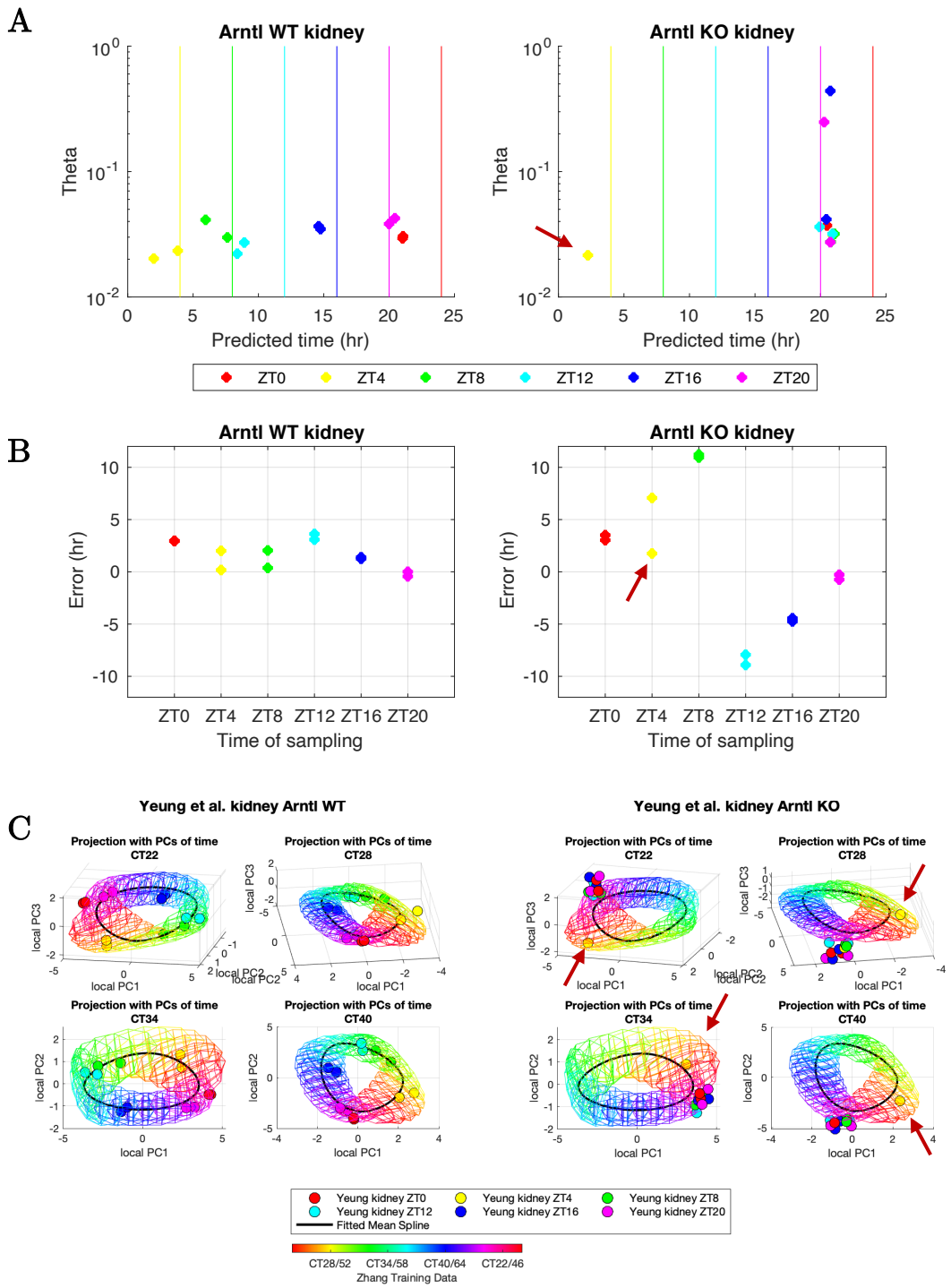


Figure 4.30 A) TimeTeller's predicted times for **Yeung et al's *Arntl* WT and KO kidney** samples, plotted against corresponding Θ values (logthresh=e^{-12}). Vertical coloured lines represent the actual time of sampling of all samples. B) Error (difference between time of sampling and TimeTeller's predicted time) of samples. C) Projected Zhang et al. training and Yeung et al. kidney *Arntl* WT and KO test data onto the first 3 principal components identified from each of the original training data timepoints (t=4). Red arrows indicate one of the ZT4 *Arntl* KO samples which behaved as a WT sample would be expected to.

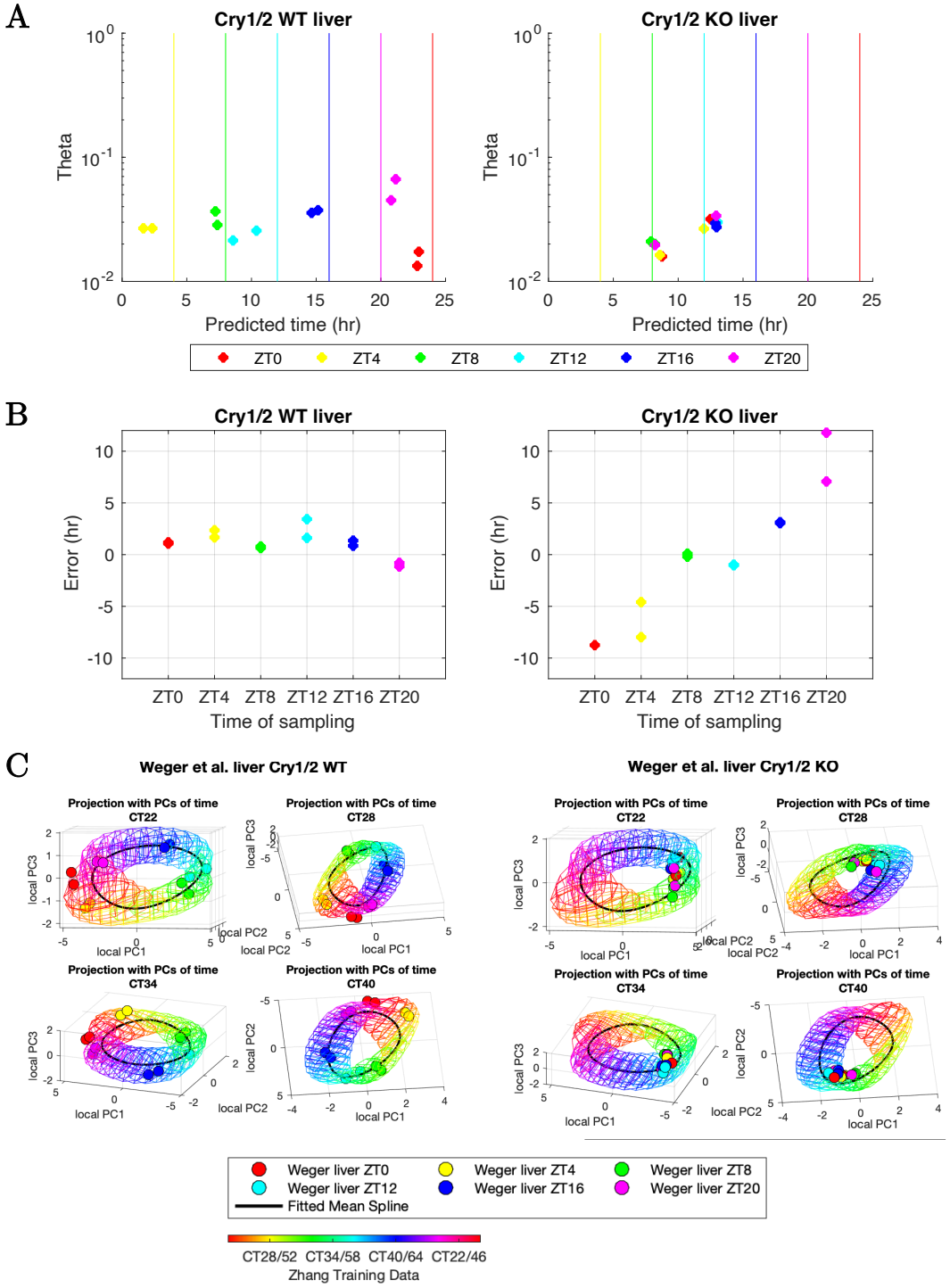


Figure 4.31 A) TimeTeller's predicted times for **Weger et al's *Cry1/2* WT and KO liver** samples, plotted against corresponding Θ values (logthresh=e^{-12}). Vertical coloured lines represent the actual time of sampling of all samples. B) Error (difference between time of sampling and TimeTeller's predicted time) of samples. C) Projected Zhang et al. training and Weger et al. liver *Cry1/2* WT and KO test data onto the first 3 principal components identified from each of the original training data timepoints (t=4).

confidence (i.e., Θ values remained < 0.08). The time predictions for the *Cry1/2* KO liver samples displayed a much higher error than the *Cry1/2* WT liver samples (Figure 4.31 A & B) because the time predictions for the *Cry1/2* KO liver samples all clustered between approximately CT8 and CT13, regardless of the original sampling time. The clustering of *Cry1/2* KO liver samples to CT8-13 was also noticeable when the data was observed in the low dimensional local projection spaces (Figure 4.31C). This clustering of samples to a particular point in time implied that when *Cry1/2* was knocked out the oscillatory clock in the liver became 'stuck or 'frozen' at a particular point in time that is approximately opposite in phase to the point at which *Arntl* KO liver clocks appear to be frozen (CT20-24).

Previously Hughey et al. demonstrated using ZeitZeiger that *Cry1/2* KO microarray samples also displayed time predictions of CT8-13 (Figure 4.32). The Weger et al. *Cry1/2* WT and KO liver samples analysed by TimeTeller support the conclusion that *Cry1/2* KO breaks the molecular oscillator such that the expression of E-box driven genes such as *Per2* and *Dbp* remains constantly high, as the Per/Cry repressor complex cannot form to repress E-box driven gene expression. Since E-box driven genes are at their highest expression at CT8-12, this is the timepoint that best represents the gene expression that results from *Cry1/2* KO.

It is interesting to note that the Weger et al. *Cry1/2* KO liver samples appeared to form a single cluster in the low dimensional projection spaces (Figure 4.31C), but the time predictions for these samples formed two discrete clusters (Figure 4.31B). Further investigation revealed that this was due to the *Cry1/2* KO liver samples displaying split likelihood curves with two discrete peaks (Figure 4.33). The phenomenon was particularly apparent for three samples ZT0B, ZT4A and ZT16B. Split likelihood curves can occur when a

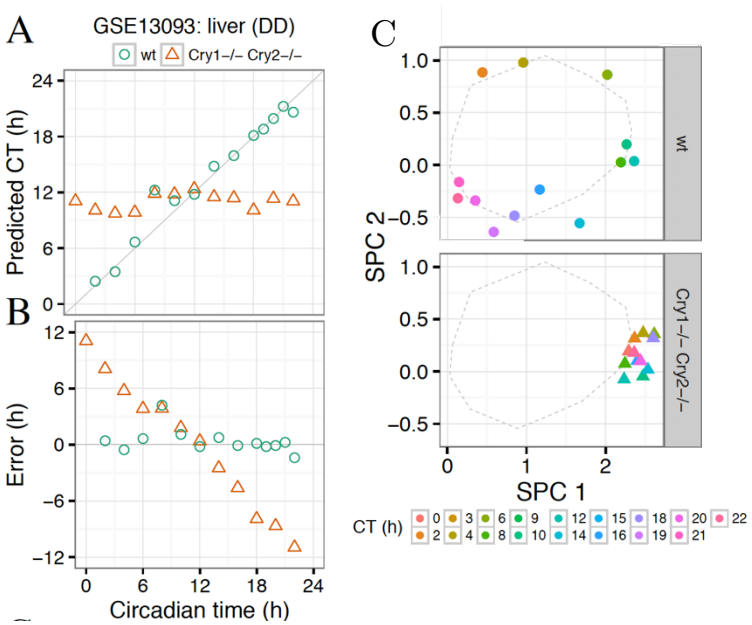


Figure 4.32 Figure reproduced from Hughey et al. 2016, supplementary Figure 11. ZeitZeiger's findings for a *Cry1/2* KO test dataset. A) Time of sampling vs ZeitZeiger predicted time. B) Error, N.B. ZeitZeiger error trends appear reversed relative to TimeTeller's error trends because ZeitZeiger calculates error as 'predicted time' – 'sampling time', rather than 'sampling time' – 'predicted time'. C) Low dimensional data projections. Dashed line represents the model built from the Zhang microarray dataset. Coloured circles/triangles indicate projected test data. *Cry1/2* KO resulted in a 'frozen' oscillator that was 'stuck' at ~CT8-13, as was the case for the Cry1/2 KO liver Weger et al. dataset processed by TimeTeller.

data point lies inside a tight curve of the projected training data, as was the case for some of the *Cry1/2* KO liver samples projected using the principal components obtained from CT22/46, CT28/52 and CT40/64 training data matrix decomposition (see Figure 4.31C for the tight curve in the model at approx. CT34 in these local projection spaces). Since each local projection results in a slightly differently shaped model, it is possible that a higher resolution training dataset, would produce a local projection where the model does not turn quite so sharply at e.g., CT34, and would result in a more continuous time prediction. On the other hand, since some of the *Cry1/2* KO samples really did appear to lie just inside the model in many of the local projections, the implication is that gene expression in the *Cry1/2* KO samples

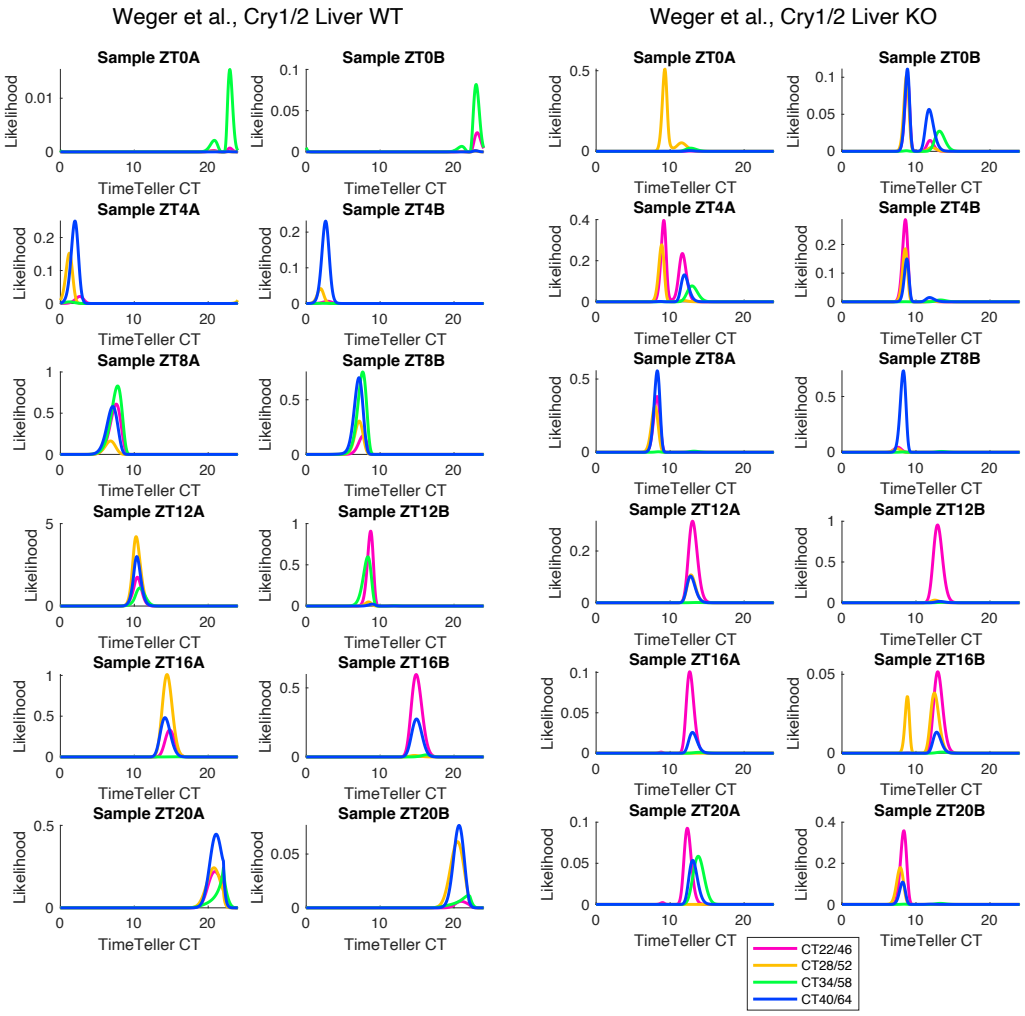


Figure 4.33 Raw likelihood curves for each Weger et al. *Cry1/2* WT and *Cry1/2* KO liver sample. Each sample has four likelihood curves, one generated from each local projection of the data. The *Cry1/2* KO samples display a phenomenon of split likelihood peaks, which was observed for the projections built from CT22/46, CT28/52 and CT40/64 data matrix decomposition, but not the CT34/58 projection.

most closely matched clock gene expression at either ~CT8 or ~CT13, and therefore the predicted times were accurate. Data such as this highlights the importance of inspecting the local projections and the raw likelihood curves for a better understanding of the outputs of TimeTeller.

4.3.4.4 Summary of quasi-time-course normalisation

The quasi-time-course normalisation approach presented here allowed TimeTeller to proficiently predict time and a clock dysfunction metric, with high accuracy, for single biological samples across a range of murine RNA-seq tissues, in a range of different genotypes. This satisfies the original key criteria that were identified in chapter 4.1.3 for a biological time-telling algorithm. Following an extensive literature search (chapter 4.1.2), TimeTeller is believed to be the first instance of a time prediction tool that has been applied to murine RNA-seq datasets that are independent of the training dataset used to build the time prediction model. TimeTeller is also the first time-prediction tool that has been used to assign predicted times to single independent RNA-seq test data samples. Lastly, TimeTeller is the only time prediction tool that has a comprehensive metric of clock dysfunction, termed Θ here.

4.4 Conclusions

The application of TimeTeller to RNA-seq data represents a step forwards for algorithmic time prediction. For this method to be useful in a human time-telling context, training datasets will be required. There are already a number of human RNA-seq time course datasets that exist for multiple individuals in either skeletal muscle or abdominal fat, which may be of interest to answer specific research questions (Perrin et al., 2018; Stenvers, Jongejan, et al., 2019). However, training data time series for deep internal tissues such as liver and kidney will not be easy to obtain from humans for obvious reasons. The particular strength of the quasi-time-course normalisation approach presented here, is that in theory a tissue-matched training dataset is not required for time prediction. If necessary, only the tissue-specific time-series means and standard deviations of gene expression are required to normalise a test sample, in order to enable time prediction using a trained multi-tissue TimeTeller. The time-dependent means and standard deviations of expressed genes in, e.g., liver samples, should be relatively easier to estimate using biopsies that are collected in the clinic for diagnostic purposes, provided that time-of-sampling information is recorded. It is worth noting here that currently time-of-sampling metadata is frequently not collected with clinical samples intended for a secondary research purpose, and that an effort should be made to improve this, for the benefit of the wider circadian research community (Stephenson et al., 2021).

In addition to TimeTeller's demonstrable use in time prediction of *ex vivo* samples, an *in vitro* application is also apparent. For instance, *in vitro* murine time series data could be mapped to the trained multi-tissue murine TimeTeller in order to produce time predictions for the *in vitro* samples. The absolute

value of the time predictions would of course not be relevant in an *in vitro* population of cells synchronised to an arbitrary phase by e.g., dexamethasone synchronisation. However, TimeTeller's relative time predictions for a number of samples could be used to estimate the period of the *in vitro* population clock. Current estimations of period generally rely on sine fits to individual clock genes, which can be error prone when a relatively small number of samples are available for fitting. The power of TimeTeller is that dimensionality reduction allows for many different clock genes to contribute to a single period estimate.

It is important to acknowledge that the quasi-time-course normalisation approach used here in combination with TimeTeller, could also be used to normalise data prior to training ZeitZeiger. A comparison of TimeTeller and ZeitZeiger is not presented here, since this is discussed at length by Vlachou et al. (2020). In short, TimeTeller's time predictions were found to be slightly less accurate than ZeitZeiger's, but TimeTeller had a more meaningful metric of confidence (Θ) in the time predictions, which was directly related to the variance structure of the underlying training data. It is exciting to consider that ultimately the strengths of both TimeTeller and ZeitZeiger could be combined to build an even better time predictor model.

The research presented here achieves the goal of single sample time-telling from a remarkably low-resolution training dataset, i.e., a sampling interval of 6 hrs. There is also no requirement for batch correction or retraining of the model when applied to novel test data sets. It is anticipated that these features will help to make TimeTeller easy for future researchers to extend. Moreover, the results presented here suggest the model works for 8 different tissues in mice, of which 3 (liver, kidney and skeletal muscle) have been independently validated with publicly available data. This is an achievement that holds real potential to help to elucidate the circadian behaviour of spontaneous or

orthotopic tumours harvested from *in vivo* murine models. It is hoped that the use of this tool will help to answer some of the questions that remain regarding the cancer circadian clock. For instance, does the tumour molecular clock behave in the same manner as the surrounding tissue clock? Are peripheral tissue clocks affected by the presence of tumours, as has been suggested by Huisman et al. (2015) using a murine liver tumour model and Masri et al. (2016) using a murine lung tumour model (Huisman et al., 2015; Masri et al., 2016)? And might the circadian function of tumours be targeted as a therapeutic avenue as suggested by Sulli et al. among others (Sulli et al., 2018)?

Finally, it is worth noting that this novel adaption of TimeTeller to time prediction of single sample murine RNA-seq data, may prove useful in other fields of circadian dysregulation research besides cancer, including but not limited to sleep/wake, metabolic and cardiovascular systems. For instance, Manella et al. (2021) have recently assessed the effect of dark-restricted feeding on the mean phase advance/delay of six core clock genes across a range of different mouse tissues by RNA-seq (Manella et al., 2021). This required that accurate sine fits be applied to each of the clock genes considered, and therefore a high number of samples were considered per tissue (n=24), resulting no doubt in high experimental costs. TimeTeller can predict clock time and function from a single sample, considers more than six genes, and accounts for amplitude/magnitude differences in addition to phase differences. Therefore, it is likely that the validated murine RNA-seq TimeTeller described in this thesis, would enable a deeper understanding of the state of the peripheral molecular oscillators under dark-restricted feeding, from many fewer samples, and therefore at lower cost.

5. Conclusions and Outlook

5.1 Summary of thesis

The body of work presented in this thesis explored circadian molecular function in a variety of different model systems, with a primary focus on circadian gene activity in an oncology context. Specifically, the three results chapters (chapters 2, 3 and 4) focused on circadian gene activity in three different model systems, which were designed to represent increasingly physiological scenarios. The selection of appropriate model systems is an important area of circadian research because it is clear from a review of the circadian-oncology literature that although there appears to be an association between circadian function and cancer, there is no universal pathway that has been identified, and findings are frequently conflicting (as reviewed in chapter 1.4). Circadian-oncology research efforts frequently overlook the importance of understanding the state of circadian clock gene expression either before or after perturbation of the model system, which is a particular focus of this thesis (Stephenson et al., 2021).

Chapter 2 explored the clock gene behaviour of the *in vitro* cell lines MCF10A and MCF7, as representatives of 'normal' and 'malignant' breast cells. The data presented in Chapter 2 does not agree with the canonical literature conclusion that MCF10A cells have a more robust circadian oscillator than MCF7 cells. In particular, this chapter emphasises the value of interrogating existing large 'omics datasets.

Chapter 3 explored the circadian behaviour of the chicken embryo, as part of an effort to extend the 3D environment CAM tumour assay to a circadian application. The relatively novel use of luciferase reporter systems *in ovo* proved challenging, as tumour bioluminescence did not appear to correlate well with tumour growth and metastasis, therefore further research will be required to address the true value of the CAM tumour assay in a circadian context.

Chapter 4 extended the mathematical model TimeTeller to murine RNA-seq datasets, in order to predict the circadian behaviour of single samples. The development and validation of this method is a key step in translating oncology samples, which are frequently not available as a time-series, to a circadian application. In particular, this chapter highlights the value of an inter-disciplinary approach to scientific research.

5.2 Areas for future research and development

5.2.1 2D vs 3D circadian research

The advantages that a 3D *in vitro* environment provides in better recapitulating a physiological environment, have already been discussed in chapter 2.4. In the field of circadian research, it has been suggested that even the stiffness of a 3D *in vitro* environment significantly affects the circadian behaviour of cultured cells (Williams et al., 2018; N. Yang et al., 2017). The research presented here did not explore an intermediary 3D *in vitro* system between 2D *in vitro* and 3D *in ovo*. In the field of circadian-oncology research, one of the primary research questions is the extent to which cancer cells interact with their 3D circadian environment in the context of local tissue clocks and the distal central oscillator. The expression of circadian genes in

cancer cells cultured in reductionist 3D *in vitro* models vs *in vivo* models may help to provide an answer to this question.

5.2.2 Single cell circadian behaviour

The methods used in this thesis all rely on a population-wide approach to the study of circadian gene activity. However, population-wide circadian activity is an average of the circadian gene activity in each of the cells within the population, and may vary from cell to cell (Nagoshi et al., 2004). Single cell circadian rhythms are interesting in the context of cancer because cancers are known to contain different populations of cells that evolve over time (Batlle & Clevers, 2017). In recent years, single cell RNA-seq has been applied to time-series samples of the SCN to demonstrate the circadian behaviour of different populations of neurons, and other cells such as astrocytes, within the central circadian oscillator (Park et al., 2016; Wen et al., 2020). Currently, the study of single cell cancer rhythms by single cell RNA-seq is limited by the expense that a time series experiment with sufficient library depth to detect clock gene oscillation would incur. The high-resolution single cell imaging technique single molecule fluorescence *in situ* hybridisation (smFISH) offers a probe-based alternative to transcript detection, such that read depth is no longer a concern (Phillips et al., 2021). Single cell techniques have the power to address the function of the circadian clock in single cells or sub-populations of established tumours, which may help to elucidate the relationship between the circadian clock and cancer.

5.2.3 Male/female circadian differences

Since the majority of experimental research uses male mice, all the studies discussed as part of chapter 4 were composed entirely of samples from male mice. This is a limitation of the work presented in this thesis – and indeed of

the wider research community – and TimeTeller should also be comprehensively validated using murine female RNA-seq samples, when such samples become available (Mauvais-Jarvis et al., 2017). It has previously been shown that there are small differences in the expression of some of the core circadian clock genes in mouse liver between male and female mice (Weger et al., 2019). For instance, the amplitudes of oscillation of *Per3*, *Npas2* and *Nfil3* were higher in the female mice than the male mice. However, Vlachou et al., (2020) noted no difference in the circadian gene expression profiles of the male and female human oral mucosa samples used to train TimeTeller. Researchers should keep sex differences in mind when training and validating future machine learning time prediction models.

5.2.4 Use of luciferin polymer patches for murine research

Since luciferin is eliminated extremely rapidly from mice (Berger et al., 2008), a non-surgical approach to luciferin delivery via a transdermal patch, would represent an improvement on the current technology. Currently longitudinal luciferin delivery to mice is achieved either via surgical implantation of a luciferin-loaded osmotic pump, or by addition of luciferin to the drinking water – which is consumed primarily nocturnally by mice and therefore does not represent constant luciferin delivery (Martin-Burgos et al., 2020). Therefore, the *in ovo* use of luciferin-loaded polymer patches demonstrated here, might also provide an advantage over current luciferin delivery systems in other *in vivo* models, such as mice.

5.3 Novel findings

5.3.1 MCF10A and MCF7 core clock gene activity

There was already some recent evidence that suggested that the MCF10A cells might have a long free-running period, and that MCF7 cells might display more core circadian oscillatory gene expression than the previous literature consensus (Lellupitiyage Don et al., 2019, 2020; Lin et al., 2019). However, this is the first time that the rhythmic gene activity of MCF7 and MCF10A cells has been directly compared under a comprehensive range of synchronisation and entrainment conditions. The luciferase reporter construct and qPCR experiments presented here consistently found that MCF10A cells were no more rhythmic than MCF7 cells. RNA-seq revealed that in fact the core molecular oscillator and wider transcriptome of MCF7 cells behaved relatively more rhythmically in response to a 12 hr 36°C / 12 hr 37°C temperature entrainment protocol than MCF10A cells. Additionally, the re-analysis of a published MCF10A and MCF7 microarray time-course dataset highlighted concerns regarding the identity of the MCF7 cells (Gutiérrez-Monreal et al., 2016). Frequently, the expression of very few circadian genes, from cells synchronised with a single entrainment method, are used to determine the circadian behaviour of a cancer cell line (see table X for instance). The data presented here cautions against such an approach, which may be misleading, and ultimately further confuse the literature. MCF10A cells cultured in a 2D *in vitro* environment do not appear to represent 'normal' molecular oscillation. This finding is important in the context of research that uses MCF10A cells as a model to try to understand how perturbations to clock gene expression affect cell proliferation (Hwang-Verslues et al., 2013).

5.3.2 Luciferin loaded polymer patches

This thesis presents the first known instance of intra-tumoural delivery of luciferin from a slow-release polymeric formulation (originally intended for transdermal drug delivery (Tombs et al., 2018)). The use of luciferin loaded polymer patches for *in ovo* bioluminescence imaging provided an improvement over luciferin delivery via a drip, the alternative method of long-term luciferin delivery that was investigated. This is because the use of a polymer patch eliminated many of the technical challenges of drip delivery, including concerns about maximum tolerated dose volume, luciferin stability at room temperature in solution, and uneven drip rate. Further investigation will be required to understand more about the release kinetics of luciferin from the polymer patch to the CAM.

5.3.3 Extension of TimeTeller method to RNA-seq data

This thesis extends the single sample prediction of time and clock dysfunction by TimeTeller to RNA-seq datasets for the first time. Additionally, quasi-time-course normalisation is newly combined with TimeTeller for accurate multi-tissue time prediction. Lastly, a 6 hr resolution RNA-seq time-series was demonstrated to be sufficient for training TimeTeller, which is excellent proof-of-concept to reduce the cost of obtaining future training datasets in either mice or humans. These are valuable developments in the field of single sample time prediction, which will enable the study of the circadian system in samples that were previously obtuse to interpretation, due to their independent nature.

5.4 Closing remarks

Taken collectively, the findings of this thesis all aimed to further our understanding of the function of the molecular circadian clock in cancer. The

data suggested that MCF10A cells cultured *in vitro* are not representative of a functional circadian oscillator, in contrast to the literature consensus. It is hoped that the validated multi-tissue murine RNA-seq TimeTeller that was presented in this thesis might be used to explore differences in circadian function between *in vitro* murine cancer cell line time series data and single sample cell-line derived tumours in an *in ovo* or *in vivo* setting.

Bibliography

Agostinelli, F., Ceglia, N., Shahbaba, B., Sassone-Corsi, P., & Baldi, P. (2016). What time is it? Deep learning approaches for circadian rhythms. *Bioinformatics*, *32*, i8–i17. https://doi.org/10.1093/bioinformatics/btw243

Aiello, I., Mul Fedele, M. L., Román, F., Marpegan, L., Caldart, C., Chiesa, J. J., Golombek, D. A., Finkielstein, C. V., & Paladino, N. (2020). Circadian disruption promotes tumor-immune microenvironment remodeling favoring tumor cell proliferation. *Science Advances*, *6*(42), 4530–4544. https://doi.org/10.1126/sciadv.aaz4530

Akasaka, K., Nasu, T., Katayama, T., & Murakami, N. (1995). Development of regulation of melatonin release in pineal cells in chick embryo. *Brain Research*, *692*, 283–286. https://doi.org/10.1016/0006-8993(95)00643-5

Alter, O., Brown, P. O., & Botstein, D. (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proceedings of the National Academy of Sciences*, *97*(18), 10101–10106. https://doi.org/10.1073/pnas.97.18.10101

Altman, B. J., Hsieh, A. L., Sengupta, A., Krishnanaiah, S. Y., Stine, Z. E., Walton, Z. E., Gouw, A. M., Venkataraman, A., Li, B., Goraksha-Hicks, P., Diskin, S. J., Bellovin, D. I., Simon, M. C., Rathmell, J. C., Lazar, M. A., Maris, J. M., Felsher, D. W., Hogenesch, J. B., Weljie, A. M., & Dang, C. V. (2015). MYC Disrupts the Circadian Clock and Metabolism in Cancer Cells. *Cell Metabolism*, *22*(6), 1009–1019. https://doi.org/10.1016/j.cmet.2015.09.003

Aluwong, T., Sumanu, V. O., Ayo, J. O., Ocheja, B. O., Zakari, F. O., & Minka, N. S. (2017). Daily rhythms of cloacal temperature in broiler chickens of different age groups administered with zinc gluconate and probiotic during the hot-dry season. *Physiological Reports*, *5*(12), e13314. https://doi.org/10.14814/phy2.13314

Anafi, R. C., Francey, L. J., Hogenesch, J. B., & Kim, J. (2017). CYCLOPS reveals human transcriptional rhythms in health and disease. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), 5312–5317. https://doi.org/10.1073/pnas.1619320114

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Ansari, N., Agathagelidis, M., Lee, C., Korf, H. W., & Von Gall, C. (2009). Differential maturation of circadian rhythms in clock gene proteins in the suprachiasmatic nucleus and the pars tuberalis during mouse ontogeny. *European Journal of Neuroscience*, *29*(3), 477–489. https://doi.org/10.1111/j.1460-9568.2008.06605.x

Arnandis, T., & Godinho, S. A. (2015). Chapter 3 - Studying centrosome function using three dimensional cell cultures. In *Methods in Cell Biology* (Vol. 129, pp. 37–50). Elsevier Ltd. https://doi.org/10.1016/bs.mcb.2015.03.010

Aryal, R. P., Kwak, P. B., Tamayo, A. G., Gebert, M., Chiu, P. L., Walz, T., & Weitz, C. J. (2017).

Macromolecular Assemblies of the Mammalian Circadian Clock. *Molecular Cell*, *67*(5), 770–782. https://doi.org/10.1016/j.molcel.2017.07.017

Auger, R. R., Burgess, H. J., Emens, J. S., Deriy, L. V., Thomas, S. M., & Sharkey, K. M. (2015). Clinical practice guideline for the treatment of intrinsic circadian rhythm sleep-wake disorders: Advanced Sleep-Wake Phase Disorder (ASWPD), Delayed Sleep-Wake Phase Disorder (DSWPD), Non-24-Hour Sleep-Wake Rhythm Disorder (N24SWD), and Irregular Sleep-Wake Rhythm Disorder (ISWRD). an update for 2015. *Journal of Clinical Sleep Medicine*, *11*(10), 1199–1236. https://doi.org/10.5664/jcsm.5100

Baker, F. C., Waner, J. I., Vieira, E. F., Taylor, S. R., Driver, H. S., & Mitchell, D. (2001). Sleep and 24 hour body temperatures: A comparison in young men, naturally cycling women and women taking hormonal contraceptives. *Journal of Physiology*, *530*(3), 565–574. https://doi.org/10.1111/j.1469-7793.2001.0565k.x

Baklaushev, V. P., Kilpeläinen, A., Petkov, S., Abakumov, M. A., Grinenko, N. F., Yusubalieva, G. M., Latanova, A. A., Gubskiy, I. L., Zabozlaev, F. G., Starodubova, E. S., Abakumova, T. O., Isaguliants, M. G., & Chekhonin, V. P. (2017). Luciferase expression allows bioluminescence imaging but imposes limitations on the orthotopic mouse (4T1) model of breast cancer. *Scientific Reports*, *7*(7715), 1–17. https://doi.org/10.1038/s41598-017-07851-z

Balsalobre, A., Brown, S. A., Marcacci, L., Tronche, F., Kellendonk, C., Reichardt, H. M., Schutz, G., & Schibler, U. (2000). Resetting of circadian time in peripheral tissues by glucocorticoid signaling. *Science*, *289*, 2344–2347. https://doi.org/10.1126/science.289.5488.2344

Balsalobre, A., Damiola, F., & Schibler, U. (1998). A Serum Shock Induces Circadian Gene Expression in Mammalian Tissue Culture Cells. *Cell*, *93*, 929–937. https://doi.org/10.1016/S0092-8674(00)81199-X

Baschieri, F., Guaraldi, P., Provini, F., Chiogna, M., Barletta, G., Cecere, A., De Scisciolo, G., Cortelli, P., & Calandra-Buonaura, G. (2020). Circadian and state-dependent core body temperature in people with spinal cord injury. *Spinal Cord*, *59*, 538–546. https://doi.org/10.1038/s41393-020-0521-8

Basti, A., Fior, R., Yalçin, M., Póvoa, V., Astaburuaga, R., Li, Y., Naderi, J., Godinho Ferreira, M., & Relógio, A. (2020). The Core-Clock Gene NR1D1 Impacts Cell Motility. *Cancers*, *12*, 853–870. https://doi.org/10.3390/cancers12040853

Batlle, E., & Clevers, H. (2017). Cancer stem cells revisited. *Nature Medicine*, *23*(10), 1124–1134. https://doi.org/10.1038/nm.4409

Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L., & Zoran, M. J. (2005). Circadian rhythms from multiple oscillators: Lessons from diverse organisms. *Nature Reviews Genetics*, *6*(7), 544–556. https://doi.org/10.1038/nrg1633

Berger, F., Paulmurugan, R., Bhaumik, S., & Gambhir, S. S. (2008). Uptake kinetics and biodistribution of 14C-d-luciferin—a radiolabeled substrate for the firefly luciferase catalyzed bioluminescence reaction: impact on bioluminescence based reporter gene imaging. *European Journal of Nuclear*

Medicine and Molecular Imaging, *35*(12), 2275–2285. https://doi.org/10.1007/s00259-008-0870-6

Blythe, S. A., & Wieschaus, E. F. (2016). Establishment and maintenance of heritable chromatin structure during early drosophila embryogenesis. *ELife*, *5*, e20148. https://doi.org/10.7554/eLife.20148

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bolin, G., & Burggren, W. W. (2013). Metanephric kidney development in the chicken embryo: Glomerular numbers, characteristics and perfusion. *Comparative Biochemistry and Physiology, Part A*, *166*, 343–350. https://doi.org/10.1016/j.cbpa.2013.07.011

Brown, S. A., Zumbrunn, G., Fleury-Olela, F., Preitner, N., & Schibler, U. (2002). Rhythms of mammalian body temperature can sustain peripheral circadian clocks. *Current Biology*, *12*(9), 1574–1583. https://doi.org/10.1016/S0960-9822(02)01145-4

Bruinsma, W., Berg, J., Aprelia, M., & Medema, R. H. (2016). Tousled-like kinase 2 regulates recovery from a DNA damage-induced G2 arrest. *EMBO Reports*, *17*(5), 659–670. https://doi.org/10.15252/embr.201540767

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2016). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, *109*, 21.29.1-21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Burden, N., Chapman, K., Sewell, F., & Robinson, V. (2015). Pioneering better science through the 3Rs: An introduction to the National Centre for the Replacement, Refinement, and Reduction of Animals in Research (NC3Rs). *Journal of the American Association for Laboratory Animal Science*, *54*(2), 198–208. www.nc3rs.org.uk

Cadenas, C., Van De Sandt, L., Edlund, K., Lohr, M., Hellwig, B., Marchan, R., Schmidt, M., Rahnenführer, J., Oster, H., & Hengstler, J. G. (2014). Loss of circadian clock gene expression is associated with tumor progression in breast cancer. *Cell Cycle*, *13*(20), 3282–3291. https://doi.org/10.4161/15384101.2014.954454

Cao, Q., Zhao, X., Bai, J., Gery, S., Sun, H., Lin, D. C., Chen, Q., Chen, Z., Mack, L., Yang, H., Deng, R., Shi, X., Chong, L. W., Cho, H., Xie, J., Li, Q. Z., Müschen, M., Atkins, A. R., Liddle, C., ... Koeffler, H. P. (2017). Circadian clock cryptochrome proteins regulate autoimmunity. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(47), 12548–12553. https://doi.org/10.1073/pnas.1619119114

Cassone, V. M. (2014). Avian circadian organization: A chorus of clocks. *Frontiers in Neuroendocrinology*, *35*(1), 76–88. https://doi.org/10.1016/j.yfrne.2013.10.002

Cermakian, N., & Boivin, D. B. (2009). The regulation of central and peripheral circadian clocks in

humans. *Obesity Reviews*, *10*(SUPPL. 2), 25–36. https://doi.org/10.1111/j.1467-789X.2009.00660.x

Chacolla-Huaringa, R., Moreno-Cuevas, J., Trevino, V., & Scott, S.-P. (2017). Entrainment of Breast Cell Lines Results in Rhythmic Fluctuations of MicroRNAs. *International Journal of Molecular Sciences*, *18*, 1499–1518. https://doi.org/10.3390/ijms18071499

Chastagner, P., Rubinstein, E., & Brou, C. (2017). Ligand-activated Notch undergoes DTX4-mediated ubiquitylation and bilateral endocytosis before ADAM10 processing. *Science Signaling*, *10*, eaag2989. https://doi.org/10.1126/scisignal.aag2989

Chen, C. Y., Logan, R. W., Ma, T., Lewis, D. A., Tseng, G. C., Sibille, E., & McClung, C. A. (2016). Effects of aging on circadian patterns of gene expression in the human prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(1), 206–211. https://doi.org/10.1073/pnas.1508249112

Cohen, M., Lippman, M., & Chabner, B. (1978). Role of pineal gland in aetiology and treatment of breast cancer. *The Lancet*, *312*(8094), 814–816. https://doi.org/10.1016/S0140-6736(78)92591-6

Comşa, Ş., Cîmpean, A. M., & Raica, M. (2015). The story of MCF-7 breast cancer cell line: 40 Years of experience in research. *Anticancer Research*, *35*(6), 3147–3154. https://ar.iiarjournals.org/content/35/6/3147

Conway-Campbell, B. L., Sarabdjitsingh, R. A., McKenna, M. A., Pooley, J. R., Kershaw, Y. M., Meijer, O. C., de Kloet, E. R., & Lightman, S. L. (2010). Glucocorticoid ultradian rhythmicity directs cyclical gene pulsing of the clock gene period 1 in rat hippocampus. *Journal of Neuroendocrinology*, *22*(10), 1093–1100. https://doi.org/10.1111/j.1365-2826.2010.02051.x

Cornelissen, G. (2014). Cosinor-based rhythmometry. *Theoretical Biology and Medical Modelling*, *11*(16), 1–24. https://doi.org/10.1186/1742-4682-11-16

Cox, C. A. (2012). *The Circadian Clock and The Cell Cycle* [University College London]. https://core.ac.uk/download/pdf/8776293.pdf

Cox, K. H., & Takahashi, J. S. (2019). Circadian clock genes and the transcriptional architecture of the clock mechanism. *Journal of Molecular Endocrinology*, *63*(4), R93–R102. https://doi.org/10.1530/JME-19-0153

Crnko, S., Du Pré, B. C., Sluijter, J. P. G., & Van Laake, L. W. (2019). Circadian rhythms and the molecular clock in cardiovascular biology and disease. *Nature Reviews Cardiology*, *16*, 437–447. https://doi.org/10.1038/s41569-019-0167-4

da Silva, L. P., & da Silva, J. C. G. E. (2011). Kinetics of inhibition of firefly luciferase by dehydroluciferyl-coenzyme A, dehydroluciferin and L-luciferin. *Photochemical and Photobiological Sciences*, *10*, 1039–1045. https://doi.org/10.1039/c0pp00379d

Dallmann, R., Viola, A. U., Tarokh, L., Cajochen, C., & Brown, S. A. (2012). The human circadian metabolome. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(7), 2625–2629. https://doi.org/10.1073/pnas.1114410109

Damiola, F., Le Minli, N., Preitner, N., Kornmann, B., Fleury-Olela, F., & Schibler, U. (2000). Restricted

feeding uncouples circadian oscillators in peripheral tissues from the central pacemaker in the suprachiasmatic nucleus. *Genes and Development*, *14*, 2950–2961. https://doi.org/10.1101/gad.183500

Davis, P. K., Ho, A., & Dowdy, S. F. (2001). Biological Methods for Cell-Cycle Synchronization of Mammalian Cells. *BioTechniques*, *30*(6), 1322–1330. https://doi.org/10.2144/01306rv01

de Assis, L. V. M., Kinker, G. S., Moraes, M. N., Markus, R. P., Fernandes, P. A., & Castrucci, A. M. de L. (2018). Expression of the circadian clock gene BMAL1 positively correlates with antitumor immunity and patient survival in metastatic melanoma. *Frontiers in Oncology*, *8*(JUN), 185. https://doi.org/10.3389/fonc.2018.00185

Dickmeis, T., Weger, B. D., & Weger, M. (2013). The circadian clock and glucocorticoids - Interactions across many time scales. *Molecular and Cellular Endocrinology*, *380*, 2–15. https://doi.org/10.1016/j.mce.2013.05.012

Dijk, D.-J., & Duffy, J. F. (2020). Novel Approaches for Assessing Circadian Rhythmicity in Humans: A Review. *Journal of Biological Rhythms*, *35*(5), 421–438. https://doi.org/10.1177/0748730420940483

Ding, J. M., Faiman, L. E., Hurst, W. J., Kuriashkina, L. R., & Gillette, M. U. (1997). Resetting the biological clock: Mediation of nocturnal CREB phosphorylation via light, glutamate, and nitric oxide. *Journal of Neuroscience*, *17*(2), 667–675. https://doi.org/10.1523/jneurosci.17-02-00667.1997

Doi, M., Hirayama, J., & Sassone-Corsi, P. (2006). Circadian Regulator CLOCK Is a Histone Acetyltransferase. *Cell*, *125*(5), 497–508. https://doi.org/10.1016/j.cell.2006.03.033

Dong, Z., Zhang, G., Qu, M., Gimple, R. C., Wu, Q., Qiu, Z., Prager, B. C., Wang, X., Kim, L. J. Y., Morton, A. R., Dixit, D., Zhou, W., Huang, H., Li, B., Zhu, Z., Bao, S., Mack, S. C., Chavez, L., Kay, S. A., & Rich, J. N. (2019). Targeting glioblastoma stem cells through disruption of the circadian clock. *Cancer Discovery*, *9*(11), 1556–1573. https://doi.org/10.1158/2159-8290.CD-19-0215

Dyer, N. P., Shahrezaei, V., & Hebenstreit, D. (2019). LiBiNorm: An htseq-count analogue with improved normalisation of Smart-seq2 data and library preparation diagnostics. *PeerJ*, *7*, e6222. https://doi.org/10.7717/peerj.6222

Eckel-Mahan, K., & Sassone-Corsi, P. (2015). Phenotyping Circadian Rhythms in Mice. *Current Protocols in Mouse Biology*, *5*(3), 271–281. https://doi.org/10.1002/9780470942390.mo140229

Edmondson, R., Jenkins Broglie, J., Adcock, A. F., & Yang, L. (2014). Three-Dimensional Cell Culture Systems and Their Applications in Drug Discovery and Cell-Based Biosensors. *Assay and Drug Development Technologies*, *12*(4), 207–218. https://doi.org/10.1089/adt.2014.573

Eelderink-Chen, Z., Bosman, J., Sartor, F., Dodd, A. N., Kovács, Á. T., & Merrow, M. (2021). A circadian clock in a nonphotosynthetic prokaryote. *Science Advances*, *7*, eabe2086. https://doi.org/10.1126/sciadv.abe2086

Elshazley, M., Sato, M., Hase, T., Yamashita, R., Yoshida, K., Toyokuni, S., Ishiguro, F., Osada, H., Sekido, Y., Yokoi, K., Usami, N., Shames, D. S., Kondo, M., Gazdar, A. F., Minna, J. D., &

Hasegawa, Y. (2012). The circadian clock gene BMAL1 is a novel therapeutic target for malignant pleural mesothelioma. *International Journal of Cancer*, *131*(12), 2820–2831. https://doi.org/10.1002/ijc.27598

Evans, C., Hardin, J., & Stoebel, D. M. (2018). Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions. *Briefings in Bioinformatics*, *19*(5), 776–792. https://doi.org/10.1093/bib/bbx008

Evans, M. S., Chaurette, J. P., Adams, S. T., Reddy, G. R., Paley, M. A., Aronin, N., Prescher, J. A., & Miller, S. C. (2014). A synthetic luciferin improves bioluminescence imaging in live mice. *Nature Methods*, *11*(4), 393–395. https://doi.org/10.1038/nmeth.2839

Fagundo-Rivera, J., Gómez-Salgado, J., García-Iglesias, J. J., Gómez-Salgado, C., Camacho-Martín, S., & Ruiz-Frutos, C. (2020). Relationship between night shifts and risk of breast cancer among nurses: A systematic review. *Medicina*, *56*(12), 1–18. https://doi.org/10.3390/medicina56120680

Farshadi, E., van der Horst, G. T. J., & Chaves, I. (2020). Molecular Links between the Circadian Clock and the Cell Cycle. *Journal of Molecular Biology*, *432*(12), 3515–3524. https://doi.org/10.1016/j.jmb.2020.04.003

Feeney, K. A., Putker, M., Brancaccio, M., & O'Neill, J. S. (2016). In-depth Characterization of Firefly Luciferase as a Reporter of Circadian Gene Expression in Mammalian Cells. *Journal of Biological Rhythms*, *31*(6), 540–550. https://doi.org/10.1177/0748730416668898

Fekry, B., Ribas-Latre, A., Baumgartner, C., Deans, J. R., Kwok, C., Patel, P., Fu, L., Berdeaux, R., Sun, K., Kolonin, M. G., Wang, S. H., Yoo, S. H., Sladek, F. M., & Eckel-Mahan, K. (2018). Incompatibility of the circadian protein BMAL1 and HNF4α in hepatocellular carcinoma. *Nature Communications*, *9*(1), 1–17. https://doi.org/10.1038/s41467-018-06648-6

Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International Journal of Cancer*, *144*, 1941–1953. https://doi.org/10.1002/ijc.31937

Fiske, C. H., & Boyden, E. A. (1926). Nitrogen Metabolism in the Chick Embryo. *Journal of Biological Chemistry*, *70*(2), 535–556. https://doi.org/10.1016/s0021-9258(18)84508-2

Fortner, B. V., Stepanski, E. J., Wang, S. C., Kasprowicz, S., & Durrence, H. H. (2002). Sleep and quality of life in breast cancer patients. *Journal of Pain and Symptom Management*, *24*(5), 471–480. https://doi.org/10.1016/S0885-3924(02)00500-6

Fujiwara, S., Baek, S., Varticovski, L., Kim, S., & Hager, G. L. (2019). High Quality ATAC-Seq Data Recovered from Cryopreserved Breast Cell Lines and Tissue. *Scientific Reports*, *9*, 516–527. https://doi.org/10.1038/s41598-018-36927-7

Gabrielli, M. G., & Accili, D. (2010). The chick chorioallantoic membrane: A model of molecular, structural, and functional adaptation to transepithelial ion transport and barrier function during embryonic development. *Journal of Biomedicine and Biotechnology*, *2010*(940741), 1–12. https://doi.org/10.1155/2010/940741

Garbarino-Pico, E., Carpentieri, A. R., Contin, M. A., Keller Sarmiento, M. I., Brocco, M. A., Panzetta,

P., Rosenstein, R. E., Caputto, B. L., & Guido, M. E. (2004). Retinal ganglion cells are autonomous circadian oscillators synthesizing N-acetylserotonin during the day. *Journal of Biological Chemistry*, *279*(49), 51172–51181. https://doi.org/10.1074/jbc.M309248200

Garcia-Bassets, I., & Wang, D. (2012). Cistrome plasticity and mechanisms of cistrome reprogramming. *Cell Cycle*, *11*(17), 3199–3210. https://doi.org/10.4161/cc.21281

Gaspar, J. M. (2018). Improved peak-calling with MACS2. *BioRxiv*, *Dec*(10.1101/496521), 1–16. https://doi.org/10.1101/496521

Gekakis, N., Staknis, D., Nguyen, H. B., Davis, F. C., Wilsbacner, L. D., King, D. P., Takahashi, J. S., & Weitz, C. J. (1998). Role of the CLOCK protein in the mammalian circadian mechanism. *Science*, *280*(5369), 1564–1569. https://doi.org/10.1126/science.280.5369.1564

Gery, S., Komatsu, N., Baldjyan, L., Yu, A., Koo, D., & 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. https://doi.org/10.1016/j.molcel.2006.03.038

Ginty, D. D., Kornhauser, J. M., Thompson, M. A., Bading, H., Mayo, K. E., Takahashi, J. S., & Greenberg, M. E. (1993). Regulation of CREB phosphorylation in the suprachiasmatic nucleus by light and a circadian clock. *Science*, *260*(5105), 238–241. https://doi.org/10.1126/science.8097062

Glaser, F. T., & Stanewsky, R. (2005). Temperature synchronization of the Drosophila circadian clock. *Current Biology*, *15*, 1352–1363. https://doi.org/10.1016/j.cub.2005.06.056

Gonzalez, M. M. C., & Aston-Jones, G. (2008). Light deprivation damages monoamine neurons and produces a depressive behavioral phenotype in rats. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(12), 4898–4903. https://doi.org/10.1073/pnas.0703615105

Grant, D., Yin, L., Collins, J. L., Parks, D. J., Orband-Miller, L. A., Wisely, G. B., Joshi, S., Lazar, M. A., Willson, T. M., & Zuercher, W. J. (2010). GSK4112, a small molecule chemical probe for the cell biology of the nuclear heme receptor rev-erbα. *ACS Chemical Biology*, *5*(10), 925–932. https://doi.org/10.1021/cb100141y

Gross, S., Abraham, U., Prior, J. L., Herzog, E. D., & Piwnica-Worms, D. (2007). Continuous Delivery of D-Luciferin by Implanted Micro-osmotic Pumps Enables True Real-Time Bioluminescence Imaging of Luciferase Activity in Vivo. *Molecular Imaging*, *6*(2), 121–130. https://doi.org/10.2310/7290.2007.00009

Guillaumond, F., Dardente, H., Giguère, V., & Cermakian, N. (2005). Differential Control of Bmal1 Circadian Transcription by REV-ERB and ROR Nuclear Receptors. *Journal of Biological Rhythms*, *20*(5), 391–403. https://doi.org/10.1177/0748730405277232

Gutiérrez-Monreal, M. A., Treviño, V., Moreno-Cuevas, J. E., & Scott, S. P. (2016). Identification of circadian-related gene expression profiles in entrained breast cancer cell lines. *Chronobiology International*, *33*(4), 392–405. https://doi.org/10.3109/07420528.2016.1152976

Gwon, D. H., Lee, W. Y., Shin, N., Kim, S. I., Jeong, K., Lee, W. H., Kim, D. W., Hong, J., & Lee, S. Y. (2020). BMAL1 suppresses proliferation, migration, and invasion of U87MG cells by

downregulating cyclin b1, phospho-AKT, and metalloproteinase-9. *International Journal of Molecular Sciences*, *21*(7). https://doi.org/10.3390/ijms21072352

Hadadi, E., & Acloque, H. (2021). Role of circadian rhythm disorders on EMT and tumour-immune interactions in endocrine-related cancers. *Endocrine-Related Cancer*, *28*(2), R67–R80. https://doi.org/10.1530/ERC-20-0280

Hadadi, E., Botelho de Souza, L. E., Bennaceur-Griscelli, A., & Acloque, H. (2018). Identification of valid reference genes for circadian gene-expression studies in human mammary epithelial cells. *Chronobiology International*, *35*(12), 1689–1701. https://doi.org/10.1080/07420528.2018.1508151

Hadadi, E., Taylor, W., Li, X. M., Aslan, Y., Villote, M., Rivière, J., Duvallet, G., Auriau, C., Dulong, S., Raymond-Letron, I., Provot, S., Bennaceur-Griscelli, A., & Acloque, H. (2020). Chronic circadian disruption modulates breast cancer stemness and immune microenvironment to drive metastasis in mice. *Nature Communications*, *11*(1), 1–17. https://doi.org/10.1038/s41467-020-16890-6

Hafner, S., Raabe, M., Wu, Y., Wang, T., Zuo, Z., Rasche, V., Syrovets, T., Weil, T., & Simmet, T. (2019). High-Contrast Magnetic Resonance Imaging and Efficient Delivery of an Albumin Nanotheranostic in Triple-Negative Breast Cancer Xenografts. *Advanced Therapeutics*, *2*, 1900084. https://doi.org/10.1002/adtp.201900084

Halstead, M. M., Kern, C., Saelao, P., Chanthavixay, G., Wang, Y., Delany, M. E., Zhou, H., & Ross, P. J. (2020). Systematic alteration of ATAC-seq for profiling open chromatin in cryopreserved nuclei preparations from livestock tissues. *Scientific Reports*, *10*(5230), 1–12. https://doi.org/10.1038/s41598-020-61678-9

Hampton, O. A., Hollander, P. Den, Miller, C. A., Delgado, D. A., Li, J., Coarfa, C., Harris, R. A., Richards, S., Scherer, S. E., Muzny, D. M., Gibbs, R. A., Lee, A. V., & Milosavljevic, A. (2009). A sequence-level map of chromosomal breakpoints in the MCF-7 breast cancer cell line yields insights into the evolution of a cancer genome. *Genome Research*, *19*(2), 167–177. https://doi.org/10.1101/gr.080259.108

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, *144*(March), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Hastings, M., O'Neill, J. S., & Maywood, E. S. (2007). Circadian clocks: Regulators of endocrine and metabolic rhythms. *Journal of Endocrinology*, *195*, 187–198. https://doi.org/10.1677/JOE-07-0378

Hatanaka, F., Matsubara, C., Myung, J., Yoritaka, T., Kamimura, N., Tsutsumi, S., Kanai, A., Suzuki, Y., Sassone-Corsi, P., Aburatani, H., Sugano, S., & Takumi, T. (2010). Genome-Wide Profiling of the Core Clock Protein BMAL1 Targets Reveals a Strict Relationship with Metabolism. *Molecular and Cellular Biology*, *30*(24), 5636–5648. https://doi.org/10.1128/mcb.00781-10

Helfer, G., Fidler, A. E., Vallone, D., Foulkes, N. S., & Brandstaetter, R. (2006). Molecular Analysis of Clock Gene Expression in the Avian Brain. *Chronobiology International*, *23*(1–2), 113–127. https://doi.org/10.1080/07420520500521871

Hernandez, R., & Brown, D. T. (2010). Growth and maintenance of chick embryo fibroblasts (CEF).

Current Protocols in Microbiology, *Supplement*(Appendix 41), 1–8. https://doi.org/10.1002/9780471729259.mca04is17

Herzog, E. D., & Huckfeldt, R. M. (2003). Circadian Entrainment to Temperature, But Not Light, in the Isolated Suprachiasmatic Nucleus. *Journal of Neurophysiology*, *90*(3), 763–770. https://doi.org/10.1152/jn.00129.2003

Hill, W. L., Bassi, K. L., Bonaventura, L., & Sacus, J. E. (2004). Prehatch entrainment of circadian rhythms in the domestic chick using different light regimes. *Developmental Psychobiology*, *45*(3), 174–186. https://doi.org/10.1002/dev.20021

Hoffman, A. E., Yi, C. H., Zheng, T., Stevens, R. G., Leaderer, D., Zhang, Y., Holford, T. R., Hansen, J., Paulson, J., & Zhu, Y. (2010). CLOCK in breast tumorigenesis: Genetic, epigenetic, and transcriptional profiling analyses. *Cancer Research*, *70*(4), 1459–1468. https://doi.org/10.1158/0008-5472.CAN-09-3798

Hogenesch, J. B., Gu, Y. Z., Jain, S., & Bradfield, C. A. (1998). The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(10), 5474–5479. https://doi.org/10.1073/pnas.95.10.5474

Holen, I., Speirs, V., Morrissey, B., & Blyth, K. (2017). In vivo models in breast cancer research: Progress, challenges and future directions. *DMM Disease Models and Mechanisms*, *10*(4), 359–371. https://doi.org/10.1242/dmm.028274

Hor, C. N., Yeung, J., Jan, M., Emmenegger, Y., Hubbard, J., Xenarios, I., Naef, F., & Franken, P. (2019). Sleep–wake-driven and circadian contributions to daily rhythms in gene expression and chromatin accessibility in the murine cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(51), 25773–25783. https://doi.org/10.1073/pnas.1910590116

Hou, L., Li, H., Wang, H., Ma, D., Liu, J., Ma, L., Wang, Z., Yang, Z., Wang, F., & Xia, H. (2020). The circadian clock gene: PER2 enhances chemotherapeutic efficacy in nasopharyngeal carcinoma when combined with a targeted nanosystem. *Journal of Materials Chemistry B*, *8*, 5336–5350. https://doi.org/10.1039/d0tb00595a

Houdek, P., & Sumová, A. (2014). In vivo initiation of clock gene expression rhythmicity in fetal rat suprachiasmatic nuclei. *PLoS ONE*, *9*(9), e107360. https://doi.org/10.1371/journal.pone.0107360

Hu, J., Ishihara, M., Chin, A. I., & Wu, L. (2019). Establishment of xenografts of urological cancers on chicken chorioallantoic membrane (CAM) to study metastasis. *Precision Clinical Medicine*, *2*(3), 140–151. https://doi.org/10.1093/pcmedi/pbz018

Hu, M. L., Yeh, K. T., Lin, P. M., Hsu, C. M., Hsiao, H. H., Liu, Y. C., Lin, H. Y. H., Lin, S. F., & Yang, M. Y. (2014). Deregulated expression of circadian clock genes in gastric cancer. *BMC Gastroenterology*, *14*, 67. https://doi.org/10.1186/1471-230X-14-67

Hua, H., Wang, Y., Wan, C., Liu, Y., Zhu, B., Wang, X., Wang, Z., & Ding, J. M. (2007). Inhibition of tumorigenesis by intratumoral delivery of the circadian gene mPer2 in C57BL/6 mice. *Cancer Gene Therapy*, *14*(9), 815–818. https://doi.org/10.1038/sj.cgt.7701061

Hua, H., Wang, Y., Wan, C., Liu, Y., Zhu, B., Yang, C., Wang, X., Wang, Z., Cornelissen-Guillaume, G., & Halberg, F. (2006). Circadian gene mPer2 overexpression induces cancer cell apoptosis. *Cancer Science*, *97*(7), 589–596. https://doi.org/10.1111/j.1349-7006.2006.00225.x

Huber, A.-L., Papp, S. J., Chan, A. B., Henriksson, E., Jordan, S. D., Kriebs, A., Nguyen, M., Wallace, M., Li, Z., Christian, M., & Lamia, K. A. (2016). CRY2 and FBXL3 cooperatively degrade c-MYC. *Molecular Cell*, *64*(4), 774–789. https://doi.org/10.1016/j.molcel.2016.10.012.CRY2

Hughes, M. E., DiTacchio, L., Hayes, K. R., Vollmers, C., Pulivarthy, S., Baggs, J. E., Panda, S., & Hogenesch, J. B. (2009). Harmonics of circadian gene transcription in mammals. *PLoS Genetics*, *5*(4), e1000442. https://doi.org/10.1371/journal.pgen.1000442

Hughes, M. E., Hogenesch, J. B., & Kornacker, K. (2010). JTK_CYCLE: an efficient non-parametric algorithm for detecting rhythmic components in genome-scale datasets. *Journal of Biological Rhythms*, *25*(5), 372–380. https://doi.org/10.1177/0748730410379711

Hughey, J. J. (2017). Machine learning identifies a compact gene set for monitoring the circadian clock in human blood. *Genome Medicine*, *9*(19), 1–11. https://doi.org/10.1186/s13073-017-0406-4

Hughey, J. J., & Butte, A. J. (2016). Differential Phasing between Circadian Clocks in the Brain and Peripheral Organs in Humans. *Journal of Biological Rhythms*, *31*(6), 588–597. https://doi.org/10.1177/0748730416668049

Hughey, J. J., Hastie, T., & Butte, A. J. (2016). ZeitZeiger: Supervised learning for high-dimensional data from an oscillatory system. *Nucleic Acids Research*, *44*(8), e80. https://doi.org/10.1093/nar/gkw030

Huisman, S. A., Oklejewicz, M., Ahmadi, A. R., Tamanini, F., Ijzermans, J. N. M., Van Der Horst, G. T. J., & De Bruin, R. W. F. (2015). Colorectal liver metastases with a disrupted circadian rhythm phase shift the peripheral clock in liver and kidney. *International Journal of Cancer*, *136*, 1027–1032. https://doi.org/10.1002/ijc.29089

Hwang-Verslues, W. W., Chang, P.-H., Jeng, Y.-M., Kuo, W.-H., Chiang, P.-H., Chang, Y.-C., Hsieh, T.-H., Su, F.-Y., Lin, L.-C., Abbondante, S., Yang, C.-Y., Hsu, H.-M., Yu, J.-C., Chang, K.-J., Shew, J.-Y., Y-H Lee, E. P., & Lee, W.-H. (2013). Loss of corepressor PER2 under hypoxia up-regulates OCT1-mediated EMT gene expression and enhances tumor malignancy. *Proceedings of the National Academy of Sciences*, *110*(30), 12331–12336. https://doi.org/10.1073/pnas.1222684110

Ieraci, A., Mallei, A., & Popoli, M. (2016). Social Isolation Stress Induces Anxious-Depressive-Like Behavior and Alterations of Neuroplasticity-Related Genes in Adult Male Mice. *Neural Plasticity*, *2016*, 6212983. https://doi.org/10.1155/2016/6212983

Imamura, Y., Mukohara, T., Shimono, Y., Funakoshi, Y., Chayahara, N., Toyoda, M., Kiyota, N., Takao, S., Kono, S., Nakatsura, T., & Minami, H. (2015). Comparison of 2D- and 3D-culture models as drug-testing platforms in breast cancer. *Oncology Reports*, *33*, 1837–1843. https://doi.org/10.3892/or.2015.3767

Ishida, A., Mutoh, T., Ueyama, T., Bando, H., Masubuchi, S., Nakahara, D., Tsujimoto, G., & Okamura,

H. (2005). Light activates the adrenal gland: Timing of gene expression and glucocorticoid release. *Cell Metabolism*, *2*(11), 297–307. https://doi.org/10.1016/j.cmet.2005.09.009

Jaksik, R., Iwanaszko, M., Rzeszowska-Wolny, J., & Kimmel, M. (2015). Microarray experiments and factors which affect their reliability. *Biology Direct*, *10*, 46. https://doi.org/10.1186/s13062-015-0077-2

Jefferies, B., Lenze, F., Sathe, A., Truong, N., Anton, M., Von Eisenhart-Rothe, R., Nawroth, R., & Mayer-Kuckuk, P. (2017). Non-invasive imaging of engineered human tumors in the living chicken embryo. *Scientific Reports*, *7*(4991), 1–9. https://doi.org/10.1038/s41598-017-04572-1

Jiang, W., Zhao, S., Jiang, X., Zhang, E., Hu, G., Hu, B., Zheng, P., Xiao, J., Lu, Z., Lu, Y., Ni, J., Chen, C., Wang, X., Yang, L., & Wan, R. (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*(2), 314–325. https://doi.org/10.1016/j.canlet.2015.12.002

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Jones, M. E., Schoemaker, M. J., McFadden, E. C., Wright, L. B., Johns, L. E., & Swerdlow, A. J. (2019). Night shift work and risk of breast cancer in women: the Generations Study cohort. *British Journal of Cancer*, *121*(2), 172–179. https://doi.org/10.1038/s41416-019-0485-7

Joseph, D., Chong, N. W., Shanks, M. E., Rosato, E., Taub, N. A., Petersen, S. A., Symonds, M. E., Whitehous, W. P., & Wailoo, M. (2015). Getting rhythm: How do babies do it? *Archives of Disease in Childhood: Fetal and Neonatal Edition*, *100*, F50–F54. https://doi.org/10.1136/archdischild-2014-306104

Jouffe, C., Cretenet, G., Symul, L., Martin, E., Atger, F., Naef, F., & Gachon, F. (2013). The Circadian Clock Coordinates Ribosome Biogenesis. *PLoS Biology*, *11*(1), e1001455. https://doi.org/10.1371/journal.pbio.1001455

Kaneko, M., & Cahill, G. M. (2005). Light-Dependent Development of Circadian Gene Expression in Transgenic Zebrafish. *PLoS Biology*, *3*(2), e34. https://doi.org/10.1371/journal.pbio.0030034

Kidd, P. B., Young, M. W., & Siggia, E. D. (2015). Temperature compensation and temperature sensation in the circadian clock. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(46), E6284–E6292. https://doi.org/10.1073/pnas.1511215112

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kinouchi, K., Magnan, C., Ceglia, N., Liu, Y., Cervantes, M., Pastore, N., Huynh, T., Ballabio, A., Baldi, P., Masri, S., & Sassone-Corsi, P. (2018). Fasting Imparts a Switch to Alternative Daily Pathways in Liver and Muscle. *Cell Reports*, *25*(12), 3299-3314.e6. https://doi.org/10.1016/j.celrep.2018.11.077

Klein, D. C., & Moore, R. Y. (1979). Pineal N-acetyltransferase and hydroxyindole-O-methyl-transferase:

control by the retinohypothalamic tract and the suprachiasmatic nucleus. *Brain Research*, *174*(2), 245–262. https://doi.org/10.1016/0006-8993(79)90848-5

Kochan, D. Z., Ilnytskyy, Y., Golubov, A., Deibel, S. H., McDonald, R. J., & Kovalchuk, O. (2015). Circadian disruption-induced microRNAome deregulation in rat mammary gland tissues. *Oncoscience*, *2*(4), 428–442. https://doi.org/10.18632/oncoscience.157

Koike, N., Yoo, S. H., Huang, H. C., Kumar, V., Lee, C., Kim, T. K., & Takahashi, J. S. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science*, *338*(6105), 349–354. https://doi.org/10.1126/science.1226339

Kojetin, D., Wang, Y., Kameneck, T. M., & Burris, T. P. (2011). Identification of SR8278, a Synthetic Antagonist of the Nuclear Heme Receptor REV-ERB. *ACS Chemical Biology*, *6*(2), 131–134. https://doi.org/10.1021/cb1002575

Komatsu, A., Matsumoto, K., Saito, T., Muto, M., & Tamanoi, F. (2019). Patient Derived Chicken Egg Tumor Model (PDcE Model): Current Status and Critical Issues. *Cells*, *8*(5), 440. https://doi.org/10.3390/cells8050440

Kourtidis, A., Jain, R., Carkner, R. D., Eifert, C., Brosnan, M. J., & Conklin, D. S. (2010). An RNAi screen identifies metabolic regulators NR1D1 and PBP as novel survival factors for breast cancer cells with the ERBB2 signature. *Cancer Research*, *70*(5), 1783–1792. https://doi.org/10.1158/0008-5472.CAN-09-1550.An

Laing, E. E., Möller-Levet, C. S., Poh, N., Santhi, N., Archer, S. N., & Dijk, D.-J. (2017). Blood transcriptome based biomarkers for human circadian phase. *ELife*, *6*, e20214-20240. https://doi.org/10.7554/eLife.20214

Lamb, J. R., Zhang, C., Xie, T., Wang, K., Zhang, B., Hao, K., Chudin, E., Fraser, H. B., Millstein, J., Ferguson, M., Suver, C., Ivanovska, I., Scott, M., Philippar, U., Bansal, D., Zhang, Z., Burchard, J., Smith, R., Greenawalt, D., ... Schadt, E. E. (2011). Predictive genes in adjacent normal tissue are preferentially altered by sCNV during tumorigenesis in liver cancer and may rate limiting. *PLoS ONE*, *6*(7), e20090. https://doi.org/10.1371/journal.pone.0020090

Lamia, K. A., Papp, S. J., Yu, R. T., Barish, G. D., Uhlenhaut, N. H., Jonker, J. W., Downes, M., & Evans, R. M. (2011). Cryptochromes mediate rhythmic repression of the glucocorticoid receptor. *Nature*, *480*(7378), 552–556. https://doi.org/10.1038/nature10700

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lee, Y.-J., Chang, P., Lu, J.-H., Chen, P.-Y., & Wang, C.-J. R. (2019). Assessing chromatin accessibility in maize using ATAC-seq. *BioRxiv*, *January*, 1–24. https://doi.org/10.1101/526079

Lee, Y., Lee, J., Kwon, I., Nakajima, Y., Ohmiya, Y., Son, G. H., Lee, K. H., & Kim, K. (2010). Coactivation of the CLOCK-BMAL1 complex by CBP mediates resetting of the circadian clock. *Journal of Cell Science*, *123*(20), 3547–3557. https://doi.org/10.1242/jcs.070300

Lellupitiyage Don, S. S., Lin, H. H., Furtado, J. J., Qraitem, M., Taylor, S. R., & Farkas, M. E. (2019). Circadian oscillations persist in low malignancy breast cancer cells. *Cell Cycle*, *18*(19), 2447–2453.

https://doi.org/10.1080/15384101.2019.1648957

Lellupitiyage Don, S. S., Robertson, K. L., Lin, H. H., Labriola, C., Harrington, M. E., Taylor, S. R., & Farkas, M. E. (2020). Nobiletin affects circadian rhythms and oncogenic characteristics in a cell-dependent manner. *PLoS ONE*, *15*(7), e0236315. https://doi.org/10.1371/journal.pone.0236315

Lericollais, R., Gauthier, A., Bessot, N., Zouabi, A., & Davenne, D. (2013). Morning Anaerobic Performance Is Not Altered by Vigilance Impairment. *PLoS ONE*, *8*(3), e58638. https://doi.org/10.1371/journal.pone.0058638

Li, Heng, Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, Huan, Zhang, S., Zhang, W., Chen, S., Rabearivony, A., Shi, Y., Liu, J., Corton, C. J., & Liu, C. (2020). Endogenous circadian time genes expressions in the liver of mice under constant darkness. *BMC Genomics*, *21*(224), 1–12. https://doi.org/10.1186/s12864-020-6639-4

Li, Q., Birkbak, N. J., Gyorffy, B., Szallasi, Z., & Eklund, A. C. (2011). Jetset: Selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*, *12*(1), 474–481. https://doi.org/10.1186/1471-2105-12-474

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Lin, H.-H., Qraitem, M., Lian, Y., Taylor, S. R., & Farkas, M. E. (2019). Analyses of BMAL1 and PER2 Oscillations in a Model of Breast Cancer Progression Reveal Changes With Malignancy. *Integrative Cancer Therapies*, *18*, 1–13. https://doi.org/10.1177/1534735419836494

Lugena, A. B., Zhang, Y., Menet, J. S., & Merlin, C. (2019). Genome-wide discovery of the daily transcriptome, DNA regulatory elements and transcription factor occupancy in the monarch butterfly brain. *PLoS Genetics*, *15*(7), e1008265. https://doi.org/10.1371/journal.pgen.1008265

Maas, M. B., Iwanaszko, M., Lizza, B. D., Reid, K. J., Braun, R. I., & Zee, P. C. (2020). Circadian Gene Expression Rhythms During Critical Illness. *Critical Care Medicine*, *48*(12), e1294–e1299. https://doi.org/10.1097/CCM.0000000000004697

Malone, M., Harris, A. L., & Luscombe, D. K. (1994). Assessment of the impact of cancer on work, recreation, home management and sleep using a general health status measure. *Journal of the Royal Society of Medicine*, *87*, 386–389.

Manella, G., Sabath, E., Aviram, R., Dandavate, V., Ezagouri, S., Golik, M., Adamovich, Y., & Asher, G. (2021). The liver-clock coordinates rhythmicity of peripheral tissues in response to feeding. *Nature Metabolism*, *3*(6), 829–842. https://doi.org/10.1038/s42255-021-00395-7

Marshall, K. M., Kanczler, J. M., & Oreffo, R. O. (2020). Evolving applications of the egg: chorioallantoic membrane assay and ex vivo organotypic culture of materials for bone tissue engineering. *Journal of Tissue Engineering*, *11*, 1–25. https://doi.org/10.1177/2041731420942734

Martin-Burgos, B., Wang, W., William, I., Tir, S., Mohammad, I., Javed, R., Smith, S., Cui, Y., Smith,

C. B., Van Der Vinne, V., Molyneux, P. C., Miller, S. C., Weaver, D. R., Leise, T. L., Harrington, M. E., & Harrington, M. (2020). Methods for detecting PER2::LUCIFERASE bioluminescence rhythms in freely moving mice. *BioRxiv*, *August*, 1–47. https://doi.org/10.1101/2020.08.24.264531

Masri, S., Papagiannakopoulos, T., Kinouchi, K., Liu, Y., Cervantes, M., Baldi, P., Jacks, T., & Sassone-Corsi, P. (2016). Lung Adenocarcinoma Distally Rewires Hepatic Circadian Homeostasis. *Cell*, *165*(4), 896–909. https://doi.org/10.1016/j.cell.2016.04.039

Mauvais-Jarvis, F., Arnold, A. P., & Reue, K. (2017). A Guide for the Design of Pre-clinical Studies on Sex Differences in Metabolism. *Cell Metabolism*, *25*(6), 1216–1230. https://doi.org/10.1016/j.cmet.2017.04.033

Mauvoisin, D., Wang, J., Jouffe, C., Martin, E., Atger, F., Waridel, P., Quadroni, M., Gachon, F., & Naef, F. (2014). Circadian clock-dependent and -independent rhythmic proteomes implement distinct diurnal functions in mouse liver. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(1), 167–172. https://doi.org/10.1073/pnas.1314066111

McCall, M. N., Bolstad, B. M., & Irizarry, R. A. (2010). Frozen robust multiarray analysis (fRMA). *Biostatistics*, *11*(2), 242–253. https://doi.org/10.1093/biostatistics/kxp059

McGlincy, N. J., Valomon, A., Chesham, J. E., Maywood, E. S., Hastings, M. H., & Ule, J. (2012). Regulation of alternative splicing by the circadian clock and food related cues. *Genome Biology*, *13*, R54. https://doi.org/10.1186/gb-2012-13-6-r54

Menaker, M. (1968). Extraretinal light perception in the sparrow, 1. Entrainment of the biological clock. *Proceedings of the National Academy of Sciences*, *59*(2), 414–421. https://doi.org/10.1073/pnas.59.2.414

Merckx, G., Tay, H., Lo Monaco, M., van Zandvoort, M., De Spiegelaere, W., Lambrichts, I., & Bronckaers, A. (2020). Chorioallantoic Membrane Assay as Model for Angiogenesis in Tissue Engineering: Focus on Stem Cells. *Tissue Engineering*, *26*(6), 519–539. https://doi.org/10.1089/ten.teb.2020.0048

Milani, P., Escalante-Chong, R., Shelley, B. C., Patel-Murray, N. L., Xin, X., Adam, M., Mandefro, B., Sareen, D., Svendsen, C. N., & Fraenkel, E. (2016). Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells OPEN. *Scientific Reports*, *6*, 25474–25484. https://doi.org/10.1038/srep25474

Mitsui, S., Yamaguchi, S., Matsuo, T., Ishida, Y., & Okamura, H. (2001). *Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism*. https://doi.org/10.1101/gad.873501

Moriyama, E. H., Niedre, M. J., Jarvi, M. T., Mocanu, J. D., Moriyama, Y., Subarsky, P., Li, B., Lilge, L. D., & Wilson, B. C. (2008). The influence of hypoxia on bioluminescence in luciferase-transfected gliosarcoma tumor cells in vitro. *Photochemical and Photobiological Sciences*, *7*, 675–680. https://doi.org/10.1039/b719231b

Mure, L. S., Le, H. D., Benegiamo, G., Chang, M. W., Rios, L., Jillani, N., Ngotho, M., Kariuki, T., Dkhissi-Benyahya, O., Cooper, H. M., & Panda, S. (2018). Diurnal transcriptome atlas of a primate across major neural and peripheral tissues. *Science*, *359*, eaao0318.

https://doi.org/10.1126/science.aao0318

Murphy, J., & Rous, P. (1911). The behavior of chicken sarcoma implanted in the developing embryo. *Journal of Experimental Medicine*, *15*, 119–141.

Nader, N., Chrousos, G. P., & Kino, T. (2009). Circadian rhythm transcription factor CLOCK regulates the transcriptional activity of the glucocorticoid receptor by acetylating its hinge region lysine cluster: potential physiological implications. *The FASEB Journal*, *23*, 1572–1583. https://doi.org/10.1096/fj.08-117697

Nagaraja, G. M., Othman, M., Fox, B. P., Alsaber, R., Pellegrino, C. M., Zeng, Y., Khanna, R., Tamburini, P., Swaroop, A., & Kandpal, R. P. (2006). Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: Comprehensive profiles by representational difference analysis, microarrays and proteomics. *Oncogene*, *25*, 2328–2338. https://doi.org/10.1038/sj.onc.1209265

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

Naik, M., Brahma, P., & Dixit, M. (2018). A cost-effective and efficient chick ex-ovo cam assay protocol to assess angiogenesis. *Methods and Protocols*, *1*(19), 1–9. https://doi.org/10.3390/mps1020019

Nakahata, Y., Yoshida, M., Takano, A., Soma, H., Yamamoto, T., Yasuda, A., Nakatsu, T., & Takumi, T. (2008). A direct repeat of E-box-like elements is required for cell-autonomous circadian rhythm of clock genes. *BMC Molecular Biology*, *9*(1), 1–11. https://doi.org/10.1186/1471-2199-9-1

Nygaard, V., Rødland, E. A., & Hovig, E. (2016). Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics*, *17*(1), 29–39. https://doi.org/10.1093/biostatistics/kxv027

O'Neill, J. S., Maywood, E. S., Chesham, J. E., Takahashi, J. S., & Hastings, M. H. (2008). cAMP-dependent signaling as a core component of the mammalian circadian pacemaker. *Science*, *320*(5878), 949–953. https://doi.org/10.1126/science.1152506

Oda, A., Katayose, Y., Yabuuchi, S., Yamamoto, K., Mizuma, M., Shirasou, S., Onogawa, T., Ohtsuka, H., Yoshida, H., Hayashi, H., Rikiyama, T., Kim, H., Choe, Y., Kim, K., Sun, H., Motoi, F., Egawa, S., & Unno, M. (2009). Clock gene mouse period2 overexpression inhibits growth of human pancreatic cancer cells and has synergistic effect with cisplatin. *Anticancer Research*, *29*, 1201–1210.

Okamoto-Uchida, Y., Izawa, J., Nishimura, A., Hattori, A., Suzuki, N., & Hirayama, J. (2019). Post-translational Modifications are Required for Circadian Clock Regulation in Vertebrates. *Current Genomics*, *20*, 332–339. https://doi.org/10.2174/1389202919666191014094349

Oshima, T., Takenoshita, S., Akaike, M., Kunisaki, C., Fujii, S., Nozaki, A., Numata, K., Shiozawa, M., Rino, Y., Tanaka, K., Masuda, M., & Imada, T. (2011). Expression of circadian genes correlates with liver metastasis and outcomes in colorectal cancer. *Oncology Reports*, *25*, 1439–1446. https://doi.org/10.3892/or.2011.1207

Oster, H., Damerow, S., Kiessling, S., Jakubcakova, V., Abraham, D., Tian, J., Hoffmann, M. W., & Eichele, G. (2006). The circadian rhythm of glucocorticoids is regulated by a gating mechanism residing in the adrenal cortical clock. *Cell Metabolism*, *4*(8), 163–173. https://doi.org/10.1016/j.cmet.2006.07.002

Palmer, T. D., Rosman, G. J., Osbornet, W. R. A., & Miller, A. D. (1991). Genetically modified skin fibroblasts persist long after transplantation but gradually inactivate introduced genes. *Proceedings of the National Academy of Sciences USA*, *88*, 1330–1334.

Papagiannakopoulos, T., Bauer, M. R., Davidson, S. M., Heimann, M., Subbaraj, L., Bhutkar, A., Bartlebaugh, J., Vander Heiden, M. G., & Jacks, T. (2016). Circadian Rhythm Disruption Promotes Lung Tumorigenesis. *Cell Metabolism*, *24*(2), 324–331. https://doi.org/10.1016/j.cmet.2016.07.001

Papp Correia, S., Chan, A. B., Vaughan, M., Zolboot, N., Perea, V., Huber, A. L., Kriebs, A., Moresco, J. J., Yates, J. R., & Lamia, K. A. (2019). The circadian E3 ligase complex SCFFBXL3+CRY targets TLK2. *Scientific Reports*, *9*, 198–207. https://doi.org/10.1038/s41598-018-36618-3

Park, J., Zhu, H., O'Sullivan, S., Ogunnaike, B. A., Weaver, D. R., Schwaber, J. S., & Vadigepalli, R. (2016). Single-cell transcriptional analysis reveals novel neuronal phenotypes and interaction networks involved in the central circadian clock. *Frontiers in Neuroscience*, *10*(October), 481. https://doi.org/10.3389/fnins.2016.00481

Paulose, J. K. (2016). *Toward the Beginning of Time: Circadian Rhythms in Development* [Texas A&M University]. https://oaktrust.library.tamu.edu/bitstream/handle/1969.1/156931/PAULOSE-DISSERTATION-2016.pdf?sequence=1

Pawlikowska, P., Tayoun, T., Oulhen, M., Faugeroux, V., Rouffiac, V., Aberlenc, A., Pommier, A. L., Honore, A., Marty, V., Bawa, O., Lacroix, L., Scoazec, J. Y., Chauchereau, A., Laplace-Builhe, C., & Farace, F. (2020). Exploitation of the chick embryo chorioallantoic membrane (CAM) as a platform for anti-metastatic drug testing. *Scientific Reports*, *10*, 16876. https://doi.org/10.1038/s41598-020-73632-w

Pazin, M. J., & Kadonaga, J. T. (1997). What's up and down with histone deacetylation and transcription? *Cell*, *89*(3), 325–328. https://doi.org/10.1016/S0092-8674(00)80211-1

Pelikan, A., Herzel, H., Kramer, A., & Ananthasubramaniam, B. (2020). Studies overestimate the extent of circadian rhythm reprogramming in response to dietary and genetic changes. *BioRxiv*, *December*, 1–26. https://doi.org/https://doi.org/10.1101/2020.12.18.423465

Perrin, L., Loizides-Mangold, U., Chanon, S., Gobet, C., Hulo, N., Isenegger, L., Weger, B. D., Migliavacca, E., Charpagne, A., Betts, J. A., Walhin, J. P., Templeman, I., Stokes, K., Thompson, D., Tsintzas, K., Robert, M., Howald, C., Riezman, H., Feige, J. N., ... Dibner, C. (2018). Transcriptomic analyses reveal rhythmic and CLOCK-driven pathways in human skeletal muscle. *ELife*, *7*, e34114. https://doi.org/10.7554/eLife.34114

Pfaffl, M. (2006). Relative quantification. In M. Tevfik Dorak (Ed.), *Real-time PCR* (1st ed., pp. 63–82). Taylor & Francis Group. https://www.gene-quantification.de/dorak-book-real-time-pcr-

2006.pdf

Phillips, N. E., Hugues, A., Yeung, J., Durandau, E., Nicolas, D., & Naef, F. (2021). The circadian oscillator analysed at the single-transcript level. *Molecular Systems Biology*, *17*(3), e10135. https://doi.org/10.15252/msb.202010135

Pineda Torra, I., Tsibulsky, V., Delaunay, F., Saladin, R., Laudet, V., Fruchart, J.-C., Kosykh, V., & Staels, B. (2000). Circadian and Glucocorticoid Regulation of Rev-erba Expression in Liver. *Endocrinology*, *141*(10), 3799–3806. https://academic.oup.com/endo/article/141/10/3799/2987804

Preitner, N., Damiola, F., Luis-Lopez-Molina, Zakany, J., Duboule, D., Albrecht, U., & Schibler, U. (2002). The orphan nuclear receptor REV-ERBα controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell*, *110*(2), 251–260. https://doi.org/10.1016/S0092-8674(02)00825-5

Prosser, R. A., & Gillette, M. U. (1989). The Mammalian Circadian Clock in the Suprachiasmatic Nuclei Is Reset in vitro by CAMP. *The Journal of Neuroscience*, *9*(3), 1073–1081.

Puram, R. V, Kowalczyk, M. S., De Boer, C. G., Schneider, R. K., Miller, P. G., Mcconkey, M., Tothova, Z., Tejero, H., Heckl, D., Järås, M., Chen, M. C., Li, H., Tamayo, A., Cowley, G. S., Rozenblatt-Rosen, O., Al-Shahrour, F., Regev, A., & Ebert, B. L. (2016). Core circadian clock genes regulate leukemia stem cells in AML. *Cell*, *165*(2), 303–316. https://doi.org/10.1016/j.cell.2016.03.015

Qin, T., Lu, X. T., Li, Y. G., Liu, Y., Yan, W., Li, N., & Sun, Y. Y. (2018). Effect of period 2 on the proliferation, apoptosis and migration of osteosarcoma cells, and the corresponding mechanisms. *Oncology Letters*, *16*(2), 2668–2674. https://doi.org/10.3892/ol.2018.8952

Qu, Y., Han, B., Yu, Y., Yao, W., Bose, S., Karlan, B. Y., Giuliano, A. E., & Cui, X. (2015). Evaluation of MCF10A as a Reliable Model for Normal Human Mammary Epithelial Cells. *PLOS ONE*, *10*(7), e0131285. https://doi.org/10.1371/journal.pone.0131285

Ramanathan, C., Khan, S. K., Kathale, N. D., Xu, H., & Liu, A. C. (2012). Monitoring cell-autonomous circadian clock rhythms of gene expression using luciferase bioluminescence reporters. *Journal of Visualized Experiments*, *67*, e4234-4243. https://doi.org/10.3791/4234

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, *42*, 187–191. https://doi.org/10.1093/nar/gku365

Ramos, C. A., Ouyang, C., Qi, Y., Chung, Y., Cheng, C. T., LaBarge, M. A., Seewaldt, V. L., & Ann, D. K. (2020). A Non-canonical Function of BMAL1 Metabolically Limits Obesity-Promoted Triple-Negative Breast Cancer. *IScience*, *23*(2). https://doi.org/10.1016/j.isci.2020.100839

Rao, M. S., Van Vleet, T. R., Ciurlionis, R., Buck, W. R., Mittelstadt, S. W., Blomme, E. A. G., & Liguori, M. J. (2019). Comparison of RNA-Seq and microarray gene expression platforms for the toxicogenomic evaluation of liver from short-term rat toxicity studies. *Frontiers in Genetics*, *10*, 636. https://doi.org/10.3389/fgene.2018.00636

Reick, M., Garcia, J. A., Dudley, C., & McKnight, S. L. (2001). NPAS2: An analog of clock operative in

the mammalian forebrain. *Science*, *293*(5529), 506–509. https://doi.org/10.1126/science.1060699

Relles, D., Sendecki, J., Chipitsyna, G., Hyslop, T., Yeo, C. J., & Arafat, H. A. (2013). Circadian Gene Expression and Clinicopathologic Correlates in Pancreatic Cancer. *Journal of Gastrointestinal Surgery*, *17*, 443–450. https://doi.org/10.1007/s11605-012-2112-2

Reske, J. J., Wilson, M. R., & Chandler, R. L. (2020). ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics and Chromatin*, *13*(22), 1–17. https://doi.org/10.1186/s13072-020-00342-y

Ribatti, D. (2016). The chick embryo chorioallantoic membrane (CAM). A multifaceted experimental model. *Mechanisms of Development*, *141*, 70–77. https://doi.org/10.1016/j.mod.2016.05.003

Riedl, A., Schlederer, M., Pudelko, K., Stadler, M., Walter, S., Unterleuthner, D., Unger, C., Kramer, N., Hengstschläger, M., Kenner, L., Pfeiffer, D., Krupitza, G., & Dolznig, H. (2017). Comparison of cancer cells in 2D vs 3D culture reveals differences in AKT-mTOR-S6K signaling and drug responses. *Journal of Cell Science*, *130*, 203–218. https://doi.org/10.1242/jcs.188102

Ripperger, J. A., & Schibler, U. (2006). Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions. *Nature Genetics*, *38*(3), 369–374. https://doi.org/10.1038/ng1738

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Robles, M. S., Humphrey, S. J., & Mann, M. (2017). Phosphorylation Is a Central Mechanism for Circadian Control of Metabolism and Physiology. *Cell Metabolism*, *25*(1), 118–127. https://doi.org/10.1016/j.cmet.2016.10.004

Rosbash, M. (2009). The Implications of Multiple Circadian Clock Origins. *PLoS Biology*, *7*(3), e1000062. https://doi.org/10.1371/journal.pbio.1000062

Rosenfeldt, M. T., Bell, L. A., Long, J. S., O'Prey, J., Nixon, C., Roberts, F., Dufès, C., & Ryan, K. M. (2014). E2F1 drives chemotherapeutic drug resistance via ABCG2. *Oncogene*, *33*, 4164–4172. https://doi.org/10.1038/onc.2013.470

Rossetti, S., Esposito, J., Corlazzoli, F., Gregorski, A., & Sacchi, N. (2012). Entrainment of breast (cancer) epithelial cells detects distinct circadian oscillation patterns for clock and hormone receptor genes. *Cell Cycle*, *11*(2), 350–360. https://doi.org/10.4161/cc.11.2.18792

Rovithi, M., Avan, A., Funel, N., Leon, L. G., Gomez, V. E., Wurdinger, T., Griffioen, A. W., Verheul, H. M. W., & Giovannetti, E. (2017). Development of bioluminescent chick chorioallantoic membrane (CAM) models for primary pancreatic cancer cells: A platform for drug testing. *Scientific Reports*, *7*(44686), 1–13. https://doi.org/10.1038/srep44686

Ruben, M. D., Wu, G., Smith, D. F., Schmidt, R. E., Francey, L. J., Lee, Y. Y., Anafi, R. C., &

Hogenesch, J. B. (2018). A database of tissue-specific rhythmically expressed human genes has potential applications in circadian medicine. *Science Translational Medicine*, *10*, 8806–8813. https://doi.org/10.1126/scitranslmed.aat8806

Ruis, J. F., Talamini, L. M., Buys, J. P., & Rietveld, W. J. (1989). Effects of Time of Feeding on Recovery of Food-Entrained Rhythms During Subsequent Fasting in SCN-Lesioned Rats. *Physiology & Behavior*, *46*, 857–866.

Ryan-Corces, M., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., Satpathy, A. T., Rubin, A. J., Montine, K. S., Wu, B., Kathiria, A., Cho, S. W., Mumbach, M. R., Carter, A. C., Kasowski, M., Orloff, L. A., Risca, V. I., Kundaje, A., Khavari, P. A., ... Chang, H. Y. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods*, *14*(10), 959–962. https://doi.org/10.1038/nmeth.4396

Saini, C., Morf, J., Stratmann, M., Gos, P., & Schibler, U. (2012). Simulated body temperature rhythms reveal the phase-shifting behavior and plasticity of mammalian circadian oscillators. *Genes and Development*, *26*, 567–580. https://doi.org/10.1101/gad.183251.111

Sato, F., Nagata, C., Liu, Y., Suzuki, T., Kondo, J., Morohashi, S., Imaizumi, T., Kato, Y., & Kijima, H. (2009). PERIOD1 is an anti-apoptotic factor in human pancreatic and hepatic cancer cells. *Journal of Biochemistry*, *146*(6), 833–838. https://doi.org/10.1093/jb/mvp126

Savard, J., Simard, S., Blanchet, J., Ivers, H., & Morin, C. M. (2001). Prevalence, clinical characteristics, and risk factors for insomnia in the context of breast cancer. *Sleep*, *24*(5), 583–590. https://doi.org/10.1093/sleep/24.5.583

Saxena, M., Stephens, M. A., Pathak, H., & Rangarajan, A. (2011). Transcription factors that mediate epithelial-mesenchymal transition lead to multidrug resistance by upregulating ABC transporters. *Cell Death and Disease*, *2*, e179-13. https://doi.org/10.1038/cddis.2011.61

Scheiermann, C., Gibbs, J., Ince, L., & Loudon, A. (2018). Clocking in to immunity. *Nature Reviews Immunology*, *18*(7), 423–437. https://doi.org/10.1038/s41577-018-0008-4

Schep, A. N., Buenrostro, J. D., Denny, S. K., Schwartz, K., Sherlock, G., & Greenleaf, W. J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Research*, *25*, 1757–1770. https://doi.org/10.1101/gr.192294.115

Schernhammer, E. S., Laden, F., Speizer, F. E., Willett, W. C., Hunter, D. J., Kawachi, I., & 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*(20), 1563–1568. https://academic.oup.com/jnci/article/93/20/1563/2519563

Schmitt, K., Grimm, A., Dallmann, R., Oettinghaus, B., Restelli, L. M., Witzig, M., Ishihara, N., Mihara, K., Ripperger, J. A., Albrecht, U., Frank, S., Brown, S. A., & Eckert, A. (2018). Circadian Control of DRP1 Activity Regulates Mitochondrial Dynamics and Bioenergetics. *Cell Metabolism*, *27*(3), 657-666.e5. https://doi.org/10.1016/j.cmet.2018.01.011

Schwartz, W. J., & Zimmerman, P. (1990). Circadian Timekeeping in BALB/c and C57BL/6 Inbred Mouse Strains. *Journal of Neuroscience*, *10*(11), 3665–3694.

Shen, W., Zhang, W., Ye, W., Wang, H., Zhang, Q., Shen, J., Hong, Q., Li, X., Wen, G., Wei, T., & Zhang, J. (2020). SR9009 induces a REV-ERB dependent anti-small-cell lung cancer effect through inhibition of autophagy. *Theranostics*, *10*(10), 4466–4480. https://doi.org/10.7150/thno.42478

Shi, C., Killoran, M. P., Hall, M. P., Otto, P., Wood, M. G., Strauss, E., Encell, L. P., Machleidt, T., Wood, K. V., & Kirkland, T. A. (2020). 5,5-Dialkylluciferins are thermal stable substrates for bioluminescence-based detection systems. *PLOS ONE*, *15*(12), e0243747. https://doi.org/10.1371/journal.pone.0243747

Shilts, J., Chen, G., & Hughey, J. J. (2018). Evidence for widespread dysregulation of circadian clock progression in human cancer. *PeerJ*, *6*, e4327. https://doi.org/10.7717/peerj.4327

Shim, H. S., Kim, H., Lee, J., Son, G. H., Cho, S., Oh, T. H., Kang, S. H., Seen, D. S., Lee, K. H., & Kim, K. (2007). Rapid activation of CLOCK by Ca2+-dependent protein kinase C mediates resetting of the mammalian circadian clock. *EMBO Reports*, *8*(4), 366–371. https://doi.org/10.1038/sj.embor.7400920

So, A. Y.-L., Bernal, T. U., Pillsbury, M. L., Yamamoto, K. R., & Feldman, B. J. (2009). Glucocorticoid regulation of the circadian clock modulates glucose homeostasis. *Proceedings of the National Academy of Sciences*, *106*(41), 17582–17587. www.pnas.orgcgidoi10.1073pnas.0909733106

Soule, H D, Vazquez, J., Long, A., Albert, S., & Brennan, M. (1973). A Human Cell Line From a Pleural Effusion Derived From a Breast Carcinoma. *Journal of the National Cancer Institute*, *51*(5), 1409–1416.

Soule, Herbert D., Maloney, T. M., Wolman, S. R., Brenz, R., Russo, J., Pauley, R. J., Jones, R. F., Brooks, S. C., & McGrath, C. M. (1990). Isolation and Characterization of a Spontaneously Immortalized Human Breast Epithelial Cell Line, MCF-10. *Cancer Research*, *50*(18), 6075–6086.

Spörl, F., Schellenberg, K., Blatt, T., Wenck, H., Wittern, K.-P., Schrader, A., & Kramer, A. (2011). A Circadian Clock in HaCaT Keratinocytes. *Journal of Investigative Dermatology*, *131*, 338–348. https://doi.org/10.1038/jid.2010.315

Stavreva, D. A., Wiench, M., John, S., Conway-Campbell, B. L., McKenna, M. A., Pooley, J. R., Johnson, T. A., Voss, T. C., Lightman, S. L., & Hager, G. L. (2009). Ultradian hormone stimulation induces glucocorticoid receptor-mediated pulses of gene transcription. *Nature Cell Biology*, *11*(9), 1093–1102. https://doi.org/10.1038/ncb1922

Stenvers, D. J., Jongejan, A., Atiqi, S., Vreijling, J. P., Limonard, E. J., Endert, E., Baas, F., Moerland, P. D., Fliers, E., Kalsbeek, A., & Bisschop, P. H. (2019). Diurnal rhythms in the white adipose tissue transcriptome are disturbed in obese individuals with type 2 diabetes compared with lean control individuals. *Diabetologia*, *62*, 704–716. https://doi.org/10.1007/s00125-019-4813-5

Stenvers, D. J., Scheer, F. A. J. L., Schrauwen, P., la Fleur, S. E., & Kalsbeek, A. (2019). Circadian clocks and insulin resistance. *Nature Reviews Endocrinology*, *15*, 75–89. https://doi.org/10.1038/s41574-018-0122-1

Stephenson, E. M., Usselmann, L. E. J., Tergaonkar, V., Virshup, D. M., & Dallmann, R. (2021). Cancer clocks in tumourigenesis: The p53 pathway and beyond. *Endocrine-Related Cancer*, *28*(4), R95–R110. https://doi.org/10.1530/ERC-20-0475

Stevens, R. G. (1987). Electric power use and breast cancer: a hypothesis. *American Journal of Epidemiology*, *125*(4), 556–561. https://doi.org/10.1093/oxfordjournals.aje.a114569

Stokkan, K. A., Yamazaki, S., Tei, H., Sakaki, Y., & Menaker, M. (2001). Entrainment of the circadian clock in the liver by feeding. *Science*, *291*(5503), 490–493. https://doi.org/10.1126/science.291.5503.490

Sulli, G., Lam, M. T. Y., & Panda, S. (2019). Interplay between Circadian Clock and Cancer: New Frontiers for Cancer Treatment. *Trends in Cancer*, *5*(8), 475–494. https://doi.org/10.1016/j.trecan.2019.07.002

Sulli, G., Rommel, A., Wang, X., Kolar, M. J., Puca, F., Saghatelian, A., Plikus, M. V., Verma, I. M., & Panda, S. (2018). Pharmacological activation of REV-ERBs is lethal in cancer and oncogene-induced senescence. *Nature*, *553*(7688), 351–355. https://doi.org/10.1038/nature25170

Sun, C. M., Huang, S. F., Zeng, J. M., Liu, D. B., Xiao, Q., Tian, W. J., Zhu, X. D., Huang, Z. G., & Feng, W. L. (2010). Per2 inhibits K562 leukemia cell growth in vitro and in vivo through cell cycle arrest and apoptosis induction. *Pathology and Oncology Research*, *16*(3), 403–411. https://doi.org/10.1007/s12253-009-9227-0

Sung, Y. K., & Kim, S. W. (2020). Recent advances in polymeric drug delivery systems. *Biomaterials Research*, *24*(12), 1–12. https://doi.org/10.1186/s40824-020-00190-7

Surjit, M., Ganti, K. P., Mukherji, A., Ye, T., Hua, G., Metzger, D., Li, M., & Chambon, P. (2011). Widespread Negative Response Elements Mediate Direct Repression by Agonist-Liganded Glucocorticoid Receptor. *Cell*, *145*(4), 224–241. https://doi.org/10.1016/j.cell.2011.03.027

Takahashi, J. S. (2017). Transcriptional architecture of the mammalian circadian clock. *Nature Reviews Genetics*, *18*(3), 164–179. https://doi.org/10.1093/nar/gks1161

Takeda, J., Adachi, K., Halprin, K. M., Levine, V., & Woodyard, C. (1983). Forskolin Activates Adenylate Cyclase Activity and Inhibits Mitosis in In Vitro in Pig Epidermis Experiments with Epidermal Slices. *Journal of Investigative Dermatology*, *81*(3), 236–240. https://doi.org/10.1111/1523-1747.ep12518219

Talamanca, L., & Naef, F. (2020). How to tell time: Advances in decoding circadian phase from omics snapshots. *F1000 Research*, *9*(1150), 1–8. https://doi.org/10.12688/f1000research.26759.1

Tanoue, S., Fujimoto, K., Myung, J., Hatanaka, F., Kato, Y., & Takumi, T. (2015). DEC2–E4BP4 Heterodimer Represses the Transcriptional Enhancer Activity of the EE Element in the Per2 Promoter. *Frontiers in Neurology*, *6*(JUL), 166. https://doi.org/10.3389/fneur.2015.00166

Tolmachov, O. E., Subkhankulova, T., & Tolmachova, T. (2013). Silencing of Transgene Expression: A Gene Therapy Perspective. In *Gene Therapy - Tools and Potential Applications* (pp. 49–68). InTechOpen. https://doi.org/10.5772/53379

Tombs, E. L., Nikolaou, V., Nurumbetov, G., & Haddleton, D. M. (2018). Transdermal Delivery of

Ibuprofen Utilizing a Novel Solvent-Free Pressure-sensitive Adhesive (PSA): TEPI® Technology. *Journal of Pharmaceutical Innovation*, *13*, 48–57. https://doi.org/10.1007/s12247-017-9305-x

Tong, H., Liu, X., Li, T., Qiu, W., Peng, C., Shen, B., & Zhu, Z. (2020). Nr1d2 accelerates hepatocellular carcinoma progression by driving the epithelial-to-mesenchymal transition. *OncoTargets and Therapy*, *13*, 3931–3942. https://doi.org/10.2147/OTT.S237804

Tosini, G., Baba, K., Hwang, C. K., & Iuvone, P. M. (2012). Melatonin: An underappreciated player in retinal physiology and pathophysiology. *Experimental Eye Research*, *103*, 82–89. https://doi.org/10.1016/j.exer.2012.08.009

Trávníčková, Z., & Illnerová, H. (1997). Melatonin entrainment of the circadian N-acetyltransferase rhythm in the newborn rat pineal gland. *Journal of Pineal Research*, *23*(3), 136–141. https://doi.org/10.1111/j.1600-079X.1997.tb00346.x

Ueda, H. R., Chen, W., Minami, Y., Honma, S., Honma, K., Iino, M., & Hashimoto, S. (2004). Molecular-timetable methods for detection of body time and rhythm disorders from single-time-point genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(31), 11227–11232. https://doi.org/10.1073/pnas.0401882101

Animals (Scientific Procedures) Act 1986, GOV.UK 1 (2014). https://doi.org/ISBN 1474100287-9781474100281

Vallabha, G. (2021). *PLOT_GAUSSIAN_ELLIPSOID*. MATLAB Central File Exchange. https://www.mathworks.com/matlabcentral/fileexchange/16543-plot_gaussian_ellipsoid

van der Horst, G. T. J., Muijtjens, M., Kobayashi, K., Takano, R., Kanno, S. I., Takao, M., De Wit, J., Verkerk, A., Eker, A. P. M., Van Leenen, D., Buijs, R., Bootsma, D., Hoeijmakers, J. H. J., & Yasui, A. (1999). Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature*, *398*(6728), 627–630. https://doi.org/10.1038/19323

Vantangoli, M. M., Madnick, S. J., Huse, S. M., Weston, P., & Boekelheide, K. (2015). MCF-7 Human Breast Cancer Cells Form Differentiated Microtissues in Scaffold-Free Hydrogels. *PLOS ONE*, *10*(8), e0135426. https://doi.org/10.1371/journal.pone.0135426

Varcoe, T. J., Boden, M. J., Voultsios, A., Salkeld, M. D., Rattanatray, L., & Kennaway, D. J. (2013). Characterisation of the Maternal Response to Chronic Phase Shifts during Gestation in the Rat: Implications for Fetal Metabolic Programming. *PLoS ONE*, *8*(1), e53800. https://doi.org/10.1371/journal.pone.0053800

Vilcacundo, R., Méndez, P., Reyes, W., Romero, H., Pinto, A., & Carrillo, W. (2018). Antibacterial activity of hen egg white lysozyme denatured by thermal and chemical treatments. *Scientia Pharmaceutica*, *86*(48), 1–17. https://doi.org/10.3390/scipharm86040048

Vitaterna, M. H., Takahashi, J. S., & Turek, F. W. (2001). Overview of circadian rhythms. *Alcohol Research & Health : The Journal of the National Institute on Alcohol Abuse and Alcoholism*, *25*(2), 85–93. http://www.ncbi.nlm.nih.gov/pubmed/11584554

Vlachou, D., Bjarnason, G. A., Giacchetti, S., Lévi, F., & Rand, D. A. (2020). TimeTeller: A New Tool for Precision Circadian Medicine and Cancer Prognosis. *BioRxiv*, *March*, 1–56.

https://doi.org/10.1101/622050

Wagner, P. M., Monjes, N. M., & Guido, M. E. (2019). Chemotherapeutic Effect of SR9009, a REV-ERB Agonist, on the Human Glioblastoma T98G Cells. *ASN Neuro*, *11*. https://doi.org/10.1177/1759091419892713

Wang, F., Li, C., Luo, Y., & Chen, L. (2016). The Circadian Gene Clock Plays an Important Role in Cell Apoptosis and the DNA Damage Response In Vitro. *Apoptosis*, *15*(3), 480–486. https://doi.org/10.1177/1533034615585433

Wang, J., Li, S., Li, X., Li, B., Li, Y., Xia, K., Yang, Y., Aman, S., Wang, M., & Wu, H. (2019). Circadian protein BMAL1 promotes breast cancer cell invasion and metastasis by up-regulating matrix metalloproteinase9 expression. *Cancer Cell International*, *19*(1), 182. https://doi.org/10.1186/s12935-019-0902-2

Wang, Yaping, Qian, R., Sun, N., Lu, C., Chen, Z., & Hua, L. (2015). Circadian gene hClock enhances proliferation and inhibits apoptosis of human colorectal carcinoma cells in vitro and in vivo. *Molecular Medicine Reports*, *11*, 4204–4210. https://doi.org/10.3892/mmr.2015.3247

Wang, Yaping, Sun, N., Lu, C., Bei, Y., Qian, R., & Hua, L. (2017). Upregulation of circadian gene "hClock" contribution to metastasis of colorectal cancer. *International Journal of Oncology*, *50*(6), 2191–2199. https://doi.org/10.3892/ijo.2017.3987

Wang, Yi lin, Jin, K., He, N. na, Cheng, S. ze, Zuo, Q. sheng, Li, D., Wang, Y. jie, Wang, F., Ji, Y. qing, Lu, Z. yu, Zhang, C., Wang, M., Zhao, R. feng, Yu, X. jian, Zhang, Y. ni, Zhao, W. ming, & Li, B. chun. (2017). Research on the appropriate way to transfer exogenous substances into chicken embryos. *Journal of Integrative Agriculture*, *16*(10), 2257–2263. https://doi.org/10.1016/S2095-3119(17)61668-X

Wang, Yongjun, Kojetin, D., & Burris, T. P. (2015). Anti-proliferative actions of a synthetic REV-ERBα/β agonist in breast cancer cells. *Biochemical Pharmacology*, *96*(4), 315–322. https://doi.org/10.1016/j.bcp.2015.06.010

Weger, B. D., Gobet, C., David, F. P. A., Atger, F., Martin, E., Phillips, N., Charpagne, A., Weger, M., Naef, F., & Gachon, F. (2021). Systematic analysis of differential rhythmic liver gene expression mediated by the circadian clock and feeding rhythms. *Proceedings of the National Academy of Sciences*, *118*(3), e2015803118. https://doi.org/10.1073/pnas.2015803118/-/DCSupplemental

Weger, B. D., Gobet, C., Yeung, J., Martin, E., Jimenez, S., Betrisey, B., Foata, F., Berger, B., Balvay, A., Foussier, A., Charpagne, A., Boizet-Bonhoure, B., Chou, C. J., Naef, F., & Gachon, F. (2019). The Mouse Microbiome Is Required for Sex-Specific Diurnal Rhythms of Gene Expression and Metabolism. *Cell Metabolism*, *29*(2), 362–382. https://doi.org/10.1016/j.cmet.2018.09.023

Weizman, E. N., Tannenbaum, M., Tarrant, A. M., Hakim, O., & Levy, O. (2019). Chromatin dynamics enable transcriptional rhythms in the cnidarian Nematostella vectensis. *PLOS Genetics*, *15*(11), e1008397. https://doi.org/10.1371/journal.pgen.1008397

Wen, S., Ma, D., Zhao, M., Xie, L., Wu, Q., Gou, L., Zhu, C., Fan, Y., Wang, H., & Yan, J. (2020). Spatiotemporal single-cell analysis of gene expression in the mouse suprachiasmatic nucleus. *Nature*

Neuroscience, *23*, 456–467. https://doi.org/10.1038/s41593-020-0586-x

Williams, J., Yang, N., Wood, A., Zindy, E., Meng, Q. J., & Streuli, C. H. (2018). Epithelial and stromal circadian clocks are inversely regulated by their mechano-matrix environment. *Journal of Cell Science*, *131*(5), jcs208223. https://doi.org/10.1242/jcs.208223

Wittenbrink, N., Ananthasubramaniam, B., Münch, M., Koller, B., Maier, B., Weschke, C., Bes, F., De Zeeuw, J., Nowozin, C., Wahnschaffe, A., Wisniewski, S., Zaleska, M., Bartok, O., Ashwal-Fluss, R., Lammert, H., Herzel, H., Hummel, M., Kadener, S., Kunz, D., & Kramer, A. (2018). High-accuracy determination of internal circadian time from a single blood sample. *Journal of Clinical Investigation*, *128*(9), 3826–3839. https://doi.org/10.1172/JCI120874

Wu, G., Anafi, R. C., Hughes, M. E., Kornacker, K., & Hogenesch, J. B. (2016). MetaCycle: An integrated R package to evaluate periodicity in large scale data. *Bioinformatics*, *32*(21), 3351–3353. https://doi.org/10.1093/bioinformatics/btw405

Wu, G., Ruben, M. D., Schmidt, R. E., Francey, L. J., Smith, D. F., Anafi, R. C., Hughey, J. J., Tasseff, R., Sherrill, J. D., Oblong, J. E., Mills, K. J., & Hogenesch, J. B. (2018). Population-level rhythms in human skin with implications for circadian medicine. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(48), 12313–12318. https://doi.org/10.1073/pnas.1809442115

Xiang, S., Mao, L., Duplessis, T., Yuan, L., Dauchy, R., Dauchy, E., Blask, D., Frasch, T., & Hhill, S. M. (2012). Oscillation of clock and clock controlled genes induced by serum shock in human breast epithelial and breast cancer cells: Regulation by melatonin. *Breast Cancer: Basic and Clinical Research*, *6*(1), 137–150. https://doi.org/10.4137/BCBCR.S9673

Xiao, L., Chang, A. K., Zang, M. X., Bi, H., Li, S., Wang, M., Xing, X., & Wu, H. (2014). Induction of the CLOCK gene by E2-ERα signaling promotes the proliferation of breast cancer cells. *PLoS ONE*, *9*(5), e95878-95891. https://doi.org/10.1371/journal.pone.0095878

Yagita, K., Horie, K., Koinuma, S., Nakamura, W., Yamanaka, I., Urasaki, A., Shigeyoshi, Y., Kawakami, K., Shimada, S., Takeda, J., & Uchiyama, Y. (2010). Development of the circadian oscillator during differentiation of mouse embryonic stem cells in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3846–3851. https://doi.org/10.1073/pnas.0913256107

Yagita, K., & Okamura, H. (2000). Forskolin induces circadian gene expression of rPer1, rPer2 and dbp in. *FEBS Letters*, *465*, 79–82.

Yamaguchi, S., Mitsui, S., Yan, L., Yagita, K., Miyake, S., & Okamura, H. (2000). Role of DBP in the Circadian Oscillatory Mechanism. *Molecular and Cellular Biology*, *20*(13), 4773–4781. https://doi.org/https://doi.org/10.1128/mcb.20.13.4773-4781.2000

Yang, D., Ma, Z., & Buja, A. (2014). A Sparse Singular Value Decomposition Method for High-Dimensional Data. *Journal of Computational and Graphical Statistics*, *23*(4), 923–942. https://doi.org/10.1080/10618600.2013.858632

Yang, N., Williams, J., Pekovic-Vaughan, V., Wang, P., Olabi, S., McConnell, J., Gossan, N., Hughes,

A., Cheung, J., Streuli, C. H., & Meng, Q. J. (2017). Cellular mechano-environment regulates the mammary circadian clock. *Nature Communications*, *8*(14287), 1–13. https://doi.org/10.1038/ncomms14287

Yao, S., Sukonnik, T., Kean, T., Bharadwaj, R. R., Pasceri, P., & Ellis, J. (2004). Retrovirus silencing, variegation, extinction, and memory are controlled by a dynamic interplay of multiple epigenetic modifications. *Molecular Therapy*, *10*(1), 27–36. https://doi.org/10.1016/j.ymthe.2004.04.007

Ye, R., Selby, C. P., Chiou, Y. Y., Ozkan-Dagliyan, I., Gaddameedhi, S., & Sancar, A. (2014). Dual modes of CLOCK:BMAL1 inhibition mediated by Cryptochrome and period proteins in the mammalian circadian clock. *Genes and Development*, *28*(18), 1989–1998. https://doi.org/10.1101/gad.249417.114

Ye, Y., Xiang, Y., Ozguc, F. M., Kim, Y., Liu, C. J., Park, P. K., Hu, Q., Diao, L., Lou, Y., Lin, C., Guo, A. Y., Zhou, B., Wang, L., Chen, Z., Takahashi, J. S., Mills, G. B., Yoo, S. H., & Han, L. (2018). The Genomic Landscape and Pharmacogenomic Interactions of Clock Genes in Cancer Chronotherapy. *Cell Systems*, *6*(3), 314-328.e2. https://doi.org/10.1016/j.cels.2018.01.013

Yeung, J., Mermet, J., Jouffe, C., Marquis, J., Charpagne, A., Gachon, F., & Naef, F. (2018). Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs. *Genome Research*, *28*, 182–191. https://doi.org/10.1101/207787

Yin, L., & Lazar, M. A. (2005). The Orphan Nuclear Receptor Rev-erbα Recruits the N-CoR/Histone Deacetylase 3 Corepressor to Regulate the Circadian Bmal1 Gene. *Molecular Endocrinology*, *19*(6), 1452–1459. https://doi.org/10.1210/me.2005-0057

Yoshida, K., Nakai, A., Kaneshiro, K., Hashimoto, N., Suzuki, K., Uchida, K., Hashimoto, T., Kawasaki, Y., Tateishi, K., Nakagawa, N., Shibanuma, N., Sakai, Y., & Hashiramoto, A. (2018). TNF-α induces expression of the circadian clock gene Bmal1 via dual calcium-dependent pathways in rheumatoid synovial cells. *Biochemical and Biophysical Research Communications*, *495*(2), 1675–1680. https://doi.org/10.1016/j.bbrc.2017.12.015

Yu, H., Meng, X., Wu, J., Pan, C., Ying, X., Zhou, Y., Liu, R., & Huang, W. (2013). Cryptochrome 1 Overexpression Correlates with Tumor Progression and Poor Prognosis in Patients with Colorectal Cancer. *PLoS ONE*, *8*(4), e61679. https://doi.org/10.1371/journal.pone.0061679

Yu, Y., Li, Y., Zhou, L., Yang, G., Wang, M., & Hong, Y. (2018). Cryptochrome 2 (CRY2) suppresses proliferation and migration and regulates clock gene network in osteosarcoma cells. *Medical Science Monitor*, *24*, 3856–3862. https://doi.org/10.12659/MSM.908596

Zamir, I., Harding, H. P., Atkins, G. B., Hörlein, A., Glass, C. K., Rosenfeld, M. G., & Lazar, M. A. (1996). A nuclear hormone receptor corepressor mediates transcriptional silencing by receptors with distinct repression domains. *Molecular and Cellular Biology*, *16*(10), 5458–5465. https://doi.org/10.1128/mcb.16.10.5458

Zeng, Z. L., Luo, H. Y., Yang, J., Wu, W. J., Chen, D. L., Huang, P., & Xu, R. H. (2014). Overexpression of the circadian clock gene bmal1 increases sensitivity to oxaliplatin in colorectal cancer. *Clinical Cancer Research*, *20*(4), 1042–1052. https://doi.org/10.1158/1078-0432.CCR-13-0171

Zeng, Z. L., Wu, M.-W., Sun, J., Sun, Y.-L., Cai, Y.-C., Huang, Y.-J., & Xian, L.-J. (2010). Effects of the biological clock gene Bmal1 on tumour growth and anti-cancer drug activity. *Journal of Biochemistry*, *148*(3), 319–326. https://doi.org/10.1093/jb/mvq069

Zhang, F., Thornhill, S. I., Howe, S. J., Ulaganathan, M., Schambach, A., Sinclair, J., Kinnon, C., Gaspar, H. B., Antoniou, M., & Thrasher, A. J. (2007). Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells. *Blood*, *110*(5), 1448–1457. https://doi.org/10.1182/blood-2006-12-060814

Zhang, R., Lahens, N. F., Ballance, H. I., Hughes, M. E., & Hogenesch, J. B. (2014). A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), 16219–16224. https://doi.org/10.1073/pnas.1408886111

Zhang, Yimao, Bressler, J. P., Neal, J., Lal, B., Bhang, H. E. C., Laterra, J., & Pomper, M. G. (2007). ABCG2/BCRP expression modulates D-luciferin-based bioluminescence imaging. *Cancer Research*, *67*(19), 9389–9397. https://doi.org/10.1158/0008-5472.CAN-07-0944

Zhang, Yuan, Devocelle, A., Souza, L., Foudi, A., Bento, S. T., Desterke, C., Sherrard, R., Ballesta, A., Adam, R., Giron-Michel, J., & Chang, Y. (2020). BMAL1 knockdown triggers different colon carcinoma cell fates by altering the delicate equilibrium between AKT/mTOR and P53/P21 pathways. *Aging*, *12*(9), 8067–8083. https://doi.org/10.18632/aging.103124

Zhang, Yuan, Giacchetti, S., Parouchev, A., Hadadi, E., Li, X., Dallmann, R., Xandri-Monje, H., Portier, L., Adam, R., Lévi, F., Dulong, S., & Chang, Y. (2018). Dosing time dependent in vitro pharmacodynamics of Everolimus despite a defective circadian clock. *Cell Cycle*, *17*(1), 33–42. https://doi.org/10.1080/15384101.2017.1387695

Zhou, L., Yu, Y., Sun, S., Zhang, T., & Wang, M. (2018). Cry 1 regulates the clock gene network and promotes proliferation and migration via the Akt/P53/P21 pathway in human osteosarcoma cells. *Journal of Cancer*, *9*, 2480–2491. https://doi.org/10.7150/jca.25213

Zhou, Y., Zhang, H. K., Liu, F., Lei, G., Liu, P., Jiao, T., & Dang, Y.-H. (2018). Altered light conditions contribute to abnormalities in emotion and cognition through HINT1 dysfunction in C57BL/6 Mice. *Frontiers in Behavioral Neuroscience*, *12*(110), 1–10. https://doi.org/10.3389/fnbeh.2018.00110