Doctor of Philosophy

TRANSCRIPTOME DYNAMICS OF THE SUPRACHIASMATIC NUCLEUS AND ASSOCIATIONS WITH PSYCHIATRIC DISEASE

February 19, 2016



William Pembroke

Department of Physiology, Anatomy and Genetics

Abstract

The mammalian suprachiasmatic nucleus (SCN) drives daily rhythmic behaviour and physiology, yet a comprehensive understanding of its coordinated transcriptional programmes is lacking. Furthermore, there has been a long-standing association between circadian disruption and psychiatric disease, whose mechanism underpinnings are unknown. To analyse in detail the circadian transcriptome of the mammalian SCN and its variation in psychiatric disease, whole transcriptome sequencing (RNA-seq) of the mouse SCN was performed over a 24-hour light / dark cycle in both wild-type and blind-drunk *(Bdr)*, a mutant strain for *Snap25* that shows both schizophrenic endophenotypes and abnormal circadian behaviours.

2,648 novel intergenic long non-coding RNAs were identified using the SCN RNA-seq data, together with novel exons of annotated protein-coding genes, including the clock gene *Cry1*. Furthermore, in addition to identifying 4,569 genes which exhibited a classic sinusoidal expression signature, co-expression network analysis identified a group of 766 genes which unexpectedly peaked twice, near to both the start and end of the dark phase; this twin-peaking group is significantly enriched for synaptic transmission genes that are crucial for light-induced phase-shifting of the circadian clock. This twin-peaking module is also dysregulated in *Bdr*, which may provide clues to the molecular pathways linking circadian disruption with psychiatric disease.

These findings offer novel insight into SCN biology and suggest that transcriptional timing in the SCN contributes to gating clock resetting mechanisms which, when disrupted, may play a role in psychiatric disease.

Acknowledgements

I would firstly like to express gratitude to my supervisors Chris Ponting and Kay Davies, and also to Pete Oliver, for providing me the opportunity to work on an exciting project and for passing down their extensive knowledge and experience. I feel I have learnt many skills and hope to use these on my path to become a better scientist. I would like to thank all the post-docs and PhD students who have offered academic help in the last three years. I know that without you my progress would have been limited. I am grateful to have shared some special memories with many of you. I hope we stay in touch. I am also thankful to those outside the PhD bubble, including those from my hometown Richings Park, my many friends from Oxford – old and new – and lastly to Lucy Warnock. Thank you all for your understanding whilst I have been largely occupied and for offering an escape from the academic life. Lastly, I would like to thank my family for their continued support and enabling me to grow into the person I am today.

Contents

1 Introduction **1**

1.1 Circadian rhythms 1

1.1.1 The SCN as a master pacemaker 1

1.1.2 The endogenous molecular clock 5

1.1.3 Circadian oscillation in SCN excitability 8

1.1.4 Resetting the clock 11

1.1.5 Circadian disruption and disease 17

1.2 Transcriptomics 33

1.2.1 What is a transcriptome? 33

1.2.2 Assessing the transcriptome 37

1.2.3 Circadian transcriptomics 43

1.3 Basis for investigation 47

1.4 Aims of this project 51

1.5 Overview of thesis 53

2 Generation and validation of the mouse SCN transcriptome over 24 hours **55**

2.1 Introduction 55

2.2 Methods 58

2.2.1 Animals 58

2.2.2 SCN RNA sample preparation 58

2.2.3 Sequencing 59

2.2.4 Acquisition of additional RNA-seq data sets 60
2.2.5 Read processing 61
2.2.6 Mapping 65
2.2.7 Duplicate removal 66
2.2.8 Expression quantification 67
2.2.9 Conditional quantile normalisation 69
2.2.10 Identification of SCN-enriched genes 70
2.2.11 Gene-annotation enrichment analysis 71
2.2.12 qRT-PCR 71
2.2.13 *In situ* hybridisation 72
2.3 Results 73
2.3.1 Generation of the mouse SCN transcriptome 73
2.3.2 Removal of batch effects 81
2.3.3 Temporal analysis of positive controls 86
2.3.4 Validation of RNA-seq data 87
2.3.5 Assessing SCN cellular subtypes and identification of SCN-enriched genes 94
2.4 Discussion 105
2.4.1 Considerations for experimental design 105
2.4.2 Collating publically available sequencing data to gain biological insight 110
2.4.3 Conclusion 113

3 Defining the circadian transcriptome of the mouse SCN 114

3.1 Introduction . . . 114

3.2 Methods . . . 121

3.2.1 Identification of sinusoidal fluctuating genes . . . 121

3.2.2 Identification of non-sinusoidal fluctuating genes . . . 123

3.2.3 Generation of SCNseq data explorer . . . 132

3.3 Results . . . 134

3.3.1 Defining the circadian transcriptome . . . 134

3.3.2 Identification of sinusoidally cycling genes and processes . . . 134

3.3.3 Comparison with other circadian transcriptomes . . . 137

3.3.4 Identification of novel twin-peaking synaptic module . . . 141

3.4 Discussion . . . 146

3.4.1 The power of co-expression and unbiased approaches in identifying circadian regulated processes . . . 146

3.4.2 Transcriptional regulation of asymmetrical twin peaks . . . 153

3.4.3 Twin-peaking expression — role in clock resetting pathways . . . 158

3.4.4 Light-induced translation . . . 162

3.4.5 Integration of sinusoidal and twin peaking genes to form differential responses . . . 166

3.4.6 Circadian oscillation of transcription, redox state and neuronal excitability . . . 169

3.4.7 Importance of web-based application . . . 173

4 Identification of novel transcripts in the wild-type mouse SCN 175

4.1 Introduction . . . 175

4.2 Methods . . . 181

4.2.1 Identification of novel exons . . . 181

4.2.2 Identification of lincRNA . . . 182

4.2.3 Comparison with other lincRNA annotations . . . 183

4.2.4 LincRNA sequence conservation . . . 183

4.2.5 Identifying syntenic human lincRNA . . . 184

4.2.6 Expression and fluctuation . . . 185

4.2.7 Expression correlation . . . 185

4.2.8 Functional annotation of non-coding regions . . . 185

4.2.9 Validation of transcripts . . . 186

4.3 Results . . . 187

4.3.1 Identification of novel and cycling SCN lincRNA . . . 187

4.3.2 Identification of novel exons, including of the clock gene *Cry1* . . . 200

4.4 Discussion . . . 207

4.4.1 Identification of novel transcripts in the SCN . . . 207

4.4.2 Assessing function of lincRNA in the SCN . . . 208

4.4.3 Regulation and function of novel *Cry1* isoform . . . 213

5 Deciphering the relationship between circadian behaviour and expression in *Bdr* 219

5.1 Introduction . . . 219

5.2 Methods . . . 227
5.2.1 Differential expression analysis . . . 227
5.2.2 Identifying dysregulated co-expression modules . . . 232
5.2.3 Disease enrichment . . . 232
5.2.4 Module preservation . . . 233
5.2.5 Activity monitoring . . . 233
5.2.6 qRT-PCR . . . 237
5.3 Results . . . 238
5.3.1 Data exploration . . . 238
5.3.2 Identification of differentially expressed genes . . . 241
5.3.3 Systematic dysregulation of the twin peaking transcriptional profile 245
5.3.4 Comparison to previous transcriptomic analysis of *Bdr* . . . 251
5.3.5 Module preservation . . . 253
5.3.6 Disease enrichment of schizophrenia implicated genes . . . 255
5.3.7 Circadian behavioural wheel-running screens . . . 256
5.4 Discussion . . . 263
5.4.1 Identification of differentially expressed genes in *Bdr* . . . 263
5.4.2 Disruption of the twin-peaking co-expression module may affect circadian function . . . 266
5.4.3 SNAP25 and synaptic function affect the timing of circadian expression . . . 269
5.4.4 Associations with schizophrenia . . . 273
5.4.5 Conclusion . . . 276

6 Conclusions and future perspectives **277**

6.1 Summary 277

6.1.1 Comprehensive examination into the mouse SCN transcriptome . . 277

6.1.2 Elucidation of the molecular links between psychiatric disease and circadian disruption 279

6.2 Discussion and future perspectives 281

6.2.1 Novel *Cry1* transcript represents novel mechanisms of circadian control 281

6.2.2 Analysing circadian time course expression data 285

6.2.3 High resolution sequencing of the SCN 288

6.2.4 Circadian isoform quantification 290

6.2.5 Transcriptional regulation of the twin-peaking transcriptional profile 291

6.2.6 Circadian-regulated translation 293

6.2.7 Pathways of plasticity in disease and circadian biology 294

6.3 Conclusion 298

7 References **299**

List of Figures

1.1 Immunostaining of coronal mouse SCN section. 3

1.2 Retinorecipient brain regions. 4

1.3 A schematic of the molecular circadian clock. 6

1.4 Molecular processes which drive electrical rhythms in the SCN. 10

1.5 The photic phase response curve. 12

1.6 The non-photic phase response curve. 13

1.7 *Snap25b* mutation in blind-drunk 30

1.8 Comparison of human and mouse activity profiles. 31

1.9 Functions of lncRNA 36

1.10 A typical RNA-seq protocol 40

1.11 Paired-end sequencing of fragments of varying insert sizes. 41

2.1 Laser capture microdissection of SCN. 59

2.2 RNA-seq experimental design. 60

2.3 Representative quality scores for sequencing data before alignment. 63

2.4 HT-seq read matching options. 68

2.5 PCA plot of WT SCN RNA-seq data points. 74

2.6 Technical biases of mapped samples between batches. 78

2.7 Assessing technical biases through PCA plots. 80

2.8 Fold change differences between timepoints pre- and post- CQN normalisation. 83

2.9 PCA plot of CQN normalised RNA-seq data points. 84

2.10 Expression profile of control genes before and after CQN. 85
2.11 Clock gene expression in the SCN across circadian time. 88
2.12 Concordance between RNA-seq and qPCR. 91
2.13 Expression of genes surveyed by RNA-seq and qPCR. 93
2.14 Expression of cell markers in the SCN. 95
2.15 PCA plot of RNA-seq data obtained from different tissues and studies. . . 98
2.16 Enrichment of SCN marker expression in the SCN. 100
2.17 146 genes with enriched expression in the SCN. 101
2.18 GO enrichments for SCN-enriched genes. 102
2.19 *In situ* hybridisation of SCN-enriched genes. 103
2.20 Current and proposed experimental design. 107
3.1 Harmonics of circadian rhythmicity. 117
3.2 The relationship between soft threshold (power) and both scale free topology and mean connectivity. 124
3.3 Normalised expression profiles of the 23 WGCNA co-expression modules. . 129
3.4 The module eigengene of the 23 WGCNA co-expression modules. 129
3.5 Dendrogram showing module membership of genes expressed in the SCN. . 130
3.6 Clustering dendrograms of WGCNA co-expression modules. 131
3.7 Screenshot of SCNseq data explorer. 133
3.8 Venn diagram showing fluctuating genes identified through various statistical approaches. 135
3.9 Identification of 4,569 cycling genes in the wild-type mouse SCN. 138

3.10 Genes involved in cytoskeletal assembly show peaks in expression in early night. . . . 139
3.11 Comparison of circadian transcriptomes. . . . 140
3.12 Gene clustering identifies a novel twin-peaking synaptic gene module. . . . 142
3.13 Validation of twin-peaking expression via qPCR. . . . 143
3.14 Functional annotation of the twin-peaking module. . . . 145
3.15 SCN microarray data not suitable for WGCNA analysis. . . . 148
3.16 Co-expression analysis captures expression relationship between genes. . . . 151
3.17 Transcriptional regulation of circadian harmonics. . . . 155
3.18 Models of twin-peaking expression through TF binding. . . . 156
3.19 Twin-peaking expression coincides with phase-shifting "susceptibility windows". . . . 160
3.20 Circadian gating of light signaling pathway. . . . 161
3.21 The coupling between circadian oscillators. . . . 170
3.22 Redox states, membrane potential and expression of 'twin-peaking' genes in the SCN. . . . 172
4.1 SCN lincRNA are lowly expressed. . . . 189
4.2 SCN lincRNA possess fewer exons than protein coding genes. . . . 190
4.3 SCN lincRNA are significantly but weakly conserved. . . . 191
4.4 Expression of SCN lincRNA does not correlate to that of the nearest protein coding gene. . . . 192
4.5 Identification of novel multiexonic lincRNA expressed in the SCN. . . . 194

4.6 Identification of multiexonic lincRNA whose expression fluctuates in the SCN. 196
4.7 NONCO7761 shows significant circadian fluctuation in the mouse SCN. . . 197
4.8 Identification of novel clock gene isoforms. 201
4.9 Novel *Cry1* exon represents an alternate transcriptional start site. 202
4.10 Novel *Cry1* isoform displays antiphase expression. 203
4.11 Multispecies alignment track displaying nucleotide sequence conservation for the protein coding region of the *Cry1* bii exon. 206
4.12 Isoform switching mechanism can produce strong protein oscillations. . . . 216
4.13 Transcriptional Regulation of *Cry1* isoforms. 218
5.1 *Bdr* mice display circadian deficits. 221
5.2 *Bdr* mice possess phase-advanced and heterogeneous phase of entrainments. 222
5.3 The core molecular clock and response to light is not altered in *Bdr* mice. . 223
5.4 Normalisation size factors generated in DESeq2. 228
5.5 Dispersion estimation in DESeq2. 229
5.6 Combinatorial approach to DE gene identification. 230
5.7 MA plots displaying DE genes for each time point assessed. 231
5.8 Skeleton photoperiod circadian paradigm. 234
5.9 Phase-shifting circadian paradigm. 234
5.10 Phase-shifting circadian behavioural analysis. 235
5.11 PCA plot of WT and *Bdr* SCN RNA-seq data points. 239
5.12 Cycling genes appear unperturbed in *Bdr* mice. 240
5.13 Examples of twin-peaking genes differentially expressed at ZT22. 244

5.14 Dysregulation of the twin-peaking co-expression module. 248
5.15 Dysregulation of additional co-expression modules. 250
5.16 Potential phase-advance of SCN-output and circadian relevant gene expression. 252
5.17 Preservation of WT co-expression modules in *Bdr* mice. 254
5.18 Actograms of WT and *Bdr* mice entrained to differing light schedules. . . . 257
5.19 *Bdr* mice display a number of circadian deficits in LD which were not altered under a SPP. 258
5.20 Actograms of mice subject to phase-shifting light pulses. 261
5.21 *Bdr* mice did not show evidence of altered phase-advance. 262
5.22 Twin peaking expression is shifted in the *Bdr* mouse model. 267
5.23 Reduction in a circadian TF can affect timing of twin-peaking transcription. 271
6.1 Genetic complementation assay in *Cry1*$^{-/-}$:*Cry2*$^{-/-}$ cells. 283

List of Tables

1.1 Mouse models of psychiatric disease with SCRD. 22

1.2 *Snap25* mouse mutants 28

2.1 RNA-sequencing data used in this thesis. 62

2.2 SCN-enriched genes. 104

3.1 Comparison of algorithms which detect rhythmic oscillations. 122

3.2 Top 20 hub genes in the twin-peaking co-expression module 144

3.3 Top 10 hub transcription factors of the twin-peaking module 164

4.1 GREAT analysis of twin-peaking lincRNA. 199

5.1 Circadian behavioural parameters assessed between genotypes. 236

5.2 Number of differentially expressed genes. 242

5.3 Examples of differentially expressed genes. 243

5.4 Co-expression modules enriched for differentially expressed genes. 246

5.5 Module enrichment for schizophrenia GWAS linked genes. 255

5.6 Module enrichment for schizophrenia risk genes determined by the CFG approach. 256

1 Introduction

1.1 Circadian rhythms

All organisms have evolved mechanisms to allow them to better adapt to their environment. Just as organisms gain a fitness advantage by adapting to features which vary in space, such as climate or terrain, organisms also benefit by adapting to features which vary in time (Paranjpe and Sharma, 2005). The rotation of the earth on its axis results in daily fluctuations of the environment, such as lighting or temperature, which encouraged the evolution of endogenous daily rhythms that are present in all eukaryotic species, from unicellular organisms to man (Sharma, 2003). Circadian (*circa*, approximately; *dian*, a day) rhythms are endogenous 24-hour oscillations in behaviour and physiology. These rhythms are set by autonomous clocks that measure time in the absence of environmental cues (Dunlap, 1999). These rhythms were first reported in 1729, when it was noted that leaves of the plant moved in a 24-hour pattern despite being in constant darkness. In mammals, many processes are found to possess a circadian component, such as locomotor activity, a measure used to survey circadian rhythms to this day (Jud et al., 2005).

1.1.1 The SCN as a master pacemaker

The circadian system is usually described as comprising of three interconnected elements: 1) the input pathway which receives and transmits environmental stimuli, 2) the pacemaker or clock, and 3) the output pathway which manifests overt rhythms (Dibner et al., 2010). In mammals, the suprachiasmatic nucleus (SCN) of the hypothalamus was identified as the master pacemaker of the circadian system. Brain lesioning studies found that

ablations in the hypothalamus caused circadian disruption described by wheel-running activity (Richter, 1960). Further studies in 1972 identified the SCN to be the crucial hypothalamic region as its ablation abolished the circadian rhythmicity of corticosterone levels, drinking behaviour and locomotor activity (Moore and Eichler, 1972, Stephan and Zucker, 1972). Furthermore, SCN-tissue transplantation studies in rodents found that the recipient animal could recover circadian rhythmicity after its own SCN was lesioned (Drucker-Colin et al., 1984) and alter its circadian period (time taken for one complete cycle) to that of the donor animal (Ralph et al., 1990).

The SCN is a paired structure and in mouse comprises of ∼20,000 neurons (Abrahamson and Moore, 2001). Each unilateral SCN can be subdivided into two distinct regions - a ventral "core" and a dorsal "shell" region which possess different roles in the control of circadian rhythms (Welsh et al., 2010) (Figure 1.1). The core receives retinal input signals which then signals to the shell which controls circadian rhythms of the organism through hormonal and neuronal outputs.

The SCN, situated above the optic chiasm, is well-positioned to receive photic input. The main SCN input pathway is the retinohypothalamic tract (RHT) which was initially identified as a simple projection linking the retina to the SCN (Hendrickson et al., 1972) but has since been shown to be more complex (Johnson et al., 1988). The RHT starts from intrinsically photoreceptive retinal ganglion cells (ipRGCs), melonopsin-containing nerve cells in the mammalian eye. These cells were shown to respond to light and be important in synchronising circadian rhythms, even in visually blind mice (Foster and Helfrich-Forster, 2001). In addition to innervating the SCN, the RHT also links the

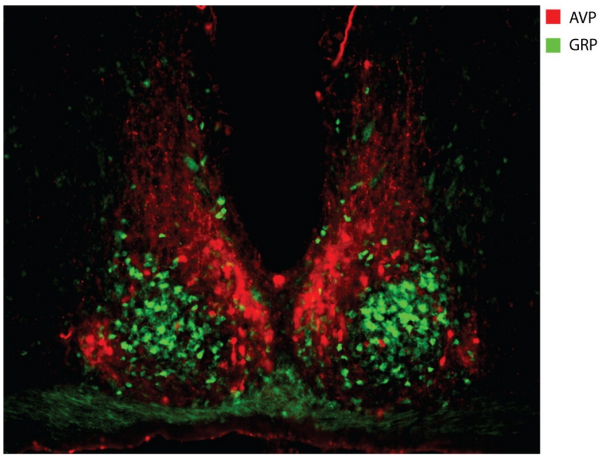


Figure 1.1. Immunostaining of coronal mouse SCN section.
Section shows the anatomical division of the SCN into a ventral core region (GRP — green) and a dorsal shell region (AVP - red). Figure obtained from (Karatsoreos et al., 2004) with permission from the Journal of Neuroscience.

retina to approximately 30 other brain regions such as the ventrolateral preoptic nucleus (VLPO) or intergeniculate leaflet (IGL) (Figure 1.2).

The SCN is also innervated by the geniculohypothalamic tract (GHT) originating from the IGL in the thalamus. The IGL is thought to play a role in circadian entrainment through both the photic and non-photic clock resetting pathways (section 1.1.4). Neuropeptide Y (NPY), a key neurotransmitter in the GHT displayed fluctuations in the SCN with peaks occurring at dawn and dusk (Pickard et al., 1987, Shinohara et al., 1993a). NPY depresses excitatory neurotransmission in the SCN, blocking photic clock resetting pathways *in vitro* and *in vivo* (Biello et al., 1997, Weber and Rea, 1997). Furthermore, despite entraining to a 12:12 LD cycle (each day consists of lights on for 12 hours followed by lights off for 12 hours), IGL lesioned rats could not entrain to a 24-hour skeleton photoperiod (1-hour light pulses signifying the start and end of the light period;

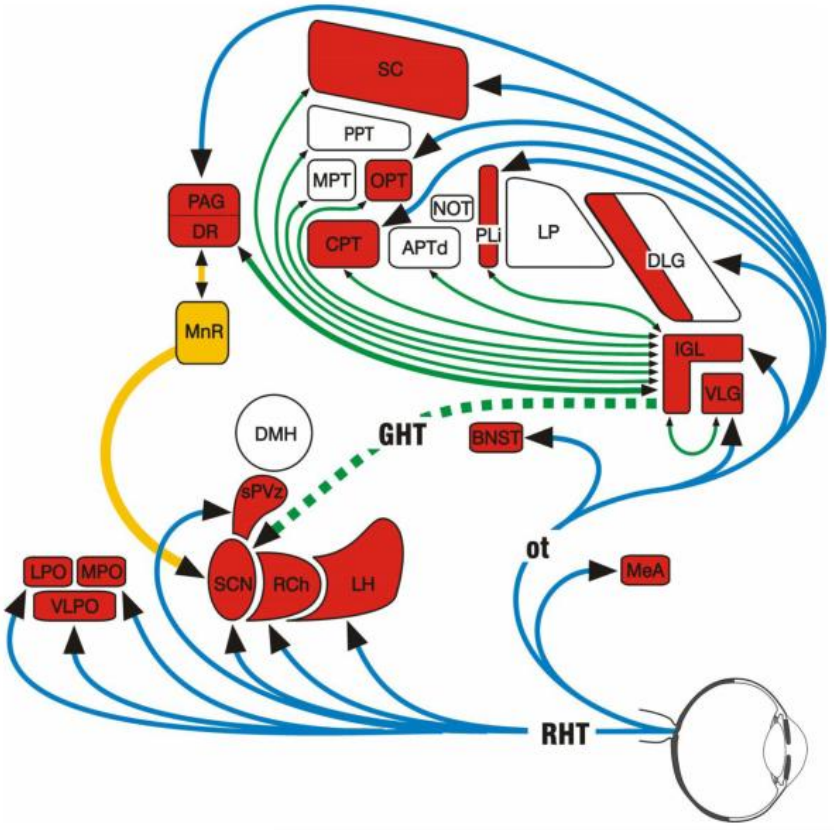


Figure 1.2. Retinorecipient brain regions.
Diagram displaying retinorecipient brain regions and SCN input from the RHT, GHT and MRN. Unfilled regions are not innervated by ipRGCs. Connections between the IGL and nuclei of the subcortical visual shell are shown in green. Figure obtained from (Morin, 2013) with permission from Experimental Neurology.

further described in section 5.1.1) (Edelstein and Amir, 1999). This suggests an importance of this brain region in entrainment to a lighting schedule that nocturnal rodents would experience in the natural environment.

The third major projection to the SCN is through serotonergic fibers originating from the median raphe nucleus (MRN) (Meyer-Bernstein and Morin, 1996). Similar to the IGL, the MRN may also play a role in the resetting of the clock. It has been reported that 5-HT (serotonin) can inhibit the phase shifting effects of light and induce phase-advances during the day (Edgar et al., 1993, Medanic and Gillette, 1992, Prosser et al., 1994, Glass et al., 1994). In line with this data, both SCN 5-HT depletion and 5-HT1A mixed agonist/antagonist application enhanced the phase-shifting potential of light during night (Bradbury et al., 1997, Gannon, 2003). Lesions of serotonin specific cells in the MRN eliminated 5-HT fibers in the SCN and induced changes in activity onset and a longer duration of nocturnal wheel-running (Meyer-Bernstein and Morin, 1996).

1.1.2 The endogenous molecular clock

SCN neurons isolated from the rest of the brain display endogenous rhythms in neuronal activity whereas the surrounding tissue does not (Inouye and Kawamura, 1979). These circadian rhythms are not derived from the multicellular network but instead proposed to be generated at the cell autonomous level because individual neurons display circadian firing rhythms, as determined by electrophysiology (Welsh et al., 1995). This endogenous circadian rhythm is set by an autonomous transcriptional-translational negative-feedback loop which drives ∼24-hour expression patterns of core clock components (Figure 1.3).

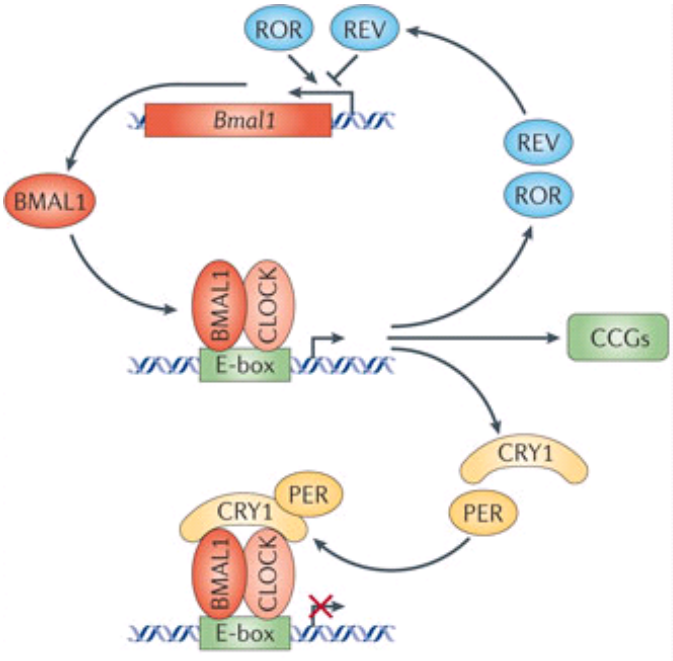


Figure 1.3. A schematic of the molecular circadian clock.
Positive elements in the loop are shown as red, negative — yellow and stabilising — blue. The final output of the negative feedback loop are 24-hour cycles of transcription through the CCGs which set the body's circadian rhythm until the next wave of transcription. Figure adapted from (Kondratova and Kondratov, 2012) with permission from Nature Publishing Group.

In the primary feedback loop, the positive elements brain and muscle ARNTL1 (BMAL1) and circadian locomotor output cycles kaput (CLOCK; or NPAS2) are expressed and combine to form a heterodimer transcription factor (TF) - BMAL1:CLOCK. This TF then binds to E-box elements which drive the transcription of clock-controlled genes (CCG), which control the circadian fluctuations around the body. The BMAL1:CLOCK TF also binds to E-box sites driving the expression of the negative elements *Period (Per) 1-3* and *Cryptochrome (Cry) 1/2* . These negative elements subsequently form a complex, become phosphorylated and after accumulating in the nucleus after a delay of many hours, directly interact with the BMAL1:CLOCK complex thereby inhibiting it. The inhibition of BMAL1:CLOCK therefore prevents the expression of the CRY/PER negative elements. The PER/ CRY inhibitory complex is then phosphorylated in the cytoplasm making it unstable and is then degraded by ubiquitination. The degradation of the negative elements releases the BMAL1:CLOCK repression and allows a new cycle of transcription through the CCG which will take nearly 24 hours to complete (Fu and Lee, 2003).

In an additional loop which provides clock robustness, BMAL1:CLOCK also directly activates nuclear receptor subfamily 1, group D, member 1/2 *(Nr1d1/2; Rev-Erba/b)* through E-box elements, which feeds back and competes with RAR-related orphan receptor (ROR) to bind to ROR response elements (RRE) in the *Bmal1* promoter and therefore inhibiting it. The REV-ERBA/B levels will then fluctuate as the *Bmal1* gene is rhythmically switched on and off (Fu and Lee, 2003).

Although these rhythms are generated at the level of a single cell, it is important that these cells in the SCN are synchronised to allow a robust generation of circadian rhythms. These neurons can be coupled synaptically through the release of neurotransmitters such

as vasoactive intestinal peptide (VIP) or gastrin-releasing peptide (GRP) (Aton et al., 2005, Maywood et al., 2006).

Although peripheral tissues also express key clock genes in a rhythmic fashion, the SCN is necessary to synchronise every cell in the body each day in order to maintain phase coherence, otherwise the cells will rapidly desynchronise (Yoo et al., 2004) which may have behavioural or pathological consequences (Bechtold et al., 2010). The timing cues for these peripheral organs are different to the SCN. Instead of light, systemic timing cues (in part, driven by SCN output CCG) such as hormone levels, daily feeding schedules or miRNA levels in the serum synchronise the organ's circadian rhythms (Damiola et al., 2000, Hara et al., 2001, Reddy et al., 2007, Shende et al., 2011).

1.1.3 Circadian oscillation in SCN excitability

The electrical activity in SCN neurons exhibits daily fluctuations and is essential for synchronising peripheral tissues to the SCN and environment. In the day, SCN neurons display increased frequency spontaneous action potentials (SAP) which then reduce at night (Brown and Piggins, 2007). The endogenous molecular clock in the SCN can drive these rhythms. Mutations or deletions of core clock genes such as *Tau* (casein kinase 1ϵ), *Clock, Cry1* and *Cry2* result in disrupted SCN neural activity and behavioural rhythms (Liu et al., 1997, Nakamura et al., 2002, Herzog et al., 1998, Albus et al., 2002).

The molecular clock may control neural activity by regulating calcium (Ca^{2+}) and potassium (K^{+}) ion channels through a variety of mechanisms (Figure 1.4). The genes which encode these ion channels may have E-box sites for CLOCK—BMAL binding, or other sequences regulated by key clock components, allowing neural activity to be driven

directly by transcription (Colwell, 2011). The clock may also drive electrical rhythms through circadian trafficking of proteins or channels which when embedded into the membrane alter synapse function (Colwell, 2011). For example, the rapid insertion and removal of glutamate AMPA receptors into the postsynaptic membrane, mediated by cyclic adenosine monophosphate- (cAMP-) protein kinase A (PKA) signalling pathways, can induce changes in excitatory synaptic transmission (Kessels and Malinow, 2009, Keifer and Zheng, 2010). Neuronal rhythms may be set by post-translational modifications of channels and proteins relating to neurotransmission. Ca^{2+} and cAMP fluctuate in the SCN and are likely to drive ion channel activity and membrane currents. These ion channels may contain secondary proteins that modulate pore forming subunits which are good targets for post-translational modifications (Colwell, 2011).

The regulation of membrane potential can be split into three classes based upon function (Colwell, 2011). Firstly, there are currents which *provide* the excitatory drive, increasing the membrane potential to a relatively depolarised state close to the threshold in generating an action potential. Sodium (Na^+) currents are thought to contribute to setting the membrane potential because blocking persistent Na^+ currents using riluzole prevents the expression of daily rhythms in neural activity (Kononenko et al., 2004).

Secondly, there are currents which *respond* to the excitatory drive, inducing sustained action potential discharge in the SCN for 4–6 hours. Fast delayed rectifier (FDR) K^+ currents have been implied in neurons which have the capacity to discharge at high rates like the SCN (Baranauskas et al., 2003, Rudy and McBain, 2001), including the SCN expressed genes *Kcnc1* and *Kcnc2*. Mice which lack both these genes exhibit both reduced

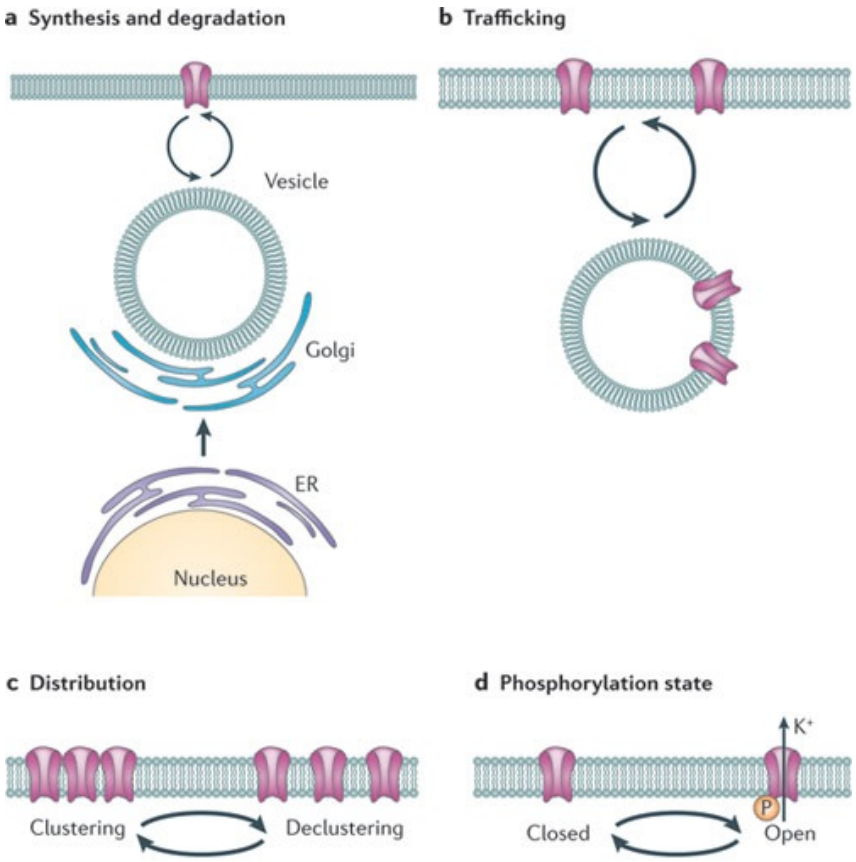


Figure 1.4. Molecular processes which drive electrical rhythms in the SCN.
(A) Altering the expression and degradation of ion channels through transcriptional and translational regulation, **(B)** Removal and insertion of receptors and ion channel in response to physiological stimulation, **(C)** Distribution of ion channels changing over time affecting the firing rhythm rates, **(D)** Post-translational modification of ion channels allowing the flow of ions through the membrane. Figure obtained from (Colwell, 2011) with permission from Nature Publishing Group.

neuronal daytime spontaneous activity and night time NMDA-evoked excitatory responses (Kudo et al., 2011).

Thirdly, there are currents which *silence* the firing at night, hyperpolarising the membrane by reducing the resting membrane potential at night. This process is mediated by hyperpolarizing K^+ channels but also by K2P, TASK and TREK channels which are active at all voltage ranges (Bayliss and Barrett, 2008).

In summary, individual SCN neurons are able to display endogenous rhythms in neuronal activity. These rhythms may be regulated by a variety of mechanisms but are ultimately established by the autonomous transcriptional-translational negative-feedback loop.

1.1.4 Resetting the clock

Since our circadian clock can measure the daily rhythm only approximately, the circadian oscillators must be corrected daily to stay in synchrony with the photoperiod. This is achieved by an environmental cue (zeitgeber — "time giver"); in mammals and most other organisms, light is the main zeitgeber for the synchronisation of the circadian clock (Gehring and Rosbash, 2003). Zeitgeber time (ZT) is a measurement of time relative to the main zeitgeber. For example, ZT0 is defined as the time of "lights on" and ZT12 defined as "lights off" in a standard experimental 12:12 LD lighting cycle. Circadian time (CT) is based upon the free-running period of the organism. CT0 generally refers to the start of the subjective day, and CT12 refers to the start of the subjective night. Diurnal animals are active during the subjective day and sleep during the subjective night, whereas the opposite is true for nocturnal animals.

If the organism's endogenous clock is out of synchrony with the environmental photoperiod, photic stimuli will induce a phase-shift — a temporal shift in its circadian rhythm. This phase-shifting capability allows organisms to align their endogenous clock to the environment, important for example, during seasonal entrainment or if the animal has entered a different timezone through migration. A phase-shift forwards in time is known as a phase-advance whereas a shift backwards in time is known as a phase-delay. The ability of light to induce a phase-shift is demonstrated by the photic phase response curve (PRC): a light impulse at early subjective night (ZT12-16) will initiate a phase-delay of the clock whereas a light impulse at late subjective night (ZT20-24) will initiate a phase-advance (Figure 1.5) (Golombek and Rosenstein, 2010, Ding et al., 1994).

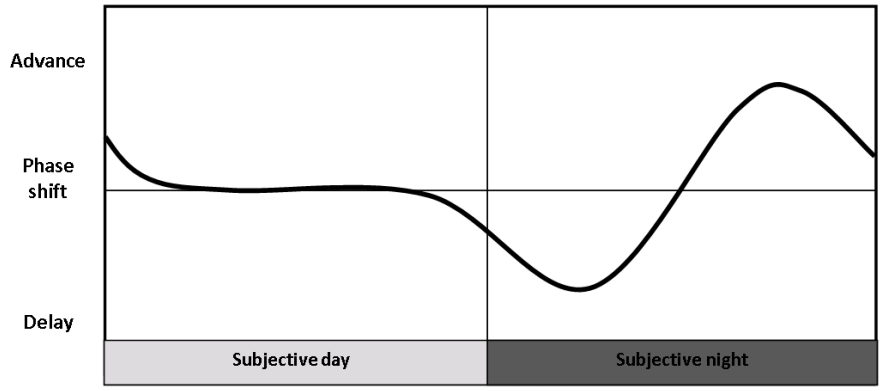


Figure 1.5. The photic phase response curve.
This curve displays the phase-shifting potential of an organism when light pulses are applied at different times of the day. The SCN respond differently to light depending on the time of day the pulse is applied. The response to light at subjective night is bifurcated (Ding et al., 1994), with phase-delays in the early night and phase-advances at late night but limited response during the subjective day. Figure reproduced from (Golombek and Rosenstein, 2010).

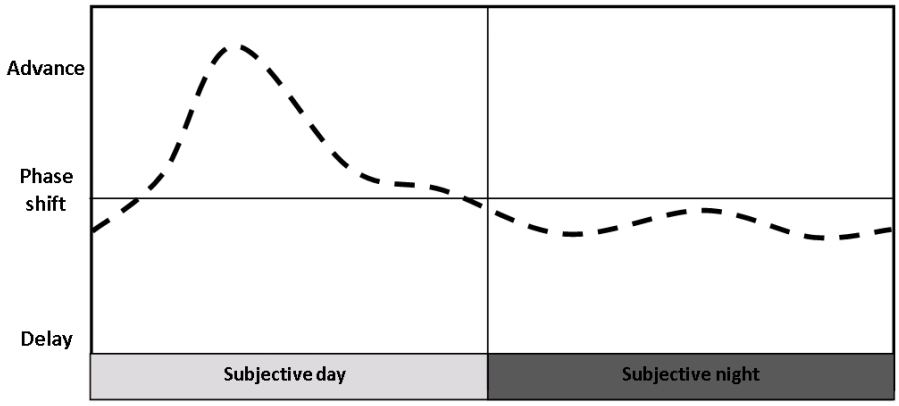


Figure 1.6. The non-photic phase response curve.
This curve displays the phase-shifting potential of an organism due to non-photic stimuli at different times of the day. Non-photic stimuli can induce a phase-advance during the subjective day and may induce small phase-delays during the subjective night. Figure reproduced from (Mrosovsky, 1995).

Photic entrainment

In mammals, light information is relayed from the retina to the SCN via the RHT (Moore, 1973, Moore and Klein, 1974) which instigates a molecular cascade which induces the expression of the repressive *Per* genes. A series of studies have sought to characterise this molecular cascade.

Photic information is transmitted to the SCN via glutamatergic signaling through the RHT. Glutamate signalling activates NMDA receptors and induces Ca^{2+} influx into the SCN (Colwell, 2001). Both glutamate and NMDA have been shown to be sufficient to activate light signalling *in vivo* and *in vitro*, altering circadian phasing and inducing immediate early gene markers, whereas NMDA antagonists prevented light-induced phase-shifts (Colwell and Menaker, 1992, Vindlacheruvu et al., 1992). Calcium influx

induces calmodulin-dependent protein kinase II (CaMKII) auto-phosphorylation which then phophorylates nitric oxide synthase (NOS). Light pulses at night activate CaMKII and its inhibition results in a reduced phosphorylation of NOS and a reduced light-induced phase-shifting response (Agostino et al., 2004, Yokota et al., 2001, Golombek and Ralph, 1994). Then, depending on the time of night, there is a differential response to light in this signalling pathway.

At early night, intracellular calcium ($Ca^{2+}{}_i$) signaling is required: thapsigargin application to SCN slices blocked the Ca^{2+}-ATPase that replenishes $Ca^{2+}{}_i$ stores in the endoplasmic reticulum and prevented glutamate induced phase-delays but also had a small inhibitory effect on phase-advances (Thastrup et al., 1990, Ding et al., 1998). $Ca^{2+}{}_i$ release is mediated through ryanodine receptors (RyRs) embedded in the endoplasmic reticulum. Activating these receptors using caffeine, FK506 or rapamycin induced phase-delays, whereas blocking them using dantrolene or ruthenium prevented phase-delays yet had no effect on phase-advances (Ding et al., 1998).

At late night, cGMP-dependent protein kinase (PKG) is proposed to be important for phase-advances as its inhibition using KT5823 prevented glutamate-induced phase-advances in peak firing, but did not affect phase-delays in early night (Ding et al., 1998). Nitric oxide (NO) is known to stimulate the production of cyclic guanosine monophosphate (cGMP). The increase of cGMP activates guanylyl cyclase (GC) and induces the production of PKG (Prosser et al., 1989).

At both early night and late night, light stimulates the phosphorylation of cAMP response element-binding protein (CREB) which is important for phase-shifts and the expression of *c-Fos* and *Per1* (Ginty et al., 1993, Gau et al., 2002).

Non-photic entrainment

The circadian clock may also be shifted through non-photic stimuli such as social stimuli, novelty or physical exercise (Mistlberger and Skene, 2004). Typically these non-photic stimuli or their relevant neurotransmitters, such as 5-HT or NPY, can induce a phase-advance during the subjective day and small phase-delays during subjective night, as displayed by the non-photic PRC (Figure 1.6). Although there are numerous non-photic stimuli which can shift the clock, they all present similar PRCs suggesting that they share similar biochemical mechanisms (Mrosovsky, 1995). Both the raphe nuclei (RN) and IGL brain regions are highly implicated in non-photic phase resetting via the neurotransmitters 5-HT and NPY, respectively.

Serotonergic activity is closely associated with both behavioural state and phase-shifting potential which suggests that 5-HT may relay behavioural non-photic stimuli to induce phase-advances (Grossman et al., 2000, Mendlin et al., 1996). Both *in vitro* and *in vivo* work supports this: application of 5-HT to SCN brain slices and microinjection of the 5-HT1A/7 receptor agonist into the SCN during the subjective day induces phase-advances (Medanic and Gillette, 1993, Prosser et al., 1993, Prosser et al., 1990, Challet et al., 1998). The MRN is the only direct source of 5-HT to the SCN but stimulation of both the medial and dorsal RN can induce 5-HT specific effects in the SCN (Meyer-Bernstein and Morin, 1999). 5-HT is also thought to act on the IGL from dorsal raphe nuclei (DRN) projections. Both novel wheel exposure and stimulation of the DRN can induce 5-HT release in the IGL (Grossman et al., 2004) and microinjection of the 5-HT1A/7 receptor agonist into the IGL induces phase-advances (Challet et al., 1998).

In addition to modulating photic responses, NPY produced by the IGL is important for inducing non-photic phase-shifts. Application of NPY to SCN slices or microinjection into the SCN induces a phase-shift response similar to the effects of wheel-running, sleep deprivation or stimulation of the GHT (Albers and Ferris, 1984, Huhman and Albers, 1994, Rusak et al., 1989, Golombek et al., 1996, Medanic and Gillette, 1993). Furthermore, preventing the action of NPY from the IGL *in vivo* through either lesioning studies or NPY antiserum attenuates non-photic induced phase-shifts (Biello et al., 1994, Janik and Mrosovsky, 1994).

Non-photic stimuli are thought to induce phase-advances during the subjective day by decreasing *Per* gene expression (Maywood and Mrosovsky, 2001, Fukuhara et al., 2001). Increasing *Per* expression at a time when its expression is already rising induces a phase-advance. Conversely, increasing *Per* expression at a time when its expression is otherwise decreasing induces a phase-delay. This explains why a light pulse can induce a phase-delay from CT10-CT16 (*Per* levels are decreasing) but a phase-advance from CT19-CT2 (*Per* levels are increasing). Light, however, is unable to shift the clock during most of the subjective day as described by the photic PRC (Figure 1.5). This is likely because *Per* transcription is already it its maximum. However, *non-photic* stimuli such as locomotor activity may induce phase-advances by decreasing *Per* expression. Interestingly, this decrease in *Per* expression was attenuated by a pulse of light just after this non-photic stimulus, suggesting *Per* transcription, having been downregulated by non-photic stimuli, can be now upregulated by light (Maywood and Mrosovsky, 2001). This may also explain why sleep deprivation induces phase-advances in dark but not light (Grossman et al., 2000). These experiments demonstrate the complex interplay between photic and

non-photic clock resetting pathways which allow the organism's circadian rhythm to be synchronous to its environment.

1.1.5 Circadian disruption and disease

Disruption to circadian rhythms result in pathological consequences. Mutating or deleting clock genes responsible for generating circadian rhythms results in metabolic, cardiovascular and neurological diseases, as well as other deficits such as reduced cognition, motor defects and irregular emotional processing (Maury et al., 2014, Bass, 2012, Sakai et al., 2004, Zueger et al., 2006, Roybal et al., 2007, Easton et al., 2003). Similar phenotypes are observed when there are defects in signalling of key SCN neuropeptides such as NPY, VIP and arginine vasopressin (AVP) (Karl et al., 2006, Hughes and Piggins, 2008, Wersinger et al., 2007). Circadian misalignments which arise from shift work affect emotional wellbeing and application of circadian principles improve employee satisfaction and health (Czeisler et al., 1982, Naismith et al., 2014). Age is highly correlated with numerous metabolic and neurological diseases (Campos Costa et al., 2013). As circadian disruption is more prevalent in older adults (Crowley, 2011), it may be that these disturbances contribute to disease in the elderly.

Circadian disruption has been implicated in metabolic and vascular disease (Maury et al., 2014, Bass, 2012). Certain pathologies have been shown to cluster about particular time windows. Heart attacks are thought to occur more frequently in the morning due to the oscillation of the protease inhibitor plasminogen activator inhibitor type 1 (PAI-1) which highly corresponds to oscillation in thrombosis risk (Schoenhard et al., 2003, Wang et al., 2006). Cardiac arrhythmia also exhibits increased morning time risk and

is associated with the circadian expression of clock-controlled potassium channels in the heart (Jeyaraj et al., 2012). Mice with mutations in core clock genes display defects in cardiovascular function. *Clock*$^{-/-}$ mice possess reduced arterial blood pressure (Zuber et al., 2009), *Cry1/2* double knock out mice suffer from hypertension (Doi et al., 2011), *Bmal1*$^{-/-}$ mice have disrupted circadian variation in heart rate and blood pressure (Curtis et al., 2007) and mice with overexpression of the *Clock*Δ*19* allele (which possesses a deletion in exon 19) in cardiac muscle cells have disrupted heart rate and contractility (Bray et al., 2008). Circadian disruption also interferes with numerous metabolic systems contributing to obesity, diabetes and cancer. Timng of feeding is important in maintaining metabolic homeostasis (Paschos, 2015). Feeding at abnormal times misaligns the circadian rhythms in peripheral tissues to that of the circadian rhythm governed by the master clock. For example, rats forced to eat at times opposite their normal feeding schedule developed obesity (Salgado-Delgado et al., 2010). In addition, mice with mutations in core clock genes display dysregulation of metabolic processes. *Bmal1*$^{-/-}$ mice show obesity and excess lipid levels in the bloodstream (Lamia et al., 2008, Rudic et al., 2004), *Clock*Δ*19* mice show an excess of glucose and lipid levels in the bloodstream (Turek et al., 2005, Marcheva et al., 2010) and *Cry1/2* double knock out mice show impaired glucose tolerance and increased glucose levels in the bloodstream (Zhang et al., 2010, Lamia et al., 2011).

Neurodegenerative diseases such as Alzheimer's (AD), Huntington's (HD) and Parkinson's disease (PD) are associated with numerous sleep and circadian rhythm disruptions (SCRD) such as sleep fragmentation, increased nocturnal activity and decreased diurnal activity (Ferreira et al., 2006, Morton et al., 2005, Volicer et al., 2001). In AD, these defects could be due to the degeneration of the SCN, as loss of AVP and VIP expressing

neurons are reported in AD patients (Stopa et al., 1999) which may prevent the circadian expression patterns of different brain regions oscillating in phase with each other. In *Drosophila*, *Per* deletion exacerbates oxidative stress and results in decreased average lifespan by up to 20% (Krishnan et al., 2008, Krishnan et al., 2009). *Bmal1* knockout mice displayed increased levels of oxidative stress in the cortex, degeneration of synaptic terminals and reduced cortical functional connectivity (Musiek et al., 2013).

SCRD are also highly prevalent in mood disorders (depression) and psychoses (schizophrenia, bipolar disorder (BP)) (Landgraf et al., 2014, Karatsoreos, 2014). Patients with major depressive disorder show weak cyclic expression of key clock genes due to shifting of expression phasing of particular circadian genes in different brain regions (Li et al., 2013). Seasonal affective disorder (SAD) is depression which tends to manifest during the winter months due to the shortening of the days. This disorder is highly associated with three circadian clock genes *Per2*, *Arntl*, and *Npas2* and SAD patients display circadian rest-activity disturbances (Partonen et al., 2007, Teicher et al., 1997). BP is a condition which consists of periods of both depression and mania. Like schizophrenia and SAD, polymorphisms of circadian related genes have been associated with BD (Mitterauer, 2000, Desan et al., 2000). Treatments of BD have also been shown to affect both circadian rhythms and the genes important for driving them (Jones, 2001). Furthermore, BD episodes can be induced by changes in lighting schedule such as time-zone travel or seasonal change (Murray and Harvey, 2010).

Up to 80% of schizophrenia patients possess SCRD; patients show sleep fragmentation, longer sleep latency and circadian biomarker phase misalignments independent of anti-psychotic treatment (Boivin, 2000, Barnard and Nolan, 2008, Wulff et al., 2012,

Karatsoreos, 2014, Cohrs, 2008, Wulff et al., 2010). In addition, patients with SCRD score poorly on numerous quality-of-life clinical subscales and schizophrenia patients often comment that one of their highest priorities during treatment is improved sleep (Cohrs, 2008, Auslander et al., 2002). A likely reason for the co-morbidity shared by SCRD and neuropsychiatric diseases such as schizophrenia is due to common brain mechanisms underlying the two. Clinical evidence supports this: the treatment of SCRD in schizophrenic patients leads to both an improvement in sleep quality and psychiatric symptoms (Kantrowitz et al., 2010, Myers et al., 2011). Furthermore, there exist many symptoms, such as impaired cognition, which are caused by SCRD, that are highly prevalent in schizophrenics, but are not linked to the disruption of sleep (Alhola and Polo-Kantola, 2007, Kyriacou and Hastings, 2010, Karatsoreos, 2014).

Mouse models of psychiatric disease with SCRD

Animal models can also help explore the hypothesis that the co-morbidity of SCRD and schizophrenia are due to common brain mechanisms. If there is a mechanistic link, SCRD would be expected to be observed in mouse models of schizophrenia and vice-versa. Sleep and circadian function has been evaluated in mutant models of genes linked strongly with schizophrenia, of which the majority have yielded significant SCRD (Table 1.1).

VIP and its receptor (VIPR2/VPAC2) play a role in the coupling of rhythms between SCN neurons: *Vipr2* $^{-/-}$ cells cannot maintain synchrony in the absence of VIP-ergic signalling (Maywood et al., 2006). *Vipr2* duplications have been shown to confer a risk for schizophrenia in human patients (Vacic et al., 2011). *Vipr2* $^{-/-}$ mice display circadian abnormalities such as a shortened circadian period, loss of rhythmicity in DD and impaired

synchronization to light (Hughes and Piggins, 2008). Light can increase phosphoprotein and immediate early gene expression in the SCN of these mice, suggesting that it is not the input, but rather the output of the clock which is disrupted.

Neuregulin 1 (*Nrg1*) encodes a growth factor with a role in synaptic plasticity. Deletion of this gene from cortical projection neurons induced inhibitory neurotransmission and reduced synaptic plasticity, whereas an over expression resulted in unbalanced excitatory-inhibitory neurotransmission and reduced synaptic plasticity (Agarwal et al., 2014). This study concluded that careful regulation of *Nrg1* is necessary to balance the excitatory and inhibitory connections in the brain. NRG1 protein and function were found to be elevated in the prefrontal cortex of patients with schizophrenia (Chong et al., 2008 , Hahn et al., 2006) potentially leading to hypofunction of the glutamatergic pathway, which is often reported in the brains of schizophrenia patients (Moghaddam, 2004, Tsai and Coyle, 2002). Mice with heterozygous disruption of *Nrg1* display disrupted rest/activity rhythms and show a loss of SCN neurons (Pritchett et al., 2012b). Furthermore, administration of NRG1 to the third ventricle in hamsters resulted in reduced wheel-running activity, likely through epidermal growth factor receptor (EGFR) signalling pathways in a similar manner to transforming growth factor-alpha (TGF-alpha) (Snodgrass-Belt et al., 2005, Zak et al., 2006).

The cholecystokinin A receptor *(Cckar)* encodes for a G-protein coupled receptor that binds members of the cholecystokinin (CCK) family of hormones (Noble et al., 1999). Polymorphisms in the promoter region may be associated with schizophrenia and PD (Tachikawa et al., 2001, Wang et al., 2003, Koefoed et al., 2009) and alterations of mRNA levels have been reported in brains of schizophrenia patients (Bachus et al., 1997,

Mouse Mutant	Link to psychiatric disease	Circadian Phenotype	Reference
Vipr2$^{-/-}$	• Duplications confer a risk for schizophrenia	• Shortened circadian period, reduced rhythmicity in DD, impaired synchronization to light	(Hughes and Piggins, 2008)
Nrg1$^{+/-}$	• Elevated protein in PFC of schizophrenia patients	• Disrupted rest/activity rhythms	(Pritchett et al., 2012)
Cckar$^{-/-}$	• Polymorphisms in the promoter region may be associated with schizophrenia and PD • Dysregulation of mRNA levels in brains of schizophrenia patients	• Significant reductions in light pulse induced *Per* expression in SCN, behavioural phase shifts and pupillary reflexes	(Shimazoe et al., 2008)
Grm2/3$^{-/-}$	• Decreased concentration of glutamate in the hippocampus of schizophrenia patients • Association of *Grm3* with schizophrenia	• Sleep fragmentation, reduced sleep time and an increased sensitivity to the circadian effects of light	(Pritchett et al., 2015)
Blind-drunk (*Bdr*) (*Snap25*$^{+/Bdr}$)	• Expression dysregulated in brains from schizophrenic patients • Associated with schizophrenia • Mutations identified in human patients of epilepsy	• Fragmented and phase advanced rest activity profiles under a 12:12 LD cycle. • Normal input pathway and core clock	(Jeans et al., 2007, Oliver and Davies, 2009)

Table 1.1. Mouse models of psychiatric disease with SCRD.

Zachrisson et al., 1999). *Cckar* knockout mice display defects to the input pathways of the circadian clock with significant reductions in: light pulse induced *Per* expression in the SCN, behavioural phase-shifts and pupillary reflexes (Shimazoe et al., 2008).

Grm2 and *Grm3* encode the metabotropic glutamate receptors mGlu2 and mGlu3 and form part of the glutamatergic pathway which has many roles in the nervous system. *Grm2* is neuronal whereas *Grm3* is expressed in glial cells where its protein can be found both pre- and post-synaptically. (Ohishi et al., 1993, Hinoi et al., 2001). Glutamatergic dysfunction has been implicated with schizophrenia: decreased concentration of hippocampal glutamate is reported in schizophrenia patients (Tsai and Coyle, 2002, Tsai et al., 1995) and a meta analysis of genome-wide association studies (GWAS) found an association between variants in the vicinity of *Grm3* and schizophrenia. *Grm2/3* double knockout (*Grm2/3* $^{-/-}$) mice displayed circadian disruptions such as sleep fragmentation, reduced sleep time and an increased sensitivity to the circadian effects of light (Pritchett et al., 2015). Both *Grm2* and *Grm3* are expressed in the SCN, and modulate NMDA-evoked calcium response, perhaps explaining the circadian disruption observed in these mouse mutants (Ebling, 1996, Cui et al., 2001, Haak, 1999).

Finally, blind-drunk (*Bdr*), a mouse model with a *Snap25* dominant negative mutation, displays both circadian disruption and psychiatric endophenotypes (Jeans et al., 2007, Oliver and Davies, 2009). This mouse model was subject to investigation and extensively discussed in this thesis to help elucidate the link between SCRD and psychiatric disease (sections 1.1.5 and 5.1).

Mouse models of *Snap25* dysfunction

Synaptososomal-associated protein of 25kDa (SNAP25) plays a central role in neurotransmitter release and has been implicated in many neuropsychiatric diseases such as ADHD (Wilson, 2000, Faraone et al., 2005), epilepsy (Zhang et al., 2004b) and schizophrenia (Thompson et al., 2003, Lewis et al., 2003). *Snap25* expression has been reported to be selectively affected in brains from schizophrenic patients (Corradini et al., 2009) and variants near the *Snap25* locus have been implicated with schizophrenia through linkage studies (Lewis et al., 2003).

SNAP25 is a t-SNARE protein which is anchored to the cytosolic face of the membrane, and with syntaxin-1A can combine with the vesicular synaptobrevin/VAMP to form a stable four-α-helix SNARE complex primed for fusion (Sollner et al., 1993). An action potential stimulates the entry of Ca^{2+} into the neuron which triggers membrane fusion allowing the vesicular contents to be released. Post-fusion, α-SNAP and N-ethylmaleimide-sensitive factor (NSF) are required to dissociate the SNARE complex allowing vesicle recycling. Postsynaptically, SNAP25 has been shown to regulate surface NMDA levels with a role in long-term potentiation (LTP) (Jurado et al., 2013). SNAP25 has also been proposed to interact with voltage-gated calcium channels (VGCCs) via a synaptic protein interaction (synprint) site on the calcium channel decreasing neuronal response to calcium (Catterall and Few, 2008, Pozzi et al., 2008, Antonucci et al., 2013). SNAP25 is thus a multifunctional protein possessing both a role in neurotransmitter release and neuronal depolarisation.

There exist two isoforms of the mouse *Snap25* gene: *Snap25a* and *Snap25b* whose proteins differ by 9 amino acids as a result of alternative splicing. *Snap25a* is expressed

at relatively low levels throughout development whereas *Snap25b* expression dramatically increases from 3 weeks of birth and is the predominant isoform in adults (Bark et al., 1995).

Homozygous *Snap25*$^{-/-}$ mice are postnatally lethal (Washbourne et al., 2002), yet heterozygous *Snap25* null mutants were largely indistinguishable to wild-type littermates. To gain further insight into *Snap25*, other mutations of this gene have been used to investigate gene function and aberrant neurotransmission.

Heterozygous deletion of *Snap25* in the coloboma mouse mutant *(Snap25*$^{+/-}$*)* appeared to induce hyperactivity, a deficit in prepulse inhibition, learning deficits and problems in motor co-ordination/ balance (Wilson, 2000, Adler et al., 2001 , Gunn et al., 2011). Although hyperkinesis was normalised by genetic rescue of a transgene encoding *Snap25*, this phenotype may depend on the loss of other genes in the deleted region. This mutation was thought to affect the dopaminergic pathway because the indirect dopaminergic agonist, amphetamine, was shown to ameliorate hyperactive behaviour. Electroencephalogram (EEG) analysis revealed that elevated thalamic low-voltage-activated currents precede the onset of absence epilepsy thought to be due to presynaptic release defects (Zhang et al., 2004b). The combination of phenotypes makes coloboma a potential model for attention deficit hyperactivity disorder (ADHD), ataxia, epilepsy and other neuropsychiatric diseases.

The homozygous *Snap25*$^{Tkneo/Tkneo}$ mouse model has a transgene insertion into the *Snap25* intron upstream of exon 6 disrupting the alternative splicing between exons 5a and 5b, increasing the expression of the *Snap25a* isoform, yet decreasing overall SNAP25 protein expression levels by 50%. These mice tended to die shortly after forebrain seizures

at 3-5 weeks of age, a time when *Snap25b* levels are meant to be at their highest. Excision of the transgene completely reversed the phenotype (Bark et al., 2004).

To address the potential different functions of the *Snap25* isoforms, a *Snap25b* deficient mutant was generated by replacing exon 5b with an additional exon 5a (Johansson et al., 2008). This allowed alternative splicing to continue as normal only producing the *Snap25a* isoform. There are two mutants, one with the *Tkneo* transgene still present, and one where this had been excised.

The mutant with the transgene still present displayed a severe phenotype which progressed with age. This phenotype also includes spontaneous seizure activity and reduced paired-pulse facilitation (PPF), a parameter of short-term synaptic plasticity. However, these mice displayed mRNA levels 50% and protein levels 80% relative to WT, which was propose to be due to the presence of the *Tkneo* transgene. The mutant with the transgene excised showed reduced short-term synaptic plasticity but lived until adulthood. SNAP25b may also play a role in long-term plasticity as the neo-excised *Snap25b* deficient animals displayed greater anxiety index and impairment in spatial learning (Johansson et al., 2008). One hypothesis is that the switch in expression from the *Snap25a* to *Snap25b* isoform matures presynaptic function. Therefore the mouse mutants with low or absent levels of *Snap25b* may not possess mature presynaptic function.

One mouse mutant was generated using "knock-in" (KI) technology. Firstly, the gene was modified to replace the Ser187 with an Ala codon. This mutation was chosen as activated protein kinase C (PKC) was found to phosphorylate Ser187 thereby enhancing the recruitment and Ca^{2+}-dependent release of dopamine (DA) and acetylcholine (ACh) (Shimazaki et al., 1996, Iwasaki et al., 2000). *Snap25* also inhibits VGCCs via phospho-

rylation of this residue therefore affecting calcium dynamics (Pozzi et al., 2008). These mice exhibited increased anxiety, a severe working memory deficit, immature dentate gyrus (DG) neurons and a reduction in the release of serotonin and dopamine in the amydala (Ohira et al., 2013, Kataoka et al., 2011). These mice also displayed increased accumulation of alpha synuclein perhaps reflecting a compensatory response to impaired synaptic vesicle release (Nakata et al., 2012). These mice also exhibited spontaneous recurrent seizures which were preceded by abnormal EEG recordings (Watanabe et al., 2015). This may be due to a failure to incorporate Ca^{2+} channels and glutamate receptors into the plasma membrane, normally regulated by *Snap25* phosphorylation. Together, these mouse models implicate *Snap25* in epilepsy, schizophrenia, and α-synucleinopathies such as PD and also suggest that PKC-dependent regulation of SNAP25 is important in maintaining normal brain function (Table 1.2).

Mouse Mutant	Phenotype	SNAP25 Isoform Gene Expression	Reference
***Snap25* Knockout Mice**			
Homozygous Mutants	Die at birth, no evoked exocytosis.	No SNAP25 expression	(Washbourne et al., 2002)
Heterozygous Mutants	No reported phenotype.	50% of normal SNAP25 expression levels, ratio between SNAP25a/SNAP25b normal	(Washbourne et al., 2002)
Heterozygous Mutants (Coloboma)	Hyperactivity, learning deficits, problems in motor co-ordination/ balance.	50% of normal Snap25 expression levels	(Wilson, 2000, Adler et al., 2001 , Gunn et al., 2011)
***Snap25*$^{Tkneo/Tkneo}$ Mutants**			
Homozygous Mutants	Most mutants die between 3–5 weeks of age, impaired short-term synaptic plasticity, seizures.	50% of normal SNAP25 expression levels, ratio between SNAP25a/SNAP25b isoforms shifted towards more Snap25a	(Bark et al., 2004)
SNAP25b-deficient Mice			
Homozygous Mutants (Neo-containing)	Developmental defects, seizures, impaired short-term synaptic plasticity, lethal in the third postnatal week.	50% of normal SNAP25 expression levels, only SNAP25a	(Johansson et al., 2008)
Homozygous Mutants (Neo-excised)	Impaired short-term synaptic plasticity, progressive morphological changes in brain, learning deficits, seizures.	Normal SNAP25 expression levels but only SNAP25a	(Johansson et al., 2008)
***Snap25*^{S187A} Knock In Mice**			
Homozygous Mutants	Increased anxiety, working memory deficit, immature DG neurons, reduced neurotransmitter release, seizures.	50% of normal SNAP25 expression levels	(Shimazaki et al., 1996, Iwasaki et al., 2000, Ohira et al., 2013, Kataoka et al., 2011 , Nakata et al., 2012, Watanabe et al., 2015)
Blind-drunk (*Bdr*) Mutants			
Heterozygous Mutants	Impaired sensorimotor gating and ataxia, impaired spontaneous and evoked release. Circadian deficits.	Normal SNAP25 expression levels	(Jeans et al., 2007, Oliver and Davies, 2009)

Table 1.2. *Snap25* mouse mutants

Blind-drunk as a mouse model of schizophrenia

A large scale, phenotype-driven dominant mutagenesis screen identified a number of mouse mutants with overt phenotypes, including one mouse mutant (GENA70) which possessed an ataxic gait (Nolan et al., 2000, Nolan et al., 2002). Subsequent breeding of this mouse line resulted in offspring of which half were ataxic suggesting the mutation is dominant. This mouse mutant, labelled blind-drunk (*Bdr*) due to its ataxic gait, was found to possess a mutation in the *Snap25* gene which is lethal in homozygous mice. This T-to-C mutation in codon 67 of *Snap25b* isoform induced an isoleucine-to-threonine amino acid substitution which interacts with the syntaxin and vesicle-associated membrane protein (VAMP) proteins (Figure 1.7). This I67T dominant substitution increases the binding affinity within the core SNARE complex impairing action potential evoked glutamate release, likely due to the inability to replenish the readily releasable pool of synaptic vesicles between stimuli (Jeans et al., 2007).

Bdr mice show a range of behavioural defects such as some aspects of apathy and anxiety and a reduction in pre-pulse inhibition (abnormal sensorimotor gating), all indicative of a schizophrenia-like endophenotype (Jeans et al., 2007). These deficits were exacerbated when *Bdr* mice were subjected to prenatal stress, which could be reversed with anti-psychotic drugs (Oliver and Davies, 2009). This suggests that *Bdr* provides an important link between behavioural abnormalities and a specific synaptic deficit caused by a point mutation in a key synaptic pathway.

Initial circadian behavioural analysis demonstrated that *Bdr* mice show atypical circadian behavioural parameters such as fragmented and phase-advanced rest/activity profiles under a light/dark cycle (12hr light / 12hr dark). This is somewhat reminiscent

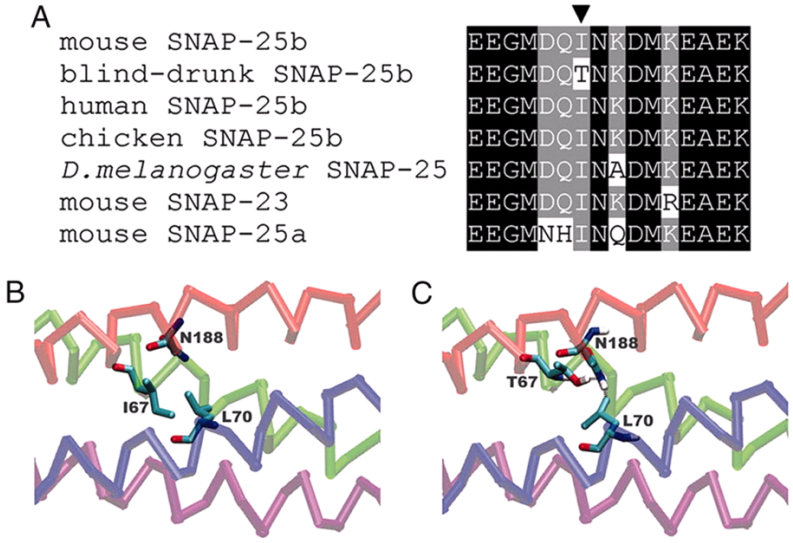


Figure 1.7. *Snap25b* mutation in blind-drunk
(A) The I67T mutation in the blind-drunk *Snap25b* mutation occurs in a highly conserved sequence. **(B & C)** Visualisation of the SNARE complex in **(B)** WT and **(C)** *Bdr*. The mutated I67T in *Snap25b* forms two additional hydrogen bonds within the SNARE complex which is predicted to increase its stability and prevent efficient recycling. Figure obtained from (Jeans et al., 2007).

of the activity profiles of schizophrenic patients and supports the link between circadian biology and psychiatric disorders (Figure 1.8).

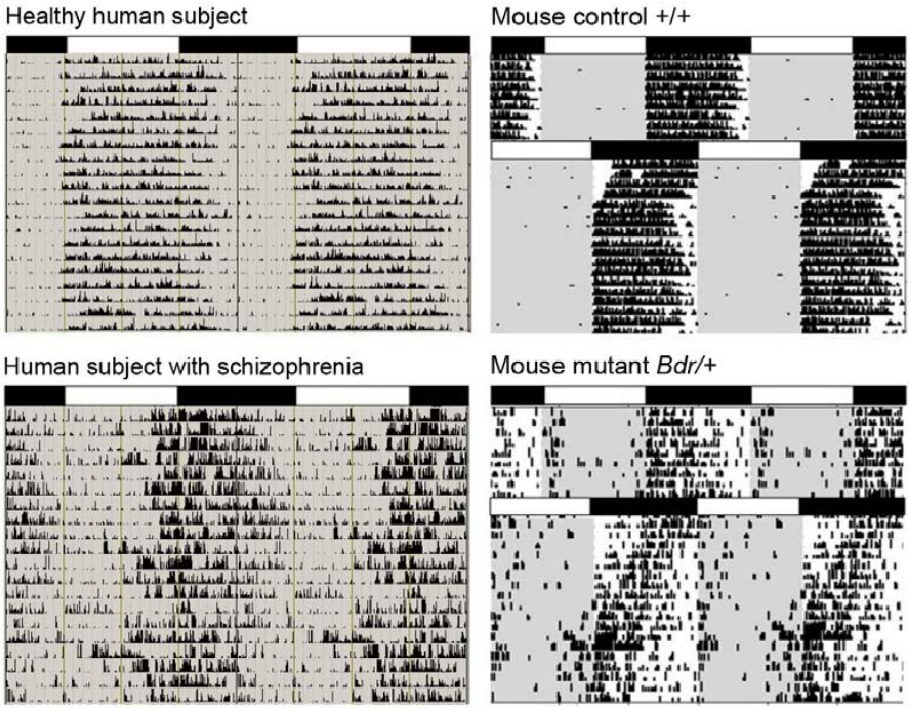


Figure 1.8. Comparison of human and mouse activity profiles.
Human schizophrenia patient displays a disrupted rest/activity profile, similar to the *Bdr* mouse mutant which displays a fragmented and phase-advance rest/activity profile. Image adapted from (Oliver et al., 2012).

To identify the genes or pathways which induced these circadian deficits, microarray expression profiling from SCN-enriched 1mm tissue punches was assessed at three time points: ZT6 (mid-light phase), ZT11 (late light phase) and ZT13 (early dark phase), as these times showed the greatest difference in locomotor activity between genotypes. A few key genes showed highly significant temporal changes including the SCN neuropeptides arginine vasopressin (*Avp*), neurotensin (*Nts*) and tachychinin1/substance P (*Tac1/Sp*).

qPCR, *in situ* hybridisation (ISH) and immunohistochemistry validated elevated *Avp* expression at ZT6 vs ZT11 in *Bdr* mice. Peripheral blood corticosterone levels - a peripheral rhythm under SCN control - was also measured and this, like *Avp* production and wheel-running behaviour, was shown to be phase-advanced (Oliver et al., 2012).

Despite unravelling some of the biology linking the *Snap25* mutation to the circadian and behavioural phenotypes, only a small number of genes were found to be differentially regulated. A more detailed analysis with greater temporal or anatomical resolution may provide additional clues to the *Bdr* phenotype and further elucidate the link between SCRD and psychiatric disease.

1.2 Transcriptomics

1.2.1 What is a transcriptome?

The transcriptome can be defined as the set of all RNA molecules present in a single or population of cells. The transcriptome is dynamic and has been shown to depend on a multitude of factors. At the simplest level, transcriptomes are stochastic due to the intrinsically random nature of mRNA synthesis: genes transition between active and inactive states, and gene activation induces burst-like expression accounting for the great variety of mRNA levels seen at a single cell level (Raj et al., 2006). These stochastic bursts of expression, however, are buffered at the protein level due to slower rates of protein degradation (Raj et al., 2006). The transcriptome also varies systematically across time whether it be developmental (Trapnell et al., 2010, Yu et al., 2014), circadian (Zhang et al., 2014, Hughes et al., 2012a) or at different stages of the cell cycle (Rustici et al., 2004). The transcriptome is also responsive to external factors; a change in environment or internal signals induce molecular changes which may be observed through the transcriptome (Kohli et al., 2015, Kogel et al., 2015, Cho et al., 2015).

The transcriptome contains a variety of different RNAs based upon structure and function. Messenger (m)RNA allows the genetic code to be carried from the DNA to the ribosomes allowing protein synthesis. These transcripts possess a start codon (specific nucleotide triplet) which, when surrounded by the ribosome, initiates protein synthesis, and a stop codon, which terminates protein synthesis (Cavener and Ray, 1991). The length of nucleotides upstream of the mRNA start codon is termed the 5' untranslated region (5' UTR) and the length of nucleotides downstream of the stop codon is termed

the 3' untranslated region (3' UTR) (Derrigo et al., 2000). RNA which does not code for protein are termed non-coding (nc)RNA and possess a variety of functions. During protein synthesis, both ribosomal (r)RNA and transfer (t)RNA are essential. tRNA are responsible for transferring a specific amino acid to the peptide chain encoded by the mRNA at the ribosome, and rRNA, the most abundant RNA in the eukaryotic cell, forms part of the ribosome and assists the catalysis of protein synthesis (Kapp and Lorsch, 2004).

Other ncRNA also form complexes with proteins. Small nuclear (sn)RNA form small nuclear ribonucleoprotein complexes (snRNPs) which play a role in splicing (Yean et al., 2000), and small nucleolar (sno)RNA form snoRNPs which mediate the modification of target RNA (Kiss, 2002).

There are also numerous ncRNA which play key roles in the regulation of mRNA. Many of these are of short length (21-31nt). Micro (mi)RNA (21-22 nt) are able to bind to complementary mRNA and prevent translation, accelerate degradation or induce cleavage through RNA interference (Ambros, 2004). Small interfering (si)RNA (20-22nt) also affect the expression of mRNA with complementary sequences but unlike miRNA which folds back onto itself to form a small hairpin, siRNA are produced from double stranded (ds)RNA (Elbashir et al., 2001). Lastly piwi-interacting (pi)RNA (26-31nt) form complexes with piwi proteins which are responsible for the silencing of retroposons in germ line cells (Siomi et al., 2011).

Non-coding RNA >200nt are defined as long non-coding (lnc)RNA which may exist in a variety of genomic contexts. LncRNA may overlap protein coding regions of DNA — these may be sense transcripts deriving from the same strand as the protein coding

gene or antisense deriving from the opposing strand. Up to 70% of sense transcripts show antisense expression which include many lncRNA (Carninci et al., 2005). LncRNA which reside in regions distant from protein-coding genes are referred to as long intergenic ncRNA (lincRNA). Finally, genes which over evolutionary time have lost their coding potential due to mutations are known as pseudogenes. Many of these pseudogenes resulted from gene duplication and have lost their selected pressure. Alternatively, pseudogenes may arise from retroposition (reverse transcription) of the parent gene's mRNA into the genome. Even though only 2-20% of pseudogenes are estimated to be expressed, many have important regulatory functions (Lafontaine and Dujon, 2010).

Unlike protein coding genes and small ncRNA, lncRNA exhibit weak conservation which is usually suggestive of non-functionality (Pang et al., 2006) and while a large proportion of lncRNA may be transcriptional noise, many important lncRNA display poor conservation (Brosius, 2005). This may be because ncRNA do not need to preserve the codon structure and are therefore less affected by substitution or frame-shift mutations. This means selection may only act on particular regions of the lncRNA depending on its functional role (Pang et al., 2006, Ponjavic et al., 2007). Poor conservation had been proposed to be a consequence of recent rapid evolution leading to the production of lineage specific ncRNA, such as X inactive-specific transcript (*Xist*) (Pang et al., 2006), however evidence suggests these effects are likely due to genetic drift in species with small effective population sizes (Haerty and Ponting, 2013). The roles of lncRNA vary substantially and act via a variety of mechanisms, although most tend to regulate mRNA levels (Figure 1.9).

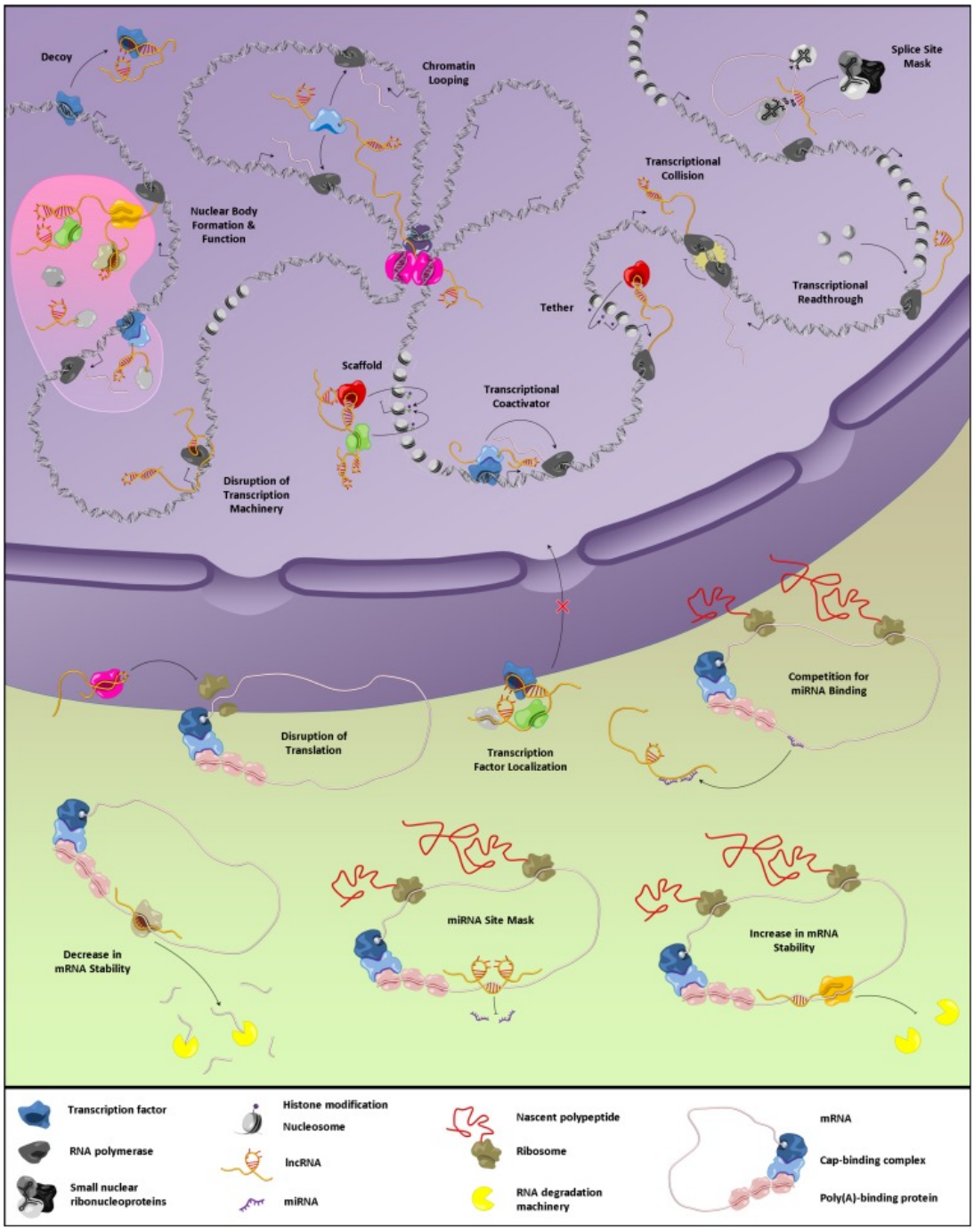


Figure 1.9. Functions of lncRNA
Schematic diagram showing the potential variety of lncRNA functions in regulating mRNA expression levels. Figure obtained from (Kung et al., 2013) with permission from the Genetics Society of America.

Increasing the complexity of the eukaryotic transcriptome, genes undergo alternative splicing, a process which allows the generation of multiple different mRNA transcripts from the same gene. Firstly, the gene is transcribed to produce a primary transcript which includes all introns and exons of the gene. Splicing normally removes introns and joins the exons together to form a mature mRNA product. Alternative splicing, however, can exclude particular exons from the final mRNA product allowing a variety of mRNA products to be produced (Black, 2003). This may result in protein products which due to their differing protein sequences have different functions, such as the *dsx* gene whose differing protein product determines the sex in *Drosophila* (Nagoshi and Baker, 1990).

1.2.2 Assessing the transcriptome

The emergence of transcriptomics occurred in the late 20th century when high throughput methods enabled a comprehensive study of the transcriptome. In 1995, the first microarrays were developed to allow the expression levels of many genes to be monitored in parallel (Schena et al., 1995). This 'probe-based' approach utilised the complementary hybridisation of mRNA to a set of sequenced nucleotide 'probes' on a chip. The relative degree of hybridisation to a probe can elucidate the expression level for that probe or gene.

The alternative to a probe-based approach is sequencing. These efforts were first initiated by Craig Venter in 1991 to generate expressed sequence tags (ESTs) from Sanger-sequencing over 600 randomly selected cDNA clones (Adams et al., 1991). These efforts have proven to be useful in processes such as discovering new genes, mapping of genomes or identifying coding regions in genomic sequence. In 1995, serial analysis of gene ex-

pression (SAGE) was developed and unlike previous methods could analyse thousands of transcripts simultaneously allowing quantification of gene expression on a global scale (Velculescu et al., 1995). In this method, the selected mRNA is used to synthesise cDNA cleaved into short "tags" of 11 nucleotides. These tags are ligated together to form cDNA concatemers to be amplified and sequenced. This method has been developed to contain fewer biases, be more robust and produce longer tags but are still based on the same protocol (Matsumura et al., 2003 , Gowda et al., 2004 , Wei et al., 2004).

The advent of next- (or second-) generation sequencing (NGS) revolutionised the field of transcriptomics by allowing the sequencing of cDNA at a massive scale (Mortazavi et al., 2008). Although the first RNA-sequencing (RNA-seq) paper published was using 454/Roche technology (Bainbridge et al., 2006), the Solexa (now Illumina) technology accelerated this revolution due to higher throughput and cheaper costs. In this process, RNA is extracted and poly(A) enriched using oligo(dT) beads. It is then fragmented and reverse transcribed into ds cDNA using random hexamers. Fragments of a particular length are selected for using gel purification. Adapters are then added to allow the cDNA to attach to the flowcell for sequencing (Figure 1.10). The cDNA nucleotide sequence is then determined by "sequencing by synthesis". In this process, a cDNA strand is used as a template to produce a complementary strand of cDNA, the dNTPs used are fluorescently tagged so when incorporated into the sequence they emit a particular colour which is received by a camera. The dNTPs are also modified to contain a terminator which prevents another base being added. When another cycle of synthesis occurs, it removes the cap and allows another fluorescently tagged base to be added, emitting another colour. The number of cycles determines the length of the read to be sequenced and the emission wave-

length and signal intensity determines the base call and confidence. In protocols which allow paired-end sequencing, the reverse strand is sequenced to determine the nucleotide sequence of the other end of the cDNA fragment (Figure 1.11). The resultant nucleotide sequences can then be mapped and analysed. A standard RNA-sequencing protocol is illustrated in (Figure 1.10).

Each sequenced read will consist of a length of nucleotides with an associated "Phred" quality score, which is a measure used to assess the reliability of a base call (Ewing et al., 1998, Ewing and Green, 1998). The larger the Phred score (Q) the more reliable the base call: Q is logarithmically related to the probability the call is incorrect (P) which can be described by the equation $Q = -10\log_{10}P$. For example, a Q of 10 (Phred10) has a base call accuracy of 90% (1 in 10 incorrect), Phred20 has a base call accuracy of 99% (1 in 100 incorrect) and Phred30 has a base call accuracy of 99.9% (1 in 1000 incorrect). If read-ends have a string of bases of mostly low quality (Q15 or below), then these quality scores are replaced with a value of 2. This value does not predict a specific error rate, but instead indicates that this segment of the read should be ignored in further analysis.

The RNA-seq method of assessing the transcriptome provides numerous advantages compared to microarrays (Wang et al., 2009, Shendure, 2008). Firstly, microarrays are necessarily tied to existing gene models. This means the portion of the transcriptome which has not been annotated, such as novel genes or exons, cannot be assessed. Furthermore, unless specifically designed, arrays are not able to capture differential splicing/ differential exon usage. RNA-seq can address these limitations and expose the full complexity of the transcriptome to a single nucleotide level. Secondly, the analogue nature of microarray signals makes it difficult to accurately detect expression differences for low

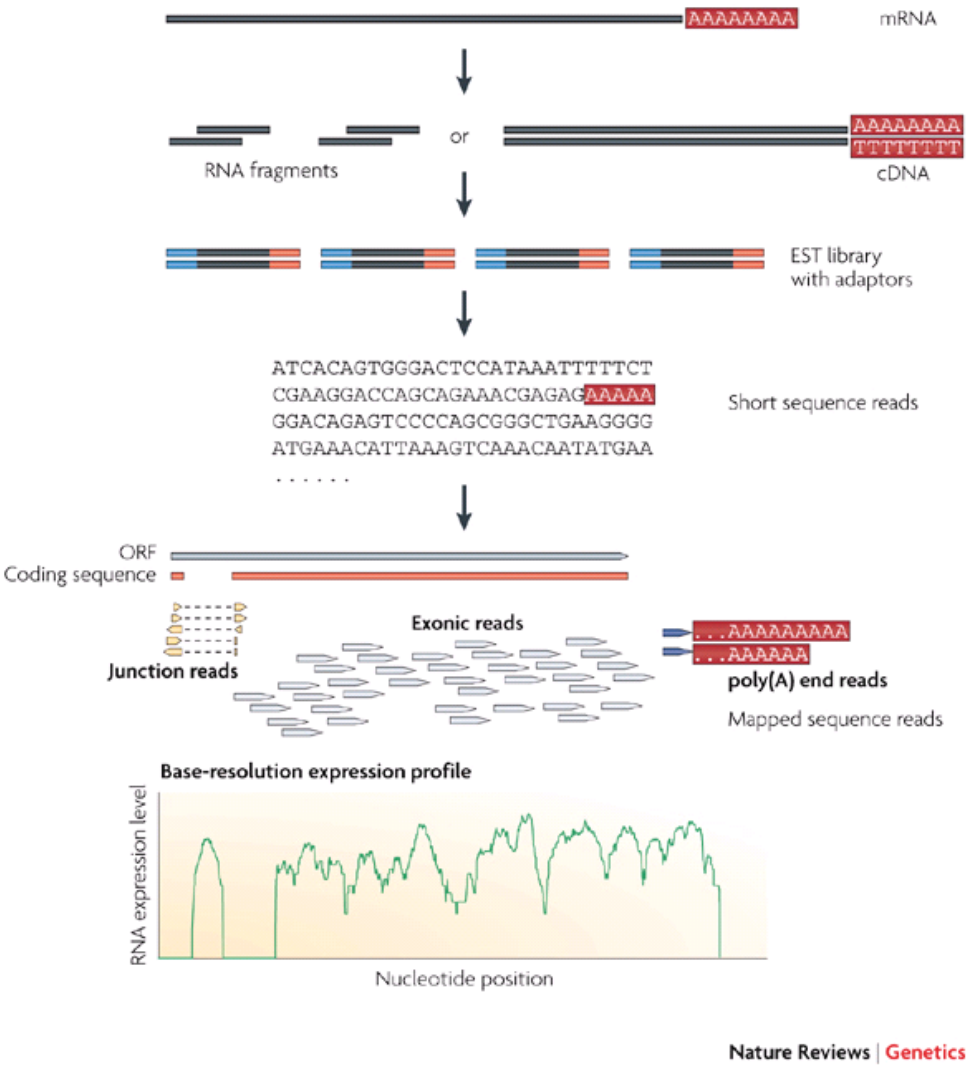


Figure 1.10. A typical RNA-seq protocol
In this process, RNA is extracted and poly(A) enriched. It is then fragmented and reverse transcribed into cDNA, or alternatively converted to cDNA before fragmentation. Adapters are added to allow the cDNA to attach to the flowcell and provide a site to allow sequencing. The resultant nucleotide sequences (reads) can then be aligned to the reference genome for analysis. Reads which align to a single exon are labeled "exonic reads" those which are aligned to multiple are known as spliced or "junction reads". The region with no reads mapping to it within the junction reads represents an intron which has been spliced out. Expression levels for the gene (or parts of a gene) can then be measured by counting the reads which map to it. Figure obtained from (Wang et al., 2009) with permission from the Nature Publishing Group.

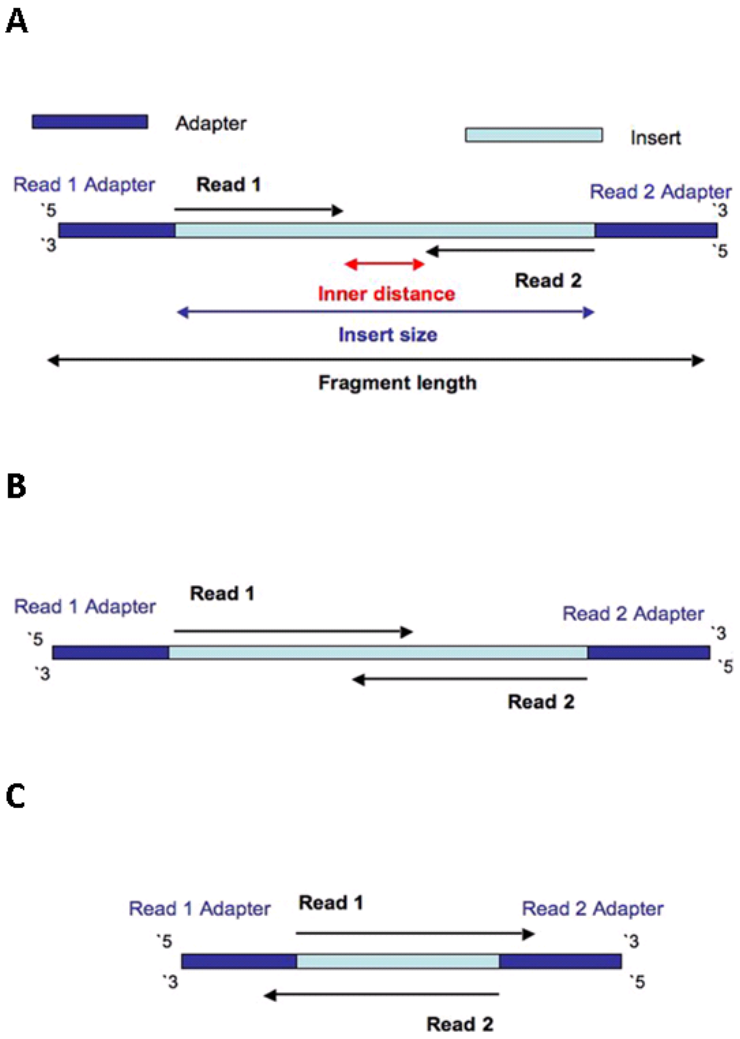


Figure 1.11. Paired-end sequencing of fragments of varying insert sizes.
(A) The insert size is greater that the total length of both reads sequenced (100bp x 2). This leaves an inner distance of the insert which has not been sequenced. **(B)** The insert size is greater than a read being sequenced but smaller than the total length of both reads (100bp $<$ x $<$ 100bp x 2). This results in an overlapping sequence from read 1 and 2. **(C)** The insert size is smaller than the length of a single read (100bp). This results in adapters being sequenced after the insert ends which may contaminate the sequencing data. Image adapted from (Turner, 2014).

abundant RNA. As RNA-seq expression is digital (based upon number of reads), it provides greater dynamic range, especially amongst the lowly and very highly expressed genes. Thirdly, probe hybridisation can be noisy which dampens the true expression profile. Furthermore, highly related sequences are difficult to accurately quantify due to cross-hybridisation, a problem better addressed using RNA-seq.

Despite the numerous advantages and popularity of RNA-seq, there still remain some caveats associated with the approach (Wang et al., 2009).

Biases may be introduced to RNA-seq data during library construction. When amplifying the RNA using PCR, particular biases may exist which preferentially amplify some fragments but not others. For example, there is a known GC bias — both GC- and AT- rich fragments are underrepresented, thought to be due to PCR poorly amplifying these fragments (Benjamini and Speed, 2012, Dohm et al., 2008). Random hexamer priming at the cDNA synthesis stage also introduces biases as an overrepresentation of reads with particular nucleotide contents is often observed for the first 13 nucleotides of the sequenced read (Hansen et al., 2010). RNA-seq also displays non-uniform coverage of expressed transcripts dependent on the fragmentation method. RNA-fragmentation shows no bias of reads across the transcript body but shows depletion at transcript ends, whereas DNA fragmentation shows a bias with more reads towards the 3' end of the transcript (Mortazavi et al., 2008, Nagalakshmi et al., 2008). These biases pose a problem when comparing between genomic regions within the same sample. However, if these biases are consistent for all samples then comparisons between samples should be valid. Unfortunately, it is possible for biases to differ between samples, therefore these factors need to be carefully assessed. If samples are processed in different groups, it is possible to introduce

"batch-effects" — systematic global differences between the batches. This is because the biases discussed (and many others) are more likely to differ when not processed together (Leek et al., 2010).

Although RNA-seq can largely overcome the array-specific problem of cross hybridisation between highly similar genes, there may still be similar issues of mapping reads to regions of the genome that are highly similar or identical. These highly similar regions in the genome may include conserved domains of duplicated genes or retroposons (repetitive DNA elements) such as small or long interspersed repeats (SINEs and LINEs). Several GAPDH retroposed pseudogenes in the mouse genome differ from the GAPDH mRNA in sequence by only 0.2%, therefore reads from the GAPDH mRNA may map to both the original gene and the pseudogene (Pepke et al., 2009). Reads that map equally well to multiple genomic locations are referred to as multireads. Furthermore, due to error rates in sequencing, it is possible for highly expressed genes to produce a shadow of artefactual expression at these pseudogenes (Pepke et al., 2009).

1.2.3 Circadian transcriptomics

The core clock proteins (eg. *Clock, Bmal1, Per1*) induce rhythmic expression of downstream clock controlled genes (CCGs) which can impose rhythmicity on cellular, physiological or behavioural processes (Hastings et al., 2003). Many of these genes have been found to be involved in rate limiting steps of various pathways and therefore allow particular molecular functions to occur at specific times of the day when it is most beneficial to the organism (Panda et al., 2002). Despite the conservation of elements of the molecular

clock between tissues, CCGs are generally tissue specific as each tissue often requires a different set of genes to be temporally regulated or expressed.

Initial microarray studies attempted to determine rhythmic CCGs in various tissues and species to better understand circadian processes and the genes responsible for driving them. In 2002, two groups attempted to identify the CCGs in the SCN. Ueda *et al.* used microarray technology to profile SCN tissue punches extracted from the mouse every 4 hours over 2 days and identified 101 genes which cycled (showed significant sinusoidal expression over 24 hours) in both LD and constant darkness (Ueda et al., 2002a). They then identified the transcription start sites of these genes to determine the transcription factor response elements responsible for generating circadian rhythms. Rev-ErbA/ROR response elements were found to be important in generating expression rhythms which peaked at night in phase with *Bmal1* expression and antiphase to *Per2* expression. Secondly, Panda *et al.* used microarrays to identify nearly 650 genes which cycled in the SCN under constant darkness to help determine the pathways in the SCN important for coordinating the physiological and behavioural rhythms of the organism (Panda et al., 2002). A large number of these rhythmic genes were found to be involved in peptide synthesis and the regulation of redox state. This suggests the importance of these processes being temporally regulated in the SCN (Panda et al., 2002).

In an effort to elucidate the mechanisms of circadian control, numerous studies have disrupted specific genes and measured how the transcriptional output is affected. For example, a *Dicer* knockout in *Drosophila* displayed disrupted circadian patterns, therefore implicating miRNA in maintaining robust circadian rhythms (Kadener et al., 2009). One study engineered a mouse strain with a conditionally active liver clock, only active in the

presence of doxycycline. 31 genes, including *Per2*, were identified as systemically driven, as they continued rhythmic expression even in the absence of a core hepatic oscillator (Kornmann et al., 2007). A similar experiment specifically restored *Clock* function in the brains of *Clock*Δ*19* mice. The rescued mice displayed normal locomotor behaviour and possessed many more cycling transcripts in the liver compared to the *Clock* mutants. However, the majority of the liver expression output was dampened or absent in the rescue mice suggesting that although the SCN can restore some rhythmicity, the majority of peripheral circadian expression depends on the local circadian oscillator (Hughes et al., 2012b).

RNA-seq also allows the profiling of circadian gene expression, providing many advantages compared to microarray technology as previously described (section 1.2.2). RNA-seq of *Arabidopsis thaliana* seedlings found that circadian clock genes often express mRNA isoforms which possess in-frame premature termination codons. The conservation of these alternative splicing events may allow the regulation of the relative abundances of these isoforms and therefore suggests a role for these non-coding alternative transcripts in clock function (Filichkin and Mockler, 2012a). A recent study sought to characterise the circadian transcriptome in 12 different mouse tissues using both microarrays and RNA-seq. 43% of all genes were found to display circadian expression somewhere in the body and many of these genes peaked at time periods near to dawn and dusk, times labelled as transcriptional "rush hours". Furthermore, they were able to identify rhythmic expression of more than 6,000 known and novel ncRNA (Zhang et al., 2014).

Next generation sequencing has also allowed the generation of additional methods which have been used to investigate circadian biology. Nascent-seq allows genome wide

sequencing of the RNA transcribed by RNA Polymerase II prior to 3' end formation. Comparison of gene expression measured by both Nascent-Seq and RNA-seq provides a method to assess the importance of post-transcriptional regulation in driving circadian mRNA expression (Rodriguez et al., 2013, Menet et al., 2012). For example, 70% of rhythmically expressed genes were found not to display *de novo* transcriptional rhythms, suggesting that post-transcriptional regulation plays an important role in driving expression levels (Menet et al., 2012). ChIP-seq, a technique used to analyze protein interactions with DNA, has been used to assess the contribution of transcription factor binding to the regulation of transcriptional rhythms. This can help to identify the genes that are under the control of a particular transcription factor, helping to better understand circadian regulation (Koike et al., 2012, Menet et al., 2012).

1.3 Basis for investigation

The mammalian SCN is the master pacemaker of the circadian system, responsible for driving the endogenous 24-hour rhythms observed in an organism (section 1.1.1). A detailed comprehension of this brain region is therefore important in truly understanding how these circadian rhythms are generated. Although proteins are often considered to be the main functional component of the cell, determining changes in gene expression on a large scale can provide a valuable measure of tissue or cellular physiology as well as indicate potential changes in protein expression. Furthermore, over the last 50 years, RNA is no longer solely seen as only a messenger between DNA and protein (Ponting et al., 2009, Amaral et al., 2008). The pervasiveness and variety of non-coding RNA transcripts suggest an additional layer of regulation with many transcripts possessing important biological functions (Hall et al., 2015, Penny et al., 1996). Measuring the temporal transcriptome of the SCN thus might identify transcripts important for circadian function, many which may be clock-controlled. Early microarray studies of the SCN identified genes and processes displaying circadian expression patterns providing insight into the SCN and circadian biology (Panda et al., 2002, Ueda et al., 2002a, Hatori et al., 2014). However, these particular studies possess some features which may limit a complete understanding of the SCN transcriptome.

Firstly, microarray technology for gene expression profiling is becoming technologically outdated with RNA-seq fast becoming the preferred option (section 1.2.2). Although Panda *et al.* identified 650 cycling genes, this number will have been limited by the number of genes surveyed. Microarrays at the time were limited by the number of features, or probe-sets, that could be accommodated, in addition to the number of suitably annotated

protein coding genes (7,000) and 3,000 expressed sequence tags (ESTs) available to survey (Panda et al., 2002). RNA-seq, as a result of more sophisticated genome annotation, is able to survey the expression of up to 35,303 protein and non-coding annotated genes (most recent Ensembl update). RNA-seq is also able to survey regions of the genome which have not yet been annotated, which may allow the identification of novel genes or alternative transcripts which play a key role in that tissue or condition. Panda *et al.* discovered that some genes that were expected to have a strong circadian component, such as *Per2*, did not fluctuate in their microarray dataset. This is likely due to the low signal-to-noise ratio of the representative probesets on the array (Panda et al., 2002). RNA-seq provides a greater dynamic range than microarrays offering a more accurate measurement of effect size (Mortazavi et al., 2008). Furthermore, this also allows RNA-seq to outperform microarrays in identifying true differentially expressed genes, especially amongst low abundance transcripts (Wang et al., 2009, Wang et al., 2014).

Secondly, previous array studies collected SCN "punches" for their experimental samples. These tissue samples, typically 1mm in diameter from a 1mm thick section, consist of perhaps only 30% SCN with the remainder made up of the surrounding hypothalamic tissue. The study of such heterogeneous samples will result in reduced amplitudes of oscillations for genes that cycle asynchronously between the SCN and surrounding regions (Zhang et al., 2014). Therefore, to improve upon these previous investigations, a combination of RNA-seq with highly anatomically precise extraction of the SCN using laser capture microdissection (LCM) was used to accurately assess the transcriptome of this tissue. LCM has previously been used to extract SCN tissue (Porterfield et al., 2007), other hypothalamic nuclei (Lee et al., 2012) and even single cells for further analysis

(Nakamura et al., 2007, Euler et al., 2012). Current RNA amplification techniques now allow the limited RNA yields from these extracts to be surveyed by RNA-seq (Tariq et al., 2011).

Finally, circadian transcriptomic surveys of mammalian systems, to date, have mostly utilized constant lighting conditions to focus on the endogenous free-running circadian clock and have not assessed more physiologically-relevant light / dark cycles (Zhang et al., 2014, Panda et al., 2002, Hughes et al., 2009 , Hurley et al., 2014). Therefore, in this study mice were housed under LD conditions (each day consists of lights on for 12 hours followed by lights off for 12 hours) to allow the identification of physiologically relevant cycling genes or light-induced transcriptional cascades that would otherwise be undetectable as cycling under constant conditions. SCN samples were extracted at six timepoints across the circadian day (ZT2, 6, 10, 14, 18 and 22) to capture these specific circadian variations.

As previously discussed (section 1.1.5), circadian rhythm disruption is associated with psychiatric diseases such as schizophrenia. To help elucidate the molecular link between circadian disruption and psychiatric disease, the *Bdr Snap25* mutant was further investigated. *Bdr* is of particular relevance to this question as it displays both schizophrenia endophenotypes and a particular pattern of rest/activity behavioural disruption. Previously, microarray analysis of SCN tissue punches from *Bdr* and littermate control mice was carried out at 3 timepoints (ZT6, 11 and 13) and surprisingly few expression differences were identified (Oliver et al., 2012). As large transcriptomic disturbances were not observed in *Bdr*, it may signify that differences were not present at these specific timepoints or that the anatomically imprecise SCN extracts were dampening SCN specific tran-

scriptomic differences. Additionally, these disturbances may also derive from non-coding RNA which were not measurable using microarrays. To overcome these limitations, SCN samples extracted by LCM were collected from *Bdr* at the same six timepoints as the non-mutant mice (ZT2, 6, 10, 14, 18 and 22) to allow a more complete investigation into the transcriptomic disruptions in this mouse model.

1.4 Aims of this project

1 To identify novel transcripts with putative function in the SCN or circadian rhythms:

(a) To identify the regions of the genome which possess robust expression but are not accounted for by current genome annotations. I will focus on the identification of novel lincRNA and novel exons of previously identified genes for reasons described later (section 4.2).

(b) To predict the functions of these novel transcripts. By investigating properties of these transcripts, including temporal expression, conservation and nucleotide sequence, I will determine whether they show features indicative of certain functions.

2 To carry out a comprehensive examination of the SCN transcriptome over 24 hours from wild-type mice:

(a) To determine the genes and processes which display temporal fluctuation. I will measure the expression for every coding and non-coding transcript in the genome across the day and apply statistical algorithms to determine whether they display significant fluctuation across time. To identify processes under temporal regulation, I will determine whether genes with particular functions or properties are more likely to display specific expression patterns.

(b) To make the sequencing data available to the scientific community via a web resource. This will provide researchers information about the expression pro-

files of all genes in the SCN over circadian time. This SCN RNA-seq project will yield an abundance of information and making it available to the common pool of knowledge will only help the progression of this field.

3 To understand the relationship between circadian behaviour and SCN expression in *Bdr*:

(a) To identify the genes and pathways disrupted in the *Bdr* model. By comparing the circadian transcriptome of the *Bdr* model to the non-mutant, I will identify the genes which are differentially expressed. To identify disrupted pathways, I will determine whether the differentially expressed genes share particular functions or properties.

(b) To confirm the pathways which play a role in both circadian rhythms and psychiatric disease. I will experimentally investigate these pathways using behavioural or pharmacological testing.

At the conclusion of this project, we wished to have gained novel insight into both SCN and circadian biology. We will have identified novel transcripts which may play key roles in SCN function and have established the circadian transcriptome of the SCN, those coding and non-coding transcripts which display specific temporal fluctuations. Furthermore, this project will provide the foundation to understand the phenotypic changes seen in the *Bdr* mouse model. Identification of the genes dysregulated in *Bdr* will provide insight into the pathways which link circadian rhythms and psychiatric disease, hopefully allowing the research and development into the treatment of human diseases such as schizophrenia.

1.5 Overview of thesis

In this thesis, I present a comprehensive transcriptomic investigation of the SCN using RNA-seq. Additionally, I propose molecular mechanisms which may underpin both psychiatric disease and circadian rhythm disruptions. Data from chapters 2-4 have been published (Pembroke et al., 2015).

In chapter 2, I discuss methods used to generate and validate the mouse SCN transcriptome over 24 hours. I then show how comparing the SCN transcriptome to publically available transcriptomic data can yield biologically important information through the identification of SCN-enriched genes.

In chapter 3, I investigate the SCN transcriptome on a temporal scale. Doing so, I identify a set of genes which display "twin-peaking" expression over circadian time which appear to provide susceptibility windows allowing light-induced phase-shifting of the circadian clock. To conclude this chapter, I discuss the "SCNseq" data explorer, a web-based application which will hopefully assist the circadian community.

In chapter 4, I demonstrate the power of RNA-seq in identifying novel transcripts expressed in the SCN. I identify 1,013 novel isoforms of previously annotated genes and discuss their potential importance using a novel isoform from a clock gene as a case study. I also present the identification of hundreds of novel (and known) lincRNA which may play important roles in SCN function.

In chapter 5, I propose mechanisms by which circadian disruption and psychiatric disease are linked at the molecular level through the blind-drunk mouse model. Using similar analyses described in chapters 2-4, I identify the genes which are disrupted in *Bdr*

and then using behavioural analysis, attempt to verify the proposed molecular mechanisms disrupted in *Bdr*.

In chapter 6, I discuss the main conclusions from this thesis and future perspectives.

2 Generation and validation of the mouse SCN transcriptome over 24 hours

2.1 Introduction

The mammalian SCN is the master pacemaker of the circadian system, responsible for driving the endogenous 24-hour rhythms observed in an organism. Although the SCN has been subject to much research in the circadian community, there is relatively little SCN transcriptomic information available, especially on a large scale. Assessing the transcriptome can provide a valuable measure of tissue or cellular physiology as well as indicate changes in protein expression. Investigating the temporal transcriptome of the SCN is thus important to identify the circadian and clock-controlled genes and processes, providing a greater understanding of how circadian rhythms are generated. Microarray technology was previously utilised to identify the circadian genes and processes in the SCN (Panda et al., 2002, Ueda et al., 2002a, Hatori et al., 2014). More recently, the SCN had been sequenced at a single time point but was not of sufficient depth or complex experimental design to allow a comprehensive analysis of the SCN transcriptome (Azzi et al., 2014). To gain a greater understanding of the SCN, this region was accurately dissected in mouse using LCM at six timepoints over 24 hours (ZT2, 6, 10, 14, 18, 22) and its transcriptional profile was measured using RNA-seq. Once the mouse SCN transcriptome was generated and validated, it formed the basis for further analyses within this thesis.

RNA-seq is increasingly being used to gain a greater understanding of tissues and cell types through analysis of their transcriptomes, providing novel biological insight not pre-

viously possible using microarray technology (Watanabe et al., 2014, Belgard et al., 2011, Filichkin and Mockler, 2012b). Although there are many advantages to RNA-seq over microarrays, there are also a number of caveats involved. As dealing with large datasets, processing of the raw data can be time consuming, computationally expensive, and require a detailed understanding of bioinformatics. Furthermore, RNA-seq may introduce distinct biases into the data which need to be considered before making biological conclusions. Each stage of the RNA-seq workflow (Figure 1.10), such as the fragmentation, amplification or sequencing of the RNA, has the potential to introduce biases not observed with microarrays. Furthermore, an understanding of how to match these sequenced reads to expression levels or other biological features of interest are important to make correct biological conclusions. Although there has been, and still is, much research into these processes and many methods to resolve biases in the data, the mantra: "garbage in; garbage out" is very applicable to RNA-seq data analysis (Tiwari and Schork, 2011, Hartley and Mullikin, 2015). Therefore a detailed understanding of the data and how it is processed is important before making robust biological conclusions.

In this chapter, I sought to generate the mouse SCN transcriptome over 24 hours using RNA-seq. The precise plans were to:

- Generate the mouse SCN transcriptome by processing the raw sequencing data and mapping reads to the genome.
- Assess and remove any technical biases incorporated into the sequencing data.
- Validate the reliability of the SCN transcriptome by comparing expression levels to positive controls and expression levels generated through qPCR
- Identify SCN-enriched genes by comparing expression levels from the generated SCN transcriptome to expression levels generated from additional publically available sequencing data.

2.2 Methods

2.2.1 Animals

The C3H/HeH mouse strain was chosen for investigation in this project because the *Bdr* mutant (Chapter 5) was backcrossed to this strain and it allowed analysis to be consistent with previous studies in our group (Oliver et al., 2012). This strain possesses strong melatonin rhythms but is homozygous for the phosphodiesterase 6B (*Pde6b*) rd1 allele which results in early onset severe retinal degeneration (Han et al., 2013). Despite this deficiency in sight, these mice are responsive to light and shift their circadian phase according to light pulses (Benloucif and Dubocovich, 1996). In addition, the Davies group have demonstrated that light:dark entrainment is maintained even at low (30 lux) light levels (Oliver et al., 2012).

The blind-drunk (*Bdr*) mutant was generated at the Mammalian Genetics Unit, MRC- Harwell (UK) via ENU mutagenesis (Nolan et al., 2000, Nolan et al., 2002) and has been backcrossed to a C3H/HeH background for over 20 generations.

2.2.2 SCN RNA sample preparation

Male WT and *Bdr* littermate mice at 10-12 weeks of age were group housed with their respective genotypes in light-tight chambers equipped with LED lighting at 150 lux at the cage floor. After 7 days of acclimatisation, mice were singly housed for 7 days prior to tissue harvesting. Brains were removed at one of six zeitgeber times (ZT2, 6, 10, 14, 18, 22) and immediately frozen on dry ice in OCT mounting media (VWR, Lutterworth, UK). Eighteen frozen coronal sections at 15 μM were cut spanning the entire rostral to

caudal region of the SCN and mounted onto polyethylene napthalate (PEN) slides (Carl Zeiss Ltd. Cambridge, UK). Laser-capture microdissection (LCM) was carried out from dehydrated, Nissl-stained sections using the PALM system (Carl Zeiss Ltd. Cambridge, UK) as previously described (Figure 2.1) (Dulneva et al., 2015). Tissue from five animals was pooled to generate one biological replicate sample and total RNA was purified using the RNeasy micro kit (Qiagen, Manchester, UK). RNA quality was determined using an RNA Picochip (Agilent, Stockport, UK), with RNA integrity number (RIN) values over 8 for all samples and average yield of approximately 10ng per replicate. Tissue collection, LCM and RNA preparation was carried out by Peter Oliver and Arran Babbs at the start of my PhD project.

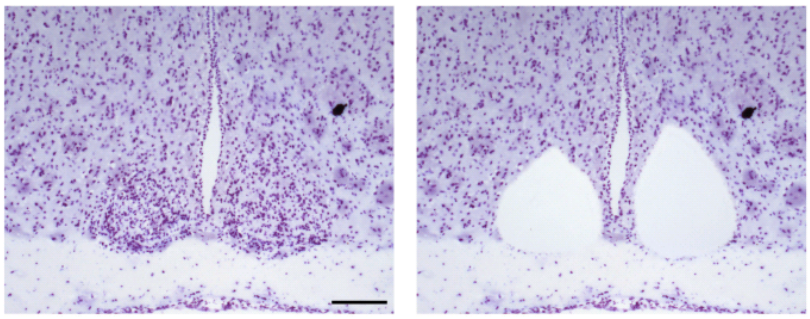


Figure 2.1. Laser capture microdissection of SCN.
Representative Nissl-stained brain sections before and after laser-capture microdissection of the SCN. Scale bar: 0.5mm.

2.2.3 Sequencing

Approximately 1ng of RNA was amplified using the SMARTer protocol (Takata Clontech Bio Europe SAS, France) and the resulting libraries were 100bp paired-end sequenced

on the HiSeq (Figure 2.2; Illumina, Fulbourne, UK; Wellcome Trust Centre for Human Genetics Sequencing Core). The average read pair count obtained was ∼35M per technical replicate (∼100M per biological replicate). RNA libraries were prepared and sequenced in three separate groups: ZT6 & 10 (∼ December 2012), ZT14 & ZT18 (∼ December 2013) and ZT22 & ZT2 (∼ June 2014).

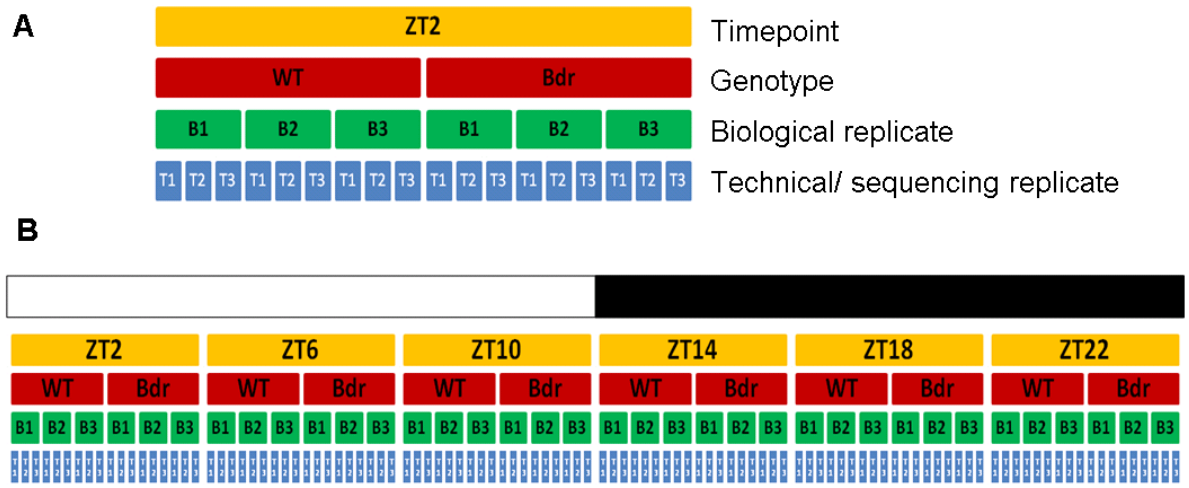


Figure 2.2. RNA-seq experimental design.
(A) Experimental design from a single timepoint. **(B)** Full experimental design. SCN samples were extracted from both WT and *Bdr* genotypes at six different timepoints over a 24-hour period (ZT2, 6, 10, 14, 18, 22). Each genotype at each timepoint has 3 'biological replicates' which comprise of pooled SCN samples from 5 mice. Each biological replicate is multiplexed allowing it to be sequenced over three separate lanes in order to increase sequencing read depth.

2.2.4 Acquisition of additional RNA-seq data sets

One SCN, six brain and five liver transcriptome datasets from six independent studies (GSE54124, GSE30352, GSE54652, GSE41637, GSE41338, and GSE36874) were down-

loaded from the Gene Expression Omnibus (GEO) database to allow comparison of expression between tissues (Table 2.1).

2.2.5 Read processing

The quality of the initial sequencing data was assessed using fastqc software (fastq — sequence file format; qc — quality control). The following observations were clear from the data:

1. The second read (the read pair of the transcript which was sequenced second) tended to possess a worse sequence quality than the first (Figure 2.3A-B).

2. The second read was more likely to possess a contaminant adapter sequence than the first.

3. The batch containing ZT6 and ZT10 had second read pairs with very low base quality (< Phred 20), especially from the 60-100bp region of the read. Sequencing accuracy is known to decrease as the number of read cycles increase causing this decrease in quality of read ends (Figure 2.3B). This sequencing batch also displayed higher levels of contaminant adapter sequences.

4. There appeared to be a correlation between poor base quality, increased adapter contamination and increased sequence duplication levels.

Group	Tissue	Additional info	Sequencing info	GEO / Reference
SCN	SCN	Brains sectioned at 15 µm and SCN dissected using LCM PolyA selected Sample pooled from 5 mice	PE 100bp	GSE72095 (Pembroke et al., 2015)
SCN	SCN	Brains isolated at CT4 and sectioned with a microtome (250 µm) SCN dissected with binocular microscope PolyA selected & rRNA depleted	SE 100bp	GSE54124 (Azzi et al., 2014)
Brain	Whole brain	PolyA selected	PE 50bp	GSE41637 (Merkin et al., 2012)
Brain	Whole brain	PolyA selected	PE 72bp	GSE41338 (Barbosa-Morais et al., 2012)
Brain	Cortex	PolyA selected Sample pooled from 3 mice	SE 76bp	GSE30352 (Brawand et al., 2011)
Brain	Cerebellum	PolyA selected	SE 76bp	GSE30352 (Brawand et al., 2011)
Brain	Cerebellum	PolyA selected	PE 100bp	GSE54651 (Zhang et al., 2014)
Brain	Hypothalamus	PolyA selected	PE 100bp	GSE54651 (Zhang et al., 2014)
Liver	Liver	PolyA selected	SE 76bp	GSE30352 (Brawand et al., 2011)
Liver	Liver	PolyA selected	PE 50bp	GSE41637 (Merkin et al., 2012)
Liver	Liver	PolyA selected	PE 72bp	GSE41338 (Barbosa-Morais et al., 2012)
Liver	Liver	PolyA selected	PE 100bp	GSE54651 (Zhang et al., 2014);
Liver	Liver	PolyA selected	PE 100	GSE36874 (Menet et al., 2012)

Table 2.1. RNA-sequencing data used in this thesis.
Comparison of expression data between the different groups (SCN, brain and liver) allowed the identification of SCN-enriched genes. PE — Paired-end, SE — Single-end.

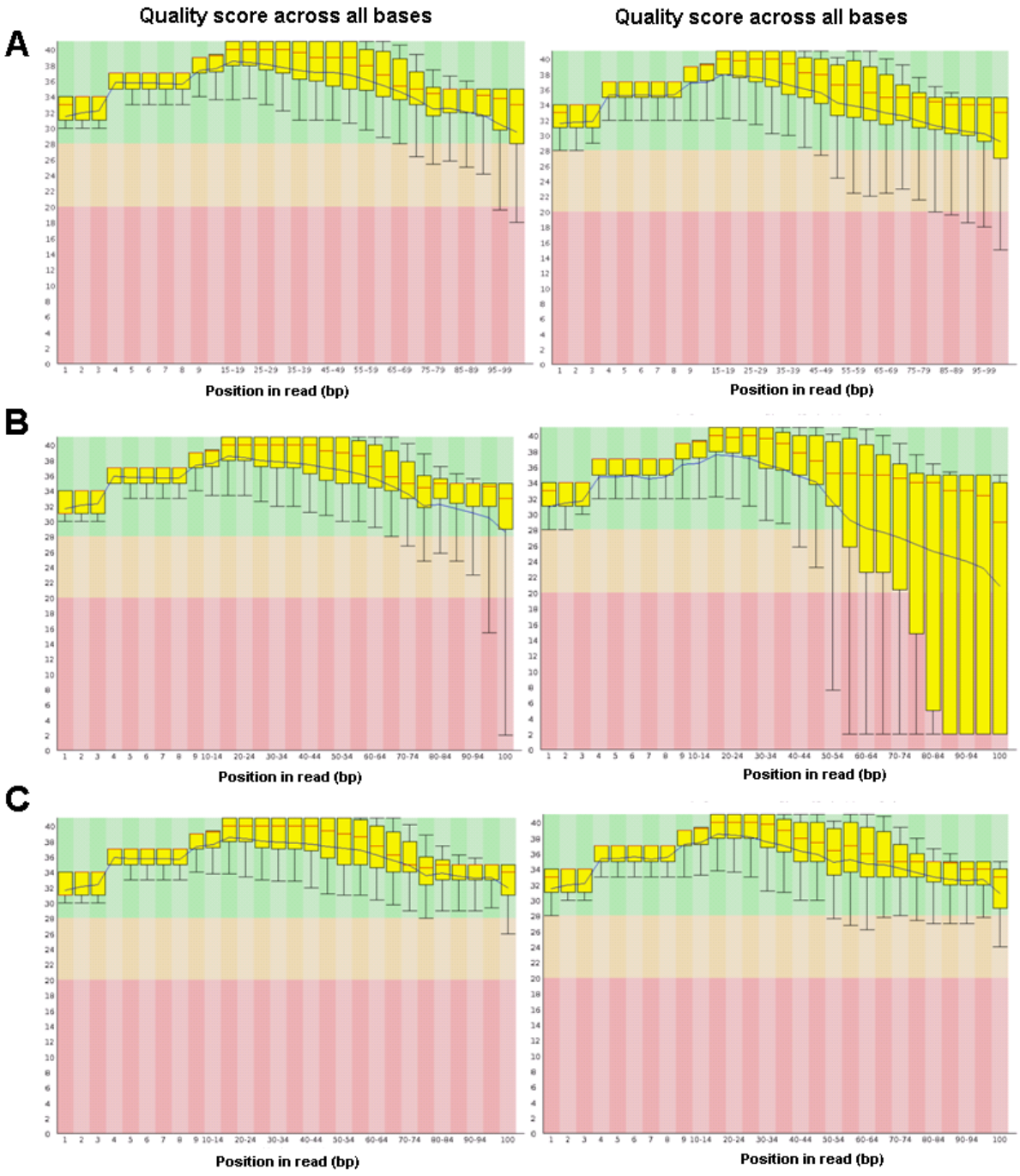


Figure 2.3. Representative quality scores for sequencing data before alignment.
Average Phred scores across the first (left) and second (right) paired-end read for: **(A)** ZT14 replicate **(B)** ZT6 replicate, and **(C)** ZT6 replicate after trimming of low quality read ends.

To improve the quality of the sequencing data before mapping the following steps were executed using an in-house script shared by Sergei Maslau.

1. Trim ends of reads with a Phred base quality score below 3 (Figure 2.3C).

2. Remove specified over-represented adapter or contaminating sequences. The specified sequences were:

   (a) AAGCAGTGGTATCAACGCAGAGTACATGGG

   (b) AAGCAGTGGTATCAACGCAGAGTAC(T+)

   - These are adapters for the SMARTer DNA amplification protocol and have been reported to contaminate RNA-seq datasets (Adiconis et al., 2013). These adapters largely contaminated the second paired-end read.

   (c) GATCGGAAGAGCACACGTCTGAACTCCAGTCACT GCCGGCTATCTCG-TATGCCGTCTTCTGCTTG(A+)

   - This is the sequence for the Illumina TruSeq Adapter 6. These adapters mainly contaminated the first paired-end read.

   (d) AAAAAAAAAAAAAAAAAAAA(A+)

   - Low complexity read which was generally over represented in the sequencing data.

3. Remove reads with a length below 50bp (or 40bp if initial sequencing length was shorter (Merkin et al., 2012, Barbosa-Morais et al., 2012)).

- The trimming of low quality sequences and adapters may have reduced the length of particular reads. Reads shorter than 50bp are removed to prevent inaccurate mapping.

4. Remove reads with an average Phred quality score below 30.

    - This decreases the number of reads which have been inaccurately called, reducing erroneous mapping.

Fastqc revealed that the processed sequencing data was of a high quality and suitable for mapping to the genome.

2.2.6 Mapping

The processed sequencing reads can either be mapped to a reference genome (Chapter 2) or assembled to create a transcriptome (Chapter 4). The purpose of a mapper is to identify the true location in the genome that the sequenced read derives from. Mapping must also consider possible polymorphisms or sequencing errors and is therefore only approximate.

In this project, Genomic Short Nucleotide Alignment Program (GSNAP; (Wu and Nacu, 2010)) was used to align the sequencing reads to the genome. This method allows reads to map over exon/intron boundaries (spliced read alignment) and is well suited for paired-end sequenced reads of this length. Furthermore, GSNAP performed consistently well when compared against other aligners for attributes such as base-wise accuracy and junction detection accuracy (Grant et al., 2011). However, GSNAP tends to be computationally expensive so can take longer than other aligners. Although there are many

mapping algorithms available, the biological outputs are often very similar and therefore mapper choice is often down to personal preference. Other popular RNA-seq alignment algorithms include Tophat (Trapnell et al., 2009), RNA-Seq unified mapper (RUM) (Grant et al., 2011), Spliced Transcripts Alignment to a Reference (STAR) (Dobin et al., 2013) and MapSplice (Wang et al., 2010).

Using Genomic Mapping and Alignment Program for mRNA and EST sequences (GMAP; version 2012-07-20; (Wu and Watanabe, 2005)) a processed mm10/GRCm38.70 reference genome was created. Processed sequenced reads were then multimapped against this reference genome using GSNAP with the option to consider novel splice sites.

For investigatory purposes, sequencing reads were also mapped using the —split-output option in GSNAP, which instead of producing a single alignment file per replicate, produces up to nine alignments based upon how the reads map to the genome. For example, reads which are uniquely mapping, multi-mapping or non-mapping are separated.

2.2.7 Duplicate removal

Duplicates are reads which are identical to each other. These may exist if a gene is expressed at a high level, inevitably leading to identical reads in the RNA pool. Alternatively, duplicates may arise due to a PCR amplification bias, when PCR selectively over amplifies a particular cDNA fragment. High duplication rates are correlated with small amounts of input RNA which results in a molecular bottleneck. Increased duplication levels for a particular gene, or region of a gene may artificially infer a high expression level when it is instead just an artefact from the PCR process. For investigatory purposes, duplication levels were assessed to determine whether they resulted in batch effects or outlier

samples. To remove duplicates, the Picards MarkDuplicates.jar program was employed, using the uniquely mapped reads which had aligned to the genome in the appropriate orientation (concordant_uniq output file in GSNAP —split-output).

2.2.8 Expression quantification

To measure the expression level for each gene in each sample, aligned reads were quantified using either HTSeq (version 0.5.4p1; (Anders et al., 2015)) or cuffNorm (Trapnell et al., 2012). Both methods used a gene transfer format (GTF) file generated from combining both the most recent Ensembl annotation and the identified set of lincRNA (identified in chapter 4). This GTF file holds information about the genes present in the genome. By matching the aligned reads to this GTF file, the expression level for a particular gene or gene structures can be quantified (Figure 2.4).

HTSeq counts the number of reads which match to a particular gene for a particular sequencing run. Collating these counts for all sequencing runs for a particular biological replicate allows a count table to be produced where each row specifies a particular gene, each column specifies a particular sample and the values in the table represent the number of reads matching to that gene in that sample. Count tables are necessary, as raw counts are the required input format for differential expression programs such as EdgeR (Robinson et al., 2010) and DESeq2 (Love et al., 2014). Count tables were normalised using DESeq2 for downstream applications. Normalisation divides each sample column by its "size factor", correcting the count values to a common scale making them comparable. The size factor is the effective library size for that sample — the larger the library, the larger the size factor. The size factor for a given sample column is computed as the

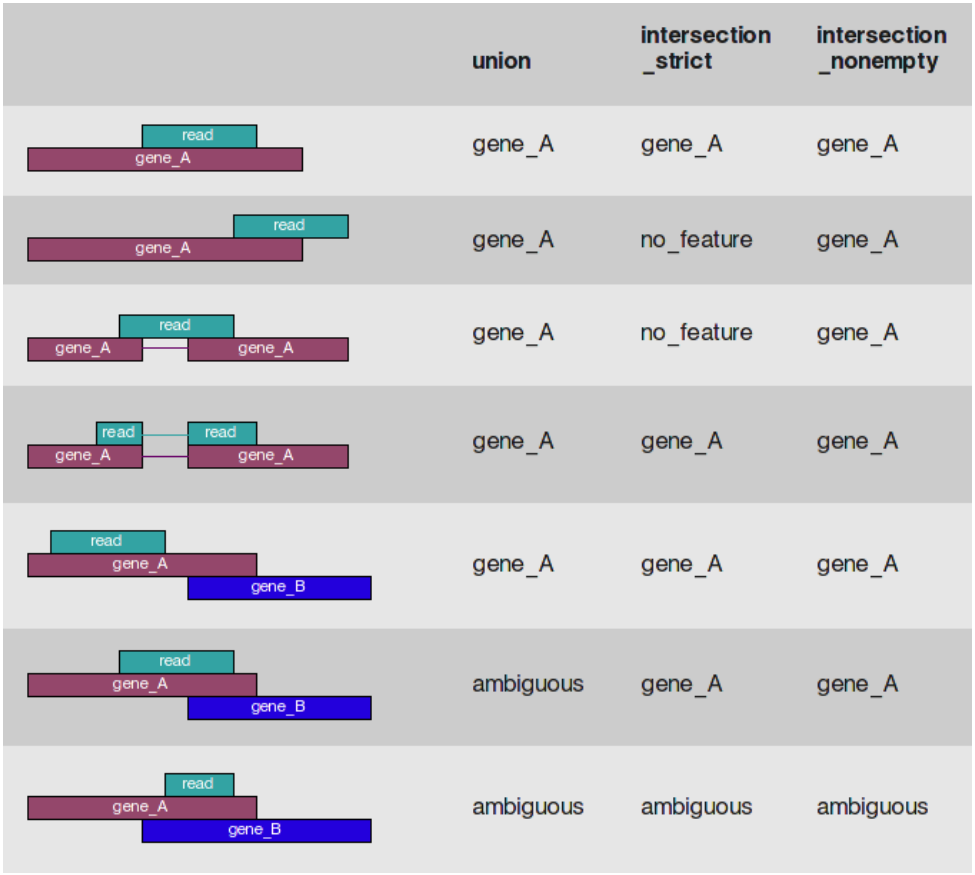


Figure 2.4. HT-seq read matching options.
The three modes of matching reads to the genetic features annotated in the genome using HTSeq. These modes affect how reads which overlap more than one feature are categorised. Intersection_nonempty was used for the main data analysis but intersection_strict was also used to test for edge effects. Image obtained from www-huber.embl.de/users/anders/HTSeq/doc/count.html (Anders et al., 2015).

"median of the ratio, for each gene, of its read count over its geometric mean across all lanes" (Dillies et al., 2013). For consistency, as count tables are required for differential expression analysis, they were also used for all temporal analyses and investigations (WGCNA, JTK Cycle — chapter 3).

CuffQuant, like HTSeq, is able to match reads to genes using the GTF file. However, instead of producing an integer count table, cuffQuant produces FPKM values. FPKM (Fragments Per Kilobase of transcript per Million) is an alternative measurement of expression and takes into account both the read length (as longer genes will normally have more reads/fragments mapping to them) and sequencing depth (the greater the depth, the more reads mapping to the gene). FPKM can generally determine how strongly a gene is expressed relative to other genes. FPKM values were normalised with cuffNorm and used for comparing gene expression between different tissues using publically available expression data.

2.2.9 Conditional quantile normalisation

Most normalisation methods assume any biases in expression are common for all genes within a library and therefore only apply a single scaling factor for that sample. However, this assumption is not necessarily true: additional factors such as gene length or GC content may affect expression levels of particular genes differently in different samples (Hansen et al., 2012). If there are GC or length biases, then failure to account for these differences could lead to incorrect biological conclusions.

Conditional quantile normalisation (CQN) is a method which aims to remove these biases (Hansen et al., 2012). By providing a matrix containing the GC content (%),

length (bp) or another covariate for each gene, CQN is able to remove these gene-specific covariates via robust generalized regression correcting global distortions. CQN removes the GC-bias by making the assumption that expression does not depend on GC content. GC content across a gene can affect the number of reads which map to it and as we expect expression to be the same across a gene (providing no alternative splicing), this effect must be due to a technical bias (Hansen et al., 2012). This suggests that CQN removes technical rather than biological variation in expression. CQN can return a matrix containing the normalisation offset required for each sample gene combination which can be incorporated into most GLM-based frameworks.

2.2.10 Identification of SCN-enriched genes

To identify genes whose expression is enriched in the SCN, genes were required to show significantly greater expression ($q < 0.05$, Benjamini-Hochberg correction) in brain relative to liver and significantly greater expression ($q < 0.05$, Benjamini-Hochberg correction) in both this study's SCN dataset and the published SCN dataset (Azzi et al., 2014) relative to brain using ANOVA in R. The mean FPKM values were calculated for each study and quantile normalised. Fold-change (FC) enrichment scores were generated by averaging (mean) the FPKM values for each study for a particular tissue and then comparing this average to the mean FPKM value for another tissue. A gene was defined as SCN-enriched if it showed a 3-fold enrichment in the SCN over brain samples.

2.2.11 Gene-annotation enrichment analysis

Enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Functional Annotation Tool v6.7 (Huang da et al., 2009). DAVID incorporates Fisher's exact test and was used to determine significant ($q < 0.05$, Benjamini-Hochberg correction) Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. GO terms were passed through Reduce and Visualize Gene Ontology (REVIGO) to prune redundant terms with a semantic similarity of 0.5, 0.7, or 0.9, dependent on GO term list length (Supek et al., 2011). Mammalian Phenotype (MP) enrichment was determined using the Mouse Genome Informatics (MGI) database (Eppig et al., 2015).

2.2.12 qRT-PCR

For validation of results from RNA-seq, qRT-PCR was performed using the BioMark system (Fluidigm) according to the two-step single-cell gene expression protocol using EvaGreen as described in the Real-Time PCR Analysis User Guide (PN 68000088, Fluidigm). 666ng of starting RNA from each sample first underwent reverse-transcription to form cDNA. The cDNA samples underwent 18 cycles of specific target amplification (STA) followed by a 10-fold dilution. SCN CT values were normalised using the housekeeping genes *ActB, Gapdh* and 8 genes which showed the most stable expression in the SCN according to the LCM SCN sequencing data (*Ankrd40, Nkiras1, Sar1a, Smarce1, Snrpn, Tpgs2, Tsn* and *Ywhaz*).

2.2.13 *In situ* hybridisation

Regions of selected transcripts were amplified by PCR from mouse genomic DNA or cDNA from adult mouse brain and cloned into the pCR4-TOPO vector (Life Technologies, Paisley, UK) for DIG-labelled riboprobe synthesis: *Avp* (1-490bp of accession number NM_009732.1), *Vip* (191-715bp of NM_011702.1), *Scg2* (844-1287bp of NM_009129.1), *Grpr* (1287-1741bp of NM_008177.1) and NONCO7761 (*3100003L05Rik*; 32-583bp of NR_045907). TOP10 cells were then transformed with pCR4-TOPO according to manufacturer's instructions. Transformed cells were plated on 50 μg/μl Kanamycin-enriched lysogeny broth (LB) agar plates and incubated overnight at 37°C. DNA was extracted from transformed cells using the Qiagen Spin Miniprep Kit, digested and run on a 1% agarose gel to confirm successful insertion of the genomic region. The successfully cloned vectors were sent for sequencing. Linearised plasmids were gel-purified and used to make digoxigenin-labelled (Roche) riboprobes as described previously (Jeans et al., 2007). Mice were individually housed for 10 days under 12:12 LD before dissection at ZT6. Brains from male mice were isolated and frozen sections at 15 μM were cut and mounted onto Superfrost slides (VWR). Slide hybridisation and riboprobe detection was carried out by Peter Oliver as described previously (Chodroff et al., 2010).

2.3 Results

2.3.1 Generation of the mouse SCN transcriptome

Mapping of the processed sequencing data resulted in ∼35M reads mapping per technical replicate. Expression data from technical replicates (same sample sequenced in different lanes) were summed together to form expression data per biological replicate as in (Mortazavi et al., 2008).

Data exploration

Before data analysis, exploration of the data is important to identify outliers or patterns of variation which may need to be addressed. One method of data exploration is to perform principal component analysis (PCA) on the samples. In this method, the data points (samples) are projected onto two dimensions so that they display the greatest amount of variation between them. Neighbouring data points are seen to be more similar, whereas more distant data points are seen to be dissimilar. We would expect the samples to separate into some structure based upon biology; samples from similar conditions should display similar expression patterns and those derived from different conditions should not. The PCA plot has two axis: PC1 (the x axis) and PC2 (the y axis). The first principal component (PC1) accounts for as much of the variability in the data as possible, and is therefore the direction which describes the greatest differences between samples. The next component (PC2) accounts for the next greatest variance possible given that it is orthogonal (and therefore completely uncorrelated) to the previous component. Although

generally only the first two PCs are displayed, the number of PCs can continue until essentially all the variation in the data is explained.

To consider outliers and batch effects and to view the substructure of the data, PCA was performed on the sequencing data derived from the WT mice SCN samples (Figure 2.5). This was executed via the prcomp R function using the top 500 expressed genes of regularized-logarithm transformed normalised counts in DESeq2.

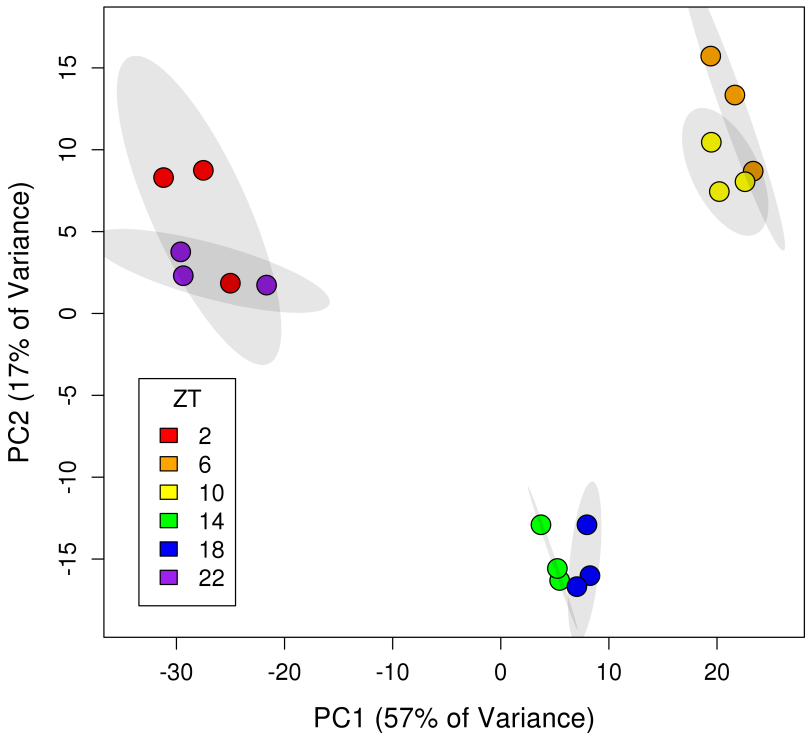


Figure 2.5. PCA plot of WT SCN RNA-seq data points.
Principal component analysis (PCA) plot showing separation of RNA-seq data according to timepoint and sequencing batch. Ellipses delineate the 95% confidence intervals for each timepoint.

The most notable substructure in the PCA plot was the separation of the data points into three separate groups, each of which comprises of samples extracted from two time points. This separation is unsurprising because these three groups were processed in separate batches: samples were extracted and sequencing libraries were generated and sequenced at three separate time points spread over the course of 18 months. PCA also revealed global expression differences between timepoints within a batch.

Assessing technical biases

To assist the identification of specific errors and biases introduced to the RNA-seq dataset as a result of batch effects, the software package Quality of RNA-Seq Toolset (QoRTs) was utilised (Hartley and Mullikin, 2015). QoRTs assesses a variety of metrics and allows systematic errors between batches to be easily identified.

Investigation into gene-body coverage revealed that all samples display a 3' bias, a tendency of reads to pile up towards the 3' end of the gene (Figure 2.6). This is often observed when RNA is poly-A selected (as poly-A tails are present on the 3' end of the transcript) and this bias is exacerbated if the RNA is degraded. The gene-body coverage appears to vary according to batch, with batch 1 (ZT6/10) displaying the greatest 3' bias. This batch may have been processed in a way that caused increased RNA degradation.

GC content also differed according to batch: samples from batch 1 were more likely to possess a lower GC content. Interestingly, for mammals, the 3' UTR has been reported to possess the lowest GC content (43% in mouse) compared to the coding region and 5' UTR (52% and 59% respectively) (Zhang et al., 2004a). This suggests a link between the reduced GC content and increased 3' bias in batch 1. Supporting this, the sample with

the greatest 3' bias (ZT6C) showed the lowest GC content and the batch with lowest 3' bias (batch 2) possessed the highest GC content (Figure 2.6).

This link could exist for two reasons. Firstly, due to the 3' bias, the sequenced RNA reads are more likely to derive from the 3' UTR, possess a low GC content and induce this shift in GC content. Alternatively, due to the GC bias in amplification (GC-depleted transcripts are more likely to be amplified), regions of high GC content, such as the 3' UTR, will display greater expression which leads to this 3' bias observed in our samples. Systematic differences in RNA-degradation or GC bias may cause problems in downstream analysis especially if confounded by the experimental variable of interest.

Batch 2 also appeared to deviate in two of the metrics surveyed: insert size and alignment clipping rate. Samples from batch 2 possess a shorter insert size (median $\sim$ 170bp) than the other batches (median $\sim$200bp) (Figure 2.6). This reduced insert size likely arises from differences in fragmentation: samples in batch 2 were perhaps left to fragment for a longer period of time. If the RNA fragment is shorter than 100bp then the attached adaptor will be sequenced for both reads (Figure 11).

A mapper will soft clip a region of a read if it has a low sequence quality or is unable to correctly map to the genome, which is to be expected if the adapters are sequenced. Due to the short length of many of these fragments, it is likely the sequenced adapters were soft-clipped creating these spikes in alignment clipping rate (Figure 2.6). For read 2, there appears to be a sudden drop in the clipping rate (from 60bp) within this batch, probably a reflection of the successful identification and removal of a particular contaminating adapter (specified in methods; $\sim$35bp) in read 2. In Figure 2.6B there is a small shelf

at around 70-100bp, which is likely to represent these reads which have had this specific adapter successfully removed.

The presence of sequenced adapters in this batch are unlikely to affect analysis because they are soft-clipped and not accounted for in alignment. However, reduced insert size may have a subtle effect on the expression levels of particular genes, or regions of genes, although this has not been described in the literature.

A variety of parameters relating to the mapping of reads to the genome and their influence upon batch effects were investigated by inspecting the corresponding count data using a PCA plot.

Removal of multi-mapping reads: Certain genes possess high sequence similarity to other genes or pseudogenes in the genome. If these genes show expression differences influenced by batch, then these sequencing artefacts may be exacerbated by multimapping (mapping of these reads to other regions of the genome which share high sequence similarity). The use of uniquely mapping reads (reads which map to only one place in the genome) may therefore reduce batch effects.

Removal of duplicates: Duplication levels differ between batches and are higher for sample outliers. Duplicate removal may therefore reduce batch specific effects if certain genes show increased duplication in certain batches.

Removal of edge-effects: Mapped reads are matched to genomic features using HTSeq which has different modes for read categorisation (Figure 2.4). If the degree of poly-A selection differs between each batch, then the number of reads which may have intron sequences as part of the transcript may differ (as there will be a differing proportion of pre-mRNA transcripts which are not poly adenylated). If the amplification

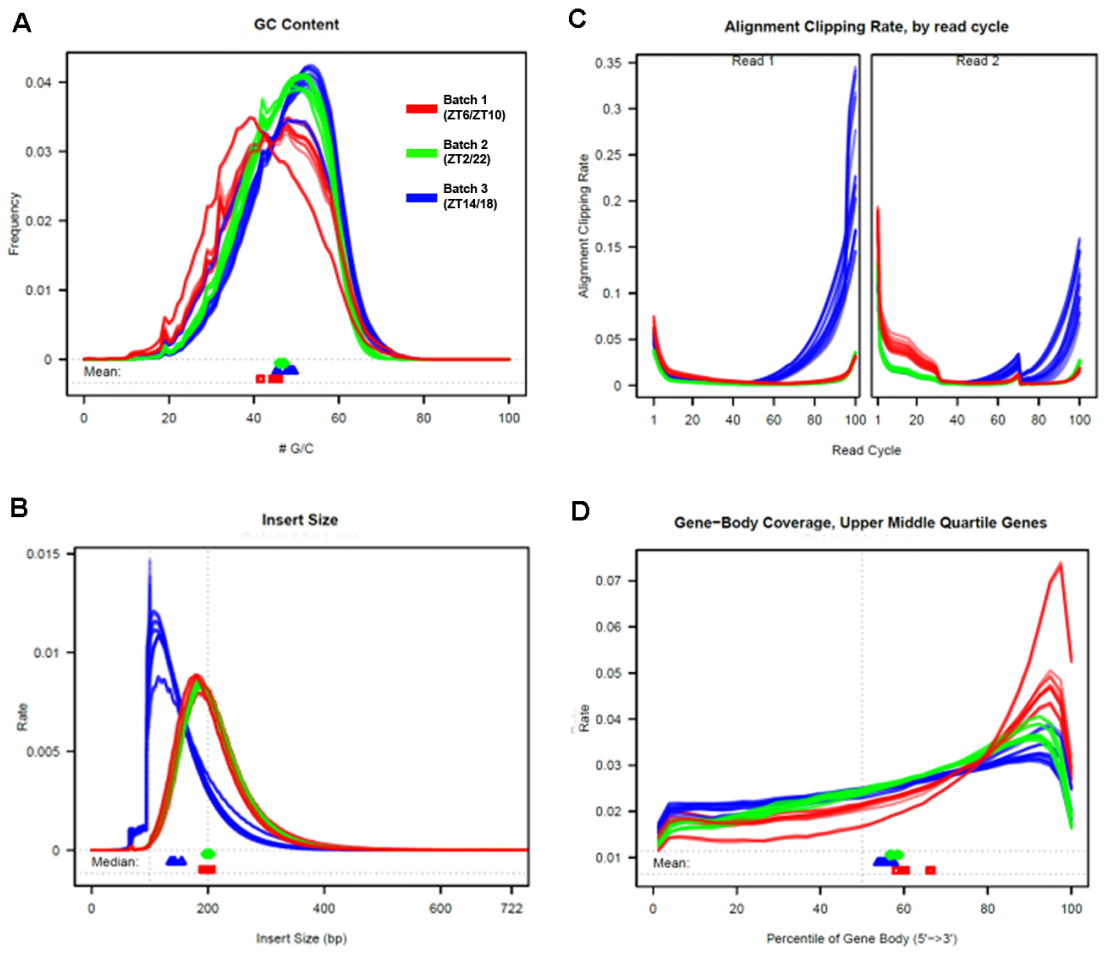


Figure 2.6. Technical biases of mapped samples between batches.
RNA-seq data was assessed for **(A)** Clipping Rate (the rate at which the GSNAP soft-clips the read with respect to read position), **(B)** Insert Size (the nucleotide length of the original RNA fragment present between the two sequencing adaptors), **(C)** Gene-body coverage (proportion of reads which map to a particular region of the gene) and **(D)** GC content (the proportions of G/C in the replicates' reads). Each plotted line represents a technical replicate and are coloured for batch (red — batch 1 (ZT6/10); blue — batch 2 (ZT14/18); green — batch 3 (ZT2/22)).

of particular RNA fragments varies based upon GC content, then we may observe a relative increase/decrease in reads with low GC content (i.e. intronic regions). Edge-effects are when reads which overlap the exon are partially derived from the intron. Genes with shorter, or a greater number of, exons are subject to these edge-effects and may have expression levels which vary according to batch. To determine whether edge-effects affected the sequencing data, count tables were generated using the 'intersection strict' option in HTSeq which does not allow the mapping of reads if they overlap an unannotated (intronic) region (Figure 2.4).

PCA of the sequencing data after removal of multimapping reads, duplicates or reads which overlap non annotated regions (Figure 2.7) did not appear to remove the batching of the data points and therefore the count tables generated considering these factors were not used for further data analysis.

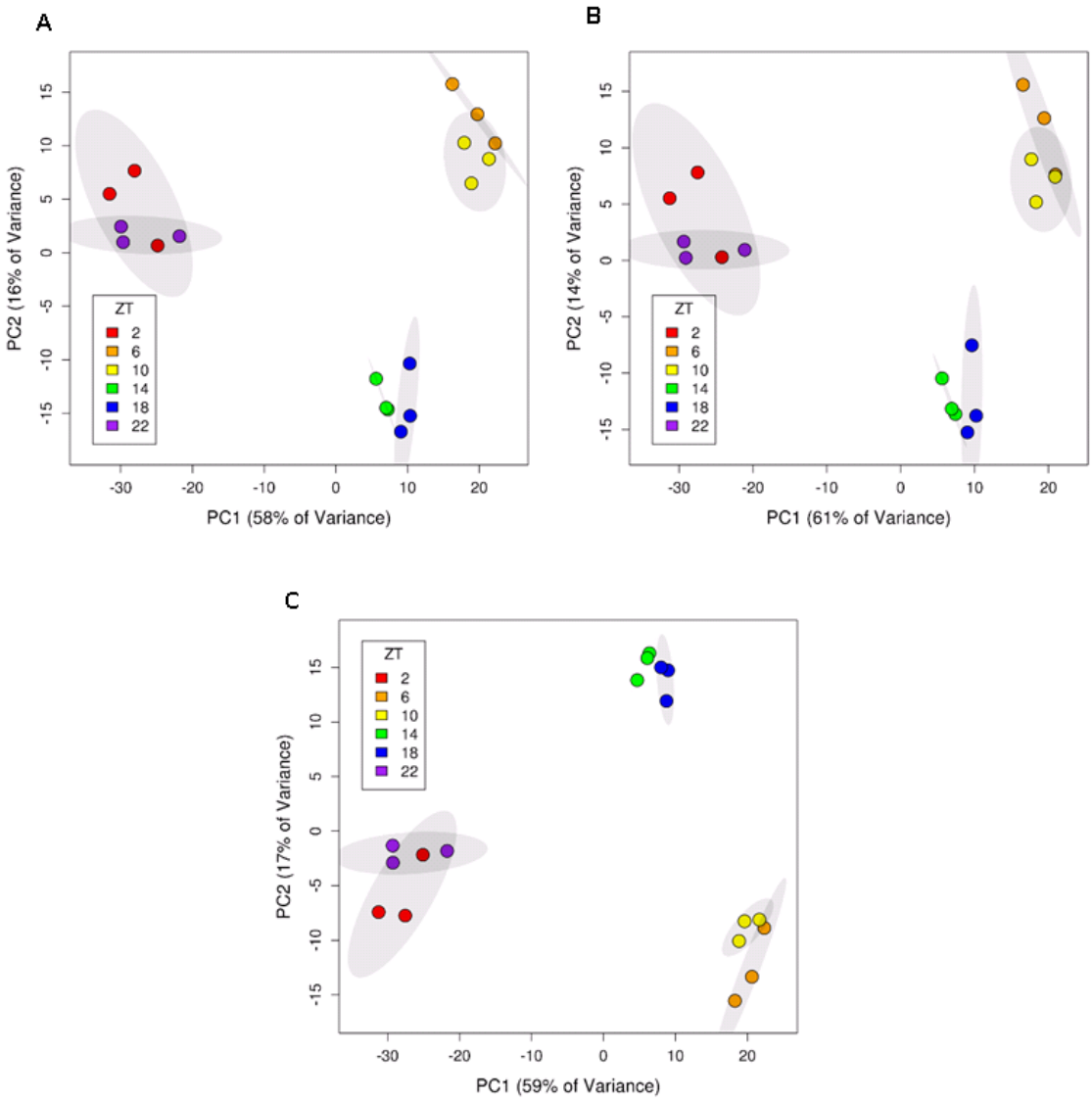


Figure 2.7. Assessing technical biases through PCA plots.
PCA plot showing separation of RNA-seq data having **(A)** removed multimapping reads, **(B)** removed duplicates, **(C)** removed edge effects. Ellipses delineate the 95% confidence intervals for each timepoint. The data points remain separated into three groups suggesting batch effects were not solely subject to these factors.

2.3.2 Removal of batch effects

Regressing out batch effects

Batch effects in large scale experiments are not uncommon and there have been attempts to produce methods capable of alleviating this problem.

If it is believed that simple surrogates, such as processing group, are able to capture all the variation seen due to batch effects, then a common method is to incorporate this variable into a linear model which compares between groups. This method accounts for differences between batches and identifies differences regarding the variable of interest. This can be done using Combat, RUVr of the RUVseq package, or any environment which allows testing of linear models such as DESeq2 (Risso et al., 2014, Leek et al., 2012, Love et al., 2014).

When the source of the batch effect is unknown or cannot be modelled by a surrogate variable, such as processing group, then other methods such as SVA (surrogate variable analysis) or PEER (probabilistic estimation of expression residuals) can be used (Leek et al., 2012, Stegle et al., 2012). These methods try to estimate the "hidden" sources of variation which are external to the variable of interest (in this case ZT), which can then be regressed out.

These approaches work best when the variable of interest (ZT) is not strongly correlated with batch. In this study, ZT is perfectly confounded by batch which means that any difference seen across time could be attributable to differences due to batch. For this reason, regressing out known and unknown sources of variation will either be impossible or largely inaccurate.

An alternative method of normalisation is to assume some properties do not change over time, and then normalise the dataset to these properties. RUVg in the RUVSeq package allows normalisation to negative control genes which are assumed to have constant expression across samples and not be influenced by the covariates of interest (ZT in this case) (Risso et al., 2014). This is useful if the sequencing samples possess ERCC spike-in sequences or if there is a known set of housekeeping genes which show constant expression over time. RUVg can then return the estimated factors of unwanted variation and the normalised counts having regressed out this variation. This experiment did not use ERCC spike-ins or possess many reliable circadian control genes. Some studies investigated housekeeping genes over circadian time but it is difficult to predict any which are stable in the SCN, therefore making this form of normalisation unfeasible (Kosir et al., 2010).

Quantile regression

Standard linear regression methods, as discussed above, only consider the average relationship between a covariate of interest and the expression data. However, these methods may not capture the full effect of the covariate upon expression. Instead, the covariate of interest may affect gene expression depending on its value — a relationship which can be corrected through quantile regression. Quantile regression allows a more complete characterisation of the data, considering the impact of the covariate on its entire distribution, not just its average. To allow this, the covariate of interest (dependent variable) should have no zeros, nor too many repeated values. Instead, the covariate should have a spectrum of continuous values so that the data can be appropriately partitioned into quantiles.

Instead of one coefficient which explains expression we have a group of coefficients depending on the quantile. At each quantile, the dependent variable has a different effect upon the independent variable. Conditional quantile normalisation (CQN) was utilised to remove the gene-specific covariates of GC content and length.

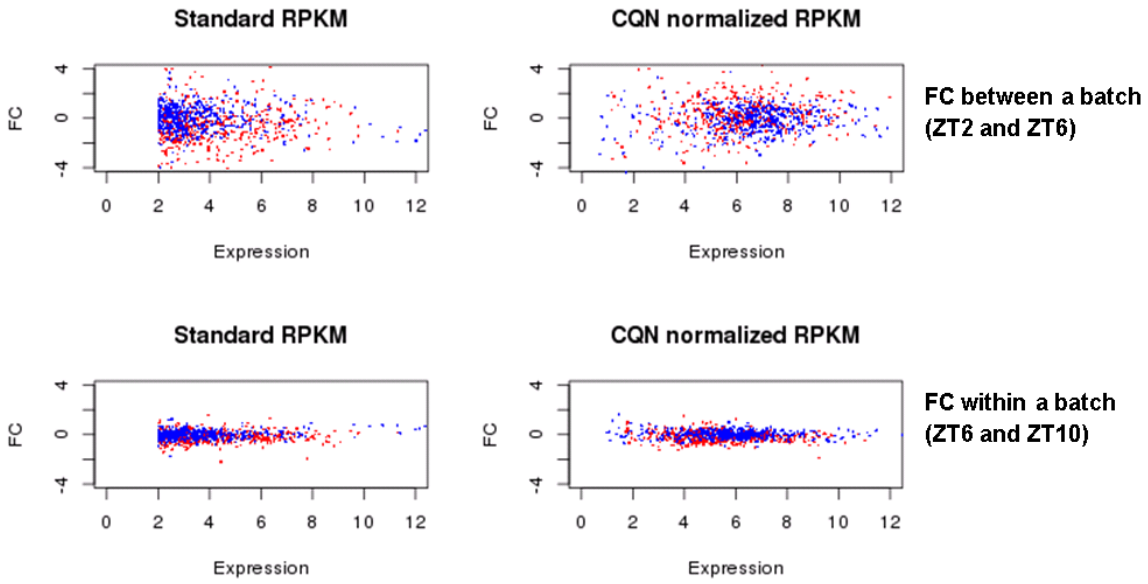


Figure 2.8. Fold change differences between timepoints pre- and post- CQN normalisation. Scatter plots which show the fold change (FC) difference between samples from ZT2 and ZT6 (top) and ZT6 and ZT10 (bottom) for both genes with a high GC content (the 15% of transcripts with highest GC content; red) and a low GC content (the 15% of transcripts with lowest GC content; blue). These show the FC before and after CQN (left and right respectively). The top panel shows that transcripts with high GC content (red) have a higher dispersion with regards to expression fold change and are more likely to be downregulated in ZT6. CQN makes the assumption that GC content should not affect the dispersion or FC between samples and corrects for this (right panel). The bottom panel displays less of a bias in FC between high and low GC content transcripts. Furthermore there appear to be fewer FC differences between ZT6 and 10 (within a batch) than between ZT2 and ZT6 (between batches).

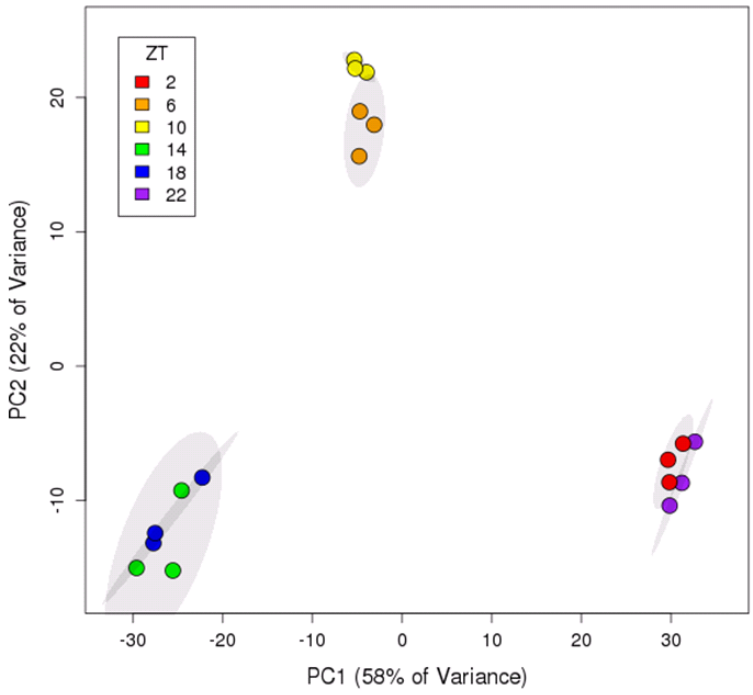


Figure 2.9. PCA plot of CQN normalised RNA-seq data points.
PCA plot showing separation of RNA-seq data having applied CQN. Ellipses delineate the 95% confidence intervals for each timepoint. The data points still separate into three batches suggesting that batch effects cannot be fully described by a length and/or GC bias.

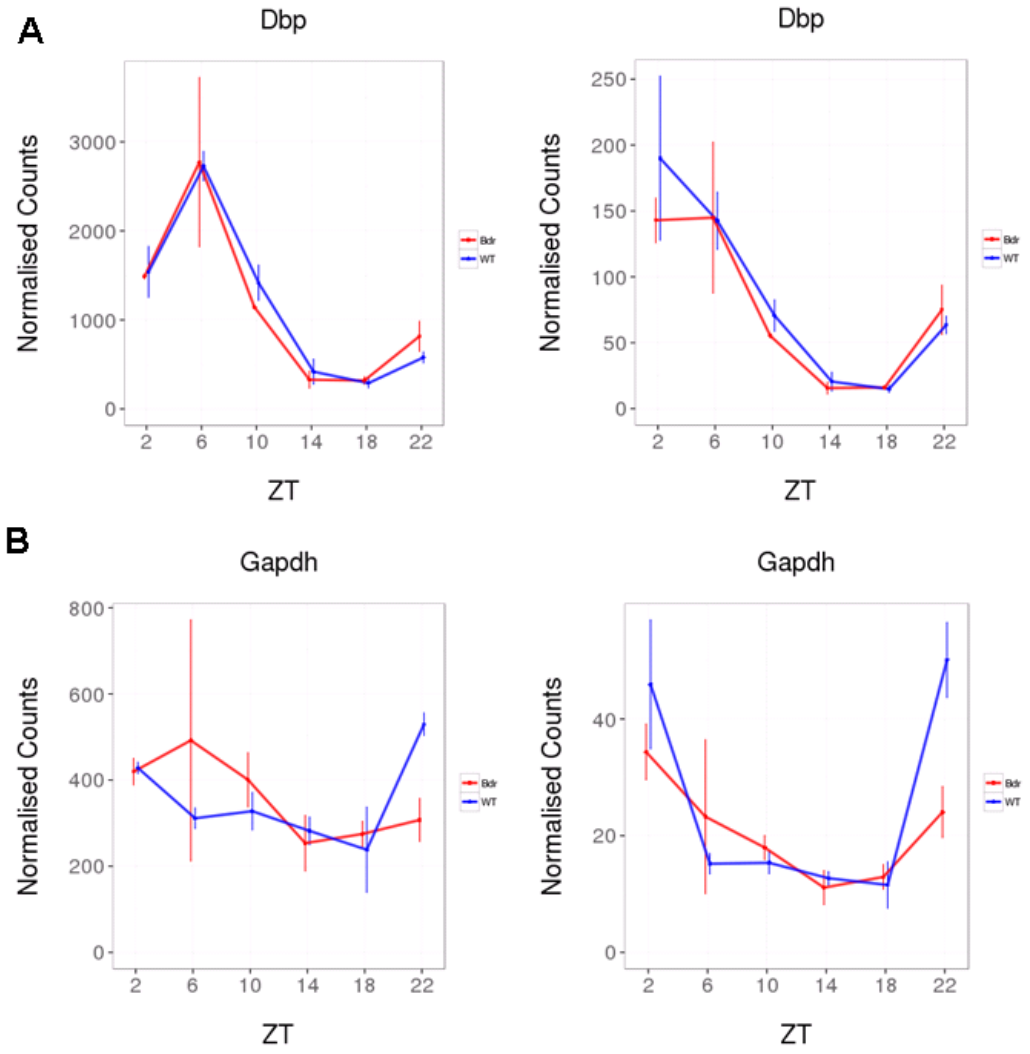


Figure 2.10. Expression profile of control genes before and after CQN.
Expression of **(A)** a 'positive' and **(B)** a 'negative' control gene which shows differing expression under both normalisation methods (standard —left, CQN — right). *Dbp* has been shown to display peak expression at ∼ZT5 and *Gapdh* expression is expected to be constant throughout — patterns which are more similar to the standard normalised genes.

PCA of the resultant CQN normalised data did not remove these batch effects as the data separated according to processing batch (Figure 2.9). CQN also failed to make the FC differences comparable when comparing within and between a batch; greater FC differences were observed between samples from different batches (ZT2 and ZT6) rather than samples from the same batch (ZT6 and ZT10; Figure 2.8). CQN normalisation either had no effect on the expression of control genes, or exacerbated batch effects. For example, *Dbp* has been reported to peak at ∼ZT5 in the SCN which is more similar to data before CQN normalisation (∼ZT6 vs ZT3; Figure 2.10A) (Yan et al., 2000). *Gapdh*, a traditional housekeeping gene would be expected to be constant across the day. However, CQN normalised data appeared to exacerbate the batching effects on temporal expression creating an unusually high expression at the timepoints ZT22 and ZT2 (Figure 2.10B).

As CQN did not remove these batch effects, and may even alter true biological expression patterns, this normalised dataset was not considered for further analysis. CQN only removes batch effects if GC or length bias is the main source of bias. It appears that there must be additional biases which cannot be removed by quantile regression of these covariates. All temporal analyses within this project used the DESeq2 normalised data as described (section 5.2.1).

2.3.3 Temporal analysis of positive controls

Although batch effects exist within the dataset, we do not know to what extent these global biases affect estimates of the temporal transcription of particular genes. To assess the strength of these biases and confirm that the sequencing data is suitable for further

analysis, the temporal expression of particular genes in the SCN was compared to their expected expression profile based upon literature values.

It was not possible to identify genes which were reported to show stable expression over circadian time in the SCN under these conditions. There are, however, many genes which have been reported to display distinct circadian fluctuations in the SCN, the most renowned being the circadian clock genes (Ko and Takahashi, 2006). To validate the SCN expression data over 24 hours, the temporal expression of key clock genes was compared to what was described in the literature.

Eight of 10 key circadian-clock genes showed the classic sinusoidal circadian profile ($q < 0.05$) with a zenith and nadir at the expected time points (Figure 2.11); this verifies the temporal expression of our LCM SCN transcriptome dataset and suggests that any sequencing batch effects are minimal.

2.3.4 Validation of RNA-seq data

Although the expression patterns of clock genes match our expectations, it is important to confirm the expression for a greater number of genes and to gain an appreciation of the impact batch effects have upon the expression data. To confirm that any novel findings were not subject to batch effects, gene expression levels should ideally be measured using an independent method that: i) does not rely on using batches, ii) can cater for small RNA concentrations, and iii) can easily assess numerous genes. To fit these criteria, Biomark qPCR was used on the same LCM SCN samples as those used for sequencing.

Biomark quantitative (q)PCR of the SCN samples showed a high degree of concordance with the RNA-seq expression data. 43/59 genes had a correlation of over 0.4 (Fig-

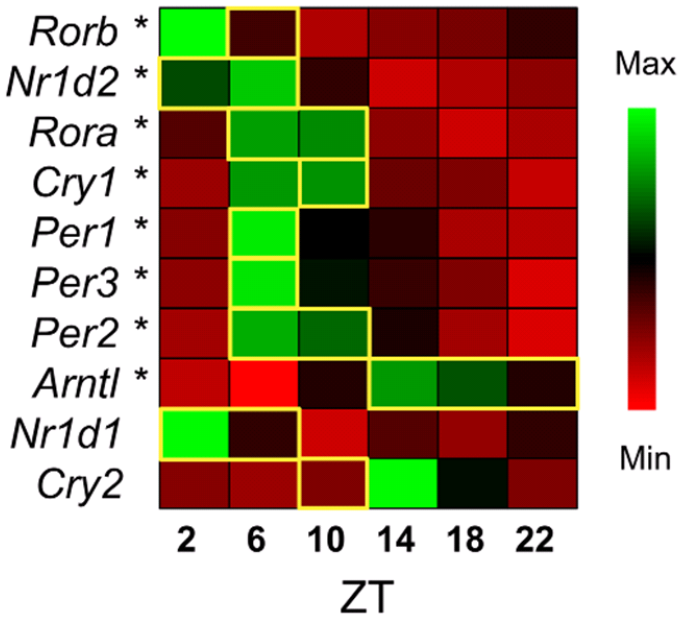


Figure 2.11. Clock gene expression in the SCN across circadian time.
Heatmap showing the mean expression of key clock genes in the SCN across circadian time. The majority (8/10) show significant cycling (DESeq2 and JTK Cycle; *q < 0.05) and peak at the expected timepoint denoted by a yellow box (Ko and Takahashi, 2006).

ure 2.12A; Figure 2.13A-B). Furthermore, there appears to be a correlation between the dynamic range of the gene's temporal expression (the fold change difference between maximum and minimum expression) and the concordance between platforms (Figure 2.12B). Gene expression values with larger dynamic ranges tend to be less susceptible to extrinsic factors such as random noise and batch effects and therefore contribute less to expression.

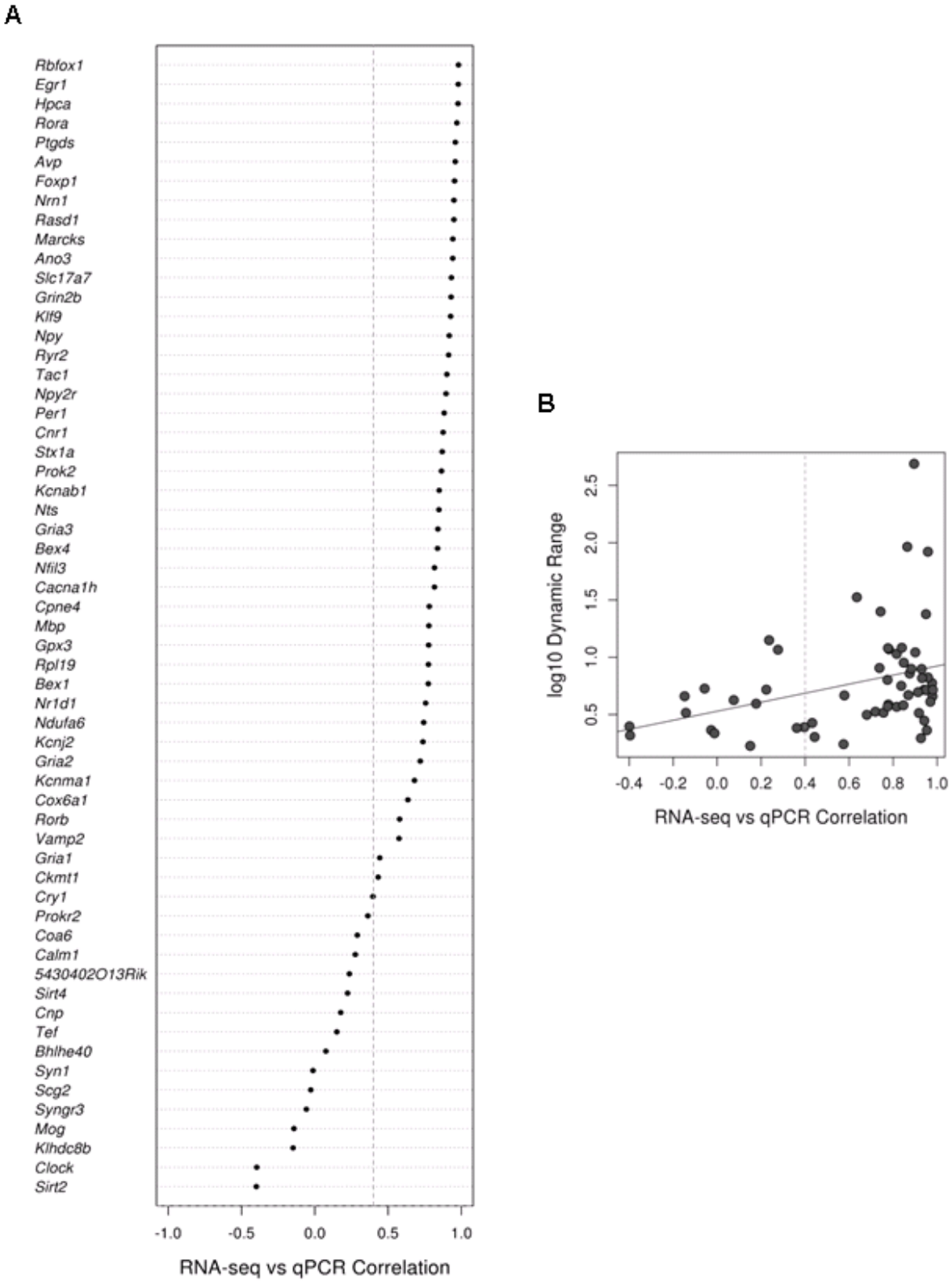

A
Rbfox1
Egr1
Hpca
Rora
Ptgds
Avp
Foxp1
Nrn1
Rasd1
Marcks
Ano3
Slc17a7
Grin2b
Klf9
Npy
Ryr2
Tac1
Npy2r
Per1
Cnr1
Stx1a
Prok2
Kcnab1
Nts
Gria3
Bex4
Nfil3
Cacna1h
Cpne4
Mbp
Gpx3
Rpl19
Bex1
Nr1d1
Ndufa6
Kcnj2
Gria2
Kcnma1
Cox6a1
Rorb
Vamp2
Gria1
Ckmt1
Cry1
Prokr2
Coa6
Calm1
5430402O13Rik
Sirt4
Cnp
Tef
Bhlhe40
Syn1
Scg2
Syngr3
Mog
Klhdc8b
Clock
Sirt2
-1.0
-0.5
0.0
0.5
1.0
RNA-seq vs qPCR Correlation
B
log10 Dynamic Range
0.5
1.0
1.5
2.0
2.5
-0.4
-0.2
0.0
0.2
0.4
0.6
0.8
1.0
RNA-seq vs qPCR Correlation

Figure 2.12. Concordance between RNA-seq and qPCR.
(A) Dot plot showing the Pearson's correlation between RNA-seq and qPCR values for all the annotated genes surveyed (excluding those housekeeping genes used to normalise the data). 43/59 genes show a high concordance between methods ($r > 0.4$). **(B)** Scatter plot showing a trend for increased correlation of cycling gene expression measured in the two different platforms for increased dynamic range. Gene expression values with larger dynamic ranges tend to be less susceptible to extrinsic factors such as random noise or batch effects.

Reviewing individual cases of poorly concordant genes provides an opportunity to assess the reasons for discrepancies between the two platforms. It is possible to distinguish the presence of batch effects for particular genes when comparing the RNA-seq and qPCR expression data. For example, at ZT2 and ZT22 (a single batch), *Sirt4* and *Scg2* (Figure 2.13C) showed very high expression compared to the qPCR data whereas *Calm1* showed very low expression at these timepoints. *Cnp, Sirt2, Klhdc8b* and *Syngr3* showed elevated expression at ZT14 and ZT18 relative to the qPCR data. *Clock, Tef, Mog, Prok2r* and *Bhelhe40* showed depleted expression at ZT6 and ZT10. RNA-seq expression of *Coa6* appeared to change according to all three batches. When considering the length and GC content for these genes, there appeared to be no correlation with any of these batch discrepancies and particular gene traits. However, this may only be one explanation for the poor concordance between the platforms for these genes. If the primers were poorly specific or only target particular isoforms of the gene, then the expression data would be expected to differ between platforms.

In summary, PCA of the RNA-seq data revealed separation into three batches according to the time that sequencing libraries were generated. Samples from these batches differed on metrics such as insert-size, gene-body coverage and GC content. These batch effects could not be removed despite interrogation of numerous potential confounders or application of a variety of statistical algorithms. qPCR of the LCM SCN samples revealed that the RNA-seq data is largely reliable in identifying temporal expression patterns as 73% (43/59) of genes showed a high concordance between platforms. Nevertheless, some genes lack concordance due to expression levels which are elevated or depleted according to batch. This may prevent the identification of genes which possess true circadian expres-

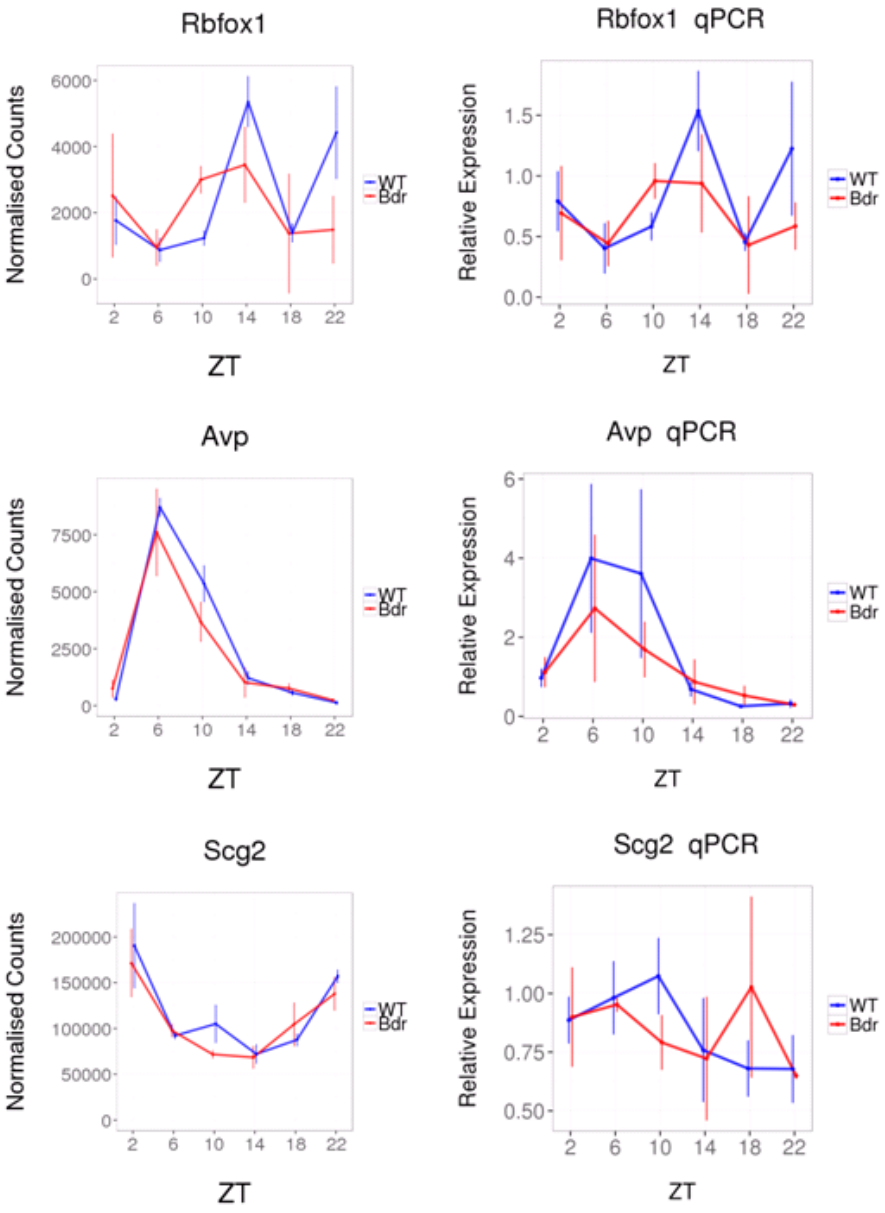


Figure 2.13. Expression of genes surveyed by RNA-seq and qPCR.
Gene expression determined through RNA-sequencing (left panels) and Biomark qPCR (right panels) showing the concordance of particular genes between the platforms. The first two genes (*Rbfox1* and *Avp*) represent highly concordant expression patterns between the platforms whereas *Scg2* represents a poorly concordant gene between the platforms.

sion patterns. 8/10 clock genes displayed appropriate expression patterns and concordance between RNA-seq and qPCR was shown to increase with dynamic range, suggesting that those circadian genes important for rhythm generation or with strong oscillations are more likely to be accurately identified.

2.3.5 Assessing SCN cellular subtypes and identification of SCN-enriched genes

Assessing the cellular subtypes within the SCN

Having generated FPKM values for the LCM SCN data (and the other brain regions for which RNA-seq data or FPKM values were available), a greater appreciation of the cellular subtypes in the SCN can be obtained. Accordingly, the FPKM values for established cell marker genes in the LCM SCN data were compared to the FPKM values generated from purified cell types derived by fluorescence-activated cell sorting (FACS) of the mouse brain (Cahoy et al., 2008).

Cell marker gene expression values suggest that the SCN largely consists of neuronal subtypes with a moderate expression of astrocytic markers (Figure 2.14). Markers from myelinated oligodendrocytes and endothelial cells also appeared to be expressed in the SCN but at relatively low levels. These results show the diversity of cell types within the SCN but also the relative importance of neurons and astrocytes within the SCN. Knowledge of the relative abundance of these subtypes can help to deduce the functional processes that are occurring within the SCN.

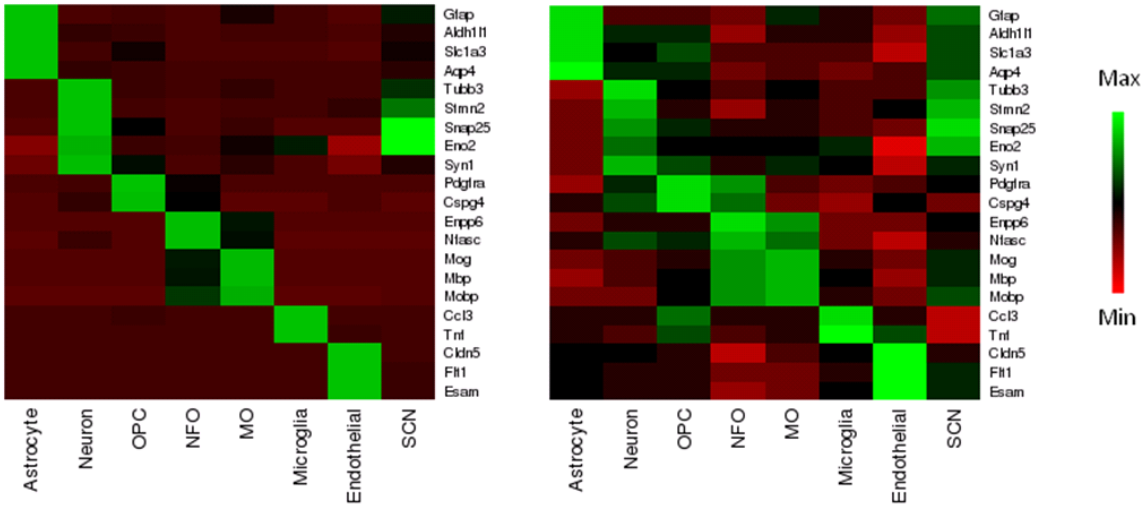


Figure 2.14. Expression of cell markers in the SCN.
Heatmaps of cell marker FPKM expression from FACS purified cell types and this SCN dataset. The SCN appears to possess a high FPKM expression for markers of neuronal cell types. The left panel shows relative expression levels and the right panel shows relative expression on log transformed FPKM values, helping to identify more subtle expression patterns.

Identification of SCN-enriched genes

The SCN plays an important role in driving circadian rhythms for which certain genes are proposed to play key roles (Reghunandanan and Reghunandanan, 2006, Ko and Takahashi, 2006). If a gene has an SCN-specific role, it is likely that its expression will be limited to, or enriched in, the SCN. For example, genes such as *Avp, Vip* and *Grp* all encode neurotransmitters essential for intra-SCN communication and display expression which is highly enriched in the SCN. Therefore the identification of other SCN-enriched genes may implicate additional processes important for circadian rhythm generation.

Currently there are no resources to determine whether a particular gene is enriched in the SCN. The Allen Brain Atlas (ABA) provides ISH data for many protein-coding genes providing some valuable spatial information on expression in the mouse brain. However, this is not a practical resource for high-throughput identification of SCN-enriched genes. Firstly, to find the brain slice containing a particular brain region for a long list of genes is practically unfeasible; and many genes are only represented by parasagittal sections and in others the SCN is missing or damaged. Secondly, the ISH staining is often over- or under- exposed therefore dampening any biologically relevant signal variation and making accurate quantification unreliable.

Therefore, it is of importance to produce a method which can identify SCN-enriched genes using genome-wide data, allowing the identification of genes or pathways important for SCN function. To identify SCN-enriched genes on a genome-wide scale with quantitative statistics, publically available sequencing data from other regions in the brain and body was acquired to allow comparison of expression (Table 2.1). Using RNA-seq data from different studies can prove to be problematic and sometimes unsuitable for com-

parison owing to the influence of laboratory batch effects. These batch effects may arise from differences in reagents, RNA extraction protocols, sequencing library preparation or computational analysis. As previously discussed, it is often useful to visualise this data using PCA to determine whether the samples cluster by tissue or by the laboratory in which they were generated.

The FPKM expression data was quantile normalised to ensure that the distribution of RNA molecules was the same in each sample or tissue; RNA derived from different tissues or sample preparations may change the distribution of expression. PCA was performed via the prcomp R function on the log2 normalised FPKM of all protein coding genes.

Importantly, PCA determined that most of the variation in expression between these RNA-seq datasets was due to tissue, but not the study, in which they were acquired (Figure 2.15). The first principal component (PC1) explains nearly 80% of the expression variance in samples and displays a large separation between liver and brain tissues signifying substantial transcriptomic differences between these disparate tissues. The second principal component (PC2; approximately 10%) in part explains variation among different regions of the brain. Samples from the same tissue (e.g. cerebellum) cluster together more closely than tissue samples from the same study (labelled by number) but different brain region. This indicates that RNA-seq batch effects are small relative to biological variation. Of the non-SCN brain tissue samples, the published hypothalamus RNA-seq dataset (Zhang et al., 2014) shows the greatest similarity to the SCN samples which is expected as the SCN is located within this brain region.

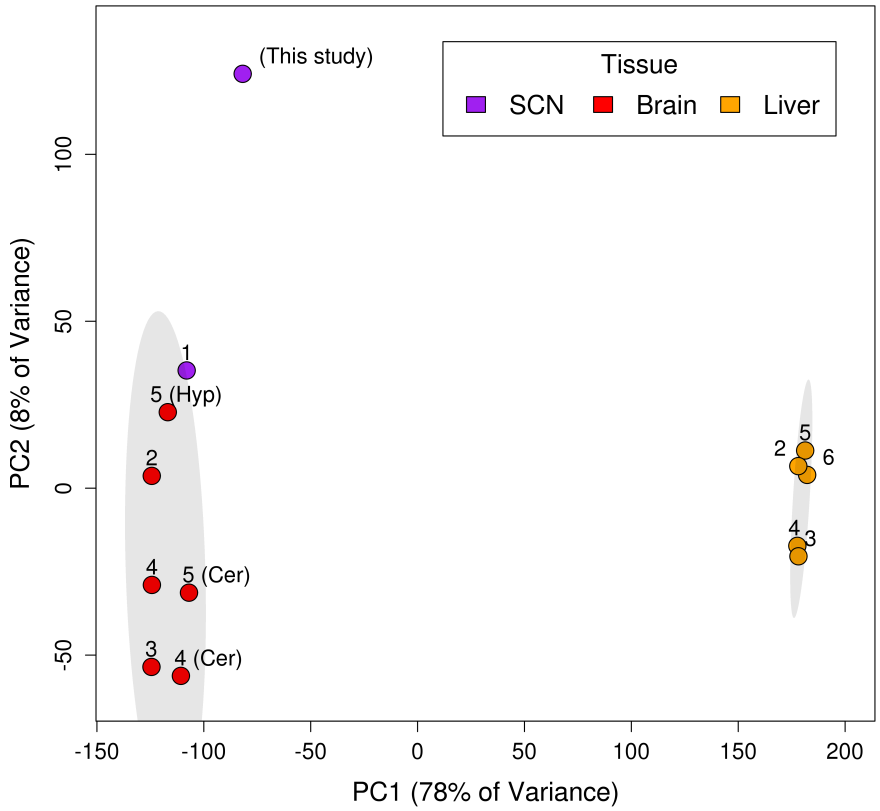


Figure 2.15. PCA plot of RNA-seq data obtained from different tissues and studies.
PCA plot showing separation of RNA-seq datasets according to the tissue they were derived. Adjacent numbers indicate the study from which data were collected (1 — (Azzi et al., 2014); 2 — (Barbosa-Morais et al., 2012); 3 — (Merkin et al., 2012); 4 — (Brawand et al., 2011); 5 — (Zhang et al., 2014); 6 — (Menet et al., 2012)). Ellipses delineate the 95% confidence intervals for each tissue.

Identification of SCN markers

PCA separated the samples according to anatomical region and not laboratory, suggesting that it is possible to compare expression levels between tissues and generate an SCN enrichment score.

The newly created method for identifying SCN-enriched genes was tested by applying it on known SCN markers. The presence of known SCN markers in the SCN-enriched gene list would provide confidence that other genes identified are also enriched. Four key SCN markers (*Avp, Vip, Grp* and *Six6*) displayed a far greater expression enrichment in the SCN over other brain regions (Figure 2.16). This matches our expectations and suggests that it is possible to identify other SCN-enriched genes without prior knowledge.

Identification of novel SCN-enriched genes and processes

146 genes were identified whose expression was highly enriched in the SCN (ANOVA, $q < 0.05$; Figure 2.17; Table 2.2) of which 4 out of 4 were confirmed using *in situ* hybridisation (ISH) on adult brain sections (Figure 2.19). These 146 genes included many known SCN markers such as *Vip, Avp* and *Six6* as well as others not previously implicated in circadian biology or SCN function (Figure 2.18; Table 2.2). Significant enrichments for the Mouse Phenotype (MP) term "abnormal circadian rhythm" ($q = 1.2 \times 10^{-4}$) and for Gene Ontology (GO) annotations relating to signalling pathways ($q < 0.05$; Figure 2.18), are consistent with the SCN's role as the master pacemaker. Enrichment for the flagellum annotation and G-protein coupled receptor (GPCR) signalling pathways (Figure 2.18) may reflect the roles of primary cilia as GPCR signalling hubs (Omori et al., 2015).

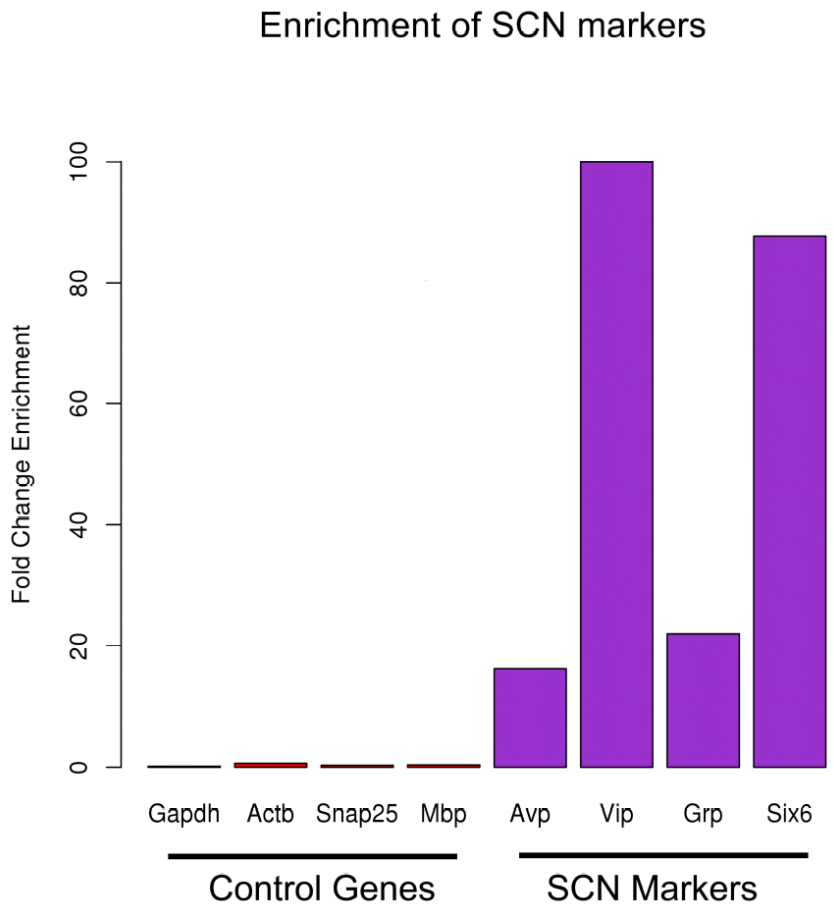


Figure 2.16. Enrichment of SCN marker expression in the SCN.
Fold-change enrichment of SCN marker (purple) and control gene (red) expression in the LCM SCN samples relative to mean brain FPKM levels from other transcriptomic studies is shown in Figure 2.15. These SCN markers display large FC enrichment compared to general markers of neurons (e.g. *Snap25*), oligodendrocytes (e.g. *Mbp*) and housekeeping genes (e.g, *ActB, Gapdh*).

These results thus suggest that GPCR-mediated neuropeptide signalling via cilia may be important for SCN function.

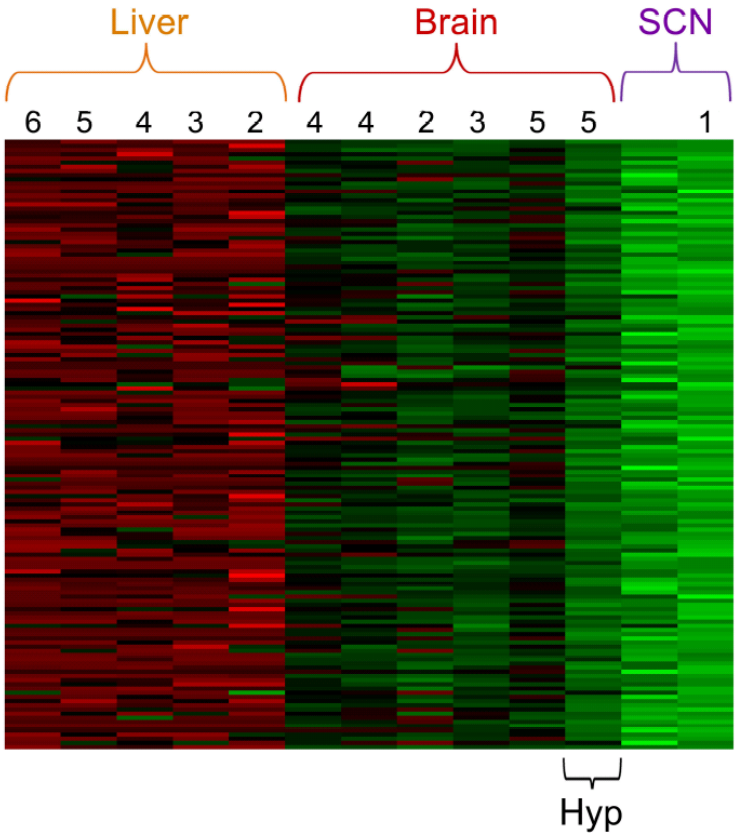


Figure 2.17. 146 genes with enriched expression in the SCN.
Heatmap showing mean expression levels of the 146 SCN-enriched genes across all other transcriptomic studies surveyed. Numbers indicate study from which data were collected as in Figure 2.15. These genes show the highest expression in the SCN followed by the hypothalamus, as expected given the location of the SCN within the hypothalamus.

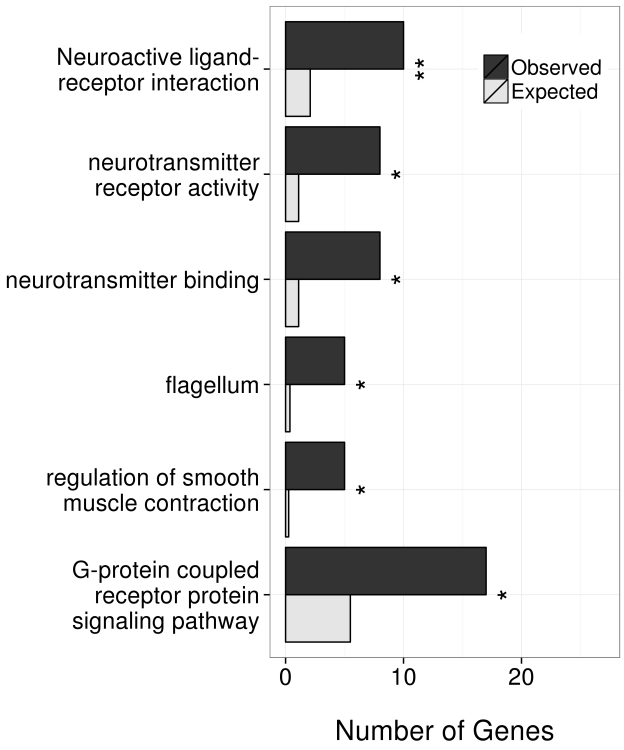


Figure 2.18. GO enrichments for SCN-enriched genes.
Statistically significant Gene Ontology (GO) enrichments for SCN-enriched genes ($q < 0.05$). Bars represent the number of genes of a given category expected by chance among SCN-enriched genes (grey) versus those observed (black). q-values: *$q < 0.05$, **$q < 0.01$, ***$q < 0.001$.

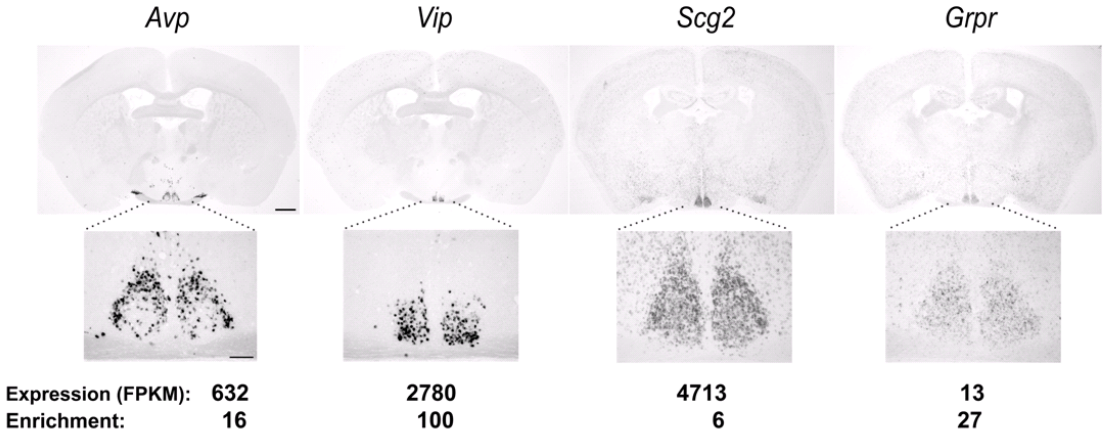


Figure 2.19. *In situ* hybridisation of SCN-enriched genes.
In situ hybridisation of four SCN-enriched genes at ZT6 in the wild-type brain. Scale bar: 0.5mm (low magnification), 0.2mm (high magnification).

Gene ID	Gene Name	FPKM	Enrichment score
Nms	neuromedin S	698.5	513.8
Npy6r	neuropeptide Y receptor Y6	35.8	285.0
Prok2	prokineticin 2	115.5	117.6
Vip	vasoactive intestinal polypeptide	2780.1	100.0
Six6	sine oculis-related homeobox 6	49.3	87.7
Gast	gastrin	35.0	54.8
D930028M14Rik	RIKEN cDNA D930028M14 gene	89.4	52.8
Gm17225	predicted gene 17225	11.4	45.9
Prokr2	prokineticin receptor 2	153.9	45.4
Nmu	neuromedin U	8.8	32.8
Syt10	synaptotagmin X	383.4	28.1
Gm20658	predicted gene 20658	5.7	27.9
Grpr	gastrin releasing peptide receptor	13.3	27.2
Chodl	chondrolectin	77.1	25.9
Slc14a2	solute carrier family 14 (urea transporter), member 2	20.1	25.7
Rln1	relaxin 1	31.0	22.5
Grp	gastrin releasing peptide	459.4	22.0
Fgf8	fibroblast growth factor 8	1.4	21.1
Kcnh8	potassium voltage-gated channel, subfamily H (eag-related), member 8	16.4	18.7
Rasd1	RAS, dexamethasone-induced 1	739.4	18.2
Ccdc109b	coiled-coil domain containing 109B	77.0	18.1
Tekt1	tektin 1	86.9	17.5
Vax1	ventral anterior homeobox containing gene 1	4.2	17.5
Vipr2	vasoactive intestinal peptide receptor 2	167.1	17.1
Avp	arginine vasopressin	632.2	16.3

Table 2.2. SCN-enriched genes.
Top 25 SCN-enriched genes ordered by enrichment score (fold-change enrichment of expression in the LCM SCN samples relative to mean brain FPKM levels from other transcriptomic studies surveyed). The full gene list can be viewed at http://www.wgpembroke.com/shiny/SCNseq/.

2.4 Discussion

In this chapter, the methods used to generate the transcriptome of the SCN over circadian time using RNA-sequencing were described. The generation of the SCN transcriptome forms the basis for further analyses within this thesis including: i) identification of SCN-enriched genes (this chapter), ii) investigation of the SCN transcriptome on a temporal scale (Chapter 3), iii) identification of novel transcripts in the mouse SCN (Chapter 4), and iv) identification of the transcriptomic differences in *Bdr*, a mouse model of psychiatric disease (Chapter 5).

2.4.1 Considerations for experimental design

Processing of batches

Exploration of the RNA-seq data using PCA revealed the presence of "batch effects" depending on the date sequencing libraries were generated. Although qPCR revealed that 43 of 59 (73%) of genes possessed a high concordance between platforms, some genes lack concordance due to expression which is elevated or depleted according to batch, preventing the identification of genes with true circadian expression patterns. This illustrates the importance of an appropriate experimental design in allowing the generation of biologically meaningful results.

PCA revealed that the RNA-seq data separated according to processing group — the set of samples which were generated and sequenced together. Each processing group shared the following procedures: source and housing of the animals, sectioning of the brains, extraction of the SCN via LCM, RNA extraction, generation of sequencing libraries

and sequencing. The library generation process is very susceptible to batch effects when performed at different times and is likely when biases were introduced to the LCM SCN dataset (Head et al., 2014). Within a batch, the slicing of the brains and SCN extraction occurred over a number of weeks, but library generation occurred on a single day. Large differences are not observed between the samples dissected or extracted on different days, but are seen between libraries generated on different days, suggesting that technical biases were introduced during the library generation step. QoRTs reveals systematic biases between the batches in relation to insert/fragment size, GC content and 3' bias, which are most likely to be introduced during library preparation. A solution would be to perform library generation and sequencing in the same batch. In this study, the samples were run in batches, principally so that the analysis could begin as soon as possible while the other tissue was being collected; this was due to the practicalities of breeding, housing and dissecting the large number of animals required for the study.

Considering these factors, if the sequencing data is to be sent in batches, it is important to ensure they are not confounded or correlated by the variable of interest. This could be achieved by separating the batches by biological replicate. This means that there will be one biological replicate per timepoint per genotype in a single batch (Figure 2.20). These changes in experimental design will not eliminate batch effects but ensures they are uncorrelated to the variable of interest. This prevents any false biological associations and allows statistical methods to regress out batch effects.

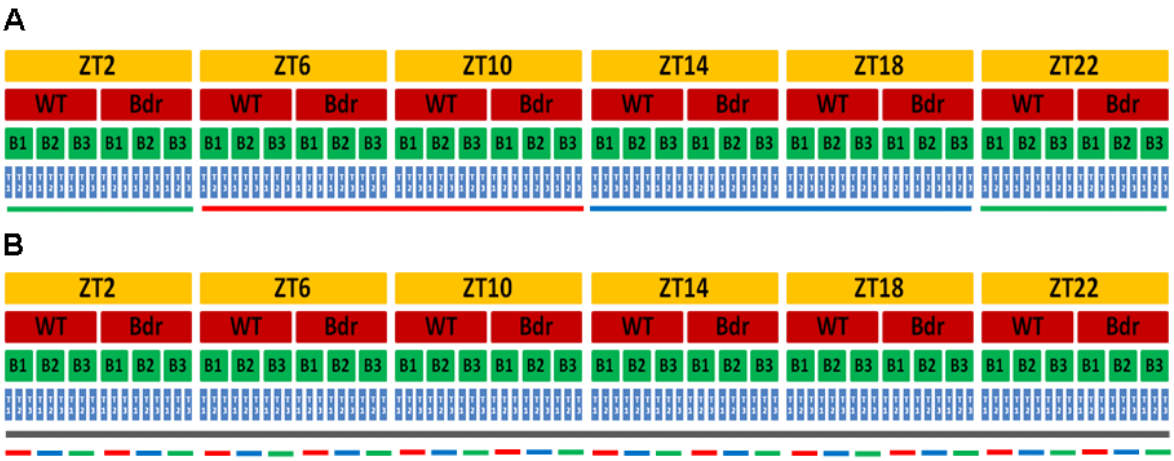


Figure 2.20. Current and proposed experimental design.
(A) Experimental design used - timepoint but not genotype confounded by batch (coloured lines). **(B)** Proposed experimental design - no biological variable of interest confounded by batch. Sample sequencing libraries are either generated and sequenced in one group (grey line) or separated into sequencing batches according to biological replicate (coloured lines).

Biomark qPCR

Mass validation of sequencing expression data was performed using Biomark qPCR. This method may be extremely useful for large scale sequencing experiments which need to be processed in batches or require technical validation. For this to be a viable option for future experiments there must be at least 500ng of extracted RNA remaining for qPCR after the majority has been sent for RNA-seq. Statistical approaches could also be developed to normalise RNA-seq expression data based upon the expression levels of up to 96 genes determined by Biomark qPCR. If this became the status quo in large scale sequencing projects the underlying sources of batch effects will soon be determined, allowing data to be correctly normalised and suggesting methods to avoid these batch effects being introduced during library preparation in the first place.

Spike-ins and normalisation

Spike-ins are RNA transcripts of a known quantity which are mixed with the experimental samples to allow normalisation of expression data. These transcripts share no similarity with sequences from the organism of interest so are not mistaken for biological transcripts during mapping. Commercially available External RNA Controls Consortium (ERCC) RNA spike-in mixes contain an assortment of these transcripts which differ by sequence and quantity and can be added to the samples before sequencing (and amplification) to allow normalisation after mapping (Jiang et al., 2011). This process is conceptually similar to using a set of endogenous "housekeeping" transcripts. However, as there is not a set of stably expressed genes known to exist in the SCN over circadian time, the use of RNA-spike-ins may be an interesting prospect.

An assumption of many statistical methods used to normalise expression levels and determine differentially expressed genes is that most genes do not significantly change their expression levels over different conditions (Love et al., 2014). Global gene expression may be subject to flux according to the time of day violating this assumption, perhaps preventing correct normalisation (Koike et al., 2012). RNA spike-ins allow detection of change in global expression allowing appropriate normalisation. For example, a study investigating the effect of the cell cycle gene *c-Myc* on gene expression identified genes which were both significantly up- and down-regulated (Loven et al., 2012). However, this assumed that global gene expression did not significantly change. Adding spike-ins to establish the absolute levels of gene expression determined that this assumption was violated and instead, *c-Myc* induced the upregulation of many genes, which supported prior beliefs (Loven et al., 2012). Therefore, the use of spike-ins could overcome any global changes in gene expression which would bias many normalisation techniques.

Standard normalisation techniques, where a single size factor is applied to each sample, are inadequate to remove batch effects. Instead a gene-/sample- normalisation offset (as in CQN) would instead be appropriate as genes with differing properties (such as GC content) might be differentially affected in different samples according to batch. Spike-ins can be designed to vary in length and GC content and be added in different quantities to allow a detailed investigation into these biases (Jiang et al., 2011, Risso et al., 2014). If the sequencing expression data possessed batch effects correlated with the biological variable of interest, spike-ins with varied properties may provide an accurate method of normalisation.

2.4.2 Collating publically available sequencing data to gain biological insight

In this chapter, publically available sequencing data was used to extract further information from the LCM SCN RNA-seq dataset. It is often problematic to compare samples generated through different protocols or within different laboratories as it is difficult to determine whether the differences are due to biology or technicalities. Global differences can be observed through PCA and can help assess the contribution of technical factors upon global expression compared to the biological variable of interest. PCA of the quantile normalised sequencing data revealed separation of the data according to the biological variable of interest suggesting that the datasets' comparison was valid (Figure 2.15).

146 genes were identified which showed enriched expression in the SCN relative to other brain regions (Figure 2.17). Confirming reliability of this method, the list contained known SCN markers and was enriched for Mouse Phenotype (MP) and Gene Ontology (GO) annotations involving circadian rhythm and signalling pathways (Figure 2.18). The enrichment of these categories of genes is expected as the SCN plays a key role in driving circadian rhythms and interacts with many regions of the brain. Enrichments were also observed for terms previously unexpected to be associated with the SCN (Figure 2.18).

This includes the enrichment of genes relating to the flagellum/ cilia. Although motile cilia are present in selected regions of the body such as the trachea, primary (or sensory) cilia are present in nearly every cell and have been proposed to act as cellular signalling hubs allowing the coordination of a large number of signalling pathways (Lancaster and Gleeson, 2009). The presence of ciliated signalling hubs in the SCN would help integrate the numerous signals received from the rest of the brain. The SCN-enriched genes also displayed an enrichment of G-protein coupled receptor (GPCR) signalling pathways

suggesting this process to be important in SCN function. Interestingly, sensory cilia are emerging as compartments for GPCR signalling (Sato et al., 2015, Mukhopadhyay and Rohatgi, 2014). For example, the SCN-enriched gene *Vipr2* encodes for a GPCR which has been found to localise to primary cilia in various brain regions including the SCN (Soetedjo et al., 2013). VIP, the ligand of this receptor plays a key role in synchronising the SCN and generating robust circadian rhythms. *Calcr* is another SCN-enriched gene and is annotated for both GPCR and cilia function. The *Drosophila* homologue of this gene is expressed by a restricted subset of circadian clock neurons and regulates sleep, showing the link between GPCR, ciliated signalling hubs and circadian control (Kunst et al., 2014). A temporal analysis of the circadian transcriptome (Chapter 3) revealed that both *Calr* and *Vip* expression peak at the same time in the SCN (ZT22), suggesting they may be involved in the same biological pathway. Together, these results suggest that GPCR-mediated neuropeptide signalling via cilia signalling hubs may be important for SCN function.

The list of SCN-enriched genes may also help identify individual genes important in SCN function. Recently a study identified a new protein coding gene *Nmf9* which possessed high expression in the SCN and when mutated, mice displayed deficits in circadian behaviour (Zhang et al., 2015). Interestingly, *Nmf9* shares many exons with another protein coding gene *Ankfn1*, which was determined to be SCN-enriched through the methods employed in this chapter. The *Drosophila* homologue (*Wake*) is expressed cyclically in arousal-promoting clock neurons and regulates sleep onset (Liu et al., 2014). This shows that the genes identified to be enriched in the SCN are likely to play key roles in circa-

dian or SCN function and may provide a candidate list of mouse mutants for circadian screening.

There is an association between neurological and psychiatric disease with circadian disruption. Genes which are implicated in these diseases and whose expression is enriched in the SCN (suggesting a role in circadian rhythm generation), may provide clues on the molecular mechanisms by which circadian disruption and neuropsychiatric disease are linked. For example, a genome-wide scan of rare copy number variant (CNVs) found a significant association of schizophrenia with duplications overlapping or within 89 Kb upstream of the *Vipr2* loci. In cultured lymphocytes with microduplications in this region, both *Vipr2* transcription and cyclic-AMP signalling were increased. Not only has cAMP signalling previously been implicated in schizophrenia, but it is known to be a major component of the central oscillator in the SCN (Millar et al., 2005, Turetsky and Moberg, 2009, O'Neill et al., 2008). cAMP displays circadian oscillations in mouse SCN tissue slices and reducing or elevating cAMP concentrations affected the circadian rhythmicity of these slices (O'Neill et al., 2008). Interestingly, peptide derivatives and small molecules have been identified which selectively inhibit VIPR2 (Moreno et al., 2000, Chu et al., 2010). Further development of these molecules may provide treatments for circadian disruption or schizophrenia.

The number of RNA-seq experiments is rapidly increasing, revealing more of the transcriptome in various tissues under different conditions. As additional relevant sequencing datasets become available, it would be of great scientific value if they could be automatically integrated with previous analyses, creating an updated list of SCN-enriched genes. When users are required to upload their raw sequencing files upon publication,

they should be required to fill in a standardised metadata file with a variety of information of the sample conditions and experimental protocol. This should allow automatic integration with the current deluge of available expression data providing the opportunity to compare between a variety of experimental traits.

2.4.3 Conclusion

For the first time, an anatomically precise and comprehensive examination of the SCN transcriptional output over 24 hours using RNA-seq was generated. Although, "batch effects" were present within my dataset, a large concordance was observed between the sequencing expression data and both positive control clock genes and qPCR expression data suggesting that these effects were limited. RNA-seq data was also used to identify SCN-enriched genes on a genome-wide scale. As well as supporting prior knowledge, these results offer novel insight into SCN biology by suggesting novel genes or processes important in its function.

3 Defining the circadian transcriptome of the mouse SCN

3.1 Introduction

Circadian rhythms are endogenous 24-hour oscillations in behaviour and physiology. These rhythms are set by autonomous clocks that measure time in the absence of environmental cues (Dunlap, 1999). Most studies have sought to define the circadian transcriptome by identifying genes and pathways which show rhythmic sinusoidal expression over 24 hours. As many of the core clock proteins are transcription factors (TFs), and display cyclic expression (sinusoidal oscillations which peak and dip once during the circadian day), cyclic expression may also be expected for their downstream clock controlled genes (CCGs) which possess E/E'-box, D-box, and ROR regulatory elements (RREs) in their promoters (Ko and Takahashi, 2006).

The advent of microarray technology facilitated genome-wide gene expression analysis across the circadian day. Many of these studies sampled from organisms which were entrained to a 12:12 LD cycle and subsequently released into constant darkness (Panda et al., 2002, Zhang et al., 2014, Hatori et al., 2014). Assessing circadian expression patterns from organisms housed in constant darkness allows identification of genes under the control of the core clock without the influence of photic input. To identify cycling transcripts in large transcriptomic datasets, a variety of algorithms have been developed such as COSOPT, Fisher's G and more recently JTK_CYCLE, all of which seek to fit expression to a sinusoidal curve (Wichert et al., 2004, Straume, 2004, Hughes et al., 2010).

Identification of these cyclic transcripts may provide novel insight into the biological processes under circadian control and the regulatory mechanisms which generate these oscillations (Panda et al., 2002, Ueda et al., 2002a).

Recently, the mouse transcriptome from 12 tissues was assessed over circadian time using microarray technology (Zhang et al., 2014). 43% of all protein-coding genes assessed possessed cyclic expression in at least one of these tissues. This percentage of genes, whose expression cycles in at least one tissue, is predicted to increase from 43% towards 55% as additional organs are sampled (Zhang et al., 2014). The SCN, focal to this thesis, has also been investigated by microarray. Using SCN tissue punches, Panda *et al.* identified nearly 650 genes whose expression cycled under constant darkness, many of which were found to be involved in peptide synthesis and the regulation of redox state suggesting that the temporal regulation of these processes in the SCN is functionally important (Panda et al., 2002). Ueda *et al.* identified 101 genes whose expression cycled in both LD and constant darkness using SCN tissue punches from mice and determined that Rev-ErbA/ROR response elements were important in generating expression rhythms which peaked at night in phase with *Bmal1* expression (Ueda et al., 2002a). Together, microarray technology has been used to identify those genes which cycle in numerous tissues, including the SCN, providing insight into circadian regulation.

In addition to these 24-hour cycles of expression, there are also second and third harmonics of circadian rhythmicity; i.e. expression patterns which display 12- or 8-hour rhythms (Hughes et al., 2009, Vollmers et al., 2009). High resolution profiling of liver samples (collected every hour for 48 hours) from mice found that genes with rhythmic expression had period lengths (time to complete an oscillation) which clustered at times

of 24 hours, 12 hours and 8 hours (Figure 3.1A-D) (Hughes et al., 2009). Genes with these shorter periods appeared to play key roles in cell division and lipid metabolism in liver (Hughes et al., 2009).

These harmonics of expression have been hypothesised to arise when a pair of rhythmic transcription factors both act on a promoter of the same gene but possess opposing circadian phases (Figure 3.1E) (Westermark and Herzel, 2013). Only at the times when both TFs are bound to the promoter will the gene be expressed. These 12-hour rhythms will display alternating peak heights if the circadian TFs are not exactly 12 hours out of phase with each other (Westermark and Herzel, 2013). In support of this model of core clock TFs regulating harmonics of circadian rhythmicity, mice with a disrupted *Clock* gene lose both their 24-hour and 12-hour rhythmic expression patterns (Hughes et al., 2012b). Harmonics of circadian expression have only been considered relatively recently (Hughes et al., 2009) and therefore previous microarray circadian datasets may have overlooked these expression patterns. Furthermore, as a high temporal resolution profiling (every hour) is often required to accurately identify these 12- or 8-hour oscillations, many circadian datasets do not have the power to accurately identify these patterns. Nevertheless, it is important to appreciate the presence of additional transcriptional patterns in circadian expression data, as well as the typical 24-hour oscillatory patterns.

Circadian transcriptional rhythms have also been assessed in a variety of organisms and tissues through sequencing. Zhang *et al.* measured the transcriptome from 12 tissues, including hypothalamus, in mice over circadian time using RNA-seq (Zhang et al., 2014). However, instead of sampling tissues every two hours over 2 days (as with their parallel array study), samples were taken every 6 hours under DD. This RNA-seq dataset from 12

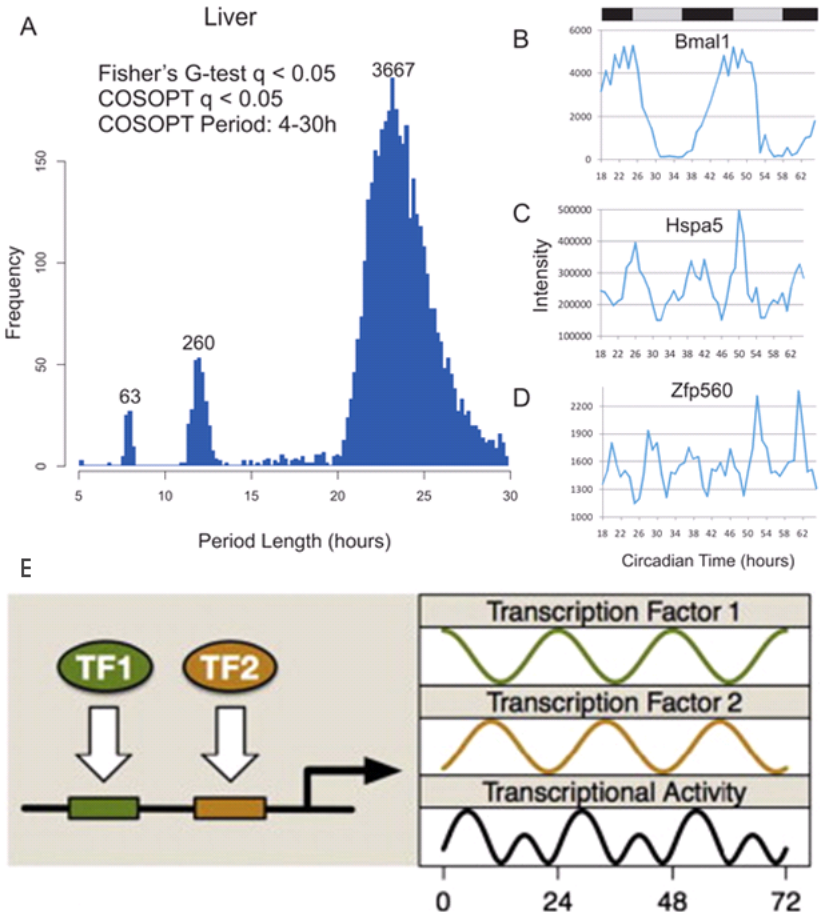


Figure 3.1. Harmonics of circadian rhythmicity.
Investigation into harmonics of circadian rhythmicity in mouse liver (Hughes et al., 2009). **(A)** Most genes with rhythmic expression in liver peaked every 24 hours, however smaller numbers of genes also displayed transcriptional rhythms with a period of 12 or 8 hours (Hughes et al., 2009). **(B-D)** Examples of genes whose expression possessed periods of 24, 12 or 8 hours. **(E)** Potential mechanism allowing generation of harmonic expression. Two circadian transcription factors (TFs) whose circadian expression are out of phase are both required to bind to the promoter to allow transcription of the downstream gene. Images obtained from (Hughes et al., 2009, Westermark and Herzel, 2013) with permission from Cell Press.

tissues also allowed the identification of 6,170 known and novel non-coding RNA of which 1,037 displayed rhythmic expression. RNA-seq can therefore be used as a method to identify those cycling coding and non-coding transcripts in a variety of tissues. Although, the SCN resides within the hypothalamus, which was assessed in this study via RNA-seq, the use of heterogeneous samples may dampen any SCN specific expression patterns. Furthermore, as the RNA-seq circadian transcriptome for each tissue was generated by sequencing one biological replicate (pooled from 3 mice) every six hours, some of the more intricate temporal expression patterns may be overlooked.

RNA-seq can also be paired with other genomic methods to gain a greater appreciation of transcriptional control. For example, analysis of expression data derived from Nascent-seq (sequencing of the RNA transcribed by RNA Polymerase II prior to 3' end formation) and RNA-seq of the liver found that 70% of rhythmically expressed genes do not display *de novo* transcriptional rhythms, indicating that post-transcriptional regulation plays an important role in driving expression level fluctuations (Menet et al., 2012).

The mammalian SCN is the master pacemaker of the circadian system and therefore a detailed comprehension of this brain region is important in understanding how circadian rhythms are generated. Microarray technology has been previously utilized to investigate the circadian transcriptome of the SCN and has helped provide an understanding of the biological processes under circadian control and the regulatory mechanisms which generate these oscillations (Panda et al., 2002 , Ueda et al., 2002a). However, at the time, microarrays could only assess up to 7,000 protein-coding genes and possessed other limitations such as the reduced signal-to-noise ratio of the representative probesets. Furthermore, as these previous array studies collected SCN "punches" to investigate the circadian transcriptome, the truly SCN cycling genes are likely to be dampened by the surrounding hypothalamic tissue and perhaps be overlooked. Lastly, previous studies largely measured the circadian transcriptome by collecting samples from mice housed under constant darkness (Panda et al., 2002, Zhang et al., 2014) and therefore those genes cycling under physiological relevant lighting conditions may not be identified.

Therefore, to gain a greater insight into the transcriptional output of the SCN across circadian time, mice were housed under physiologically relevant lighting conditions (12:12 LD) and SCN samples were accurately dissected using laser capture microdissection (LCM) at six timepoints across the circadian day (ZT2, 6, 10, 14, 18 and 22). The RNA from the SCN extracts were subsequently sent for sequencing to enable the profiling of circadian expression patterns for all coding and non-coding transcripts.

In this chapter, I sought to define the circadian transcriptome of the mouse SCN using RNA-seq, allowing a more comprehensive investigation than previous array technology. The precise plans were to:

- Identify coding and non-coding genes with cycling expression in the SCN using sinusoidal fitting algorithms typically used in the circadian field.
- Investigate the potential for non-sinusoidal circadian transcriptional patterns in the SCN using a co-expression analysis approach which does not assume a single mode of temporal variation.
- Determine which biological processes are circadian regulated by determining if there is an enrichment of genes with particular processes in particular circadian transcriptional profiles.
- Make the sequencing data available to the scientific community via an online web resource.

3.2 Methods

3.2.1 Identification of sinusoidal fluctuating genes

The R packages JTK_CYCLE (v2.1; (Hughes et al., 2010)) and DESeq2 (v 1.6.3) were used together to identify genes whose expression significantly altered over time with a sinusoidal oscillatory pattern.

DESeq2 was used to perform a likelihood ratio test (LRT) to determine the genes whose expression significantly altered over time ($q < 0.05$, Benjamini-Hochberg correction). The DESeq2 methodology is described in greater detail in chapter 5. The LRT tests for differences over multiple timepoints by examining two models for the count values observed. The full model includes the variable of interest (ZT/ Time) whereas the reduced model does not. The LRT determines whether the observed count values fit the full model better than expected compared to when the variable (ZT) was removed from the formula. LRT is similar to an ANOVA, except that it instead uses an analysis of deviance (ANODEV) where the deviance represents the difference between the full and reduced models.

JTK_CYCLE was used on the DESeq2 normalised count data to identify genes whose expression followed 24-hour periodic waveforms ($q < 0.05$, Bonferroni correction). The Jomckheere-Terpstra-Kendall (JTK) cycle applies principles from both the Jomckheere-Terpstra (JT) test and Kendall's tau. The JT test is a non-parametric algorithm which is most often used for identifying monotonic ordering of data across ordered independent groups. Kendell's tau is a non-parametric rank correlation which considers the correlation between two variables. JTK searches for the combination of period and phase that min-

imizes the p-value of Kendell's correlation. JTK_CYCLE is highly resistant to outliers because tau only depends on the sign differences between the groups thereby preventing artificially high/low datapoints that would otherwise greatly affect the correlation. As a result, fewer false negatives with a similar number of false positives are called compared to other algorithms used to detect cyclic oscillations (Table 3.1). JTK_CYCLE also predicts the phase (time of peak expression) for each gene surveyed. These phases were binned into 2- or 4-hour windows for visualization in Figure 3.9B and to group together those genes with similar peak phases for GO enrichment.

	Threshold	False-positives	False-negatives
JTK_CYCLE	$p < 0.05$	5	3
	$p < 0.01$	2	9
	$p < 0.001$	0	17
COSOPT	$p < 0.05$	4	135
	$p < 0.01$	0	152
	$p < 0.001$	0	187
Fisher's G	$p < 0.05$	10	131
	$p < 0.01$	1	156
	$p < 0.001$	0	187

Table 3.1. Comparison of algorithms which detect rhythmic oscillations.
Comparison of the sensitivity and specificity between JTK_CYCLE, COSOPT, and Fisher's G-test, algorithms utilised for detecting sinusoidal oscillations. COSOPT tests the goodness-of-fit between expression data and a series of cosine curves of varying phases and period lengths whereas Fisher's G determines the periodicity of the expression data and determines whether it is statistically periodically expressed or subject to random processes. On a test dataset, JTK_CYCLE displayed fewer false negatives with the same number of false positives. Table reproduced from (Hughes et al., 2010).

3.2.2 Identification of non-sinusoidal fluctuating genes

To identify genes with non-sinusoidal oscillatory patterns, Weighted Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath, 2008) was used on DESeq2 normalised count data to identify modules based upon gene co-expression. Instead of simply identifying genes which significantly change across conditions, WGCNA attempts to capture a greater depth of information by considering the relationships between transcripts using the pair-wise correlation score between gene expression profiles. WGCNA uses a network approach where each gene represents a node and the edges connecting them signify whether the genes are linked by co-expression.

To create a weighted gene network, the soft thresholding power needs to be selected, which is the degree to which co-expression similarity is raised to calculate adjacency (Zhang and Horvath, 2005). The choice of soft thresholding power determines the specificity and sensitivity of the connection strength between two genes. If the soft thresholding value is too high then the resulting network may be too sparse to allow detection of co-expression modules, whereas if the value is too low then there will be increased node connections which make the network more noisy (Zhang and Horvath, 2005).

Gene co-expression networks which do not satisfy scale-free topology should not be considered for further analysis as results are likely to be biologically meaningless (Zhang and Horvath, 2005). The soft threshold value of 10 was chosen as the R^2 (a metric used to quantify how well a network approximates a scale free topology) was relatively high (0.7; Figure 3.2). The scale free topology starts to plateau towards 0.8 from this point and therefore a gain in soft thresholding power does not provide substantially additional benefit. Furthermore, a power of 10 allows the generation of networks which have a

mean connectivity large enough to allow detection of co-expression modules. There is often a trade-off between maximizing the scale-free topology (R^2) and mean connectivity. Biological significance is often seen to be greatest when R^2 values are between 0.7 and 0.8 (Zhang and Horvath, 2005).

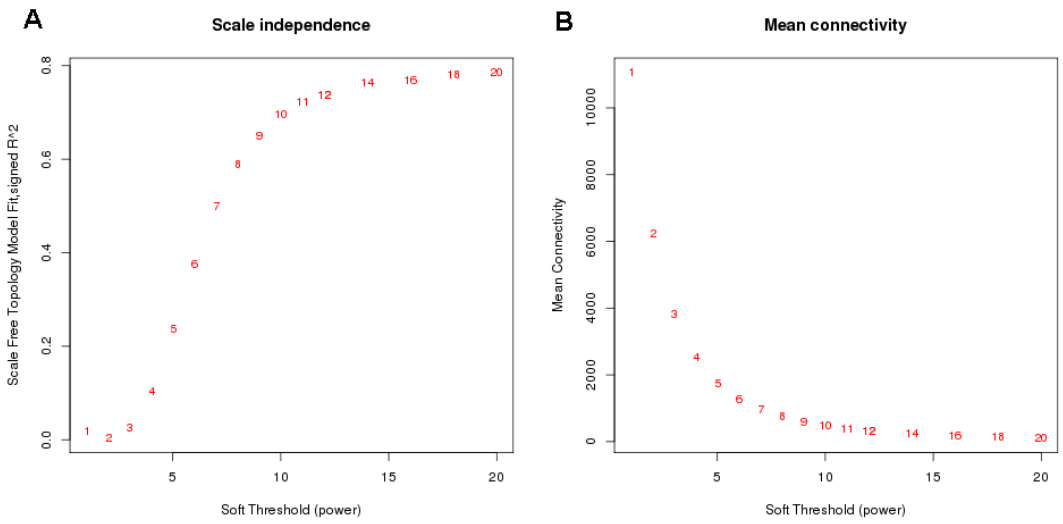


Figure 3.2. The relationship between soft threshold (power) and both scale free topology and mean connectivity.
Criteria for choosing a soft threshold. **(A)** The scale free topology as a function of soft-thresholding power. The larger the R^2 value the more the network approximates a scale free topology and therefore the greater the reliability of this co-expression network. **(B)** Mean connectivity as a function of the soft-thresholding power. The smaller the soft threshold value selected, the more node connections in the resultant network (which may make the network noisy). Conversely, the larger the threshold value selected, the fewer the connections in the resultant network (which may make the network too sparse to detect modules).

WGCNA was then utilised to create an adjacency matrix which corresponds to a co-expression network (Zhang and Horvath, 2005). A 'signed' adjacency matrix was created, meaning it only considered positively correlated genes, opposed to also considering

negatively correlated genes as co-expressed. A signed matrix therefore provides a clear and unambiguous comprehension of the specific circadian expression patterns that are occurring in the SCN. Considering negatively correlated genes prevents the generation of clusters with defined expression patterns and can therefore be more difficult to interpret.

If the network is unweighted, the adjacency matrix is encoded by 1 or 0 entries indicating whether or not a pair of genes is connected. In a weighted network, the adjacency matrix encodes the connection strength between each gene. The adjacency matrix was created by transforming co-expression similarities into connection strength. The adjacency matrix was then transformed into a Topological Overlap Matrix (TOM) which reflects their relative interconnectedness and minimizes the effects of noise and misleading associations, allowing the generation of more distinct gene modules than when considering only the correlation coefficient (Zhang and Horvath, 2005). The TOM matrix was then subtracted from 1 to create a dissimilarity matrix which was subsequently clustered to identify 54 network modules comprising of co-expressed groups of genes (Figure 3.5,3.6).

Modules were then merged if their module eigengenes possessed a correlation of over 0.7 (height cut at 0.3 according to dendrogram; Figure 3.5,3.6). The threshold of 0.7 was chosen by observing whether expression patterns of related modules were sufficiently similar to be merged. Of the resultant 23 modules, only the "lightsteelblue1" module was investigated further because it was the only non-sinusoidal module whose expression pattern was not influenced by likely anomalous expression of a single sample at a particular timepoint (Figure 3.3,35).

The lightsteelblue1 module comprised of 766 genes which were only defined as "fluctuating" if their expression significantly altered over time ($q < 0.05$, Benjamini-Hochberg

correction; DESeq2). The lightsteelblue1 module is described as “twin-peaking” later in the text.

blue module (n= 1380)
brown module (n= 2078)
brown4 module (n= 435)
darkgreen module (n= 2231)
darkmagenta module (n= 269)
darkorange module (n= 287)
darkred module (n= 338)
greenyellow module (n= 596)
ivory module (n= 190)
lightcyan module (n= 1180)
lightsteelblue1 module (n= 862)
lightyellow module (n= 393)
midnightblue module (n= 3883)
orange module (n= 296)
paleturquoise module (n= 907)
saddlebrown module (n= 564)
salmon4 module (n= 775)
sienna3 module (n= 118)
thistle2 module (n= 71)
turquoise module (n= 4701)
violet module (n= 179)
white module (n= 235)
yellowgreen module (n= 111)
Normalized Expression
ZT

blue eigengene vs. ZT
Module Eigengene (59.3%)
darkred eigengene vs. ZT
Module Eigengene (65.05%)
midnightblue eigengene vs. ZT
Module Eigengene (49.58%)
thistle2 eigengene vs. ZT
Module Eigengene (67.78%)
brown eigengene vs. ZT
Module Eigengene (57.26%)
greenyellow eigengene vs. ZT
Module Eigengene (54.91%)
orange eigengene vs. ZT
Module Eigengene (58.97%)
turquoise eigengene vs. ZT
Module Eigengene (49.32%)
brown4 eigengene vs. ZT
Module Eigengene (57.93%)
ivory eigengene vs. ZT
Module Eigengene (58.99%)
paleturquoise eigengene vs. ZT
Module Eigengene (48.4%)
violet eigengene vs. ZT
Module Eigengene (61.86%)
darkgreen eigengene vs. ZT
Module Eigengene (52.07%)
lightcyan eigengene vs. ZT
Module Eigengene (59.33%)
saddlebrown eigengene vs. ZT
Module Eigengene (54.64%)
white eigengene vs. ZT
Module Eigengene (54.98%)
darkmagenta eigengene vs. ZT
Module Eigengene (58.81%)
lightsteelblue1 eigengene vs. ZT
Module Eigengene (50.09%)
salmon4 eigengene vs. ZT
Module Eigengene (57.07%)
yellowgreen eigengene vs. ZT
Module Eigengene (59.16%)
darkorange eigengene vs. ZT
Module Eigengene (62.37%)
lightyellow eigengene vs. ZT
Module Eigengene (56.95%)
sienna3 eigengene vs. ZT
Module Eigengene (57.57%)
ZT

Figure 3.3. Normalised expression profiles of the 23 WGCNA co-expression modules.
Normalised expression of the genes in each of the 23 merged co-expression modules over the circadian day. A solid black line represents the mean expression of the module eigengene.

Figure 3.4. The module eigengene of the 23 WGCNA co-expression modules.
The module eigengene (first principal component) of the 23 merged co-expression modules over the circadian day. Coloured dots represent each biological replicate. Solid lines represent the mean value from the three biological replicates at each timepoint. The module eigengene accounts for between 48-68% of the variation in each of these co-expression modules.

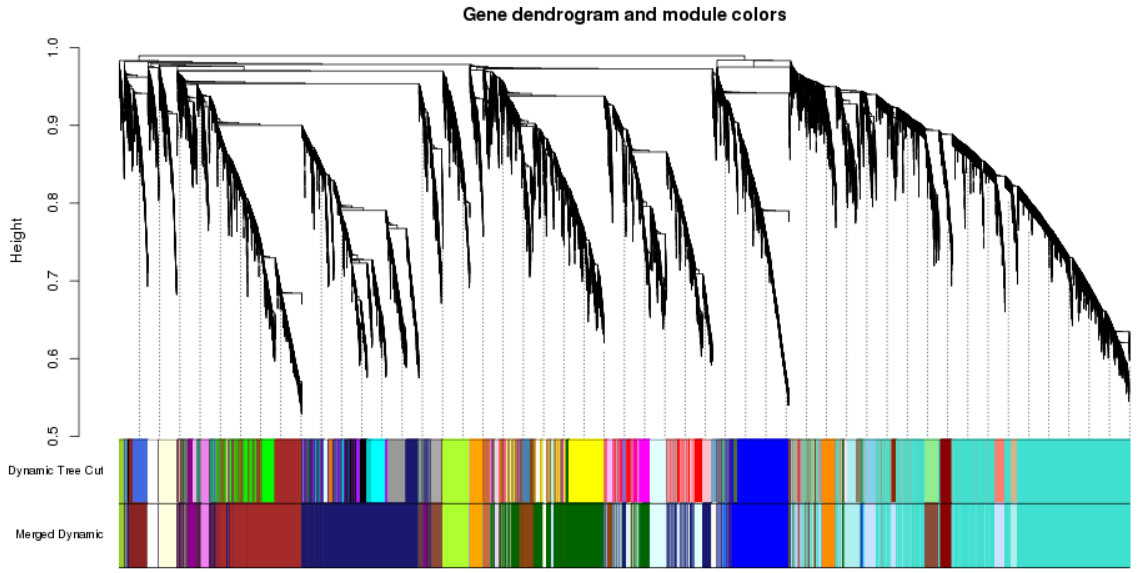


Figure 3.5. Dendrogram showing module membership of genes expressed in the SCN. Clustering dendrogram of 18,889 SCN expressed genes with assigned module colours. The genes were initially partitioned into 54 co-expression modules through dynamic tree cutting (upper coloured block), but were subsequently merged to form 23 modules (lower coloured block).

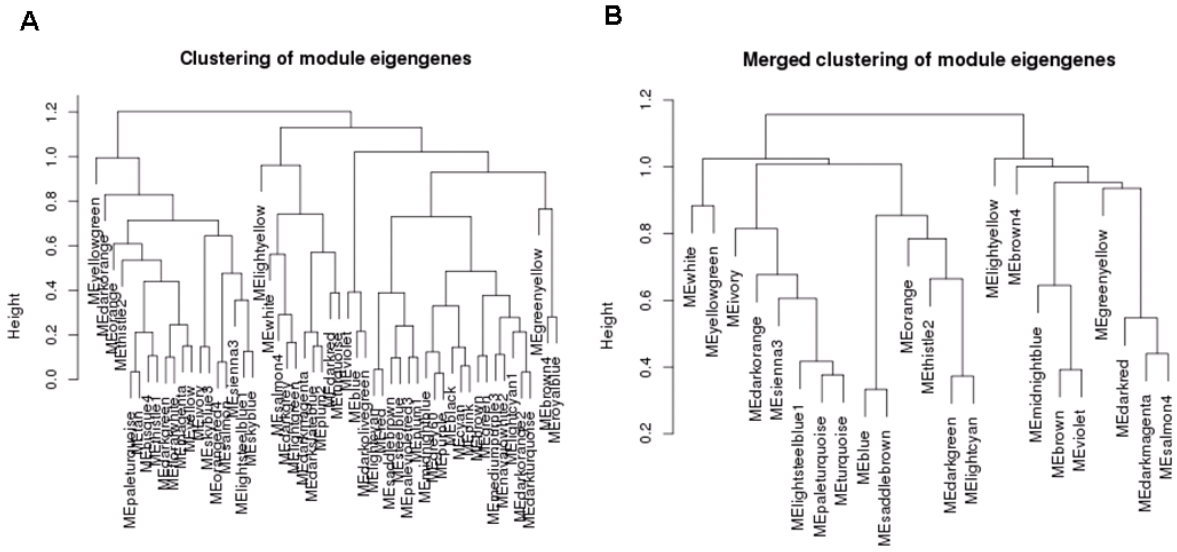


Figure 3.6. Clustering dendrograms of WGCNA co-expression modules.
(A) Clustering dendrogram of the 54 co-expression module eigengenes. Modules below a height of 0.3 were subsequently merged due to high correlation in expression. **(B)** Clustering dendrogram of the resulting 23 merged co-expression modules.

3.2.3 Generation of SCNseq data explorer

The SCNseq app was built using "Shiny" - a web application framework for R. The web app is hosted on a personal Ubuntu 14.04 x64 server with Rstudio and shiny server installed. The SCNseq data explorer presents data which have been generated via methods discussed in this thesis (Figure 3.7).

Further information can be viewed on the webpage - www.wgpembroke.com/shiny/SCNseq/.

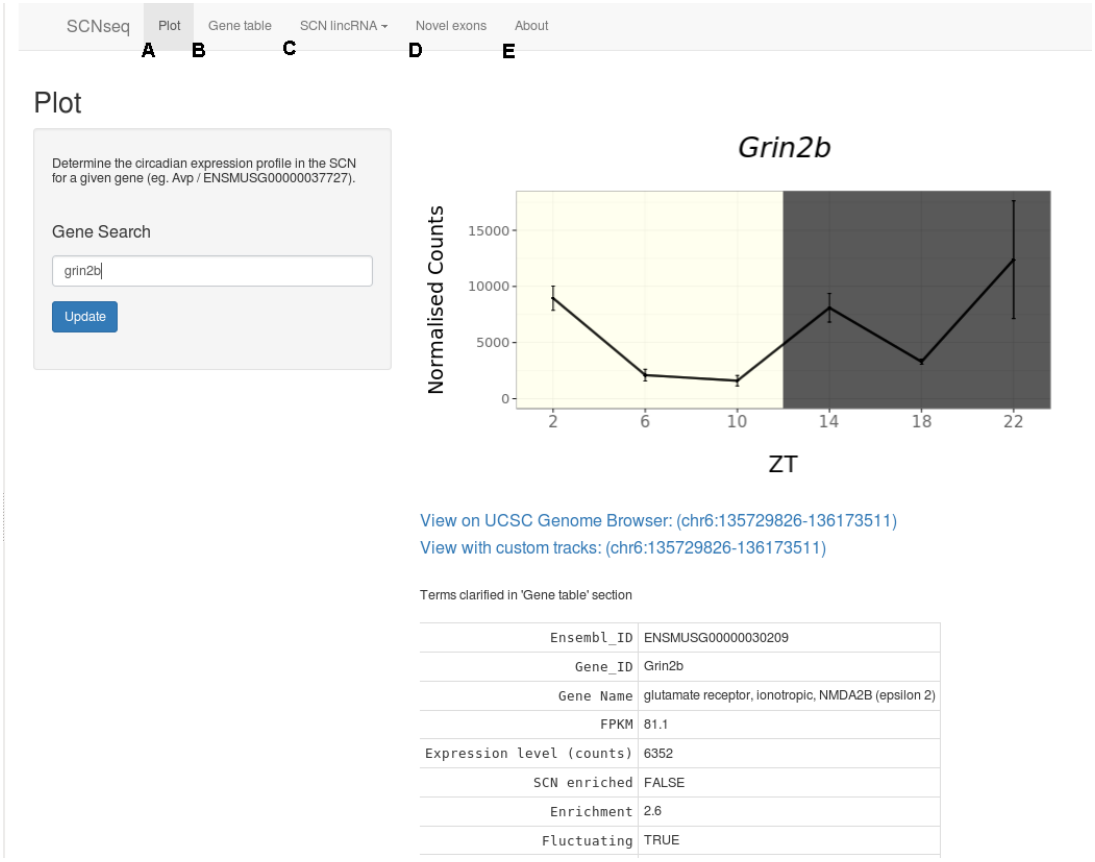


Figure 3.7. Screenshot of SCNseq data explorer.
Web interface offers: **(A)** Visualisation tool showing the circadian expression pattern of any gene which is expressed in the SCN. Gene name is inserted and generates graph and statistics table corresponding to that gene. **(B)** Gene table for all genes expressed in the SCN. Genes can be sorted or filtered based upon various statistics generated within this thesis. **(C)** Visualisation tool to view the circadian expression pattern of lincRNA expressed in the SCN (Chapter 4) or a lincRNA table to allow for filtering based on generated statistics. **(D)** Coordinates of the novel exons identified (Chapter 4) with hyperlinks to their position in the UCSC genome browser. **(E)** General information on background and experimental design. Only screenshot for **(A)** provided.

3.3 Results

3.3.1 Defining the circadian transcriptome

The expression of a gene was described as fluctuating if it displayed significant variation over time ($q < 0.05$; DESeq2) and if it either showed significant sinusoidal (cycling) expression ($q < 0.05$; JTK) or is part of the twin peaking co-expression module (WGCNA). Genes part of the twin peaking module whose expression also 'fluctuate' are referred to as 'significantly fluctuating twin-peaking genes'. Genes which 'fluctuate' and show significant sinusoidal oscillations are here referred to as 'cycling'. These definitions are illustrated by a Venn diagram (Figure 3.8). In total, 5,017 fluctuating (or circadian) genes were identified. Previous transcriptomic studies only considered sinusoidal expression patterns, therefore in these studies 'circadian' is synonymous with 'cycling'.

3.3.2 Identification of sinusoidally cycling genes and processes

In chapter 2, the SCN circadian transcriptome was generated of which 18,889 Ensembl-annotated coding or non-coding models were expressed ≥ 1 read) in over 6 samples. 4,569 genes (24% of 18,889 expressed genes) displayed a sinusoidal oscillatory expression pattern according to both JTK Cycle and DESeq2 approaches (Figure 3.9A-B). The peak phases for these cycling genes appeared to peak during transcriptional "rush hours" near to light transitions as described previously (Figure 3.9B) (Zhang et al., 2014). Importantly, this number of 4,569 cycling genes is 7-fold higher than previously reported in the SCN by microarrays (Panda et al., 2002). This large increase is likely to reflect four factors:

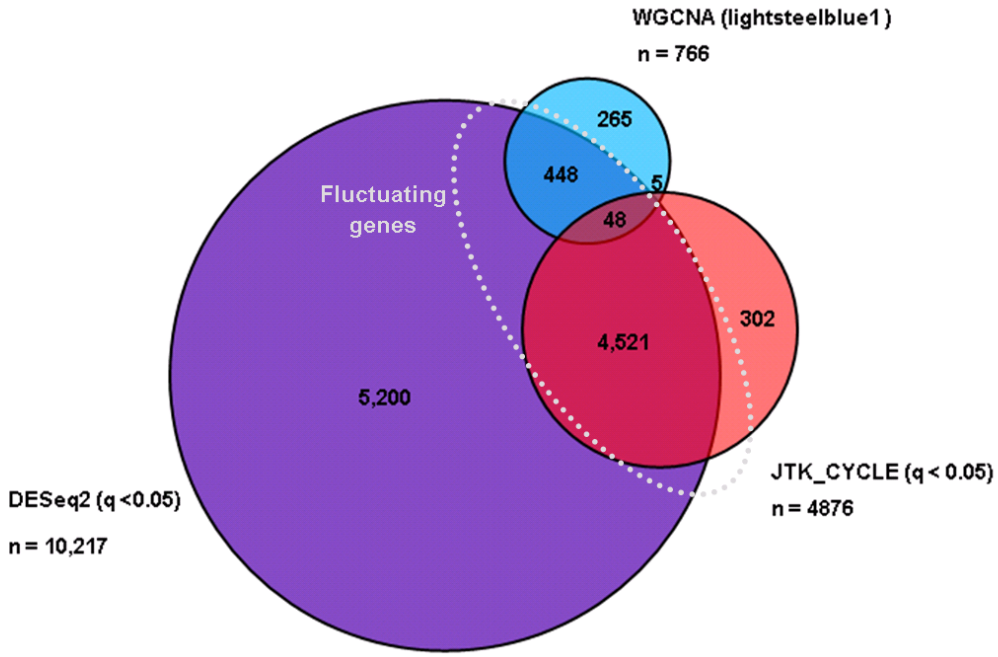


Figure 3.8. Venn diagram showing fluctuating genes identified through various statistical approaches.
Identification of 5,017 fluctuating genes (4,521 + 48 + 448) of which 4,569 (4,521 + 48) display significant sinusoidal oscillation (cycling) and 496 (448+48) displayed significant twin-peaking fluctuation.

First, the significantly improved anatomical enrichment of the SCN tissue using laser capture microdissection (LCM). Previous studies used SCN 'punches' (typically $1mm^3$ hypothalamic sample of which 30% comprises the SCN); such heterogeneous samples will result in reduced amplitudes of oscillations for genes that cycle asynchronously between the SCN and surrounding regions (Zhang et al., 2014). Second, the greater sensitivity and specificity of the methodological and statistical approaches (Hughes et al., 2010). RNA-seq not only can survey more genes than microarrays (18,889 (number of expressed genes in LCM SCN RNA-seq study) vs 7,000 (number of probes in SCN array study (Panda et al., 2002))), but it is also more sensitive, providing a greater dynamic range at both high and low expression levels (Wang et al., 2014). Furthermore, the JTK_CYCLE algorithm has been shown to be more sensitive and specific in detecting cycling genes than other sinusoidal based algorithms (Table 3.1) (Hughes et al., 2010). Third, the presence of batch effects may create artificially cycling genes. As time is perfectly confounded by batch it is difficult to distinguish between the effects due to the variable of interest (time) or the effects due to technical processes (batch). Although a large proportion of the fluctuating transcripts were shown to validate via qPCR, a number did not, perhaps for reasons relating to batch effects (Figure 2.12). Finally, our sampling approach under LD differed to the constant conditions (DD) utilised in earlier studies (Panda et al., 2002). By housing mice in LD conditions, we will be able to detect cycling genes that may be physiologically relevant to light-exposure or light-induced transcriptional cascades that are otherwise undetectable as cycling under DD (Ueda et al., 2002b, Leming et al., 2014).

Our new collection of SCN cycling genes was enriched for GO terms relating to key processes such as translation and mitochondrion-related genes, in line with previous

findings (Figure 3.9; (Panda et al., 2002)) which suggests that the temporal regulation of these processes in the SCN is functionally important (section 4.1.5). To identify processes which may induce a differential response to light at ZT14 and ZT22, a GO enrichment analysis was employed for genes with peak expression phases of either ZT14 or ZT22. Although there were no GO term enrichments for genes peaking at ZT22, genes which peaked at ZT14 were enriched for terms relating to actin cytoskeleton organization and small GTPase regulator activity (Figure 3.10A). Visualization of their peak phases shows that cycling genes with these GO terms are expressed at a time period between ZT12 and ZT14 (Figure 3.10B).

3.3.3 Comparison with other circadian transcriptomes

To identify the genes which exclusively cycle in the SCN and those which cycle across all tissues, the 4,569 genes which cycled in the SCN were compared to those which were called as cycling in up to 12 other tissues based on published circadian expression data (JTK Cycle; $q < 0.05$) (Zhang et al., 2014). Approximately half of the SCN cycling genes (2451 / 4,569) showed no evidence for sinusoidal expression in 12 other tissues, including the whole hypothalamus (Figure 3.11A) (Zhang et al., 2014). These genes included those essential for normal SCN function, such as *Prok2, Rorb* and *Nts*, which were also validated by qPCR (Figure 23). This indicates that many genes cycle exclusively in the SCN, or only cycle under a 12:12 LD lighting schedule rather than under the constant dark conditions used previously (Zhang et al., 2014). Comparing our results under LD with published DD datasets may inflate the perceived number of genes cycling solely in the SCN. Those genes that only cycle in LD conditions due to light-induced transcriptional cascades will be

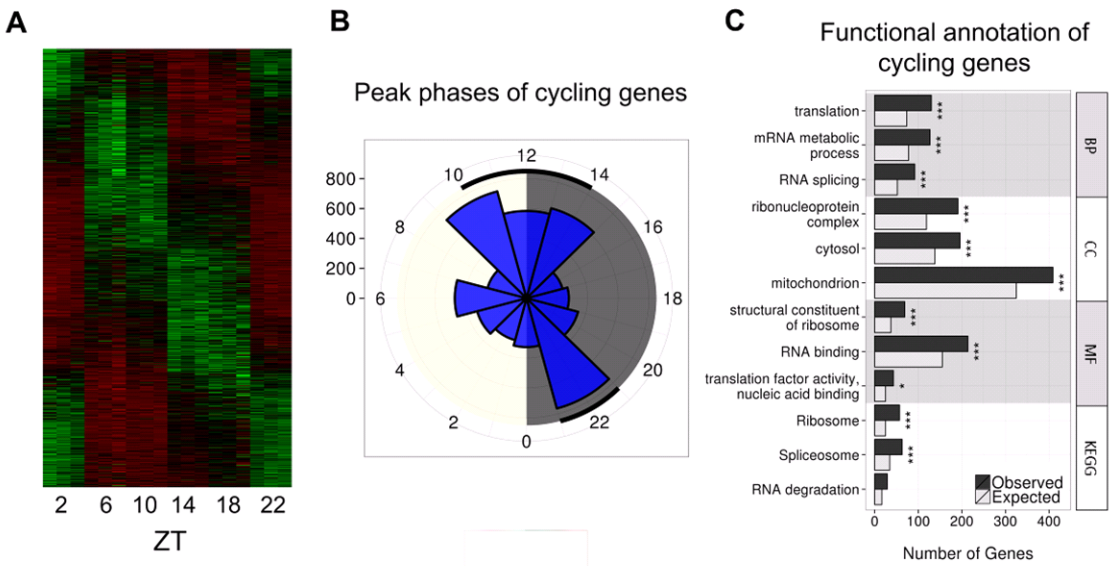


Figure 3.9. Identification of 4,569 cycling genes in the wild-type mouse SCN.
(A) Heatmap showing the expression levels in each biological replicate for the 4,569 significantly cycling genes in the SCN across circadian time ordered by peak phase (DESeq2 and JTK Cycle, $q < 0.05$). **(B)** Polar plot showing the distribution of peak phases for the identified 4,569 cycling genes binned in two-hour windows. Cycling genes tend to peak during transcriptional "rush hours" near to light transitions (shown by black lines), as described previously (Zhang et al., 2014). **(C)** Top 3 statistically significant GO (Biological Process (BP), Cellular Component (CC), Molecular Function (MF)) and KEGG enrichments for the 4,569 cycling genes. q-values: *$q < 0.05$, **$q < 0.01$, ***$q < 0.001$.

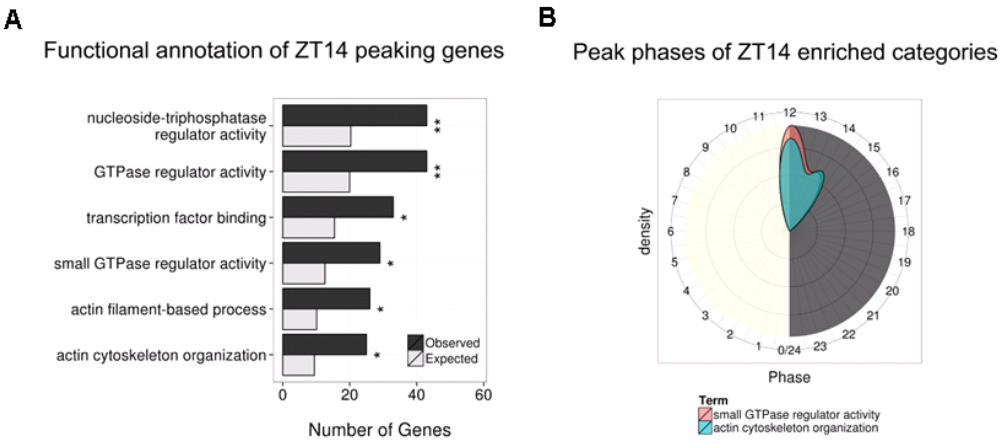


Figure 3.10. Genes involved in cytoskeletal assembly show peaks in expression in early night. **(A)** Statistically significant GO and KEGG enrichments for the 1,034 cycling genes peaking at ZT14. q-values: *q $<$ 0.05, **q $<$ 0.01, ***q $<$ 0.001. **(B)** Distribution of peak phases for the 25 and 29 significantly cycling genes with the GO annotations "actin cytoskeleton organisation" and "small GTPase regulator activity" respectively.

detected in the SCN LCM study but unlikely to cycle in DD for other tissues (Ueda et al., 2002b, Leming et al., 2014). Microarray studies in lower organisms identified those genes which cycle under LD and DD conditions. Comparing the cycling gene lists obtained from both lighting conditions revealed that a greater number of genes cycled exclusively in LD (787) than in DD (639) and that the circadian transcriptomes greatly differed between lighting conditions — of the 1,035 genes which cycled under LD, only 248 also cycled under DD (Leming et al., 2014).

Six of the 9 genes which cycled significantly across all 12 other tissues (q $<$ 0.05) based on published circadian expression data also cycled in the LCM SCN samples (*Arntl, Dbp, Nr1d2, Per2, Per3* and *Tsc22d3*) suggesting a core set of universally cycling genes (Figure 3.11A). In support of this notion, our LCM SCN cycling gene list significantly

overlapped with those cycling in 8 of 12 other tissues ($q < 0.05$), as well as those cycling in an SCN microarray study (Zhang et al., 2014, Panda et al., 2002) (Figure 3.11B). As expected, the SCN, hypothalamus and cerebellum showed the greatest overlap of their circadian transcriptomes compared to other organs (Figure 3.11B). This suggests that anatomically similar regions are likely to possess similar circadian transcriptomes.

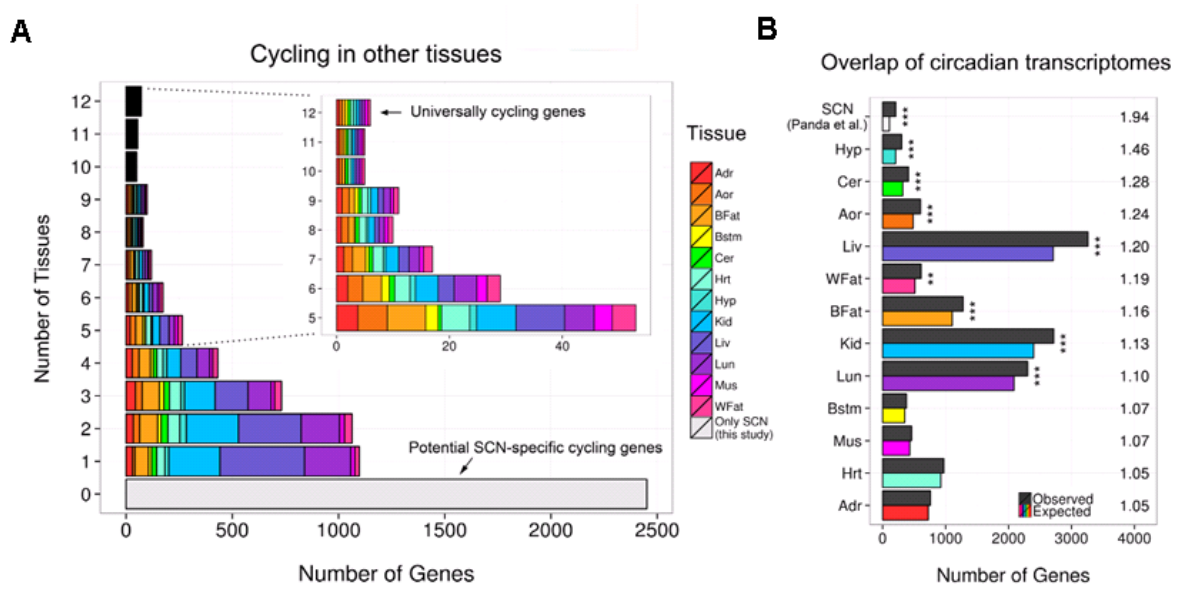


Figure 3.11. Comparison of circadian transcriptomes.
(A) Number of other tissues in which the 4,569 cycling genes also cycle. Colours indicate the distribution of these cycling genes across the different tissues (Adr — adrenal; Aor — aorta; BFat - brown fat; Bstem — brainstem; Cer — cerebellum; Hrt - heart, Hyp — hypothalamus; Kid — kidney; Liv — liver; Lun — lung; Mus - skeletal muscle; WFat - white fat). **(B)** Number of the 4,569 cycling genes that are also significantly cycling in other tissues from other studies (Zhang et al., 2014) ordered by fold-change enrichment (right). Tissues which show significantly greater LCM SCN cycling genes are denoted with an asterisk (*$q < 0.05$, **$q < 0.01$, ***$q < 0.001$); as expected, the greatest enrichment in overlap was with the SCN results from the microarray study of (Panda et al., 2002).

3.3.4 Identification of novel twin-peaking synaptic module

Previous circadian transcriptomic studies only considered circadian gene expression following a sinusoidal pattern by using sinusoidal fitting algorithms such as JTK_CYCLE (Hughes et al., 2010, Zhang et al., 2014). To identify alternate transcriptional profiles, WGCNA — which does not assume a single mode of temporal variation - was employed on the DESeq2 normalised expression data. Interestingly, a single co-expression module of 766 genes was identified, which instead of adopting a sinusoidal profile, displayed two peaks of expression at ZT14 and ZT22 (Figure 3.12; Table 3.2). 496 genes from this 'twin-peaking' module individually exhibited statistically significant temporal fluctuations (DESeq2, $q < 0.05$), of which 17 of 17 were validated by qPCR (Figure 3.13).

Functional annotations related to synaptic transmission, calcium signaling and gated channel activity were greatly and significantly enriched among these twin-peaking genes ($q < 0.05$; Figure 3.14), suggesting a circadian component to establishing the electrophysiological environment of the SCN.

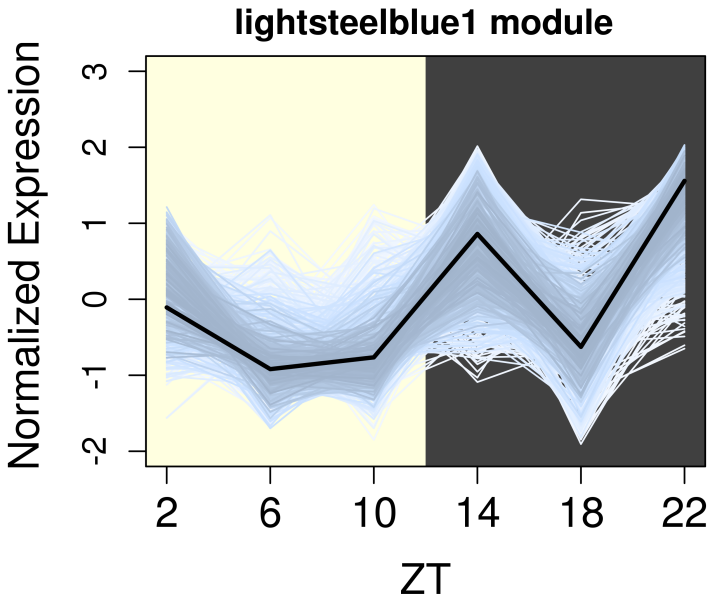


Figure 3.12. Gene clustering identifies a novel twin-peaking synaptic gene module.
Relative expression of each gene present in the identified twin-peaking module over 24 hours. The black line represents the module eigengene — the first principal component of the gene expression matrix for this module.

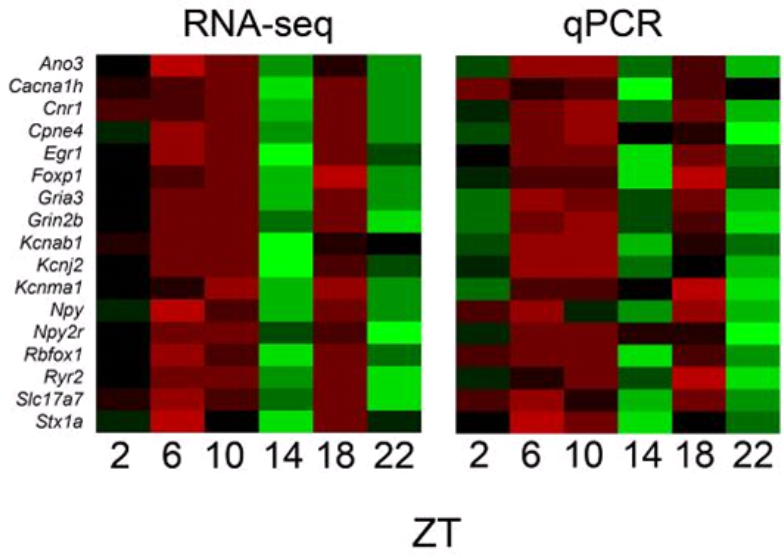


Figure 3.13. Validation of twin-peaking expression via qPCR.
Mean expression of twin-peaking genes as assessed by RNA-seq and qPCR. High concordance between methods suggests that the novel transcriptional pattern is a true biological signal.

Ensembl ID	Module membership	Gene ID	Gene name
ENSMUSG00000053477	0.927	*Tcf4*	transcription factor 4
ENSMUSG00000026872	0.912	*Zeb2*	zinc finger E-box binding homeobox 2
ENSMUSG00000005583	0.909	*Mef2c*	myocyte enhancer factor 2C
ENSMUSG00000037984	0.903	*Neurod6*	neurogenic differentiation 6
ENSMUSG00000032402	0.888	*Smad3*	SMAD family member 3
ENSMUSG00000051177	0.885	*Plcb1*	phospholipase C, beta 1
ENSMUSG00000041324	0.877	*Inhba*	inhibin beta-A
ENSMUSG00000038418	0.875	*Egr1*	early growth response 1
ENSMUSG00000010175	0.871	*Prox1*	prospero-related homeobox 1
ENSMUSG00000049107	0.870	*Ntf3*	neurotrophin 3

Table 3.2. Top 20 hub genes in the twin-peaking co-expression module
The top 20 twin-peaking genes in relationship to module membership. Genes with high module membership are most representative of the co-expression module due to their high correlation to the module eigengene.

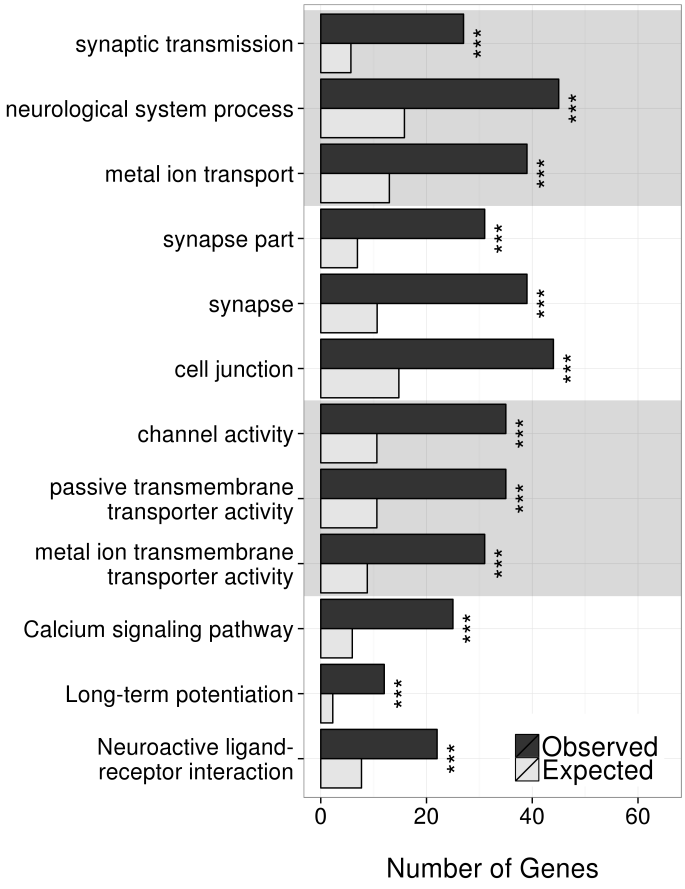


Figure 3.14. Functional annotation of the twin-peaking module.
Top 3 statistically significant ($q < 0.05$) GO (Biological Process (BP), Cellular Component (CC), Molecular Function (MF)) and KEGG enriched terms for the 496 significantly fluctuating twin-peaking genes. q-values: *$q < 0.05$, **$q < 0.01$, ***$q < 0.001$.

3.4 Discussion

3.4.1 The power of co-expression and unbiased approaches in identifying circadian regulated processes

Using analyses described in this chapter, 5,017 genes were identified that displayed significant fluctuation over circadian time.

Previous circadian transcriptomic studies only considered circadian gene expression following a sinusoidal pattern. This seems intuitive as most circadian processes display a single peak and trough over 24 hours. Furthermore, as the core clock proteins display cyclic expression, their transcriptional targets (clock controlled genes - CCGs) will also be likely to show cyclic expression (Kondratova and Kondratov, 2012). However, as these studies did not investigate the potential for non-sinusoidal patterns of expression, alternative and potentially functionally important expression profiles may have been overlooked. Using WGCNA - which does not assume a single mode of temporal variation - a single module of 766 genes was identified, which instead of adopting a sinusoidal profile, displayed two peaks of expression at ZT14 and ZT22 (Figure 3.12). This profile is also distinct from previously described 12-hour harmonics (Hughes et al., 2009), and prior to this, only single genes from individual studies were shown to possess similar expression patterns (Shinohara et al., 1993b, Cagampang et al., 1998a, Cagampang et al., 1998b). For example, isoforms of protein kinase C (*Prkcb* and *Prkcg*) exhibit a twin-peaking profile in rat SCN similar to that observed in our RNA-seq data (Cagampang et al., 1998a). This twin-peaking transcriptional profile, particularly for this large number of genes, was thus unexpected.

To obtain meaningful results using WGCNA, the study must be well powered possessing both a large sample size (at least 15 samples) and biological replicates for each condition (Langfelder and Horvath, 2008). My study possessed 18 SCN samples and three replicates per timepoint which allowed me to determine whether circadian expression was driven by an outlier. Unfortunately, many transcriptomic datasets do not have this power. For instance, the first microarray of SCN tissue punches over circadian time comprised of 22 samples but had only two replicates per timepoint (Panda et al., 2002). Furthermore, the array only surveyed the expression for 8,000 protein-coding genes (vs. 22,079 in my RNA-seq dataset) making it difficult to generate a detailed co-expression network. Inspection of the SCN array data through WGCNA determined that it was not suitable to identify co-expression networks for further investigation (Figure 3.15). Recently, the circadian transcriptome of 12 tissues was surveyed using array and RNA-seq. Samples for microarray were extracted every 2 hours (24 timepoints) and samples for array were extracted every 6 hours (8 timepoints). Although these samples were pooled from 3 separate mice to average out biological and technical variance, there was only one biological replicate per timepoint preventing reliable identification of robust co-expression patterns. Therefore, the LCM SCN study which possessed 18 samples and 3 biological replicates per timepoint provides a robust platform to generate co-expression networks across circadian time or identify novel transcriptional patterns.

In this chapter, only the lightsteelblue1 (twin-peaking) module of the 23 produced by WGCNA was chosen for further investigation. This is because it was the only robust non-sinusoidal expression pattern that was not driven by outlier expression — a single biological replicate with differing expression to the other two biological replicates (Figure

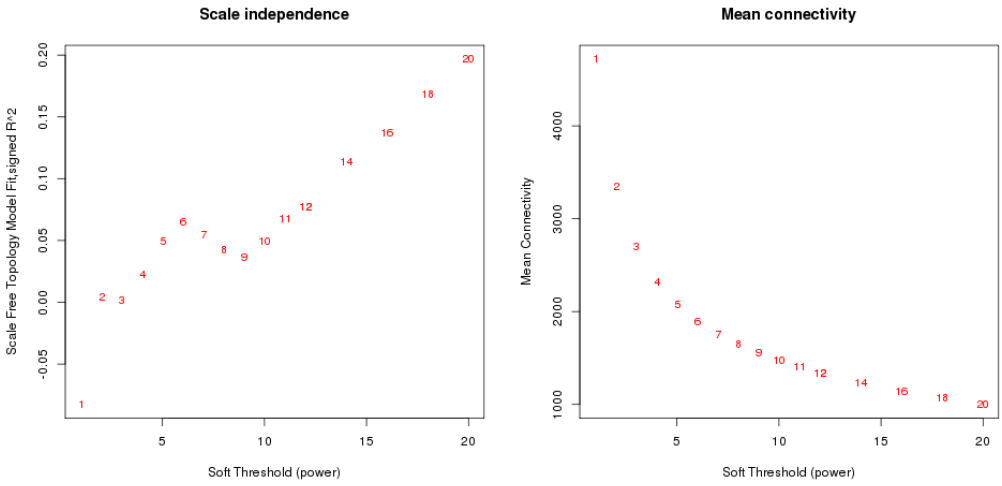


Figure 3.15. SCN microarray data not suitable for WGCNA analysis.
The SCN microarray gene co-expression network does not satisfy a scale-free topology ($R^2 > 0.7$) at any soft-thresholding power. The resultant network will be too noisy and not yield biologically relevant results.

3.4). All other modules appeared to have a single biological replicate which was influencing the mean expression for that timepoint. For example, the normalized expression of the "violet" module appeared to show 12-hour rhythms, peaking at both midday and midnight, but was likely to be driven by outliers (Figure 3.3,3.4). Interestingly, co-expression modules representing harmonics of circadian rhythmicity were not identified (Hughes et al., 2009). Perhaps those genes which peak every 8 or 12 hours possess a continual phase distribution (peak throughout the day) and are not "co-expressed", preventing them from being partitioned into their own separate module. Furthermore, harmonics of circadian expression have usually been identified when samples are profiled at a higher temporal resolution (every hour) (Hughes et al., 2009).

There were also robust co-expression patterns which displayed sinusoidal-like expression. However, as discussed in chapter 1, these transcriptional profiles are extremely susceptible to batch effects. If expression is determined solely by batch, perceived expression for some of these genes could approximate a sinusoidal profile and be clustered in the same co-expression module as true sinusoidal circadian genes. Thus, JTK Cycle was chosen to identify sinusoidal genes because this method is more robust against genes whose expression changes only according to batch.

Co-expression modules which contain known clock genes may provide clues on circadian regulation. For example, the binding of the CLOCK-ARNTL heterodimer to E-box elements drives the expression of the clock genes *Per1-3* and *Cry1* as well as other CCGs (Kondratova and Kondratov, 2012). The E-box-driven clock genes (*Per1-3* and *Cry1*) are present in two co-expression modules (brown and midnightblue). Consequently, the other genes also present in these modules may also possess E-box sites and be driven by the CLOCK-ARNTL heterodimer. It would be interesting to determine whether these genes contain significantly more E-box sites in their promoters than expected by chance, therefore revealing a set of CCGs. In addition, as these CLOCK-ARNTL driven CCGs reside within separate co-expression modules (brown module - *Per1, Per3, Rev-erbB*; midnightblue module - *Cry1, Per2, Rora, Avp, Prok2*), it suggests that these two modules may be differentially regulated.

In summary, I believe that previous sinusoidal methods of assessing circadian expression profiles are outdated and circadian studies should transition towards co-expression analysis and unbiased statistical methods (such as ANOVA) to identify circadian genes. Although JTK_CYCLE is an accurate and robust method for identifying expression pat-

terns which possess a sinusoidal profile, this method has many limitations. For example, JTK_CYCLE assumes that circadian expression patterns are sinusoidal (Hughes et al., 2010). Although this may seem intuitive given the rhythmic expression of core clock proteins, the circadian regulation of genes is more complicated than the binding of a single activating TF complex as described in the core-clock loop (Figure 1.3) (Kondratova and Kondratov, 2012). Instead, a variety of TFs binding at different times combined with a variety of post-transcriptional mechanisms are likely to produce a variety of transcriptional responses (Westermark and Herzel, 2013, Xue et al., 2014). The co-expression analysis approach applied in this chapter makes no assumptions, and detects the main transcriptional patterns in the data which statistically fluctuate over time. The incorporation of these unbiased methods allowed the identification of an unexpected transcriptional profile which may play a key role in synaptic transmission pathways.

Secondly, co-expression analysis has strong implications in biological regulation: genes which show similar expression and are co-expressed across samples are likely to be transcriptionally (or post-transcriptionally) regulated in a similar manner and therefore likely share similar biological functions (Azuaje, 2014). As illustrated in Figure 3.16, although JTK_CYCLE is proficient at identifying robust sinusoidal expression patterns, it fails to capture the richness of RNA-seq data. For example, even if JTK_CYCLE identifies two genes whose expression peaks at the same time, it does not consider how their expression relates to each other. Gene expression may oscillate in a similar manner but not be highly co-expressed across all samples, suggesting differing underlying regulatory mechanisms (Figure 3.16).

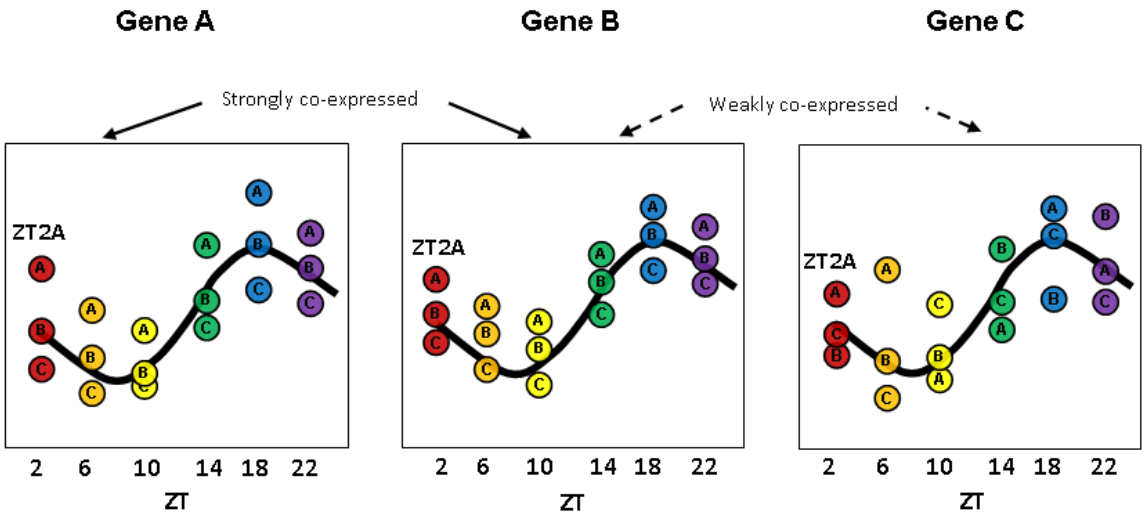


Figure 3.16. Co-expression analysis captures expression relationship between genes.
The expression of three hypothetical genes across circadian time. Each timepoint possesses three biological replicates (labeled A-C). Samples from each timepoint are coloured differently. WGCNA will likely partition Gene A and Gene B in the sample co-expression module; not only do they show similar circadian expression patterns but the expression levels of their samples are in the same order at each timepoint — they are highly co-expressed. Although both Gene A and Gene B share the same phase as gene C, they do not appear to be highly co-expressed as the ordering of samples within a timepoint is inconsistent. By comparing pairwise similarity for all genes (not just three), complex regulatory networks (Figure 3.3) are generated which can be very biologically informative.

WGCNA co-expression analysis can be further exploited by identifying modular 'hub' genes — genes which possess high intramodular connectivity and therefore likely drive the expression of its module (Chou et al., 2014, Azuaje, 2014). This approach of hub gene identification has been employed to implicate genes in many neurological disorders such as Alzheimer's disease and autism (Forabosco et al., 2013, Parikshak et al., 2015, Parikshak et al., 2013). However, for circadian studies, this approach must assume that there is minimal lag between the transcription of a hub gene and its transcriptional regulation of the other genes in that module. Over 40% of proteins in liver display a lag of >6 hours relative to their transcript levels, including many TFs (Robles et al., 2014). Therefore, the TF hubs do not necessarily regulate the transcriptional module it is part of, unless there are mechanisms allowing immediate translation. Nevertheless, genes within the same co-expression module as genes whose transcriptional regulation is well defined are likely to be regulated in a similar manner. As *Per* and *Cry* expression is driven through the binding of CLOCK-ARNTL to E-box sites, other genes in these co-expression modules would also be likely to possess E-box sites in their promoters.

There are situations, however, where the JTK_CYCLE or other sinusoidal fitting algorithms may be preferred. First, if the study is underpowered for co-expression analysis. There should be over 15 samples with at least 3 biological replicates per timepoint (Langfelder and Horvath, 2008). For instance, some co-expression modules identified in this study appeared to be driven by outliers, which would not have been illuminated with fewer than 3 biological replicates per timepoint. Secondly, if the study is subject to batch effects (as in this study), searching for sinusoidal fluctuating genes, instead of significantly

fluctuating genes part of specific co-expression modules (which may be changing purely to batch), will yield fewer false positives.

3.4.2 Transcriptional regulation of asymmetrical twin peaks

As discussed, this study identified 4,569 cycling genes as well as a novel transcriptional profile consisting of 766 genes which peaked twice during the subjective night. Although this twin-peaking transcriptional profile had not been previously characterised, genes with harmonics of circadian expression have (Hughes et al., 2009). As the endogenous molecular clock consists of protein TFs which fluctuate with a period of 24 hours, these harmonics of circadian expression are unlikely to be generated through binding of a single circadian TF. Instead, these harmonics may arise when a pair of rhythmic TFs both act on a promoter of the same gene but possess differing phases (Figure 3.17) (Westermark and Herzel, 2013). This mechanism has been experimentally validated: insertion of a RRE into an E-box promoter altered the circadian transcriptional pattern from 24-hour rhythms into 12-hour harmonics (Ukai-Tadenuma et al., 2011). As the circadian TFs which bind to these elements possessed differing phases, the reporter construct showed greatest expression every 12 hours at those times when there was sufficient binding from both TFs.

This *cis*-regulatory mechanism of generating 12-hour harmonics is applicable if both circadian TFs, which bind to the promoter motifs of the downstream gene, are either both activators or both repressors. However, harmonics of circadian expression can also be generated from an activator-repressor TF combination - but only if they possess similar phases in expression. In this situation, the transcriptional response of the regulated gene

is lowest when the repressor peaks in expression (maximal repression) and when the activator dips in expression (minimal activation) (Westermark and Herzel, 2013).

Circadian TFs exactly 12 hours out of phase with each other will induce transcriptional peaks of similar heights (Figure 3.17A). In practice, biological variability may prevent circadian TFs cycling in antiphase which results in 12-hour rhythms which display differing peak heights (Figure 3.17B). Previous studies failed to identify alternative expression patterns such as the described twin-peaking transcriptional profile, and therefore have not attempted to understand the transcriptional dynamics required to produce them. The twin-peaking expression is similar to 12-hour harmonics except instead of expression peaking every 12 hours, peaks appear to be asymmetric, with a gap of 8 and 16 hours between each peak in expression. These asymmetric circadian harmonics may perhaps be generated if the circadian TFs differ in expression level or strength in activation/repression (Figure 3.17C). Altering TF expression level/ strength can affect the degree of symmetry: the larger the difference in TF expression levels/ strength, the greater the asymmetry, with transcriptional peaks occurring closer together (Figure 3.17C).

The twin peaking expression appears to be modelled by dual circadian activators with differing strengths (Figure 3.18A). Although dual repressors generate asymmetrical harmonics of expression, they generate "double dips" (two quick dips in expression) rather than "twin-peaks" (two quick peaks in expression) and therefore are unlikely to contribute to the described twin-peaking transcriptional profile (Figure 3.17C). Twin-peaking expression may also be modelled using an activator-repressor TF combination (Figure 3.18B).

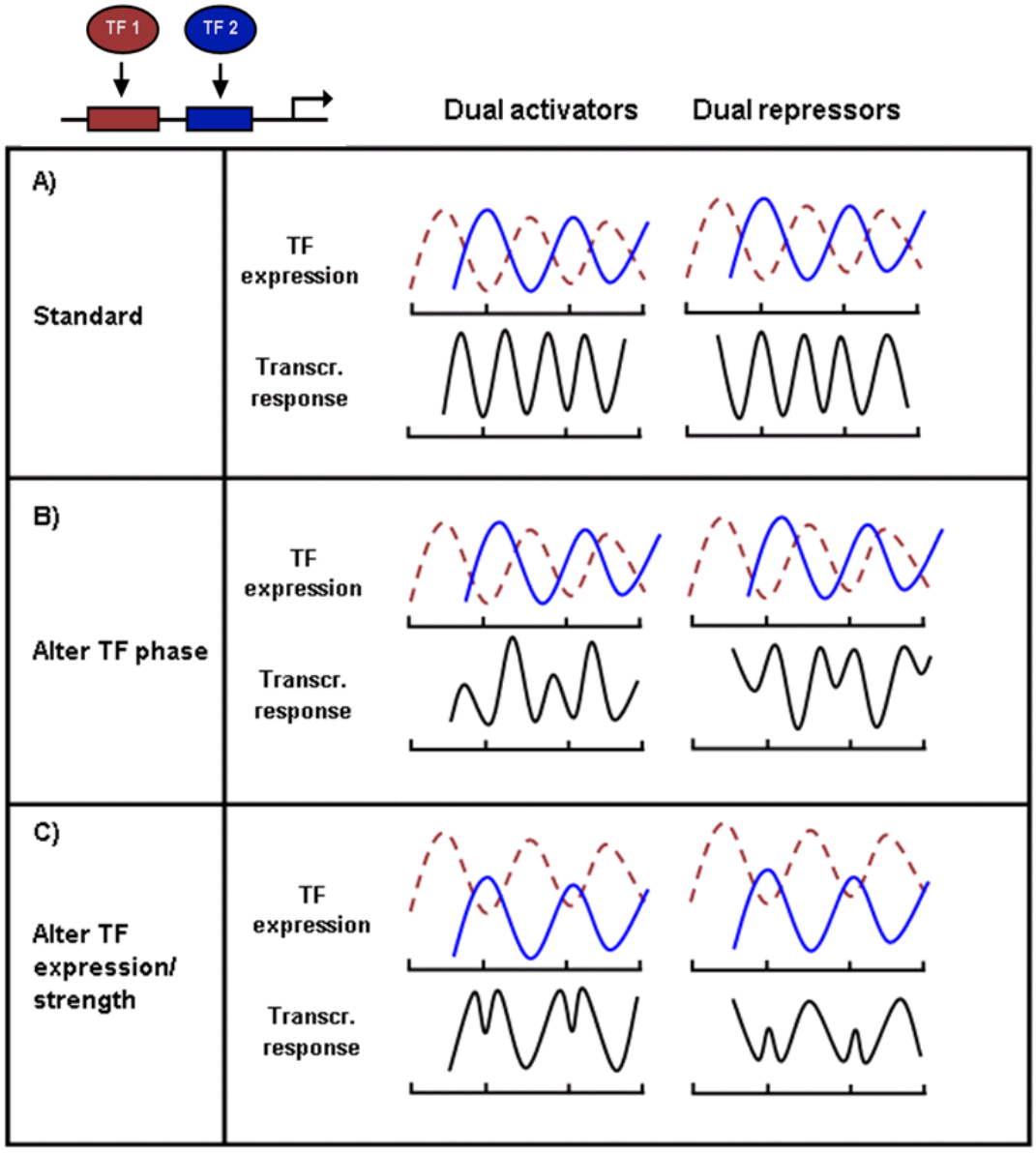


Figure 3.17. Transcriptional regulation of circadian harmonics.
Transcriptional response from a gene (black) regulated by two circadian TFs (blue and brown). **(A)** TFs which cycle in antiphase can generate peaks of expression every 12 hours. **(B)** TFs which are not phase-shifted by 12 hours produce peaks of expression every 12 hours but with alternating heights. **(C)** If the expression level or strength of induction of each TF differs, the gene is able to produce asymmetrical transcriptional peaks. Dual activators produce twin-peaks, dual repressors produce double-dips.

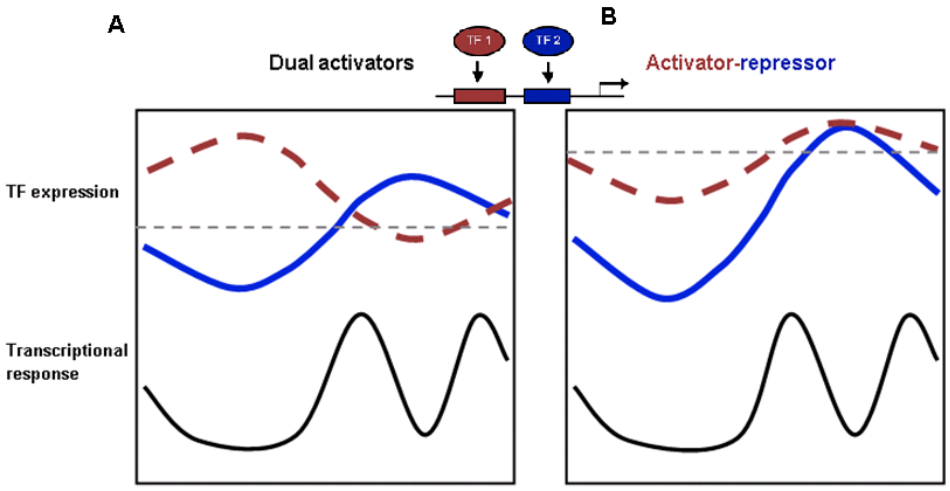


Figure 3.18. Models of twin-peaking expression through TF binding.
Transcriptional response from a gene (black) regulated by two circadian TFs (blue and brown). **(A)** Binding of two antiphase activating TFs. Gene is only expressed when both TFs are of a sufficient concentration (dashed grey line). **(B)** Binding of an activator and a repressor. The gene can only be expressed when the activating TF is above, and when the repressing TF is below, a threshold level (dashed grey line).

Interestingly, the generation of these harmonics of circadian expression may not be solely due to core clock mechanisms. There is evidence that feeding schedules may generate 12-hour rhythms because restricted feeding in mice abolished 12-hour harmonics (Vollmers et al., 2009, Hughes et al., 2009). Hughes *et al.* displayed that when food is only available during an 8-hour time window during subjective day when mice are normally asleep, the expression of these 12-hour harmonics dramatically changes with most gaining 24-hour rhythmicity. The expression of the TF NR2F2 is repressed by feeding (Perilhou et al., 2008) and in the mouse liver NR2F2 expression peaks at ZT6 and dips at ∼ZT18, the time where ingested food is most likely to enter the stomach and induce metabolic effects (Eckel-Mahan et al., 2012). NR2F2 may therefore either be regulated by the core clock or external feeding schedules. As NR2F2 binding sites were found to be strongly enriched in promoters of harmonic genes, feeding may play a key role in the generation of harmonic expression patterns (Westermark and Herzel, 2013). *Nr2f2* also shows a significant peak in expression at ZT6 in the LCM SCN expression data which suggests feeding schedule may also influence expression patterns in the SCN.

In summary, the transcriptional regulation of the circadian system is complex, driven by TF binding. TFs may bind to the promoters of a number of different genes at specific times, which interact to produce a variety of transcriptional profiles. A pair of circadian TF whose expression cycles in antiphase may bind to a promoter of a particular gene driving harmonics of circadian expression. Although asymmetrical harmonics of circadian expression have not been described previously, if circadian TFs possess differing expression levels or strength in inducing a transcriptional response, then their binding across the day may result in a transcriptional profile similar to the identified twin-peaking co-expression

module. These circadian TFs may either drive expression for the whole module or for only a small set of TFs which subsequently regulates the twin-peaking transcriptional module. It is also feasible that these TFs may not be entirely clock regulated. Instead, these TFs responsible for inducing these asymmetrical harmonics of expression may be temporally regulated by external factors such as light or feeding schedule (Vollmers et al., 2009, Hughes et al., 2009). As the twin-peaking transcriptional pattern was not previously described in the SCN when mice were subject to constant darkness, it may be regulated by light. However, there are several other reasons why the twin-peaking pattern was not described previously, such as the anatomical specificity of SCN dissection or the use of statistical algorithms applied to identify fluctuating genes (Panda et al., 2002). Additional mechanisms contributing to generate these patterns cannot be discounted. Instead of two circadian TFs, a gene may be driven by the combination of a TF and a co-regulator, where alternative phases of these components may induce circadian harmonics in gene expression. Alternatively, post-transcriptional mechanisms such as miRNA mediated decay may be important in patterning the twin-peaking expression (Xue et al., 2014).

3.4.3 Twin-peaking expression — role in clock resetting pathways

The twin-peaking co-expression module not only possessed a novel circadian transcriptional profile, it also appeared to be biologically important. Functional annotations related to synaptic transmission, calcium signaling and gated channel activity were greatly and significantly enriched among these twin-peaking genes ($q < 0.05$; Figure 3.14), suggesting a circadian component to establishing the electrophysiological environment of the SCN.

Interestingly, the peaking of expression at ZT14 and ZT22 coincides with phase-shifting "susceptibility windows"; time periods in which the photic input permits phase-shifting (or 'resetting') of the clock (Ding et al., 1994 , Iyer et al., 2014). This is demonstrated by the phase response curve (PRC): a light impulse at ZT14 will initiate a phase-delay of the clock whereas a light impulse at ZT22 will initiate a phase-advance (Figure 3.19) (Golombek and Rosenstein, 2010 , Ding et al., 1994). At all other timepoints assessed under a LD cycle the twin-peaking genes exhibited low expression levels exactly when the clock is resistant to phase-shifts. This suggests a possible transcriptional gating mechanism in which a threshold level of gene expression needs to be exceeded for light to activate this phase-shifting pathway (Ding et al., 1994 , Gillette and Mitchell, 2002 , Iyer et al., 2014). Transcriptional profiling at a higher temporal resolution would be required to confirm the elevated expression of twin-peaking genes above this critical threshold at timepoints when light-induced phase-shifts are possible (for example, ZT12; Figure 3.19).

To confirm that the twin-peaking transcriptional profile has a potential role in the light resetting pathway and represents a transcriptional gating mechanism shaping the times when light can induce a phase-shift, the genes previously implicated in the light resetting pathway were identified and their transcriptional profile in the SCN assessed. Interestingly, 7/9 of the genes central to this resetting pathway displayed a twin-peaking expression pattern (Figure 3.20), and their pharmacological inhibition (for example, for *Grin2a/b, Camk2a, Nos1, Ryr2*), block light-induced phase-shifts (Ebling et al., 1991 , Yokota et al., 2001 , Ding et al., 1998 , Ding et al., 1994 , Brown and Piggins, 2007 , Golombek and Rosenstein, 2010).

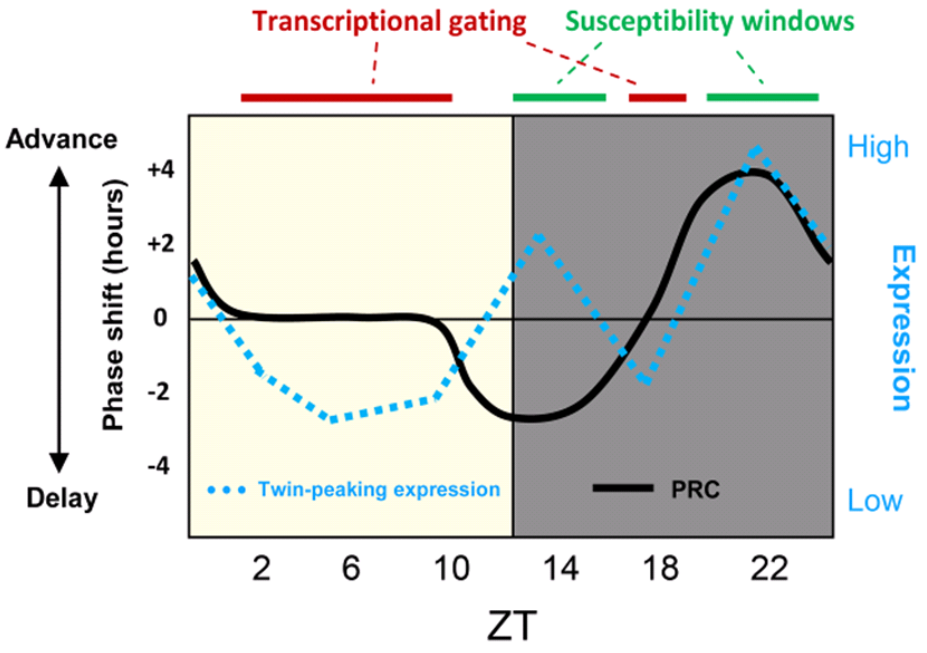


Figure 3.19. Twin-peaking expression coincides with phase-shifting "susceptibility windows". Illustration of the theoretical phase response curve (PRC) in mouse, which demonstrates the phase-shifting potential when a light pulse is applied at different times (black line) (Golombek and Rosenstein, 2010). Our module eigengene for the twin-peaking module (blue dotted line) indicates that when its expression lies above the threshold level (black horizontal line), the phase of the organism is able to shift. The greater the expression over this threshold, the larger the phase-shift. Conversely, when expression is low and below this threshold, the SCN is resistant to phase-shifting induced by light pulses. This suggests the presence of susceptibility windows (shown by green bars) when light impulses are able to shift the phase of the SCN.

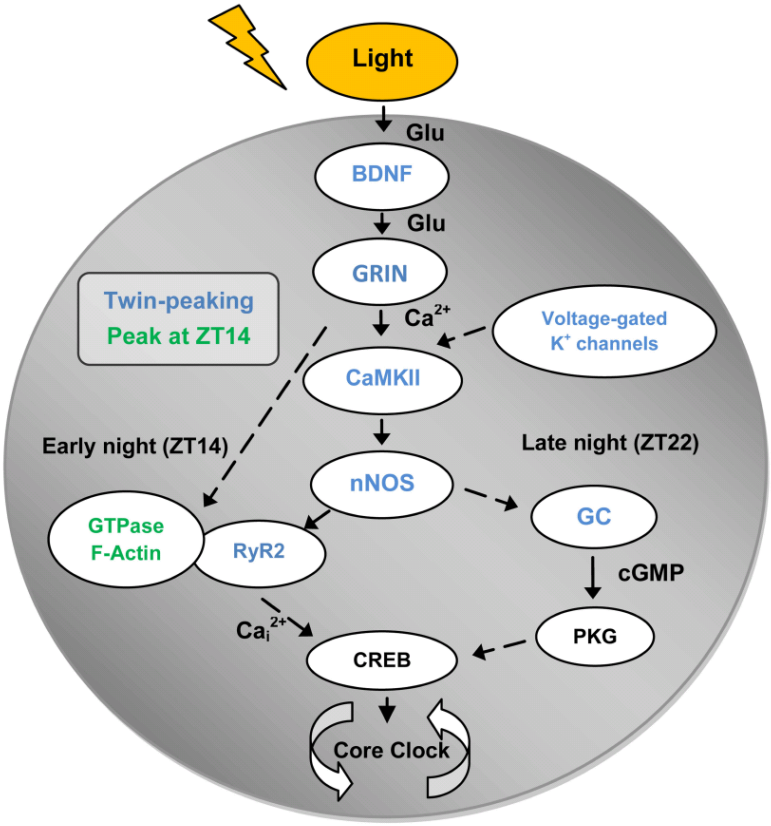


Figure 3.20. Circadian gating of light signaling pathway.
Twin-peaking gene-encoded proteins identified by LCM SCN RNA-seq are indicated in blue. Photic information is transmitted to the SCN via glutamatergic signaling through the retinohypothalamic tract (RHT). i) Brain-derived neurotrophic factor (BDNF) gates glutamatergic synaptic transmission by regulating either presynaptic release or membrane channel activity. ii) Voltage-gated potassium channels hyperpolarise the SCN neurons allowing photic input from the RHT to elicit a significant response. iii) Glutamate binding to GRIN induces calcium influx inducing CaMKII auto-phosphorylation. iv) Active pCaMKII phophorylates nNOS, which generates NO, producing differential effects according to time of day. v) In early night (ZT14), genes relating to small GTPase activation and actin-based cytoskeletal remodelling peak in expression. At this timepoint, F-actin levels are at their highest and an influx of Ca^{2+} depolymerises the F-actin affecting components of the MAP kinase pathway. Ca^{2+} also induces intracellular Ca^{2+} release though the opening of the endoplasmic reticulum calcium channel RyR2. vii) In late night (ZT22), NO activates GC, which increases cGMP levels activating PKG. viii) Phosphorylation of CREB leads to the transcription of PER1, which contributes to the shifting of the circadian phase. Image adapted from (Iyer et al., 2014).

Although the gating of the light-induced resetting pathway in the SCN has been known for years, the exact mechanism underpinnings remained unclear (Iyer et al., 2014). The identification of this twin-peaking transcriptional profile suggests that a transcriptional gating mechanism may contribute to the SCN's differential response to light depending on the time of day. Only at the times when the expression of these twin-peaking genes in the light signaling pathway is above a threshold level will light-induce a phase-shift. This transcriptional profile may also implicate additional genes or mechanisms which have a role in this pathway.

3.4.4 Light-induced translation

To further investigate the regulation of this twin-peaking module and the gating of the light-resetting pathway, it is important to consider the protein levels for these genes across the day. RNA is often not the functional state of the cell and requires translation before it can induce its effect. To assess the circadian proteome of the SCN and determine whether the corresponding proteins of the twin-peaking genes displayed similar expression levels, published proteomic data was interrogated.

Although RNA-seq can survey the expression of up to 35,303 protein and non-coding annotated genes (most recent Ensembl update), it is technically difficult to gain a high coverage of the proteome (currently only up to 5,000 proteins can be identified). This is largely due to the difficulty in protein extraction, for instance, many extraction protocols fail to capture transmembrane proteins (Nilsson et al., 2010). In addition, the detection threshold of mass spectrometry may prevent the detection of low abundance or low-molecular-weight proteins (Nilsson et al., 2010).

Proteomic profiling of the mouse SCN determined that 421 proteins (of the 2,112 detected) displayed significant fluctuation over the day in constant darkness (Chiang et al., 2014). Many of these fluctuating proteins (373/421) did not show sinusoidal expression but instead displayed peaks in expression every 8 or 12 hours across the circadian day with 321 (73%) showing abrupt fluctuation at the light transitions. However, hierarchical clustering identified a module which displayed peaks in expression at CT14 and CT22. Of the 2,112 proteins identified, 73 corresponded to those encoded by genes in the SCN LCM twin-peaking module. Of these, only 14 displayed significant fluctuation (Chiang et al., 2014). Inspecting the expression pattern of these 14 proteins, only three (SV2B, MYO5A, HPCAL4) approximated the twin-peaking expression profile. The lack of concordance between the RNA and protein expression patterns suggests that the clock resetting pathway may be gated at the transcriptional, but not translational level. This proteomic study did not, however, capture the protein expression levels for any of the top 10 "hub" twin-peaking transcription factors such as EGR1 or TCF4 (Table 3.3), suggesting that these TFs may still drive the expression of this twin-peaking module. Furthermore, this proteomic study was performed in constant darkness and it is possible that the twin-peaking transcriptional or translational patterns may be specific to LD conditions as in the LCM SCN study. In addition, some translation has been shown to be synapse specific, therefore, these proteomic studies are unable to capture these synapse specific translational profiles, but instead assess the protein levels within the whole cell which may not be representative of its functional properties (Schuman et al., 2006). It would therefore be of interest to assess the synapse specific circadian profile for these proteins (Bai and Witzmann, 2007).

Ensembl ID	Module membership	Gene ID	Gene name
ENSMUSG00000053477	0.927	*Tcf4*	transcription factor 4
ENSMUSG00000026872	0.912	*Zeb2*	zinc finger E-box binding homeobox 2
ENSMUSG00000005583	0.909	*Mef2c*	myocyte enhancer factor 2C
ENSMUSG00000037984	0.903	*Neurod6*	neurogenic differentiation 6
ENSMUSG00000032402	0.888	*Smad3*	SMAD family member 3
ENSMUSG00000051177	0.885	*Plcb1*	phospholipase C, beta 1
ENSMUSG00000041324	0.877	*Inhba*	inhibin beta-A
ENSMUSG00000038418	0.875	*Egr1*	early growth response 1
ENSMUSG00000010175	0.871	*Prox1*	prospero-related homeobox 1
ENSMUSG00000049107	0.870	*Ntf3*	neurotrophin 3

Table 3.3. Top 10 hub transcription factors of the twin-peaking module
The top 10 TFs in relationship to module membership to the twin peaking (liststeelblue1) coexpression module. Genes were labeled as TFs if they possessed any of the following GO terms: GO:0010628 (positive regulation of gene expression), GO:0045893 (positive regulation of transcription, DNA-templated), or GO:0045944 (positive regulation of transcription from RNA polymerase II promoter).

If twin-peaking expression is gated at the transcriptional, but not translational level, mRNA must be translated when the SCN is exposed to light. This is because there is often a lag between the peak in RNA and protein levels, for example, over 40% of proteins in liver display a lag of >6 hours relative to their transcript levels. In order for RNA levels to gate the SCN's response to light, there must be a mechanism which stimulates the immediate translation of RNA. Interestingly, light pulses at early and late subjective night (CT15 and CT22) have been reported to induce a 3-fold increase in the phosphorylation of the translation factor eIF4E (eukaryotic translation initiation factor 4E) in the SCN, which promotes the translation of *Per1/2* mRNA (Cao et al., 2015). Conversely, a light pulse in the subjective day (CT6) had no effect on eIF4E phosphorylation. As light only stimulated eIF4E phosphorylation at the times the SCN can be entrained, it suggests that eIF4E phosphorylation has a role in photic entrainment. Indeed, mouse mutants which possessed non-phosphorylatable eIF4E displayed a reduced ability to phase-shift to light impulses (Cao et al., 2015). The proposed transcriptional gating mechanism for these genes is a valid hypothesis given the light-regulated translational control in the SCN which acts in a time-dependent manner. There may also be other mechanisms which allow mRNA to be translated when a light pulse is applied: perhaps the mRNA is sequestered in the nucleus and is released into the cytoplasm to be translated on the receipt of a light pulse.

Taken together, our data indicate that oscillations in the SCN transcriptome play an important role in gating the differential responses to light, a fundamental circadian process. As there is no evidence to date for twin peaking protein expression in the SCN,

RNA from the twin-peaking genes may only be translated on the receipt of a light pulse. The translated protein can then contribute to the resetting of the clock.

3.4.5 Integration of sinusoidal and twin peaking genes to form differential responses

4,569 genes were identified which displayed a sinusoidal oscillatory pattern using methods concordant with previous studies (Figure 3.9) (Zhang et al., 2014, Hughes et al., 2010). Our collection of SCN cycling genes was enriched for GO terms relating to key processes such as translation and energy production (Figure 3.9C). As determined by microarray technology, genes which cycled in the SCN under constant darkness also showed enrichment in similar processes (Panda et al., 2002). Mitochondrial genes may also be under circadian regulation in the SCN, as they are responsible for setting the redox state, which is central to all circadian time keeping mechanisms from algae to humans (O'Neill and Reddy, 2011, Loudon, 2012). Robust oscillations of redox state are observed in the mouse SCN, but abolished in the *Bmal1*$^{-/-}$ mouse mutant, suggesting redox state is under clock control (Wang et al., 2012). Translation is energetically expensive and therefore likely to be allocated to certain circadian windows when energy supply is higher (Jouffe et al., 2013). Interestingly, terms relating to translation appear to peak at ∼ZT6, at a similar time to mitochondrial related genes, which supports this proposition.

The 4,569 cycling genes also display novel enrichments in GO terms relating to RNA splicing and ribonucleoprotein (RNP) complexes (Figure 3.9C). As *de novo* transcription has been shown to drive rhythmicity of approximately 20-30% of all cycling genes (Koike et al., 2012, Menet et al., 2012), there is a great importance of post-transcriptional regulation

in shaping the circadian transcriptome. These novel GO enrichments suggest that the considerable post-transcriptional regulation establishing oscillatory expression patterns may be mediated by RNP complexes such as splicing or control of mRNA polyadenylation (Wang et al., 2013). It would therefore be interesting to investigate these processes in greater detail over circadian time. However, other post-transcriptional mechanisms, such as miRNA level fluctuation, which were not assessed in this experiment, could also contribute to initiating these patterns. For example, in *Neurospora*, the transcription of *qrf* represses the clock gene *frequency (frq)* through antisense interference. *qrf* expression is required for circadian rhythmicity but is also regulated by light and regulates clock resetting pathways (Xue et al., 2014).

As described above, the twin-peaking expression pattern for certain genes may represent a transcriptional gating mechanism in which a threshold level of expression must be exceeded to allow light to induce a phase-shift. However, there appears to be a bifurcation in the SCN's response to nocturnal light (Figure 3.20). In early night (ZT14), Ca^{2+}-induced-Ca^{2+}release from ryanodine receptors (RyRs) is proposed to induce these shifts; inhibition of RyRs blocks light-induced phase-shifts at early night but not at late night (Ding et al., 1998). In late night (ZT22), NO activates GC, which increases cGMP levels thereby activating PKG. This is thought to affect phase-advances only; inhibition of PKG at late night prevents phase-advances but does not prevent phase-delays in early night (Weber et al., 1995, Mathur et al., 1996, Ferreyra and Golombek, 2001). Furthermore, protein kinase A (PKA) and pituitary adenylate cyclase-activating polypeptide (PACAP) have been shown to suppress light-induced phase-advances at late night but amplify phase-delays in early night (Gillette and Mitchell, 2002). These results suggest a

divergence in mechanisms inducing phase-shifts after NO production, depending on the time of night. Surprisingly, *RyR2* and *GC* possess twin-peaking expression suggesting they may play roles in both early and late night (Figure 3.20). Therefore, there must be additional genes or processes which are selectively expressed to induce a diverging signal at either early or late night.

To identify processes which may induce a differential response to light at ZT14 and ZT22, GO enrichment was applied for genes with peak expression phases at either ZT14 or ZT22. Genes whose expression peaked at ZT14 were enriched for terms relating to actin cytoskeleton reorganization and to small GTPase signaling (Figure 3.20). These enrichments support the hypothesis that actin depolymerisation is important for phase-delays but not phase-advances (Gerber et al., 2013, Cabej, 2013). At this timepoint (ZT14), F-actin levels are at their highest and an influx of Ca^{2+} depolymerises the F-actin into G-actin which affects components of the MAP kinase pathway (Arnold, 2013). Stabilisation of F-actin with jasplakinolide prevents light-induced phase-shifts in early night, whereas exposure to Lat A, which destabilizes F-actin, can induce a phase-delay. The circadian expression of genes relating to the actin cytoskeleton or Rho GTPases, which play a role in actin assembly, may therefore contribute to the increase in F-actin at early night, which contributes to light-induced phase-delays but not phase-advances. This shows that a variety of transcriptional profiles can converge on a pathway to induce specific effects.

In summary, 4,569 genes were identified which displayed significant sinusoidal expression. These genes were enriched for processes previously implicated in SCN circadian biology, such as mitochondrial and ribosomal related processes which peaked in the sub-

jective day. Furthermore, cycling genes were enriched for other processes such as RNP related posttranscriptional regulation, such as splicing, which may generate oscillatory expression patterns from transcripts which do not oscillate through *de novo* transcription. In addition, genes responsible for actin assembly, a process essential in inducing phase-delays in the early night were found to peak in expression at ZT14.

3.4.6 Circadian oscillation of transcription, redox state and neuronal excitability

The generation of the circadian transcriptome of the SCN using RNA-seq also allows us to re-interpret related hypothesis in the circadian field. For example, although the core transcriptional translational feedback loop is thought to drive circadian rhythmicity, it also works in synchrony with other non-transcriptional oscillators (Figure 3.21) (Gillette and Wang, 2014). It is therefore of interest to determine whether this novel transcriptional data from the SCN may provide additional insight into the non-transcriptional oscillators.

SCN neurons oscillate according to redox state, which can be viewed as the availability of free electrons. If the neuron is in a reduced state, it has an excess of free electrons, but if in an oxidized state, it has a deficit. Measuring glutathiolation (the ability for proteins to incorporate the reducing agent glutathione) and the ratio of the oxidized amino acid dehydroascorbic acid (DHA) versus its reduced counterpart ascorbic acid (AA) can determine redox state in the SCN (Wang et al., 2012). From these measurements, the authors concluded that the SCN is oxidized during the day, and reduced at early night (Figure 3.22) (Wang et al., 2012). However, these conclusions appear to be misguided in light of the discovery of the twin-peaking transcriptional profile. The DHA/AA ratio

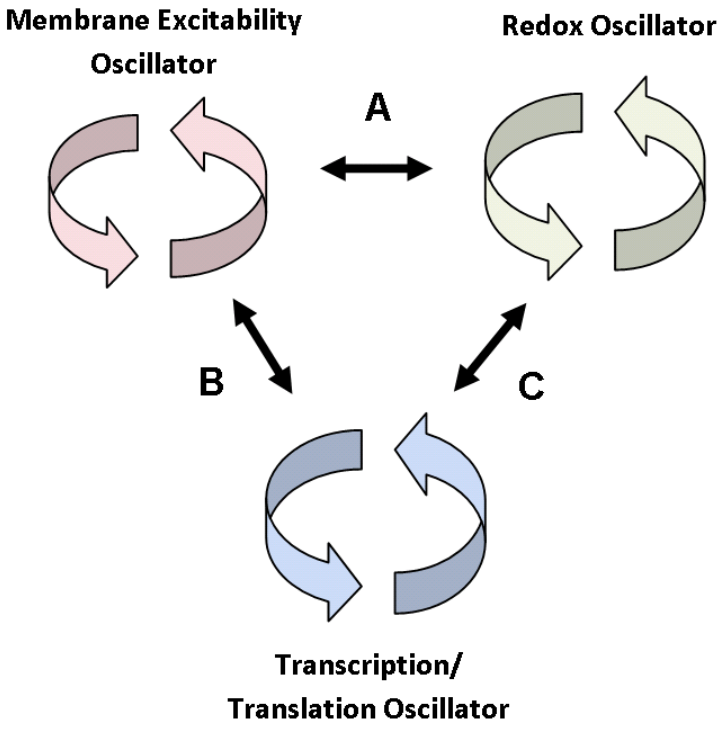


Figure 3.21. The coupling between circadian oscillators.
The three main circadian oscillators are tightly coupled **(A)** Redox state can modulate K^+ conductance influencing membrane potential. Neuronal activity can increase bloodflow and energy availability affecting metabolism important in setting redox state (Wang et al., 2012, Rutter et al., 2002). **(B)** Gating of membrane excitability affects light input to the SCN affecting core clock expression (Gillette and Mitchell, 2002). Ion channels which affect membrane excitability are under the control of the core clock (Colwell, 2011). **(C)** A functional core clock is essential for oscillations of redox state (Wang et al., 2012). Redox state can also modulate clock gene expression (Dioum et al., 2002, Rutter et al., 2001). Figure reproduced from (Gillette and Wang, 2014).

appears to display peaks in both early and late night, with dips in the subjective day and at CT18 (Figure 3.22B). Measurements of glutathiolation show a similar pattern although measurements were not taken at late night (Figure 3.22A). It is likely that the previous results of redox state across the circadian day have been misinterpreted and instead the redox state varies in a similar manner to the described twin-peaking expression module. Measurements of the resting membrane potential (Vm) also display peaks at ZT6 and ZT18 which appears to mirror that of redox state (Figure 3.22C) (Wang et al., 2012). In their paper, the authors again appear to misinterpret these findings: results are inappropriately binned into 5 groups and therefore overlook the variation in Vm observed across the subjective night. Nevertheless, the raw data reveals that the oxidised state at early and late night align with the hyperpolarized Vm and the reduced state at daytime/midnight aligns with the depolarized Vm (Figure 3.22).

Redox state can affect neuronal excitability; acute exposure to oxidizing and reducing agents changes Vm. It may accomplish this change in the membrane through post-translational modification of K^+ channels (Wang et al., 2012). Therefore, during early and late night, neurons become more oxidized, enhancing K^+ conductance which inactivates the neuron.

The coupling of these processes suggests that the light-induced phase-shifting pathway is gated at multiple levels. Membrane potential, potentially controlled by redox state, is gated only allowing significant response from light at particular times. Additionally, the light signaling pathway is gated at the transcriptional level. Instead of the twin peaking transcriptional profile being regulated by two core clock TFs, it may be under the control of the fluctuating redox state. For example, the TF EGR1 is "redox sensitive" with ROS

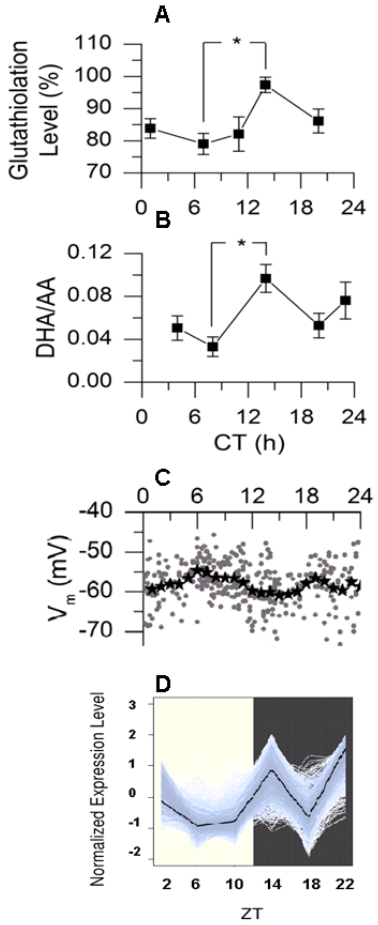


Figure 3.22. Redox states, membrane potential and expression of 'twin-peaking' genes in the SCN.
(A-B) Two measures of redox state. SCN appears to be more oxidized in early and late night (ZT14 + ZT22) but reduced at other timepoints (Wang et al., 2012). **(C)** The membrane potential (Vm) appears to be higher at midday and midnight (depolarized state) but lower at early and late night (hyperpolarized). The hyperpolarisation of SCN neurons allows photic input from the RHT to elicit a significant response (Wang et al., 2012). **(D)** The twin-peaking coexpression module as described previously. Redox state, membrane potential and twin-peaking expression appear to be closely linked, exhibiting deviations at time periods susceptible to light-induced phase-shifts. Panels A-C obtained from (Wang et al., 2012) with permission from Science Magazine.

rapidly inducing *Egr1* gene and protein expression (Pagel and Deindl, 2012). Redox state failed to oscillate in the SCN of in *Bmal*$^{-/-}$ mice and in embryonic fibroblasts derived from *Cry1 Cry2* double-knockout mice (O'Neill and Reddy, 2011, Wang et al., 2012). This suggests that transcriptional oscillations are tightly coupled and may be essential in setting these redox rhythms which help drive circadian rhythms.

In summary, the transcriptional profiles of the twin peaking genes, the redox state and the membrane potential (neuronal excitability) appear to be closely interlinked in the SCN. It is therefore difficult to determine cause and effect between these oscillators. One potential mechanism could be that the core clock may control the redox state in the SCN, through circadian control of mitochondria and antioxidants. The redox sensitive EGR1 may then drive the twin-peaking transcriptional profile important for gating the light-induced phase-shifting pathway. The redox state also affects the activity of K^+ channels and thus controls membrane excitability. These mechanisms work together to create susceptibility windows which allows photic input to reset the clock. However, this is only one of many potential mechanisms which could explain how these circadian oscillators interact to generate robust circadian rhythms.

3.4.7 Importance of web-based application

The investigation of the mouse SCN transcriptome using RNA-seq has yielded an abundance of information (Chapters 2-4) which may be of use to researchers based in the fields of neuroscience and circadian biology. Although the main biological conclusions from my analyses, such as the enrichment of biological processes to particular transcriptional patterns, can be readily interpreted, much of the richness in RNA-seq data is

often lost to researchers with limited experience in working with large datasets. This creates a boundary between scientific disciplines and may prevent the generation of new ideas or hypothesis. The aim of this project was also to create a web-app interface which encourages researchers to explore different aspects of the RNA-seq dataset and to allow easy visualization of gene expression patterns. The web-based application (http://www.wgpembroke.com/shiny/SCNseq/) is now live (from Nov. 2015).

For instance, if a circadian biologist is interested in identifying genes which play key roles in the circadian system, they could filter for genes which are SCN-enriched and significantly fluctuate (amongst many other statistics). The expression of *Creb3l1* significantly fluctuates peaking at ZT8, is SCN-enriched and displays significant fluctuation in 7 (of 12) other tissues surveyed (Zhang et al., 2014). It would therefore be of interest to circadian screen a mouse knockout for this gene. When sorting these SCN-enriched fluctuating genes by enrichment score (the degree to which their expression is enriched in the SCN relative to the rest of the brain), there are many genes which have known roles in circadian rhythms (e.g. *Nms, Npy6r, Prok2, Vip*). Identifying those whose function is unknown may help to reveal unknown aspects of circadian biology. If a researcher is interested in identifying the roles of lincRNA in brain function, they may filter for those highly conserved lincRNA which also display a twin-peaking expression pattern, as this pattern is associated with synaptic function. If a neuroscientist is investigating the role of a particular gene implicated in a neurological or psychiatric disease, visualization of its expression pattern in the SCN may determine whether it plays a role in the circadian system. This may help form new hypotheses on the link between circadian disruption and neurological disease.

4 Identification of novel transcripts in the wild-type mouse SCN

4.1 Introduction

One advantage of using RNA-sequencing over array technology is that the data are not limited to annotated genomic regions. Therefore, it is also possible to identify unannotated genomic regions which display robust expression. Although identification of these expressed regions is more important in species which do not have a reference genome assembly, it can also be used to improve existing genome annotations and reveal new potentially important genes or isoforms.

If sequencing reads map to both an unannotated region of the genome and to a region encoding an exon of a known gene, then the unannotated region likely encodes a 'novel' exon from the same known gene. This novel exon may extend the pre-existing mRNA annotation by representing additional 3' or 5' UTRs. Alternatively, the novel exon may represent an alternative isoform of that gene which is generated through alternate splicing. If reads map to an unannotated region of the genome but do not also map to a region of the genome encoding a known gene, this region may represent a novel unannotated gene.

The ENCODE project has shown that up to 75% of the mammalian genome is capable of being transcribed with under 2% encoding protein coding transcripts (Alexander et al., 2010, Djebali et al., 2012). LncRNA (> 200nt) resemble mRNA in terms of transcription and processing; the majority of lncRNA are transcribed by RNA-pol II and capped (Kung et al., 2013). Furthermore, lncRNA can be both spliced and polyadenylated, although to a

lesser degree than mRNA (Dinger et al., 2008). The roles of lncRNA are proposed to vary substantially and to act through a variety of mechanisms, although most are proposed to regulate mRNA levels (Morris and Mattick, 2014, Roberts et al., 2014, Kung et al., 2013). In most genomic studies, long ncRNA are often subject to investigation because RNA-seq only considers fragments >200nt due to the size selection step during library preparation. Small RNA sequencing is required to study short ncRNA such as miRNA (Szeto et al., 2014). Intergenic lncRNA models are also subject to increased investigation in genomic studies because lncRNA models which overlap or occur within the introns of protein coding genes on the same strand may often represent incomplete splicing of protein coding mRNA (Sowalsky et al., 2015). Furthermore, it is often difficult to assess certain features for non-intergenic lncRNA such as sequence conservation, as selection may be acting on either the lncRNA or the protein coding gene; reads which derive from the lncRNA but also overlap sites coding exons will be subject to increased conservation even if this is unrelated to the lncRNA. Additionally, many genomic studies focus on only multiexonic lincRNA. This is because a larger proportion of monoexonic transcripts are likely to represent transcriptional noise or mapping artefacts, despite the existence of known monoexonic lincRNA such as *Malat1* or *Dali* (Tripathi et al., 2010, Chalei et al., 2014).

Many lncRNA are suggested to possess roles in differentiation and development, some of which important in the central nervous system (CNS). Knock out of *linc—Brn1b* in mice can induce a decrease in proliferation of cortical progenitors in vivo and prevent proper generation of cortical neurons (Sauvageau et al., 2013). However, knock-outs of many highly conserved transcripts such as *Neat1* do not show any detrimental effect on

the organism (Nakagawa et al., 2011, Ahituv et al., 2007). Given that most mammalian lncRNA are expressed in the brain, more complex and cognitive specific behavioural screens may be necessary to delineate their function (Mercer et al., 2008). For example, mice with the lncRNA *brain cytoplasmic RNA 1 (BC1)* knocked-out showed no overt anatomical phenotype but displayed changes in behaviour which would be lethal in the wild (Lewejohann et al., 2004). Although function is known for a number of lncRNA, very often the phenotype can be subtle and would not be captured unless subject to specific behavioural paradigms.

Some lncRNA have been shown to be important in the regulation of sleep and circadian rhythms in a number of organisms. The lincRNA HULC (Highly Upregulated in Liver Cancer) was found to affect the circadian clock in mouse liver (Cui et al., 2015). Knock-down of HULC siRNA reduced CLOCK mRNA and protein levels whereas HULC upregulation was induced PER and CRY expression which is expected if it were to also upregulate CLOCK. In *Neurospora*, the transcription of the antisense ncRNA *qrf* represses the clock gene *frequency (frq)* through antisense interference. *qrf* expression is required for circadian rhythmicity but is also regulated by light and regulates clock resetting pathways (Xue et al., 2014). Drosophila with knock-out of the lncRNA *yar* display no overt morphological phenotypes, yet display fragmented and reduced sleep (Soshnev et al., 2011). Despite the identification of some lncRNA with roles in circadian biology, this number is limited, especially in mammalian species. In addition, there has been no reported lncRNA with importance in the mammalian SCN, therefore it would be of interest to identify these transcripts or prioritise known lncRNA based upon their expression in the SCN to determine whether they are likely to play a role in the circadian system.

There has been much effort to identify and characterise lincRNA in a variety of tissues and conditions using RNA-seq, including identification of lincRNA expressed in the CNS. Belgard *et al.* identified 1,055 lincRNA loci, of which 67% were novel, including 66 whose expression were patterned in the adult mouse cortex across cortical layers (Belgard et al., 2011). The authors found that protein-coding genes patterned in their expression across the cortex were 39% more likely to be proximal to a lincRNA locus than by chance ($p < 10^{-4}$). Furthermore, patterned lincRNA loci were 63% more likely to be proximal to patterned protein-coding genes than expected by chance ($p < 0.01$) suggesting that patterned lincRNA may regulate the patterning of protein coding gene expression across the cortex.

Ramos *et al.* used sequencing data from three brain regions at different stages in the neuronal lineage (the subventricular zone, olfactory bulb, and dentate gyrus) as well as from embryonic stem cells (ESCs) to identify 5,731 lincRNA loci potentially involved neural development (Ramos et al., 2013). Assessing the chromatin state of these lincRNA loci through neural development helped generate hypothesis regarding their function. Genes which are "bivalent" for both H3K4me3 and H3K27me3 are thought to be transcriptionally silent but "poised" for either transcriptional activation or repression. The genes which were bivalent in ESCs, but active in neural stem cells and repressed in non-neural cell types were therefore likely to play roles in neural development. 302 protein coding genes matched these criteria and were strongly enriched for GO terms relating to brain development (Ramos et al., 2013). 100 lincRNA displayed similar patterns of chromatin modification suggesting they may also play a role in neural development.

Sequencing LCM SCN samples provides a unique opportunity to survey the unannotated portion of the transcriptome of this tissue. This may include SCN-specific isoforms of protein coding genes or SCN-specific lincRNA which, given their localisation in this region, may play key roles in circadian or SCN function.

LncRNA exhibit greater tissue-specific expression patterns than mRNA, suggesting that lncRNA may have a very precise role in the organism and may not be expressed at appreciable levels in most tissues (Derrien et al., 2012, Cabili et al., 2011). Furthermore, as lincRNA expression is much lower than protein coding genes they will often not be detected, unless the samples investigated are highly cell-type specific (which increases the lincRNA's relative abundance) or when using a high sequencing depth (Sims et al., 2014b, Tarazona et al., 2011). As there are no reported lncRNA which play specific roles in the SCN, it will be interesting to identify those novel transcripts in the SCN, which due to their regional specificity, may have key roles in SCN function. By assessing the circadian expression pattern of these lincRNA and their genomic proximity to other known genes (such as clock genes), we can prioritise those which may be important in circadian biology. For example, if a lincRNA displays a highly correlated or anti-correlated expression to a neighbouring clock gene, it may have a role in regulating its expression and therefore a knock-out of this lincRNA may induce specific circadian deficits.

Identification of novel isoforms of protein coding genes using the LCM SCN RNA-seq data may imply roles of these isoforms in circadian or SCN biology given their localisation in this specific brain region. In addition, assessing the circadian expression pattern and other properties of these isoforms may help elucidate their potential role in the SCN.

In this chapter, I sought to identify novel transcripts expressed in the mouse SCN which may play key roles in circadian or SCN function. The precise plans were to:

- Identify novel exons of previously identified genes which are expressed in the SCN through transcriptome assembly methods.
- Investigate properties of these novel isoforms to help determine whether they play key roles in circadian or SCN function.
- Identify multi-exonic lincRNA expressed in the SCN through transcriptome assembly methods.
- Investigate properties of these lincRNA to help generate hypothesis regarding their function.

4.2 Methods

4.2.1 Identification of novel exons

Reference guided *de novo* transcript assembly considering known transcript annotations was conducted using cufflinks v2.0.2 (Trapnell et al., 2010). Cufflinks is able to identify transcripts from the sequencing data without any prior knowledge. Novel exons were defined as expressed regions of the genome which did not overlap, but were within a 10kb window, of any feature in the most recent Ensembl or UCSC gene annotation sets.

Novel exons were further filtered to increase their confidence. Only those: i) with a reciprocal overlap of $< 25\%$ with any retroposed pseudogene (ucscRetroAli6) and, ii) with a reciprocal overlap of $< 50\%$ with any transposable element (TE) were kept. This was to filter out the 'novel exons' which may have been a by-product of multimapping, which is where an RNA read is mapped to numerous regions across the genome due to high sequence similarity, often observed amongst transposable elements or pseudogenes. Furthermore, only those: iii) with over 20 spliced reads in the novel exon, iv) with over 10 spliced reads per 100 bp of the novel exon, and v) with over 20 reads that splice into a known transcript were kept. This was to ensure that these reads were spliced into a known gene structure and were expressed at a robust level. Finally: those lincRNA vi) longer than 3,000bp were removed as this is well above the usual threshold accepted for an exon length and will remove any assembly artefacts (Sakharkar et al., 2005).

4.2.2 Identification of lincRNA

Identification of lincRNA was conducted through the use of the CGAT NGS pipelines rnaseqtranscripts.py and rnaseqlncrna.py (Sims et al., 2014a). The first pipeline identified transfrags using cufflinks in a process similar to that described above and retained those transfrags present in at least two samples.

The second pipeline predicted lncRNA by removing the assembled transfrags (transcribed fragments) which overlapped protein coding exons on the same strand. These lncRNA were assessed for coding potential using the coding potential calculator (CPC; (Kong et al., 2007)) and removed if annotated as 'coding' (CP score >1). CPC assesses the protein coding potential of these transcripts based on six criteria with an accuracy of over 95%. Some of these criteria assess the presence of open reading frames (ORFs): a true protein coding sequence is likely to have a long and high quality ORF whereas a lncRNA will not. CPC also scans the transcript for known protein coding sequences and assesses whether these sequences reside within a single frame to help determine coding potential. Transcripts with a CPC score below -1 are considered to be non-coding, and transcripts above 1 are considered to be coding. Transcripts scoring between -1 and 1 vary from weak-noncoding to weak-coding potential.

Only the intergenic (>2kb from any protein coding gene), multiexonic lncRNA with expression ($\geq$1 read) in each of at least 7 samples were chosen for further investigation as these are less likely to represent transcriptional noise or mapping artefacts. LincRNA loci which overlapped retroposed pseudogenes or TEs were flagged as these regions may possess reads as a by-product of multimapping. Further details of all lincRNA surveyed can be viewed on www.wgpembroke.com/shiny/SCNseq/.

4.2.3 Comparison with other lincRNA annotations

LincRNA coordinates were compared with: i) transcripts in the ensembl_78 annotation, ii) lincRNA identified in adult mouse subventricular zone neural stem cell lineage (Ramos et al., 2013), and iii) lincRNA identified in mouse neocortex (Belgard et al., 2011). A lincRNA was considered to be concordant with those previously identified if there was a reciprocal fraction overlap of at least 25% using intersectBed. That is $> 25\%$ of an exon from my lincRNA overlapped an exon from an annotated lincRNA and $> 25\%$ of the annotated lincRNA overlapped an exon from my lincRNA.

4.2.4 LincRNA sequence conservation

Sequence conservation of identified lincRNA was assessed using the mm10 phastCons conservation scores. The phastCons score represents the likelihood that a particular nucleotide is part of a conserved element and is based upon the sequence alignment against 59 different vertebrate genomes. The conservation score for the lincRNA loci was calculated by selecting the mean phastCons score for the exonic regions of the lincRNA.

Conservation scores of each lincRNA were compared to different proxies of neutral evolution: Ancestral repeats (ARs; transposable element sequences present in both human and mouse), and flanking sequence (FS) to the lincRNA loci. To predict the conservation of the lincRNA loci, the mean conservation score of the exonic regions of the lincRNA were compared to the mean conservation score of: i) all AR sequences within 50kb either side of the lincRNA loci, ii) 500b of sequence on either side flanking the lincRNA loci starting from 2kb, and iii) 500b of sequence on either side flanking the lincRNA loci by

a random distance (random sequences will be within a 1-5kb window both sides of the lincRNA loci).

A z-score statistic was calculated against each neutral sequence to show the degree of sequence conservation. LincRNA with an average z-score > 1.5 were called as being conserved.

Each identified lincRNA was tested to assess whether they were significantly more conserved than proxies of neutral evolution by using the Kolmogorov—Smirnov (KS) test. This is a non-parametric test which determines whether the samples come from the same distribution. The KS test was also employed for a variety of other lincRNA properties.

4.2.5 Identifying syntenic human lincRNA

Each lincRNA locus was assessed whether it was present in the same syntenic region of the human genome. Having downloaded the 1:1 mouse human orthologues from Ensembl Biomart, the closest 1:1 mouse:human orthologues flanking the identified mouse lincRNA locus were identified and their locations were determined in the human genome. These two regions on the human genome were not described as syntenic if they were: a) on different chromosomes or b) over 1MB away from each other. If the mouse lincRNA locus occurs within a syntenic region, it was assessed whether there were any human lincRNA loci present in this syntenic region which may be orthologous to the identified mouse lincRNA.

4.2.6 Expression and fluctuation

To determine whether these lincRNA fluctuate in expression, methods described in chapter 3 were employed. To measure the expression levels of the *Cry1* isoforms, the number of mapped reads for the genomic regions encoding the exons 1a and 1b were calculated at each timepoint and normalised using the size factors generated using DESeq2 (Chapter 5).

4.2.7 Expression correlation

The temporal expression of the lincRNA was assessed to determine whether it correlated with the expression of the genomically most proximal protein coding gene across time by performing a Pearson's correlation for all samples. The Pearson's correlation (r) sits between -1 (total negative correlation) and 1 (total positive correlation). A lincRNA whose correlation score lies between -0.5 and 0.5 was considered not to correlate in expression with the genomically most proximal protein coding gene.

4.2.8 Functional annotation of non-coding regions

GREAT (Genomic Regions Enrichment of Annotations Tool) is a computational tool which provides a method for testing the enrichment of annotations in non-coding regions (McLean et al., 2010). First, every annotated gene is assigned a regulatory domain: a proximal regulatory domain is defined as up to 5kb upstream and 1kb down stream of the TSS and a distal regulatory domain is up to 1,000 kb away. Then, each lincRNA locus is associated with those genes whose regulatory domain it overlaps. Statistics were generated using the hypergeometric test, as in other gene enrichment tools such as DAVID (Huang

da et al., 2009). The p-value equals the probability of selecting the number of observed test regions (or more) when randomly drawing genes from the genome (or background test set). The loci for all 3,187 identified SCN lincRNA was used as a background test set in the analysis. To calculate a False Discovery Rate (FDR) q value, the p values were sorted and ranked (most significant value is assigned 1, the least significant value is assigned N (the total number of annotations). Then each p value was multiplied by N and divided by its rank. This resulting q value now accounts for the false discovery rate.

4.2.9 Validation of transcripts

For RT-PCR of novel exons of *Cry1*, total RNA was purified from dissected SCN tissue and used to prepare cDNA (Revertaid, Fermentas). Primers representing previously annotated (5' GTGAGGAGGT TTTCTTGGAAG 3') and novel (5' CTTCTAGGGA ATTGCGACTG 3') exons were used with a common reverse primer (5' CTGGGAAATA CATCAGCTGG 3') to amplify products by RT-PCR followed by visualisation on an agarose gel and sequencing. The genomic coordinates of the novel exon are: chr10: 85,183,342 to 85,183,832 (mm10) with splicing into the following exon at 85,183,342 or 85,183,510.

qRT-PCR and ISH were also used to confirm the expression of some lincRNA and novel isoforms using methods described in Chapter 2.

4.3 Results

4.3.1 Identification of novel and cycling SCN lincRNA

Identification of multiexonic lincRNA expressed in the SCN

As the SCN is very small, it is normally difficult to obtain sufficient tissue to detect rare lowly expressed transcripts. Due to the anatomically precise dissection of the SCN and high sequencing depth, transcriptome assembly using cufflinks will have a high sensitivity for transcript discovery (Sims et al., 2014b). This may allow the identification of SCN-specific transcripts which may play key roles in SCN or circadian function. In total, 3,187 multiexonic long intergenic non-coding RNA (lincRNA; >200 nucleotides) loci which displayed expression ($\geq$1 read) in each of at least 7 samples were identified from the SCN transcriptome data.

Properties of these identified multi-exonic SCN lincRNA were compared to those from lincRNA in other mouse tissues to assess whether they may share similar functions. LincRNA have been previously reported to be more lowly expressed than protein-coding genes, an effect also observed for the identified SCN lincRNA (Figure 4.1) (Ravasi et al., 2006, Derrien et al., 2012, Cabili et al., 2011, Ponjavic et al., 2007). LincRNA are reported to possess fewer exons than protein coding genes (Ravasi et al., 2006, Derrien et al., 2012). This does not appear to be due to incomplete assembly of lincRNA loci due to their low expression or poor annotation, as this effect is also observed for lincRNA loci which have experimental support for their 5' and 3' boundaries (Harrow et al., 2012). In line with these findings, the identified SCN lincRNA loci possess fewer exons than protein coding genes (Figure 4.2). Sequence conservation is often used as a proxy for

functionality (Rands et al., 2014). If a sequence is functionally important its sequence is likely to be more conserved than neutrally evolving sequence. Unlike protein coding genes and small ncRNA, lncRNA exhibit weak sequence conservation which is usually suggestive of non-functionality (Pang et al., 2006). However, many important lncRNA display poor sequence conservation (Brosius, 2005) which may be because ncRNA do not need to preserve the codon structure and are therefore less affected by substitution or frame-shift mutations. This means selection may only act on particular regions of the lncRNA depending on its functional role (Pang et al., 2006, Ponjavic et al., 2007). As described for other mammalian lincRNA, the SCN lincRNA display significant sequence conservation relative to proxies of neutral evolution, yet significantly lower conservation than protein coding genes (Figure 4.3). Some lincRNA have also been reported to possess correlated expression with that of the nearest protein-coding gene (Guttman et al., 2011). This may be because these lincRNA represent enhancer (e)RNA which have an active role in transcriptional regulation (Lam et al., 2014). However, unlike previous studies, the SCN lincRNA did not show significantly correlated expression over circadian time with that of the nearest protein-coding gene (Figure 4.4), perhaps indicating that eRNA might not play significant roles in the SCN circadian system.

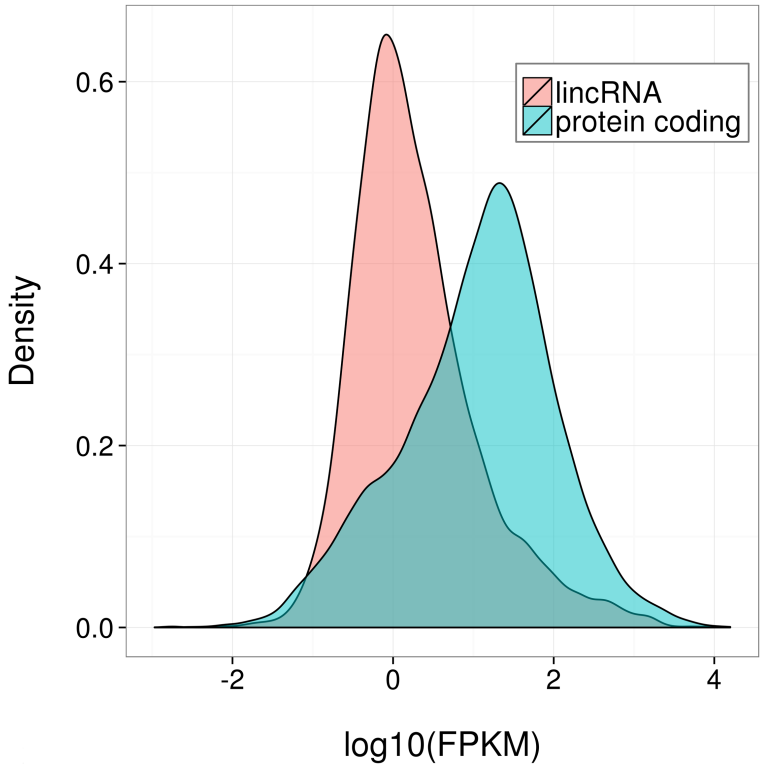


Figure 4.1. SCN lincRNA are lowly expressed.
Density plot comparing the distribution of expression levels of the 3,187 identified lincRNA (red) and protein coding genes (blue) in the SCN; lincRNA tend to be significantly lower expressed than protein coding genes ($p < 2.2x10^{-16}$, KS test).

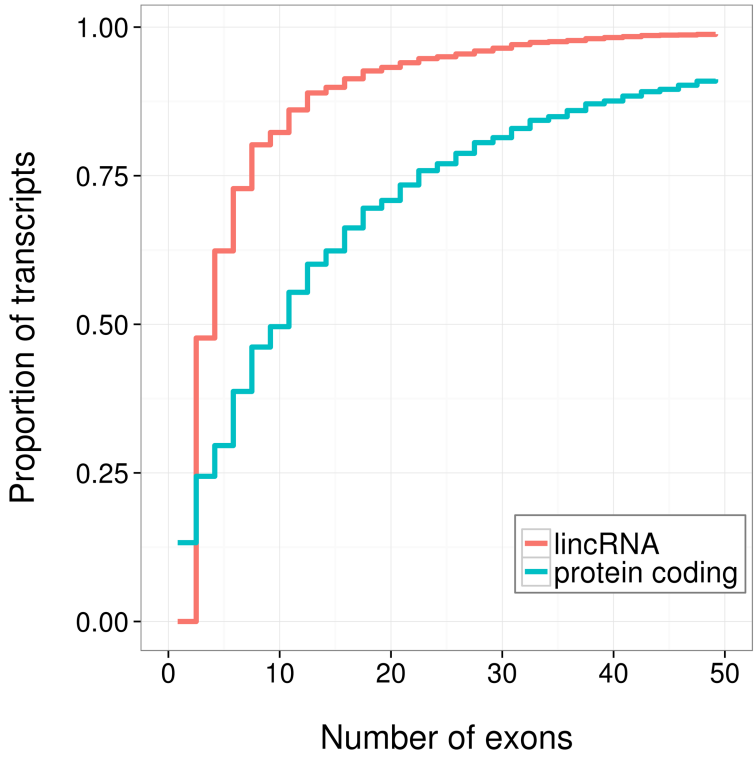


Figure 4.2. SCN lincRNA possess fewer exons than protein coding genes.
Cumulative frequency graph showing the number of exons possessed by protein coding genes and the identified 3,187 SCN lincRNA loci. Despite this study only considering multiexonic lincRNA, protein coding genes tend to possess significantly more exons ($p < 2.2x10^{-16}$, KS test).

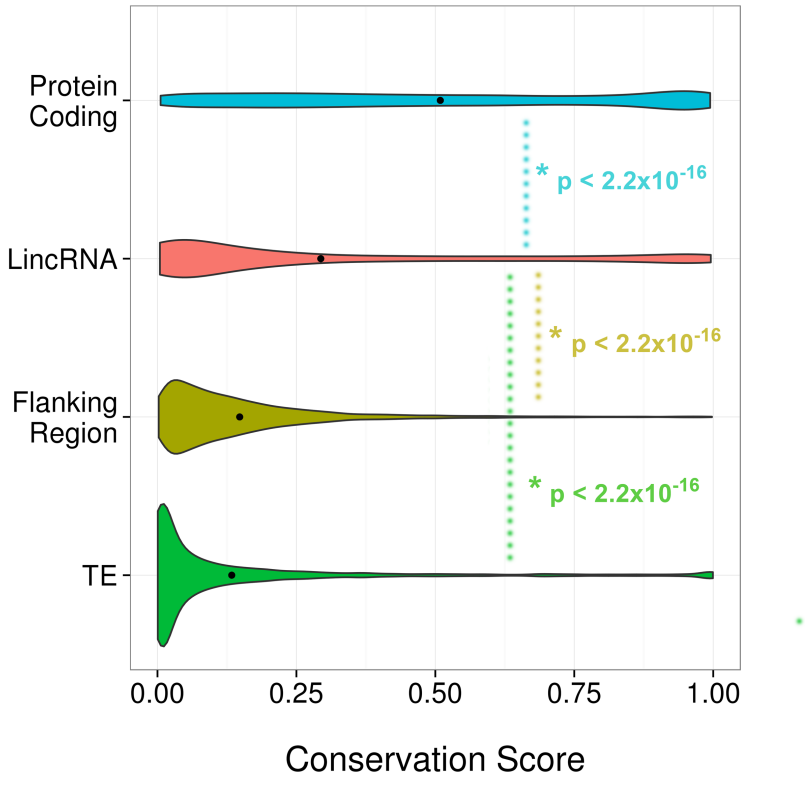


Figure 4.3. SCN lincRNA are significantly but weakly conserved.
Violin plot showing phastCons conservation scores for the exonic regions of the 2,796 identified lincRNA with available conservation scores. Black dots represent mean scores. The identified lincRNA loci showed significantly less sequence conservation than protein coding genes but significantly greater conservation than two proxies for neutral evolution: i) transposable elements within a 50kb window of the lincRNA locus ($p < 2.2x10^{-16}$, KS test), and ii) 500bp of sequence up- or downstream of the lincRNA locus selected randomly within a 1-5kb window ($p < 2.2x10^{-16}$, KS test).

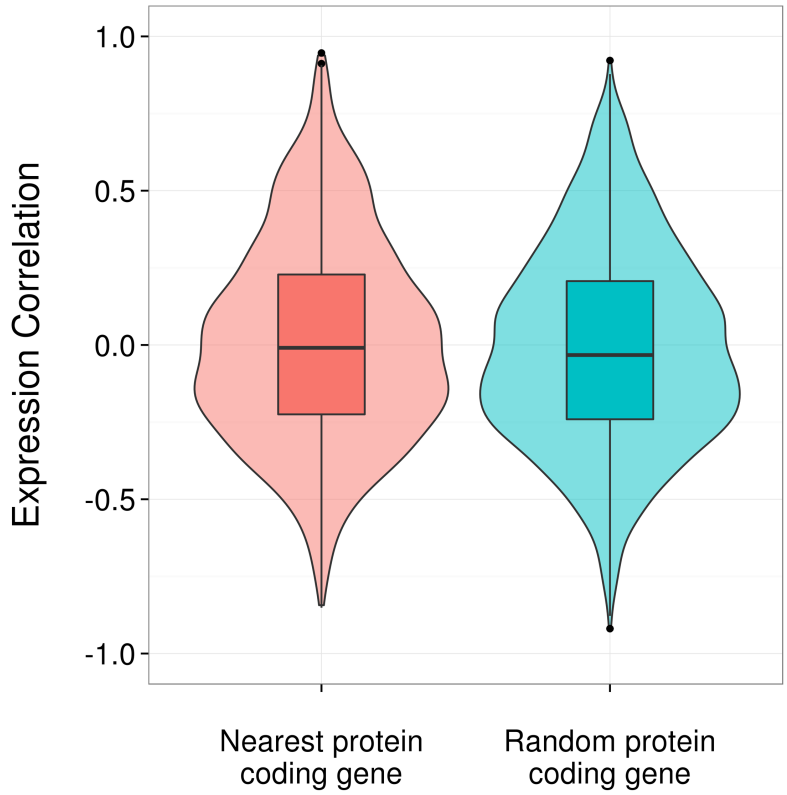


Figure 4.4. Expression of SCN lincRNA does not correlate to that of the nearest protein coding gene.
Violin and box plot showing no significant difference ($p = 0.45$, KS test) in the Pearson's correlation of a lincRNA's expression with that of its genomically most proximal protein coding gene compared to a randomly selected protein coding gene. Expression correlation of a lincRNA and its closest gene was not possible if the closest gene was not expressed, therefore only 2,261 lincRNA were surveyed.

To determine what proportion of these identified SCN lincRNA are novel, lincRNA coordinates were compared with previously identified lincRNA loci. Only 230 (7%) were Ensembl or UCSC annotated with an additional 309 (10%) identified by preceding RNA-seq studies of mouse brain tissues (Figure 4.5) (Belgard et al., 2011, Ramos et al., 2013). This leaves 2,648 lincRNA, a large proportion (83%) of identified lincRNA that had not been identified previously (Figure 4.5). This indicates that transcriptome assembly from sequencing reads derived from the anatomically precise dissection of the SCN can successfully identify novel and potentially SCN-specific lincRNA transcripts (Derrien et al., 2012, Cabili et al., 2011).

This method of lincRNA discovery does not allow identification of monoexonic lncRNA such as *Malat1* and *Neat1*. However, as these monoexonic have been previously annotated, their expression levels will have been previously assessed in chapter 3. For instance, the lncRNA *Malat1* was found to possess the fourth highest expression of all annotated genes with an FPKM of 10,581. This means that some unannotated SCN-specific monoexonic lincRNA may be overlooked.

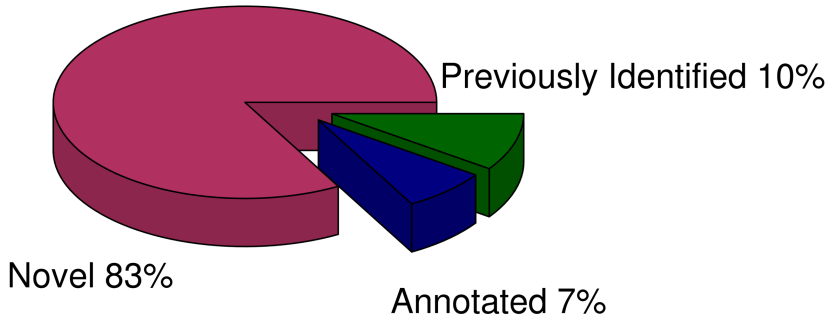


Figure 4.5. Identification of novel multiexonic lincRNA expressed in the SCN.
Proportion of the 3,187 identified lincRNA which i) have been previously annotated (according to Ensembl or UCSC), ii) are not annotated in these resources but have been identified in other studies (Belgard et al., 2011 , Ramos et al.) or iii) are unique to this study.

Functional annotation of lincRNA

To prioritise these lincRNA on importance in the SCN and to allow speculation on their function, it was assessed which lincRNA significantly fluctuate in expression over time (Chapter 3). Those lincRNA which cycle over time may play key roles in circadian rhythm generation and those which are highly co-expressed with a group of protein coding genes are more likely to possess shared functions (Cogill and Wang, 2014, Guo et al., 2015). 11% of the identified lincRNA were seen to display significant fluctuation in expression compared to 27% of protein coding genes (Figure 4.6A). This reduced proportion may reflect the lower expression and greater variation of lincRNA which may prevent them from being called as significantly fluctuating. Of these 341 fluctuating lincRNA, 28 lincRNA were part of the twin-peaking co-expression module and 317 lincRNA showed sinusoidal expression (4 common to both sets) (Figure 4.6B). The fluctuating lincRNA included lincRNA NONCO7761 (*3100003L05Rik*) whose presence in the SCN was confirmed by *in situ* hybridization at ZT6 (Figure 4.7A) as well as by qPCR to peak at both ZT6 and ZT18 (Figure 4.7B).

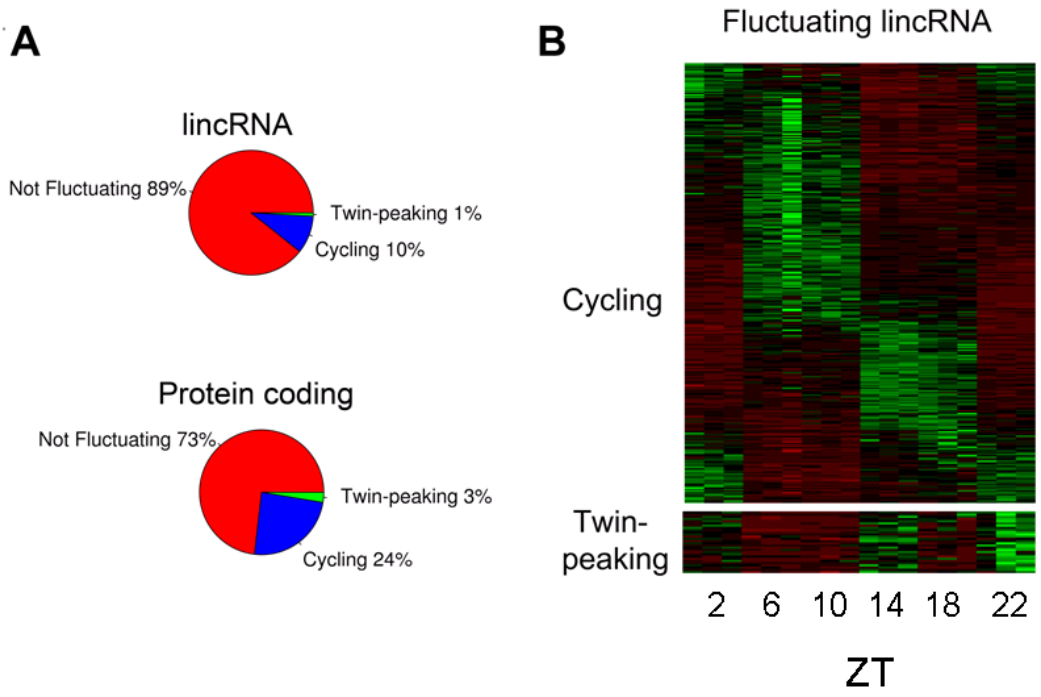


Figure 4.6. Identification of multiexonic lincRNA whose expression fluctuates in the SCN. **(A)** Proportion of the 3,187 identified lincRNA which show significant fluctuating expression patterns (twin-peaking module and DESeq2 q < 0.05, or JTK Cycle q < 0.05 and DESeq2 q < 0.05); protein coding genes are shown for comparison. **(B)** Heatmap showing the expression levels in each biological replicate for the 341 fluctuating lincRNA.

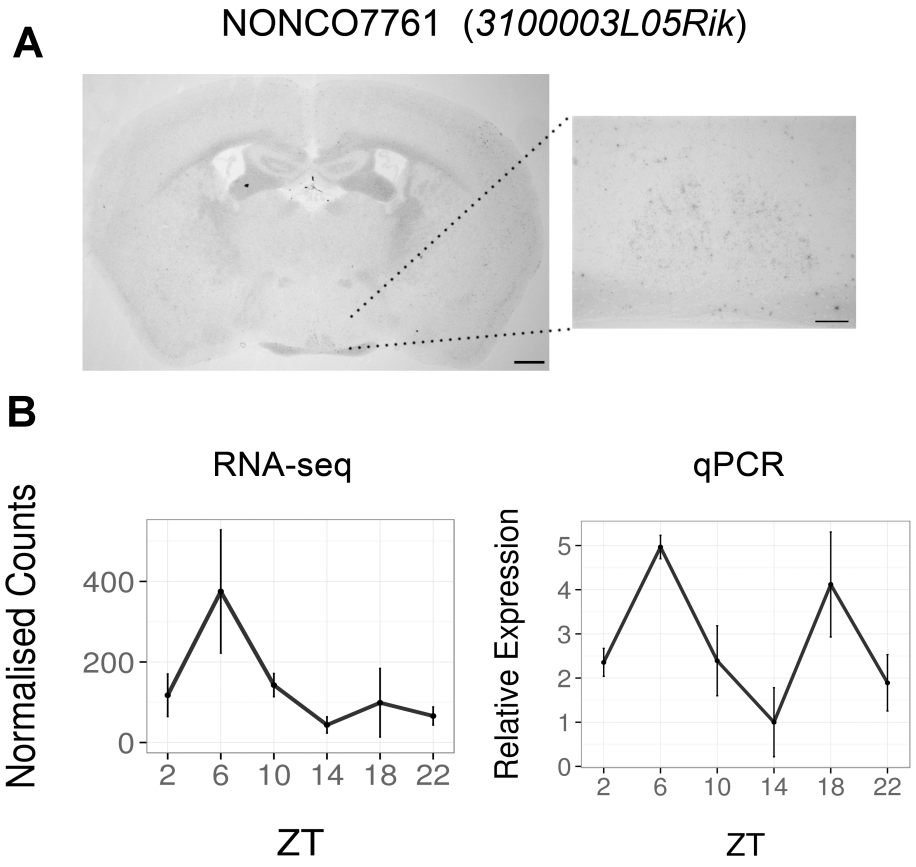


Figure 4.7. NONCO7761 shows significant circadian fluctuation in the mouse SCN. **(A)** *In situ* hybridisation of NONCO7761 (*3100003L05Rik*) at ZT6 in the mouse brain showing low yet distinct expression in the SCN. Scale bar: 0.5mm (low magnification), 0.2mm (high magnification). **(B)** RNA-seq and qPCR show fluctuating mean expression ± SD of the lincRNA NONCO7761 (*3100003L05Rik*) over 24 hours.

Although the computational tool GREAT is usually applied to DNA binding events across the genome (ChIP-seq), it may also be employed for similar functional datasets, in this case identified lincRNA loci. To gain a greater appreciation of the potential roles of these lincRNA, GREAT was employed against the genomic regions of the 100 lincRNA which formed part of the twin-peaking expression module to determine whether they were preferentially located near to genes with a particular function (McLean et al., 2010). This 'twin-peaking' module was the functionally most interesting group of fluctuating genes and was enriched for genes related to synaptic transmission (Chapter 3). GREAT can identify associations of these lincRNA loci with annotations such as gene ontology (GO) terms of nearby genes (Table 4.1). GREAT analysis revealed that genes annotated for glutamate receptor activity and DNA binding (e.g. TFs) were significantly ($q < 0.05$) associated with these lincRNA loci (Table 4.1). This suggests these lincRNA were not randomly located in the genome but instead preferentially located near to genes which possess these functions.

As some lincRNA have been described to regulate genes in close genomic proximity to its loci, it is possible that lincRNA located genomically proximal to clock genes may affect their expression and therefore affect circadian regulation. However, none of the 3,187 identified SCN lincRNA loci were within a 10Kb window of any gene implicated in core clock function. As the method for lincRNA identification only considered for loci >2Kb away from any gene, it could be possible for some unidentified lncRNA to be located within this window regulating clock gene expression.

A

Term Name	Hyper FDR Q-Val	Hyper Fold Enrichment	No of lincRNA	Set Coverage
N-methyl-D-aspartate selective glutamate receptor activity	0.025	15.94	4	4%
HMG box domain binding	0.020	12.26	5	5%
glutamate receptor activity	0.020	8.69	6	6%
core promoter proximal region DNA binding	0.015	8.26	7	7%
core promoter proximal region sequence-specific DNA binding	0.015	8.26	7	7%
RNA polymerase II regulatory region DNA binding	0.019	5.79	10	10%
RNA polymerase II regulatory region sequence-specific DNA binding	0.020	5.31	9	9%
transcription regulatory region sequence-specific DNA binding	0.013	4.23	13	13%

B

Term Name	Hyper FDR Q-Val	Hyper Fold Enrichment	No of lincRNA	Set Coverage
skeletal muscle cell differentiation	0.035	6.10	9	9%
learning or memory	0.027	5.56	11	11%
cognition	0.029	5.16	11	11%

C

Term Name	Hyper FDR Q-Val	Hyper Fold Enrichment	No of lincRNA	Set Coverage
N-methyl-D-aspartate selective glutamate receptor complex	0.031	14.16	4	4%
neuronal postsynaptic density	0.040	12.75	4	4%
dendritic spine head	0.040	12.75	4	4%
ionotropic glutamate receptor complex	0.012	8.58	7	7%
synaptic vesicle	0.017	6.22	8	8%
coated vesicle	0.021	5.03	9	9%
clathrin-coated vesicle	0.034	4.81	8	8%

Table 4.1. GREAT analysis of twin-peaking lincRNA.
Annotation enrichments of genes genomically proximal to the 100 lincRNA loci which were part of the twin-peaking expression module (Chapter 4) using GREAT. A significant ($q < 0.05$) association of these lincRNA was observed with particular GO **(A)** Molecular Function, **(B)** Biological Process and **(C)** Cellular Component terms. Tables ordered by fold enrichment (the observed / expected number of regions with a particular annotation near to the lincRNA set). Only annotations with a false discovery rate (FDR) q val < 0.05 are displayed.

4.3.2 Identification of novel exons, including of the clock gene *Cry1*

Identification of novel isoforms of annotated genes

Transcriptome assembly was used to identify SCN-specific novel isoforms of previously annotated genes which may play key roles in SCN or circadian function. 6,610 novel exons were identified from the assembled transcriptome, which were further filtered to generate a robust set of 1,013. Further details of all novel exons surveyed can be viewed on www.wgpembroke.com/shiny/SCNseq/.

Novel clock gene isoforms display antiphase cycling

To prioritise these novel isoforms based on their potential role in circadian function, it was assessed whether there were novel isoforms of core clock genes (Figure 1.3). Of the 1,013 novel exons, two appear to display robust splicing into known core clock genes (*Rora* and *Cry1*) and appear to represent alternative transcriptional start sites (Figure 4.8).

Due to the large size (322bp) and high expression, it was surprising that the novel *Cry1* exon had not been described previously. RT-PCR was used to confirm the expression of this novel exon in SCN tissue having designed primers to regions of *Cry1* (Figure 4.9A). PCR using primers designed to the known first exon and to the second exon (F1 & R) produced a transcript commensurate with the novel exon not being spliced in, suggesting that if the novel exon exists, it does not represent an internal exon despite being situated between the two known exons; this was confirmed by sequencing the PCR product. PCR using primers designed to the novel first exon and to the second exon (F2 & R) also generated a product (exon 1bi) of the expected size of the novel exon, confirming it was

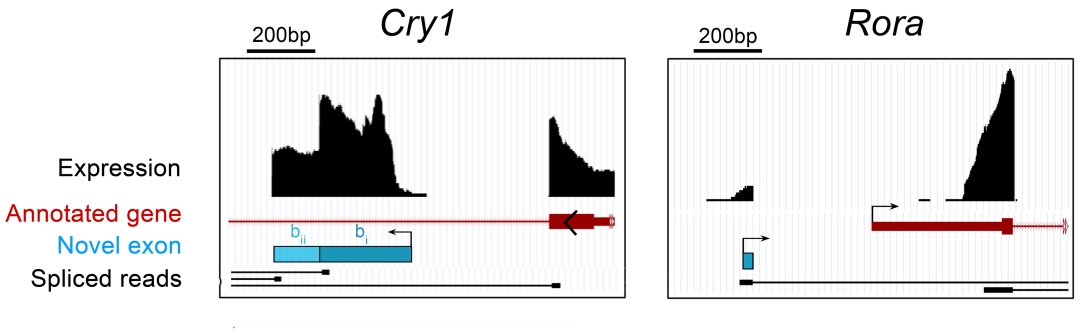


Figure 4.8. Identification of novel clock gene isoforms.
Genomic regions of mouse chromosome 10 and 9 displaying the novel exons (blue bar) for the clock genes *Cry1* and *Rora*, which appear to represent alternate transcriptional start sites (TSS). Peaks (black) signify total expression of the exon whereas representative RNA-seq spliced reads (black) indicate exonic splice junctions.

transcribed. Surprisingly, two transcripts of differing lengths appeared to be amplified. Referring to the reads which map to this exon it was predicted that this novel *Cry1* TSS was capable of producing two different isoforms (exon 1bi and 1bii) depending on where it was spliced into exon 2 (Figure 4.9). Both these transcripts were confirmed by sequencing the corresponding PCR products.

Having calculated the expression levels for both the novel and known alternative *Cry1* start sites from RNA-seq data over 24 hours, it appears that their expression cycles in antiphase, which was also confirmed by qPCR from LCM SCN samples (Figure 4.10). This may suggest different mechanisms of regulation for these two isoforms, with each isoform potentially possessing a different function.

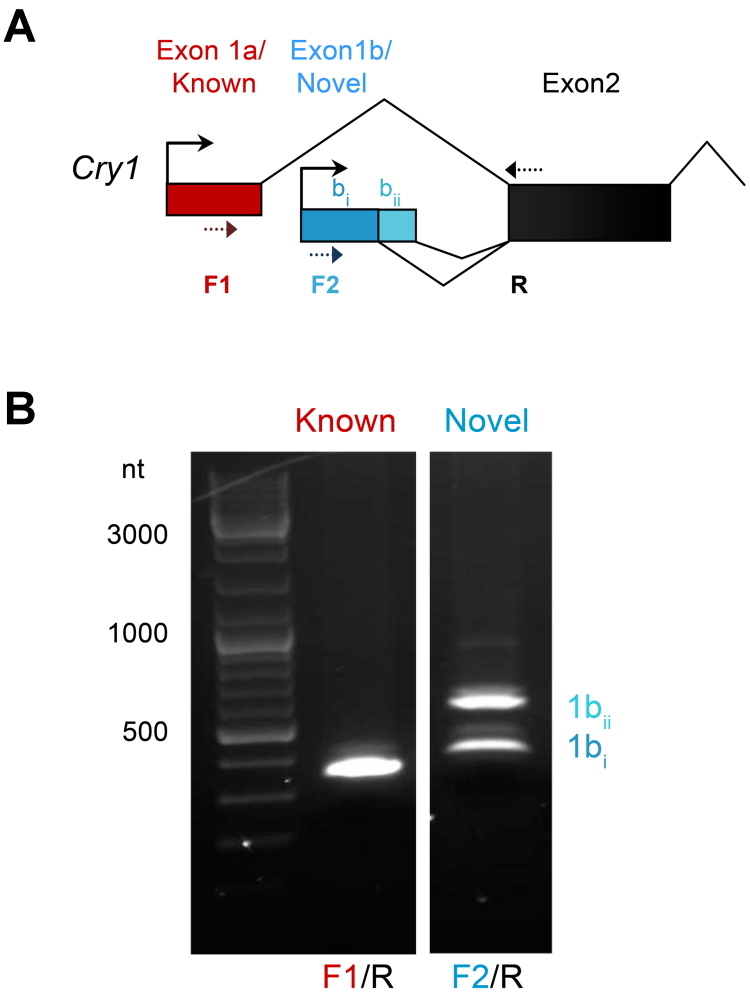


Figure 4.9. Novel *Cry1* exon represents an alternate transcriptional start site. **(A)** Schematic diagram showing the 5' gene structure of the canonical (red) and novel (blue) *Cry1* isoforms. The novel exon (exon 1b) splices into the subsequent second exon using two different exon boundaries. Dotted arrows denote the position of PCR primers used to amplify *Cry1* by RT-PCR. **(B)** RT-PCR from SCN RNA at ZT6 confirms the presence of a novel *Cry1* TSS capable of producing two different isoforms as labeled. The identity of all PCR products was confirmed by sequencing.

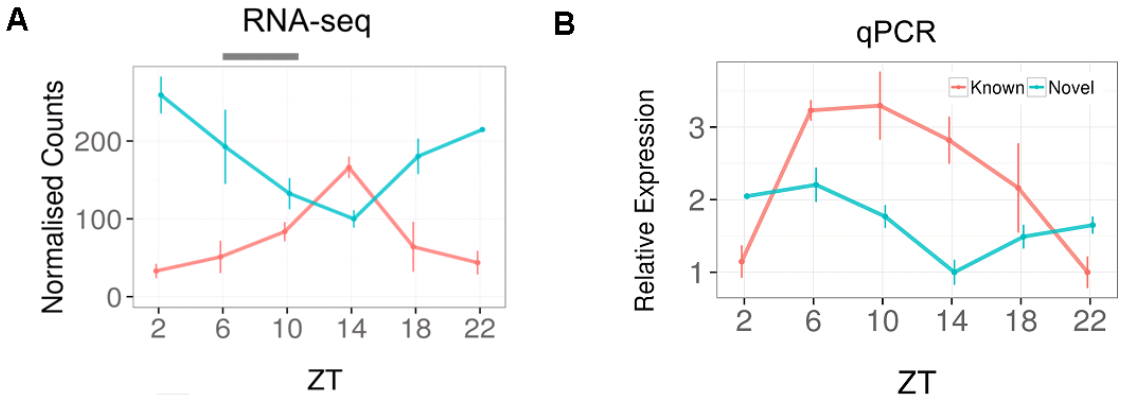


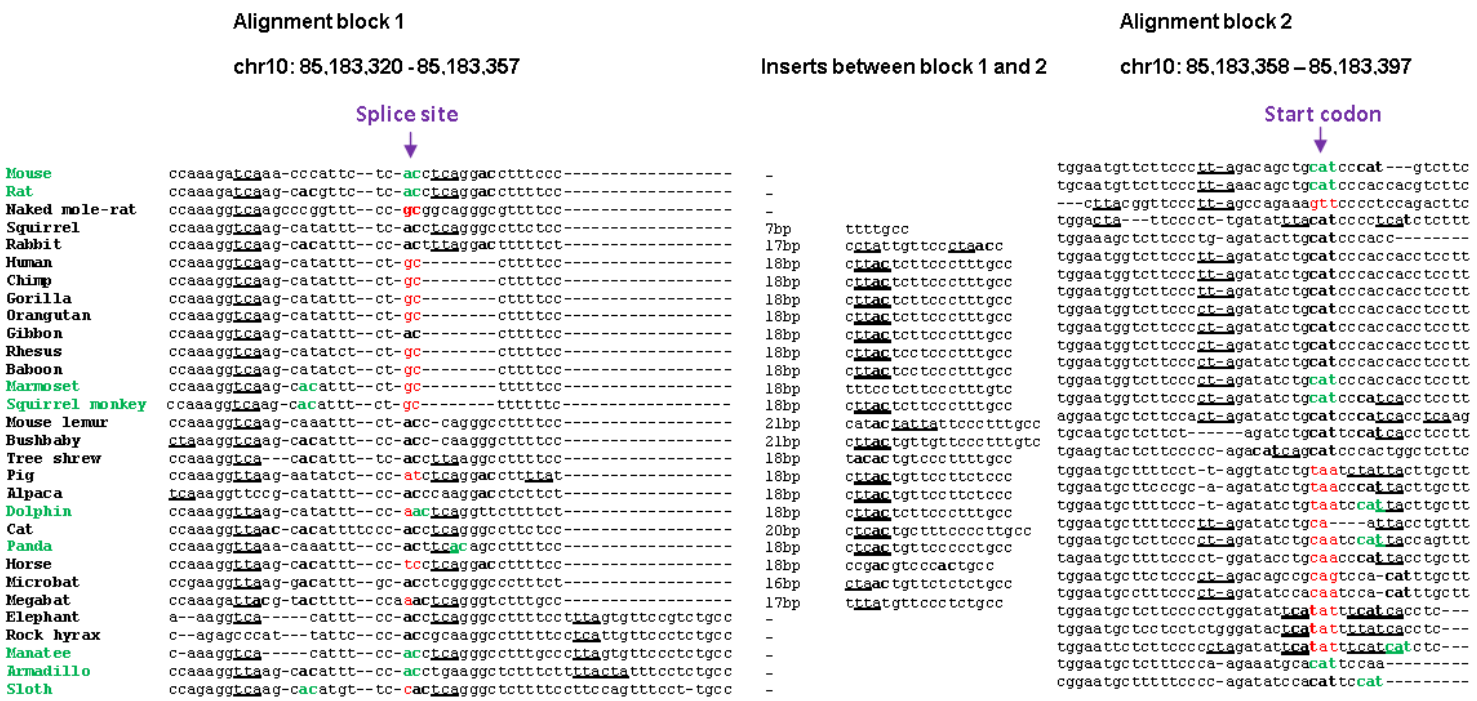
Figure 4.10. Novel *Cry1* isoform displays antiphase expression.
(A) RNA-seq expression of the canonical and novel *Cry1* TSS show antiphase cycling expression (mean ± SD) patterns. Grey line indicates potential batch effects accounting for discrepancy between RNA-seq and qPCR expression data. **(B)** qPCR of the canonical and novel *Cry1* TSS shows antiphase cycling expression (mean ± SD) patterns.

Coding potential of the novel *Cry1* exon

To further help elucidate the function of the novel *Cry1* isoform, the protein coding potential of this transcript was assessed. The novel isoforms produced by this novel TSS does not include the known TSS exon 1a which contains the initiating methionine/ start codon. Consequently it is of interest to investigate whether the novel exons1 bi and 1bii produced from this novel *Cry1* TSS contain a start codon which preserves the open reading frame and therefore coding potential. Although the shorter of the two exons (exon 1bi) from the novel TSS does not retain the ORF, the longer exon (exon 1bii) does. The resultant protein would be a truncated form of the normal CRY1 protein which is shorter by 38 amino acids (52 amino acids lost, 14 gained). This new protein isoform loses the first two beta- sheets and the first alpha- helix and alters the second (Czarna et al., 2013).

Conservation of translational machinery

To better gauge whether this transcript is protein coding or even functionally relevant, the key sequences required for translation were assessed for conservation across species. Using the UCSC multispecies alignment track for the protein coding region of the *Cry1* bii exon, it appears as though the translational features (start codon/ splice sites / reading frame) are poorly conserved in numerous other mammalian and vertebrate species (Figure 4.11). Of the 30 species surveyed, only 4 (mouse, rat, manatee and armadillo) were capable of producing a transcript with protein coding potential from the same start and stop site, whereas 9 were capable of producing the transcript when allowing for other upstream/ downstream ORFs or splice sites. There may also be additional species which possess upstream start codons or downstream in frame splice sites which allow transcription of a protein coding isoform, but were not considered in this analysis. As the novel *Cry1* isoform did not appear to have its translational machinery conserved across other mammalian and vertebrate species, it may therefore represent a non-coding transcript, despite exon 1bi possessing an in-frame start codon. This anti-phase non-coding *Cry1* isoform may represent novel mechanisms of circadian transcriptional regulation important for the functioning of the core circadian clock.



Bold – Potential start codons (**cat** - reverse complement of the atg start codon) in alignment block 2 and potential splice sites (**ac** - reverse complement of the gt splice site) in both alignment block 1 and insert block. The reverse complement is used as the ORF is on the reverse strand.

Underlined – Potential stop codon (tta, cta, tca - reverse complement of the uaa, uac, uga stop codons).

Red – Unconserved start codon / splice site.

Green (nucleotide) – Start codon / splice site which allows the production of a protein coding *Cry1* isoform.

Figure 4.11. Multispecies alignment track displaying nucleotide sequence conservation for the protein coding region of the *Cry1* bii exon.
Analysis reveals that only 4 species (mouse, rat, manatee and armadillo) are capable of producing a transcript with protein coding potential from the same start and stop sites, whereas an additional 5 (9 in total - marked in green) are capable of producing a protein coding transcript when allowing for other ORFs or splice sites. Species not capable of producing a protein coding *Cry1* isoform (marked in black) are not conserved for features of translation (start codon/ splice site/ reading frame) or have gained a mutation encoding a stop codon, without gaining compensatory mechanisms allowing protein coding potential. This analysis makes the assumption that there are not additional start sites/ splice sites up or downstream of the alignment block. It may be possible for some of the species (in black) to display coding potential from an upstream ORF or if the transcript is spliced further downstream but was not considered in this analysis.

4.4 Discussion

4.4.1 Identification of novel transcripts in the SCN

RNA-seq possesses the advantage that it can survey transcribed regions of the genome which have not been previously annotated. Due to the tiny size of this tissue and the tissue specificity of particular transcripts such as most lincRNA (Derrien et al., 2012, Cabili et al., 2011), it is unlikely that SCN-specific transcripts will have been detected from whole brain RNA-seq investigations. Instead, the high sequencing depth and highly accurate dissection of this region allow the identification of many novel transcripts which may play key roles in circadian or SCN function (Sims et al., 2014b).

Of 3,187 identified multiexonic SCN lincRNA, 2,648 (83%) had not been previously annotated or identified in two similar studies assessing mouse neuronal tissue (Belgard et al., 2011, Ramos et al., 2013). This indicates that while there are lincRNA which are known and expressed in a variety of tissues, the catalogue of lincRNA is still incomplete. It is also important to note that there are other ncRNA not detected using the approach described in this chapter. For example, miRNA cannot be detected due to the size-selection step in the library generation and both non-intergenic or monoexonic lncRNA were not assessed, even though they may play roles in circadian rhythm generation (Koike et al., 2012, Xue et al., 2014).

Furthermore, 1013 novel exons of protein coding genes were identified, including one from the clock gene *Cry1*. It was surprising that this exon had been overlooked given its large size and high expression. This illustrates the power of using RNA-seq to interrogate the transcriptome of a specific anatomical region.

4.4.2 Assessing function of lincRNA in the SCN

Although the identified SCN lincRNA share properties similar to lincRNA identified in other mammalian tissues they did not display correlated expression with that of the nearest protein-coding gene. This suggests that these SCN lincRNA are often not functioning as enhancer (e)RNA whose expression is expected to be highly correlated with neighbouring loci due to their active role in transcriptional regulation (Lam et al., 2014). eRNA transcripts may function by opening the chromatin at defined genomic loci allowing RNA pol II access (Lam et al., 2013, Melo et al., 2013, Mousavi et al., 2013). Alternatively, chromatin accessibility may be affected by the act of enhancer transcription, rather than the eRNA transcript itself (Kaikkonen et al., 2013). The third possibility is that enhancer transcription may just be "noise" from the engagement of RNA pol II due to the open chromatic environment of the enhancers (Lam et al., 2014). As this subclass have not been identified in our dataset, it could imply that eRNA do not possess poly(A) tails, as previously suggested (Kim et al., 2010), or are located <2.5kb from the nearest protein coding gene and therefore are not captured by the sequencing and analysis protocol employed in this thesis. Alternatively, eRNA might not play significant roles in the SCN circadian system.

To prioritise the identified lincRNA on their importance in the SCN, their expression was assessed in the SCN over 24 hours (Chapter 4). 300 (11%) of the lincRNA were seen to display significant fluctuation. However, this level is less than that seen for protein coding genes (27%). This could reflect the generally lower expression and greater variation of lincRNA which may prevent them from being called as significantly fluctuating.

Other studies have employed similar methods to identify circadian fluctuating lncRNA. RNA-seq of the rat pineal gland in day and night allowed the identification of 137 lncRNA of which 112 showed over 2-fold night/day differential expression (Coon et al., 2012). This number of identified lncRNA is significantly smaller than identified in this chapter. Perhaps these lowly expressed transcripts were not identified because samples were only extracted from two timepoints and were sequenced at a lower depth. In addition, although these lncRNA were described as undergoing "circadian changes", they were only measured at two timepoints and therefore could not be observed as varying with more sophisticated expression patterns.

More recently, RNA-seq was used to characterise expression levels for protein-coding and non-coding RNA over circadian time in 12 mouse tissues (Zhang et al., 2014). This study identified 1,016 ncRNA which were highly conserved between human and mouse and an additional 5,154 putative multi-exonic novel ncRNA. The increased detection of novel transcripts was expected as the authors surveyed a greater number of tissues.

Similar to the findings in this chapter, only 14% of novel and non-conserved ncRNA were found to display rhythmic expression, significantly lower than the 35% described for protein coding genes (Zhang et al., 2014). Interestingly, 32% of conserved ncRNA were seen to display rhythmic expression suggesting that sequence conservation (a proxy for functionality) rather than expression levels may correlate with circadian rhythmicity (Zhang et al., 2014). Unfortunately, the genomic coordinates for these novel lincRNA were not supplied therefore it could not be determined whether these lincRNA were also present in the SCN RNA-seq dataset.

As genes that contribute to the same pathway are more likely to be coexpressed, those 100 lincRNA which form part of the same transcriptional module as the twin-peaking protein coding genes could play a role in synaptic function (Chapter 4). These may be modulating the transcriptional or post transcriptional regulation of these protein coding genes though many different mechanisms (Figure 1.9).

Further properties can be assessed to help determine the mechanisms by which these lincRNA function. This may allow prioritisation of these lincRNA for functional studies potentially confirming lincRNA involved in key circadian processes. For example, the lincRNA HULC (Highly Upregulated in Liver Cancer) was found to affect the circadian clock in mouse liver (Cui et al., 2015). HULC siRNA dose-dependently reduced CLOCK mRNA and protein levels and HULC upregulation induced PER and CRY expression which is expected if it were to also upregulate CLOCK (Cui et al., 2015).

Properties which might aid prioritisation include: i) distance to the genomically most proximal protein coding gene, ii) expression correlation to that gene, iii) expression level of that gene, iv) sequence conservation of that gene and v) identification of syntenic human lincRNA. These may help determine the mechanism through which the lincRNA acts.

LincRNA which act in *cis* affect the expression of neighbouring loci and will therefore likely display correlated expression (Cabili et al., 2011) and may potentially act in a distant-dependent manner; for instance, lincRNA loci genomically more proximal to the nearby gene may display a greater correlation in expression. A lincRNA positively correlated with its neighbouring gene may enhance expression (such as through RNA Pol II recruitment or chromatin opening) whereas those which are negatively correlated may repress it. Although the identified SCN lincRNA were determined to not often be

functioning as eRNA (Figure 4.4), it is possible that some lincRNA may act in *cis* and affecting the expression of neighbouring loci.

Expression levels of particular lincRNA may also be important to decipher function. It has been suggested that cytoplasmic ceRNA are too lowly expressed to titrate miRNA preventing their binding to mRNA (Denzler et al., 2014). Therefore ceRNA may only be a viable hypothesis if either the lincRNA is highly expressed or the corresponding miRNA for which they bind are lowly expressed. CeRNA expression levels are also expected to positively correlate with those genes which it regulates; increased expression of the ceRNA soaks up the miRNA repressing the mRNA, thereby preventing mRNA degradation and increasing its level within the cell (Khvorova and Wolfson, 2012).

Conservation of a lincRNA's exons suggests that its sequence is important for function and is conserved across species. The conservation score of the lincRNA locus can be compared to proxies of neutral evolution to determine conservation strength. Lastly, as mouse lincRNA are generally poorly annotated, if the same lincRNA locus is conserved and annotated in humans, it can be hypothesised that its function is similar in both organisms.

The computational tool GREAT was employed to help reveal the function of a particular group of lincRNA loci based upon their genomic proximity to other regions in the genome. Twin-peaking lincRNA whose expression was highly correlated with the protein coding genes of the same module were more likely to be located near to genes which were annotated for glutamate receptor activity or transcription factors (Table 4.1). Although the identified SCN lincRNA were determined to not often be functioning as eRNA (Figure

4.4), some of the twin-peaking lincRNA might be involved in those processes described by GREAT.

In summary, 3,187 multiexonic SCN lincRNA were identified and a framework was created which allows prioritisation of these lincRNA for further investigation. These lincRNA should be experimentally validated before conducting functional experiments. It would be of interest to determine the function of the twin-peaking lincRNA, which are highly co-expressed with the twin-peaking protein coding genes which play roles in synaptic function and whose expression is important in clock resetting pathways (Chapter 3). As co-expressed genes are likely to be under similar transcriptional regulation and possess shared functions, these lincRNA may also be important for synaptic function. Furthermore, these twin-peaking lincRNA were more likely to sit near to genes relating to glutamate signalling suggesting they may act in *cis* to regulate the expression of these genes. Investigation of cycling lincRNA may also provide novel insight into circadian biology. For example, NONCO7464 (*Gm29683*) shows a strong cycling expression pattern peaking at ZT6. This lincRNA appears to be strongly conserved with respect to proxies of neutral evolution (z-scores of 13.88 (AR) and 6.61 (FS)) and highly correlated (r = 0.86) in expression with the closest protein coding gene *Nr2f2*. This gene encodes a ligand-inducible transcription and can affect the regulation of many other genes. Both the arrhythmic Clock$^{\Delta 19/\Delta 19}$ mouse mutant and *in-vitro* models of circadian dysfunction displayed significant reduction and alteration of *Nr2f2* expression suggesting it as a candidate CCG (Heckman, 2009). It would therefore be interesting to determine whether NONCO7464 affects the expression of *Nr2f2*, and if disrupted, how it may affect the circadian system or other aspects of physiology.

4.4.3 Regulation and function of novel *Cry1* isoform

The newly identified *Cry1* exon is capable of producing a protein coding transcript starting from exon 1bii. The resultant CRY1 protein will be slightly truncated having lost its first two beta- sheets and first alpha- helix, regions which appear to be dispensable for its function in the SCN (Czarna et al., 2013, Khan et al., 2012). However, as the key sequences for translation were not evolutionarily conserved, this transcript may only be functional in specific species, or alternatively, its protein coding potential may not be important for function. Even if non-coding, this *Cry1* transcript may still play an important role in the circadian system or represent complex layers of regulation upon core clock genes.

For example, unproductive alternative splicing is able to produce transcripts which incorporate in-frame premature stop codons. Although the functions of these non-coding mRNA are largely unknown they may play an important role in the circadian system (Filichkin and Mockler, 2012b). In order to create oscillations in core clock proteins there needs to be rapid changes in mRNA levels which may be difficult through only relatively slow *de novo* transcription. Unproductive alternative splicing or isoform generation may also contribute.

RNA-seq of the *Arabidopsis* transcriptome over circadian time found that many clock genes displayed extensive alternate splicing (AS) events (Filichkin and Mockler, 2012b). For example, the clock gene *CCA1* was found to retain an intron which contained a premature stop codon, a splicing event conserved across phyla (Filichkin et al., 2010). The accumulation of this nonsense *CCA1* transcript could result in a down regulation of this gene at particular times or under specific conditions. For example, experimentally

subjecting *Arabidopsis* to cold temperatures decreased the nonsense *CCA1* transcript expression suggesting it plays a role in temperature compensation mechanisms (Filichkin and Mockler, 2012b).

As the novel *Cry1* isoform was not evolutionarily conserved for features of translation, its protein coding potential is unlikely to be important for its function or regulation of this clock gene. Instead, this transcript may be a result of an unproductive alternative splicing event necessary to alter the expression of the canonical protein coding *Cry1* isoform in order to regulate protein levels.

Nucleotide sequences of the novel *Cry1* isoform may reveal other features of transcriptional and translational regulation. 5' UTRs have an important role in translational control by harbouring translation regulatory elements such as upstream open reading frames (uORFs) which usually inhibit translation (Wethmar et al., 2014). Ribosome profiling allows the sequencing of mRNA transcript regions which are bound by ribosomes. A recent study found that transcripts with ribosomes bound to their 5' UTR through uORFs were able to decrease translational efficiency by >30% (Janich et al., 2015a). Experimental mutation and deletion of these uORFs increased translation of that gene. Furthermore, when multiple uORFs of the same gene were mutated, this effect upon translation was shown to be additive (Janich et al., 2015a). Interestingly, these uORFs play a role in the circadian system as loss of function of *Denr*, a regulator of ribosome recycling and reinitiation after uORF usage, induced a shortening of circadian period in fibroblast cell culture (Janich et al., 2015a).

The newly identified novel *Cry1* exon possessed 10 uORFs upstream of its potential start site. uORFs were defined as start codons (ATG) which were out of frame with the

canonical *Cry1* transcript or produced a stop codon in the resulting protein sequence. This number (10) was considerably more than the two upstream of the start site of the canonical protein-coding *Cry1* isoform, despite having similar length UTRs (386nt vs. 583nt). This indicates that even in the species with protein coding potential for this isoform, there are numerous uORFs which reduce the translation efficiency of this transcript and therefore CRY1 protein levels. In addition, western blots of CRY1 protein do not show two separate bands (which would indicate different length isoforms) (Kaasik and Lee, 2004). This may be because the protein lengths are very similar (differ by only 38 amino acids) and cannot be distinguished through this method, the novel isoform is SCN-specific, or perhaps more likely because this novel *Cry1* transcript is not translated.

This novel isoform was expressed in antiphase to the canonical protein coding isoform suggesting additional levels of regulation at the transcriptional level. It is possible these data may represent an isoform-switching mechanism between the *Cry1* isoforms allowing the proportion of the coding and non-coding transcript to fluctuate relative to each other potentially allowing production of steep CRY1 protein oscillations (Figure 4.12).

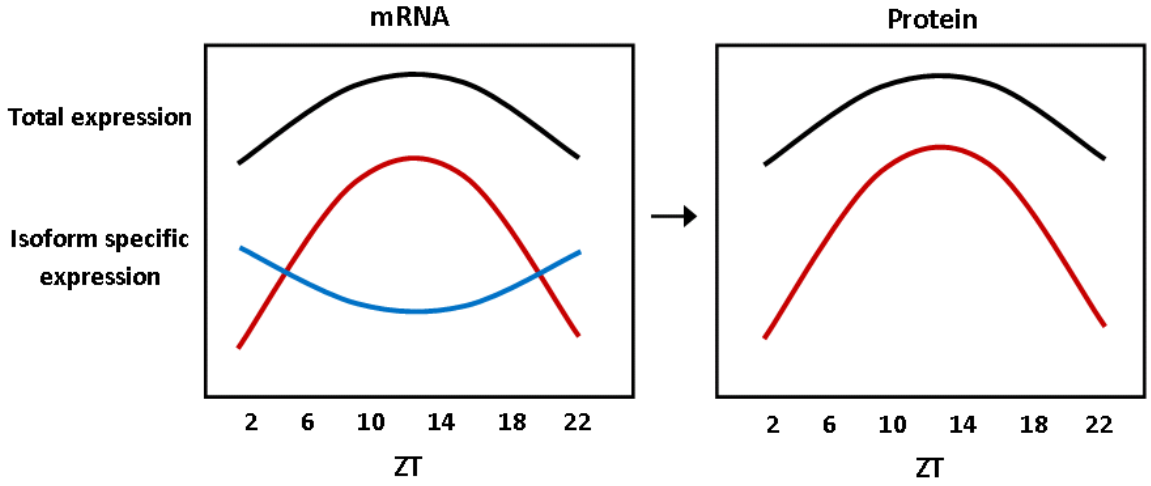


Figure 4.12. Isoform switching mechanism can produce strong protein oscillations.
Schematic diagram showing that isoform switching between a coding (red) and non-coding (blue) transcript across circadian time can produce strong protein oscillations (red). If a gene encodes only a protein coding mRNA (black — left) then the resultant protein (black — right) is largely dependent on the total mRNA level which may be regulated by TF binding or other regulatory processes. Isoform switching allows the production of a stronger oscillation in the coding mRNA transcript and therefore corresponding protein levels. For visualisation purposes, diagram assumes no delay between transcription and translation.

Regulation of the timing of clock gene expression depends on the presence of protein binding sites which regulate expression. The transcription of *Cry1* depends on these elements which include E'/E-boxes and D-boxes found in the promoter, but also RORA/Rev-ErbA response elements (RREs) which have been identified in a highly conserved region of 336 bp within the first *Cry1* intron (Ukai-Tadenuma et al., 2011). Interestingly, this RRE acts as an enhancer to delay the phase conferred by the *Cry1* promoter; a D-box driven luciferase reporter construct showed a 5-hour delay in expression in mouse fibroblasts when the 336bp *Cry1* intron sequence was included in the construct (Ukai-Tadenuma et al., 2011). Furthermore, asynchronous *Cry1* double knock out cells could only regain synchronicity when an inserted *Cry1*-expressing plasmid was driven by a *Cry1* promoter with RRE-containing 336bp of *Cry1* intron. When *Cry1* expression was driven solely by the *Cry1* promoter without this intronic sequence, circadian rhythmicity could not be rescued (Ukai-Tadenuma et al., 2011). This suggests that this delay of *Cry1* expression is essential for circadian rhythmicity.

Both the RRE and the identified novel *Cry1* exon are situated in the first *Cry1* intron. If the RRE is situated too close to the novel exon it may be unable to regulate its transcription (i.e., through a DNA looping mechanism) and therefore unable to induce a phase-delay. This may explain the ∼5-hour phase-advance observed for this transcript relative to the canonical isoform (Figure 4.10). There may therefore be a RORA/ Rev-ErbA-dependent mechanism which is able to affect the switching between these two isoforms in a time of day dependent mechanism to regulate CRY1 protein levels (Figure 4.13).

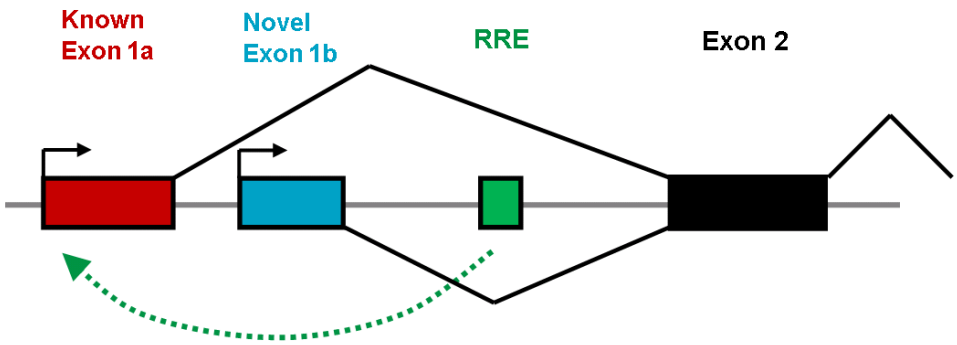


Figure 4.13. Transcriptional Regulation of *Cry1* isoforms.
Schematic diagram showing potential looping mechanism of the RRE to induce a delay in the canonical but not novel *Cry1* isoform.

In summary, identification of novel exons and isoforms, using the novel *Cry1* exon as a case study has exposed the complexity of gene regulation in the circadian system. Temporal expression profiles and sequence information of the *Cry1* isoforms provide insight into this gene's regulation in producing robust CRY1 protein oscillations. The expression of this novel isoform appears to cycle in antiphase to the canonical transcript which may be dependent on RORA/Rev-ErbA binding. Although this transcript does possess protein coding potential, it contains numerous uORFs and does not possess conserved features of translation. This suggests that it may be important for this transcript to produce very little or no functional protein product. This process may have evolved to allow the rapid generation of CRY1 protein oscillations which may not be possible through solely *de novo* transcription of a single isoform.

5 Deciphering the relationship between circadian behaviour and expression in *Bdr*

5.1 Introduction

As previously discussed (section 1.1.5), circadian rhythm disruption is associated with psychiatric diseases such as schizophrenia. An aim of this study was to further investigate the blind-drunk (*Bdr*) *Snap25* mouse mutant to establish the molecular pathways which when disrupted contribute to both circadian disruption and psychiatric disease. *Bdr* is of particular relevance to this question as it displays both schizophrenia endophenotypes and fragmented rest/activity profiles.

The *Bdr* mouse model has an I67T dominant negative mutation in the *Snap25b* isoform which increases the binding affinity within the core SNARE complex impairing action potential evoked glutamate release (Jeans et al., 2007). *Bdr* mice show a range of behavioural defects such as aspects of apathy and anxiety: mutant mice were less likely to interact with the toys in the "playground" novelty paradigm and displayed fewer transitions between the two compartments in the two-compartment light/dark test (Jeans et al., 2007). In addition, *Bdr* mice displayed a reduction in pre-pulse inhibition (abnormal sensorimotor gating) suggesting they are less able to filter out redundant or unnecessary stimuli, a measure often used to help diagnose schizophrenia patients (Jeans et al., 2007, Ludewig et al., 2003). These schizophrenia-like endophenotypes were shown to be exacerbated when *Bdr* mice were subjected to prenatal stress, which could be reversed with anti-psychotic drugs (Oliver and Davies, 2009). Together, these data suggest that the *Bdr*

model possesses numerous phenotypes which make it suitable for investigating psychiatric disease.

A circadian behavioural wheel-running screen was carried out in *Bdr* mice to determine whether there were deficits in any key circadian behavioural parameters (Oliver et al., 2012). *Bdr* mice possessed a fragmented and phase-advanced rest/activity profile under a 12:12 light/dark cycle: mutants were more likely to display arrhythmic circadian behaviour or have an earlier phase of entrainment compared to littermate controls (Figure 5.1,5.2). However, *Bdr* displayed similar tau (free-running period) under constant dark (DD) or light (LL) conditions to wild-type controls (Figure 5.3), suggesting that the core molecular clock is not disrupted in *Bdr*. This was supported by the fact that the core clock genes show similar expression patterns between the genotypes assessed at ZT6, ZT11 and ZT13 using qPCR from SCN tissue punches (Oliver et al., 2012). In addition, sleep was determined not to be greatly affected in *Bdr* mice, as determined by video tracking (Oliver et al., 2012). The failure of *Bdr* mice to entrain to 12:12 LD cycles suggests that there may be deficits in the light input to the circadian clock. However, *Bdr* mice responded normally to negative masking of activity by light and to phase-shifting paradigms such as the 6-hour phase-advance and the phase-delay inducing light pulse at CT16, suggesting that the circadian deficits were not due to the direct effects of light (Figure 5.3). As the input and endogenous core clock appear to be unaltered, the *Bdr* circadian phenotype was proposed to be due to disruption of SCN output signals.

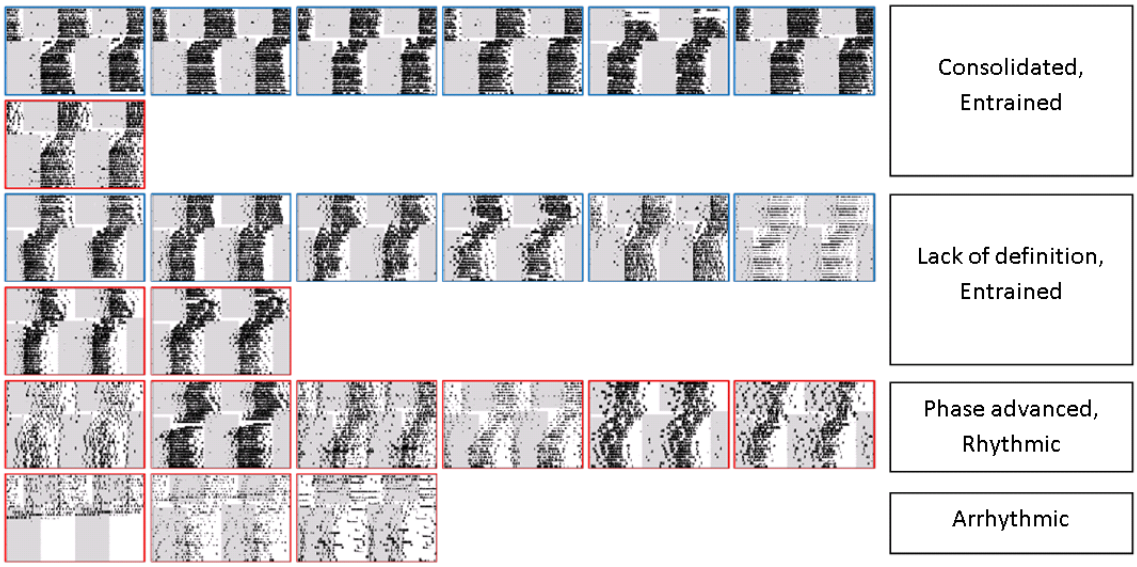


Figure 5.1. *Bdr* mice display circadian deficits.
Actograms from WT and *Bdr* under a 12:12 LD cycle and a 6-hour phase-advance. There is considerable heterogeneity amongst the *Bdr* mice actograms (red borders): most appear to have a rhythmic phase-advance profile, but also span from robustly entrained to completely arrhythmic. The WT mice (blue borders) show clear entrainment to the LD cycle. Image obtained from (Oliver et al., 2012).

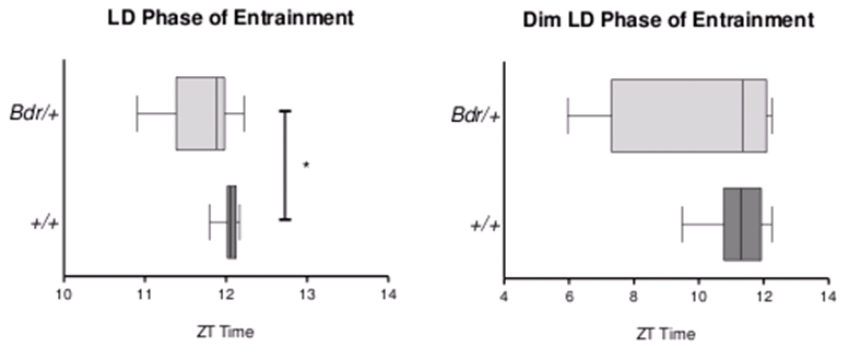


Figure 5.2. *Bdr* mice possess phase-advanced and heterogeneous phase of entrainments. Phase of entrainment (time of activity start) for both WT and *Bdr* mice in a standard (150 lux) LD cycle and in a dim (30 lux) LD cycle. In both conditions, *Bdr* mice display an increase in heterogeneity, for example some *Bdr* mice under 30 lux displayed a rhythmic 6-hour phase-advanced profile whereas some displayed no phase difference to the WT. Entrainment is significantly earlier in the *Bdr* animals under 150 lux LD cycle. Image obtained from (Oliver et al., 2012).

To identify the genes or pathways which induced the circadian deficits observed in the *Bdr* mutant, microarray expression profiling from SCN-enriched 1mm tissue punches was assessed at ZT6, ZT11 and ZT13. Analysis revealed the dysregulated expression of the genes encoding SCN neuropeptides *Avp*, *Nts* and *Tac1/Sp*. The elevated expression of *Avp* at ZT6 vs ZT11 in *Bdr* mice was validated by qPCR, *in situ* hybridisation (ISH) and immunohistochemistry. Peripheral blood corticosterone levels - a peripheral rhythm under SCN control - was also measured and this, like *Avp* production and wheel-running behaviour, appeared to be phase-advanced (Oliver et al., 2012). Together, the *Bdr* mice show schizophrenia-like endophenotypes and aspects of circadian disruption which may be a result of disruption to the SCN output pathway.

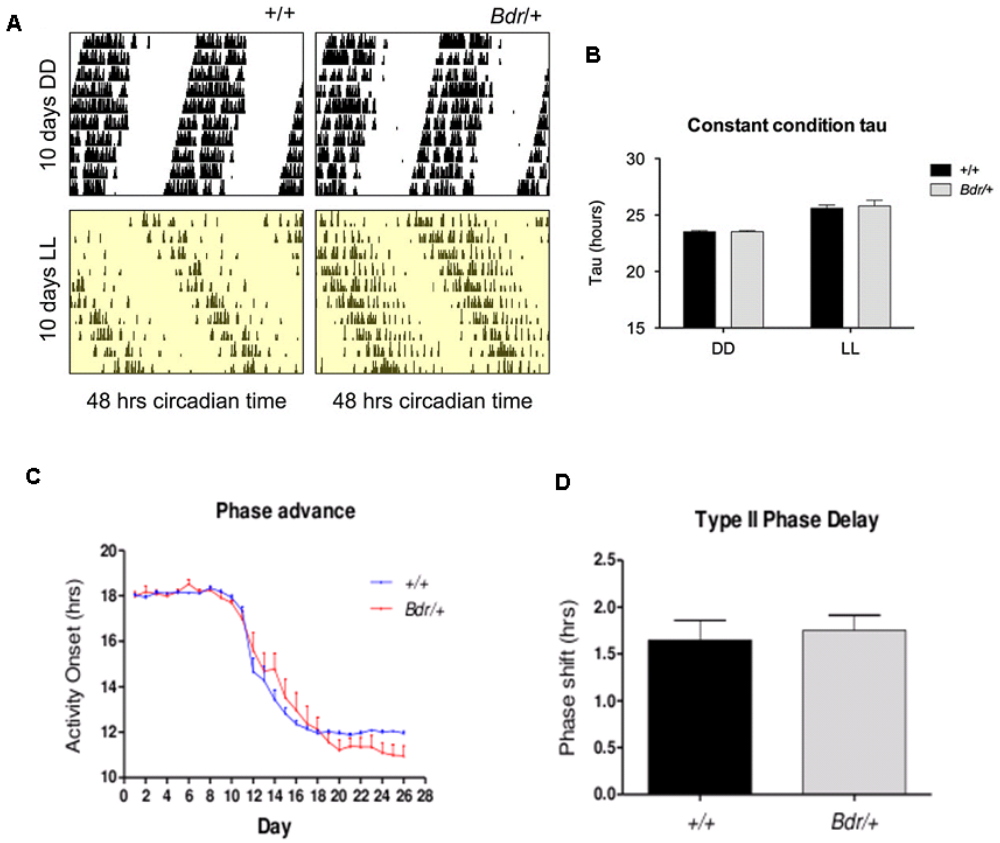


Figure 5.3. The core molecular clock and response to light is not altered in *Bdr* mice. **(A)** Representative actograms of mice running under constant dark (DD) and constant light (LL) conditions. **B)** The free running period (tau) in DD and LL show no significant differences between genotypes. **(C)** A 6-hour phase-advance jet-lag paradigm and **(D)** a 1-hour light pulse at CT16 determined no significantly differences in phase-shifting potential between genotypes. Image obtained from (Oliver et al., 2012).

However, as only a very small number of protein coding genes were differentially expressed in *Bdr*, a more detailed transcriptomic analysis over additional time points may provide additional clues to the *Bdr* rest/activity phenotype and further elucidate the link between SCRD and psychiatric disease. Firstly, the transcriptomic differences which induce circadian deficits may be SCN-specific and therefore genes which are truly differentially expressed between genotypes may be obscured by the surrounding hypothalamic tissue and not be identified. Secondly, any transcriptomic differences may have been missed due to the low temporal resolution of SCN profiling. The samples were only extracted at 3 timepoints (ZT6, ZT11 and ZT13) therefore mechanistically important transcriptomic differences at specific times in the day, such as during late subjective night, may have been overlooked. Lastly, as previously discussed (section 1.1.5), use of microarray technology may prevent identification of truly differentially expressed genes. It is possible that the gene(s) inducing the behavioural phenotypes may encode a lincRNA or a coding transcript not represented on the microarray. Additionally, some microarray probes may possess a low signal-to-noise ratio and therefore subtle differences in expression cannot be captured. Microarrays have been shown to be inferior to RNA-seq in detecting differentially expressed genes, especially amongst low abundance transcripts (Wang et al., 2009, Wang et al., 2014).

To overcome these limitations and allow a more complete investigation into the transcriptomic disruptions in this mouse model, SCN samples were extracted by laser capture microdissection (LCM) from *Bdr* at the same six timepoints (ZT2, 6, 10, 14, 18 and 22) in parallel with the wild-type mice (Chapters 2-4). Importantly, it was hypothesised that housing *Bdr* in LD conditions would assist in the identification of the transcriptomic dis-

turbances in *Bdr* as these were the lighting conditions which resulted in atypical circadian behaviours (Figure 5.1,5.3).

In this chapter, I sought to understand the relationship between circadian behaviour and SCN expression in *Bdr*. The precise plans were to:

- Identify those genes whose expression is dysregulated in the *Bdr* mouse model compared to WT using differential expression analysis tools on the RNA-seq count data.
- Identify the biological pathways or processes dysregulated in the *Bdr* SCN by assessing whether the genes differentially expressed between genotypes are enriched or implicated in particular biological processes or diseases.
- Confirm the pathways which may play a role in both circadian rhythms and psychiatric disease through behavioural paradigms.

5.2 Methods

5.2.1 Differential expression analysis

DESeq2 (and DEseq) are methods for differential expression analysis of count data (e.g. the number of reads which match to a gene), attempting to identify those genes which are significantly altered between two (or more) conditions. DESeq2 first normalizes the count data by estimating 'size factors' for each sample as described in chapter 2. The raw counts for all protein coding and lincRNA loci (identified in chapter 4) which were expressed in the WT mice were divided by the size factor associated with that sample to normalise the count table (Figure 5.4).

To test for differential expression, the dispersion parameter for each gene needs to be estimated for biological replicates of the same condition. DESeq2 differs from standard General Linear Models (GLM) which treat the dispersion for each gene separately which, when sample size is small, may fail to identify differentially expressed (DE) genes as a consequence of the reduced power due to the uncertainty on the variance estimates. DESeq2 overcomes this lack of power by pooling variance information for genes of similar expression levels (Love et al., 2014). DESeq2 therefore makes the assumption that genes of a similar expression level are assumed to possess similar variances. Genes which appear to have a dispersion estimate far from the expected level will have their variation corrected towards this expected level (Figure 5.5). However, if the test is well powered (large number of replicates), the genes' variance is adjusted to a lesser extent. The original version of DESeq selected the maximum of the fitted curve and the gene-wise dispersion estimate

when testing for differential expression and therefore tended to overestimate dispersion leading to a higher number of false negatives.

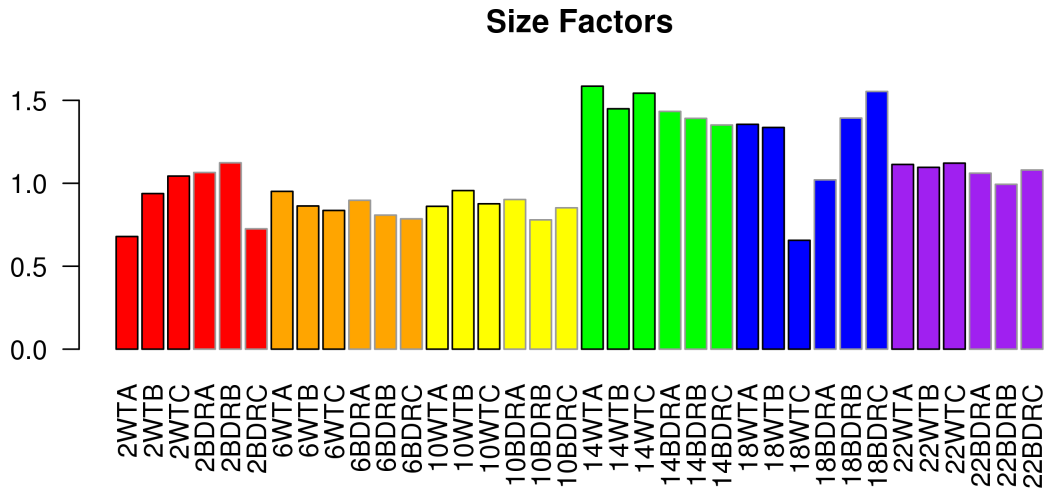


Figure 5.4. Normalisation size factors generated in DESeq2.
The size factors determined by DESeq2. The gene count values for a particular sample are divided by the corresponding size factor. Each timepoint is signified by a different colour and genotypes are signified by different coloured borders (WT — black, *Bdr* — Grey)

After a GLM was applied to each gene, DESeq2 performed a Wald test which tests individual coefficients or contrasts of coefficients to identify significant DE genes.

DESeq (in addition to DESeq2) was used to help identify DE genes. DESeq2 determined whether there was a significant FC difference (between mutant and wild-type) at any timepoint given the dispersion of that gene's expression. However, DESeq2 did not account for the possibility that variation at each timepoint may differ. Therefore, a gene would be called as DE if it possessed low variation at all timepoints except for one, where there may be a discrepancy in expression levels between the genotypes, perhaps due to

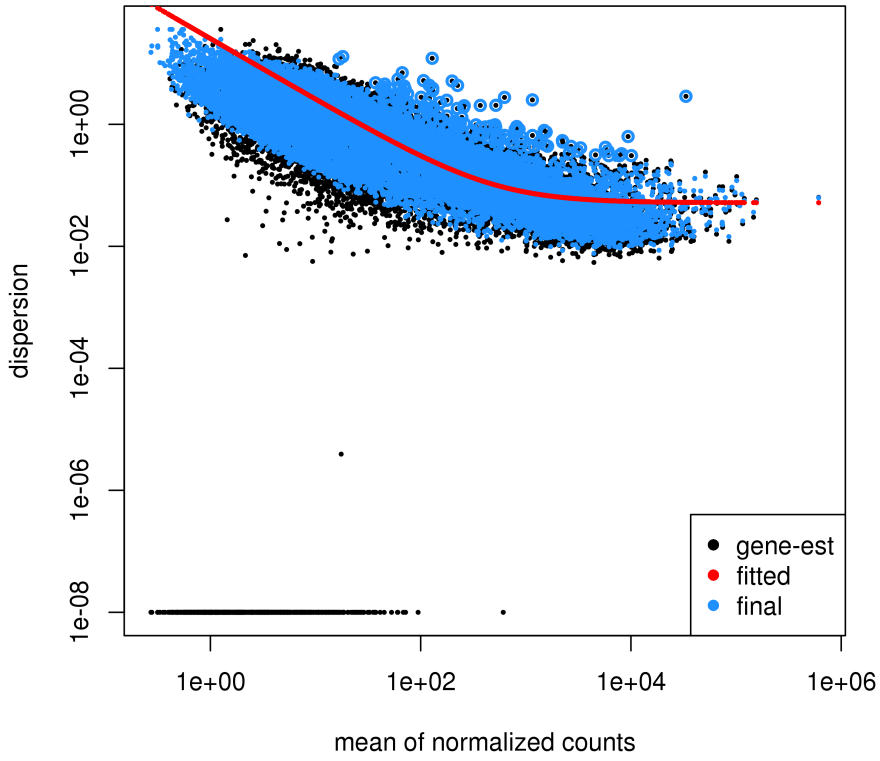


Figure 5.5. Dispersion estimation in DESeq2.
A fitted variance value (red) is calculated for each gene according to its expression level. This value is generated by pooling variance of genes with a similar expression level. The estimated variance for a single gene is estimated (black) but is then corrected towards the fitted variance resulting in a final dispersion estimate (blue). There is a tendency for dispersion to be the highest at low expression levels. DESeq selects the maximum of the fitted curve and the gene-wise dispersion when testing for differential expression.

an outlier (Figure 5.6). To combat this, a pairwise comparison of the genotypes was performed using DESeq, but considering the variation only from that timepoint. The caveat with this approach is that very noisy genes, which by chance, show very little variation but a large expression difference at a single timepoint, are called as DE (Figure 5.6). To combat the limitations in both approaches, a combinatorial approach was used to detect truly DE genes. DESeq removed the bottom 30% of lowly expressed genes before DE testing to alleviate the problem of multiple testing.

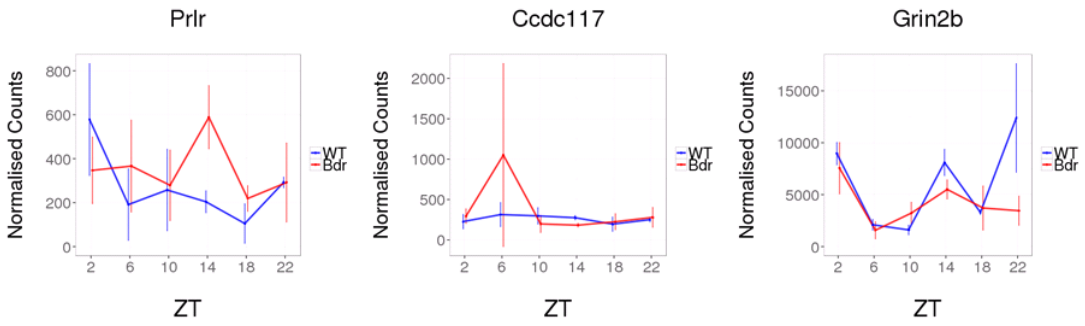


Figure 5.6. Combinatorial approach to DE gene identification.
A pairwise approach (which only considers the variation at a single timepoint) will identify *Prlr* as DE but *Ccdc117* as not. A GLM approach which considers the variation for that gene (for all timepoints) and determines whether the FC at a single timepoint is sufficiently large to account for that dispersion identifies *Ccdc117* as significant but not *Prlr*. Therefore, using both approaches prevents noisy genes which may appear to be DE at a single timepoint due to chance (e.g *Prlr*) and genes with a large fold change difference driven by outlier expression (e.g *Ccdc117*) to be called as DE. Genes which pass both criteria, such as *Grin2b* were called as DE.

A gene was therefore called DE at a particular timepoint if it was called DE according to the pairwise approach in DESeq (p.adj<0.05) and according to the GLM approach in DESeq2 (p.val<0.05). The DE genes identified through the pairwise DESeq approach (as

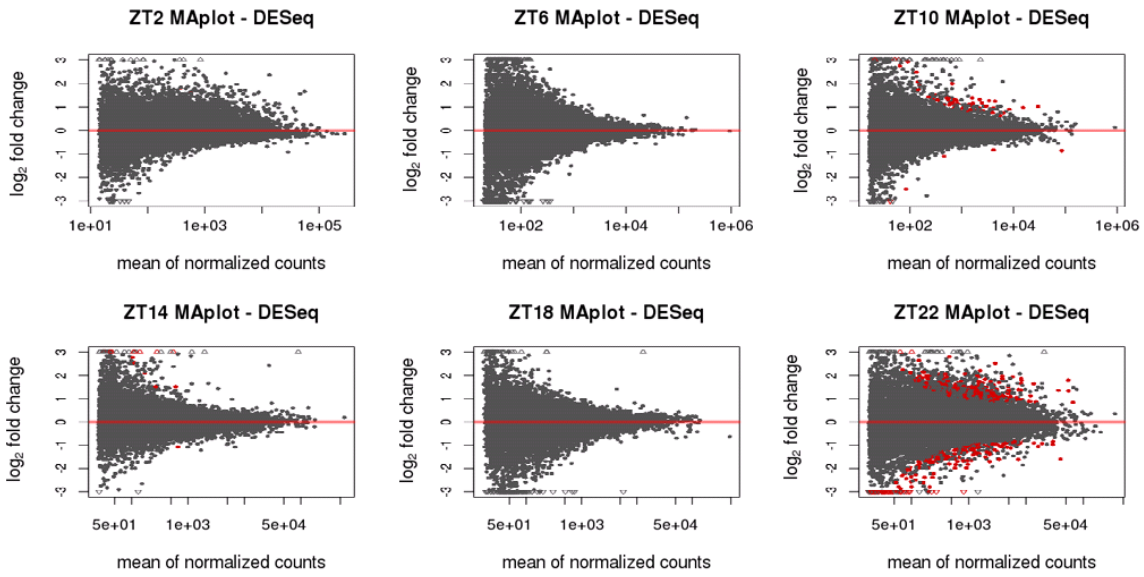


Figure 5.7. MA plots displaying DE genes for each time point assessed.
The log2 fold change between the genotypes is displayed for each gene assessed at each timepoint. Those significantly DE genes are coloured red. Horizontal lines represent no FC difference between the genotypes (red). Genes above the red line are upregulated in *Bdr* and those below it are downregulated.

this was the main filtering criterion) were visually represented using MA plots (Figure 5.7).

5.2.2 Identifying dysregulated co-expression modules

To identify dysregulated co-expression modules in *Bdr*, the userListEnrichment R function was utilised to identify the modules which displayed significantly more DE genes (as determined from the combinatorial DESeq/DESeq2 approach) than expected by chance. This statistical procedure is based upon the hypergeometric test and applies a Bonferroni correction for each pairwise comparison.

5.2.3 Disease enrichment

The userSetEnrichment R function as described above was also employed to determine whether specific co-expression modules were enriched in genes implicated in specific neurological diseases. Enrichment testing was performed for two lists separately.

The first list was based upon a recent schizophrenia genome-wide association study (2014). This study identified 108 loci which met genome-wide significance. These 108 loci intersected 350 human protein-coding genes based on GENCODE gene models which corresponded to 286 protein-coding genes in mouse to be tested for enrichment.

The second list comprised of schizophrenia risk genes, generated through a translational convergent functional genomics (CFG) approach (Ayalew et al., 2012). This method prioritised schizophrenia risk genes by integrating GWAS data with other genetic and gene expression studies in both humans and animal models. 153 schizophrenia risk genes with a CFG score of 3 or over were selected for enrichment testing.

5.2.4 Module preservation

To assess whether particular co-expression modules as identified in chapter 3 were preserved in the *Bdr* mouse model, the modulePreservation function in WGCNA was employed. This can determine whether there are differences in the organisation of co-expression networks between the genotypes, changes in connectivity patterns or whether there are species specific co-expression modules.

5.2.5 Activity monitoring

Locomotor activity of mice was determined by wheel-running and used to assess circadian behaviour and response to light. When mice are active they will often run on a wheel which is connected to a centralised computer which monitors the number of wheel rotations per minute. Age matched male mice were individually housed in large cages (44x26x12 cm) fixed with running wheels, maintained at constant temperature, with *ad libitum* food and water. The lighting conditions were set by 40W L.E.D tubes and set to a light intensity of 150lux as assessed by a lux meter. A variety of circadian paradigms were applied to the WT and *Bdr* mice by altering the lighting schedule while collecting wheel-running data to assess locomotor activity. Wheel running data was subsequently analysed using the Actimetrics Clocklab analysis software for Matlab. To test for specific deficiencies in the circadian system of the *Bdr* mutant, a number of circadian paradigms were applied (Figure 5.8,5.9) and a number of circadian behaviours were assessed (Figure 5.10, Table 5.1). A two-tailed student's t-test was applied to identify significant differences between the genotypes for each condition (two-sample t-test) and between each condition for each genotype (paired t-test).



Figure 5.8. Skeleton photoperiod circadian paradigm.
Bdr mice and littermate counterparts were first entrained to 15 days in 12:12LD before subject to a skeleton photoperiod. Instead of 12 hours of light, two 1-hour light pulses were applied at ZT0 and ZT11 to signify the start and end of the subjective day. Locomotor activity was assessed to determine whether there was a differential response to this paradigm between the genotypes.

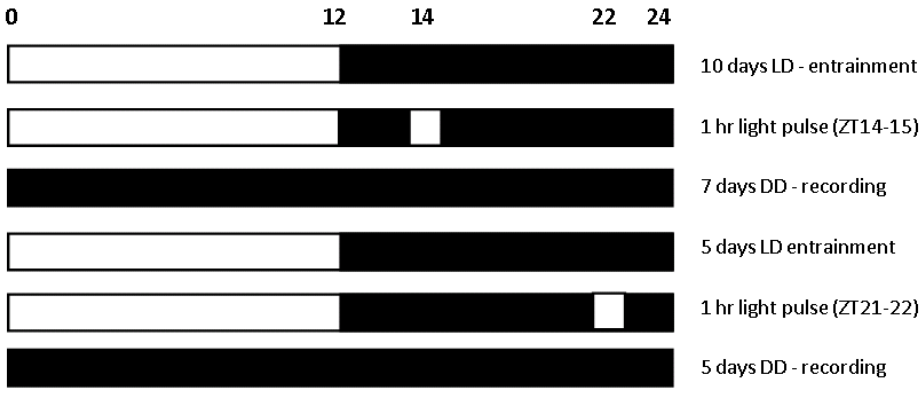


Figure 5.9. Phase-shifting circadian paradigm.
Bdr mice and littermate counterparts were entrained to 10 days in 12:12LD before subject to a 1-hour light pulse at either ZT14 or ZT22 in order to induce a phase-delay and a phase-advance respectively. After the 1-hour light pulse, the locomotor activity of mice was assessed in constant darkness for 7 days.

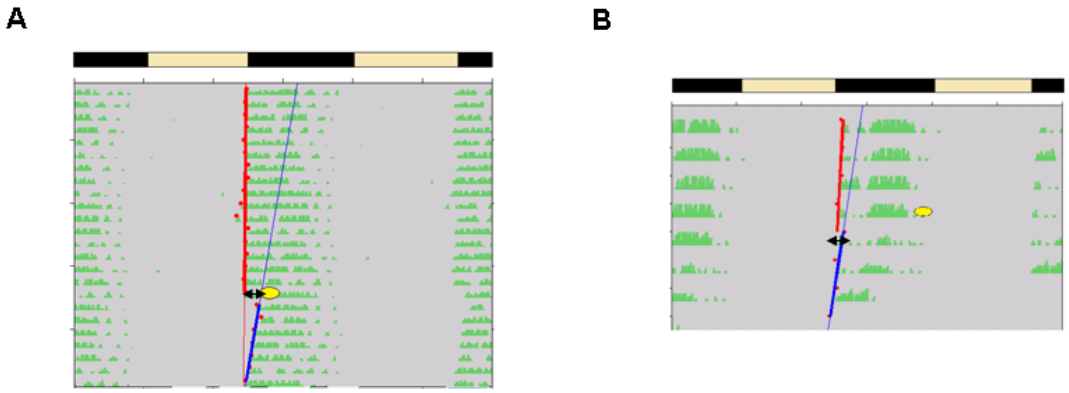


Figure 5.10. Phase-shifting circadian behavioural analysis.
Mice were housed in 12:12 LD conditions before the receipt of a light pulse and were subsequently released into constant darkness. A one-hour light pulse was applied at **(A)** ZT14 to induce a phase-delay and **(B)** ZT21 to induce a phase-advance. Regression lines were fitted to the activity onsets (time activity starts) preceding (red) and following (blue) the light pulse. The phase-shift — which is the amount of time between the regression lines — was then calculated. A phase-shift after a light pulse at ZT14 is the amount of time between the onset time in LD (red line) and the proposed onset time after the light pulse (determined from the blue line). The light pulse at ZT14 appeared to induce a ∼2-hour phase-shift, whereas in this example, the light pulse at ZT21 did not induce a shifting effect on circadian behaviour.

Circadian phenotype	Unit	Definition
Tau	Hours	Time it takes to complete a cycle of activity.
Mean activity onset	Clock time	Average clock time of activity onset, this is a robust indicator of entrainment to light. Note: lights on (ZT0) at 17.00 GMT and lights off (ZT12) at 5.00 GMT under 12:12 LD conditions.
Mean activity offset	Clock time	Average clock time of activity onset.
Activity length	Hours	Average length of activity period (time between mean activity offset and mean activity onset).
Daily activity	Counts/min	Average daily activity.
Relative light activity	%	Percentage of activity bouts that occur during the day.
Number of bouts	Bouts/day	A bout is defined as a period of activity (> 2 wheel revolutions per minute) > 10 minutes which is preceded and followed by a rest period of >10 minutes.
Phase advance	Hours	The time difference between the original and subsequent phase of entrainment following a light pulse at ZT22. Illustrated Figure 5.10.
Phase delay	Hours	The time difference between the original and subsequent phase of entrainment following a light pulse at ZT14. Illustrated Figure 5.10.

Table 5.1. Circadian behavioural parameters assessed between genotypes.

5.2.6 qRT-PCR

Validation of expression patterns and variation between the genotypes was assessed using Biomark qPCR as discussed in chapter 2 and 3. qPCR was performed on the same 36 samples sent for RNA-sequencing.

5.3 Results

5.3.1 Data exploration

To identify outliers or patterns of variation within the *Bdr* SCN RNA-seq data, PCA was applied on all 36 SCN samples from WT and *Bdr* mice (Figure 5.11). As expected, the *Bdr* samples clustered into the same batches as the WT samples they were processed with. Although WT and *Bdr* samples of the same timepoint clustered close together, there appeared to be increased heterogeneity within the *Bdr* samples.

The circadian wheel-running phenotype of the *Bdr* mice has been reported to be phase-advanced and fragmented (Figure 5.1) (Oliver et al., 2012). Genes which display circadian patterns of expression in the SCN of WT mice may therefore be phase-advanced or dysregulated in *Bdr*. To assess this hypothesis, the sinusoidally expressed genes in WT were compared between genotypes (Figure 5.12).

The cycling genes in the WT mouse appear to be largely unperturbed in *Bdr*, displaying no global changes in phase or strength of circadian oscillations as determined by basic data visualisation. To identify the specific genes which are dysregulated in *Bdr* a more detailed differential expression analysis approach upon all SCN expressed genes was employed.

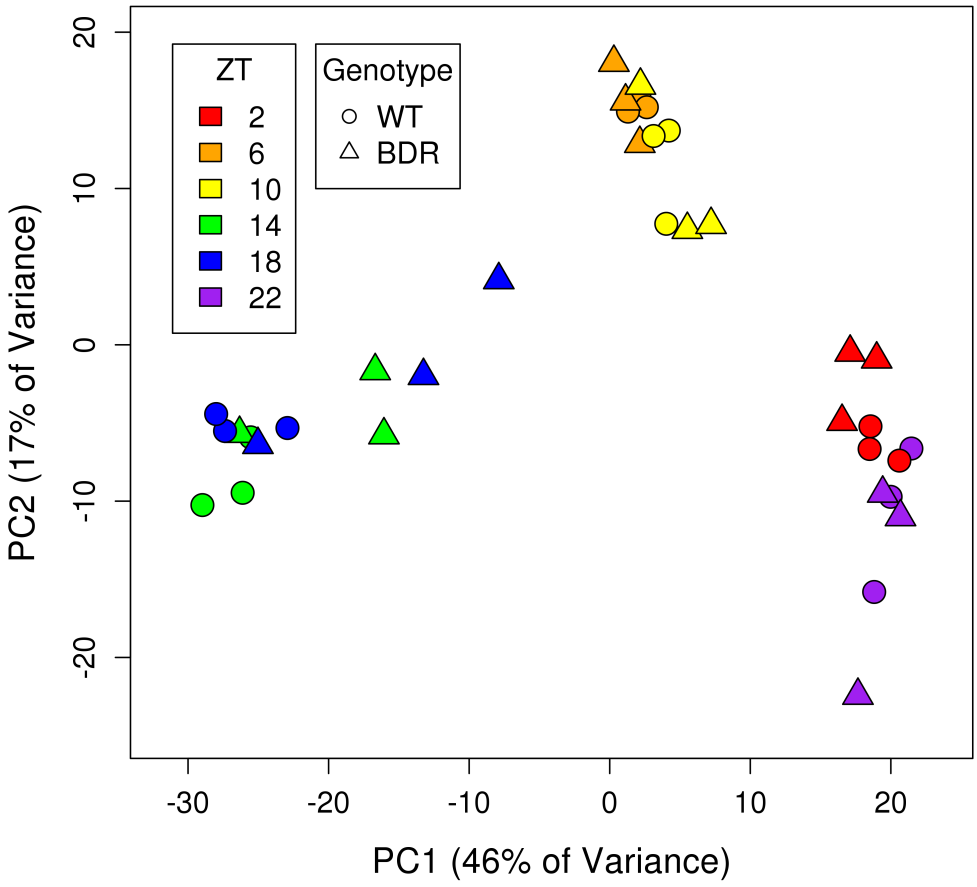


Figure 5.11. PCA plot of WT and *Bdr* SCN RNA-seq data points.
Principal component analysis (PCA) plot showing separation of RNA-seq data according to timepoint and sequencing batch as in Figure 2.5. In addition, PCA revealed only small differences between the WT and *Bdr* genotypes although there appears to be increased heterogeneity within the *Bdr* cohort.

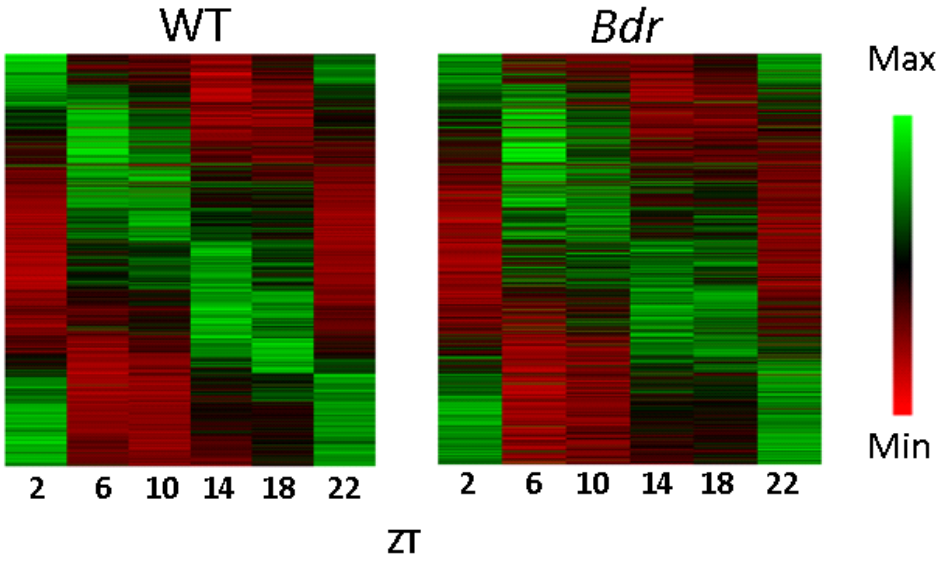


Figure 5.12. Cycling genes appear unperturbed in *Bdr* mice.
Heatmap showing the mean expression levels for the 4,569 significantly cycling genes (DESeq2 and JTK Cycle, q < 0.05) in the SCN across circadian time for both WT and *Bdr* mice. Genes are ordered by their peak-phase in WT.

5.3.2 Identification of differentially expressed genes

Using a combinatorial pairwise (DESeq) /GLM (DESeq2) comparison approach, a total of 374 genes were identified to be differentially expressed (DE) between genotypes for at least one timepoint (Table 5.2,5.3). The pairwise approach (DESeq padj <0.05) was used as the main filtering criterion because the GLM approach at a similar padj threshold filtered out many genes which appeared to be DE at a single timepoint but displayed increased variation at other timepoints. Therefore, a relaxed filtering approach (DESeq2 pval < 0.05) was used to remove only those transcripts which were called as DE using the pairwise approach but which displayed a very large variation over circadian time with only a small FC at the specified timepoint.

There appeared to be numerous genes (1,984) called as DE at ZT6 using the GLM approach even though there were none called as DE using the pairwise approach (Table 5.2). This likely reflects the increased expression variation at this timepoint either due to biological variation or the presence of a sample outlier, which prevented it being called as DE using the pairwise approach (Figure 5.6). The GLM approach makes the assumption that variation should be equal across all timepoints which may overestimate DE genes at those times where expression variation is greater. Gene expression variation has been shown to differ over developmental time and therefore variation may also vary across circadian time (Garfield et al., 2013).

To identify the processes that may be differentially expressed, GO enrichment analysis was performed on these 374 genes. However, there was no significant enrichment for any GO or KEGG annotations.

	ZT2	ZT6	ZT10	ZT14	ZT18	ZT22	Total
PW - DESeq (padj <0.05)	2	0	60	16	0	449	507
GLM - DESeq2 (pval <0.05)	314	1984	916	372	903	1961	3606
Union	0	0	47	5	0	330	374

Table 5.2. Number of differentially expressed genes.
Differentially expressed genes were identified using a pairwise (PW; DESeq)/ general linear model (GLM; DESeq2) combinatorial approach. Only those genes which were identified through both approaches at a particular timepoint were considered as DE. In total, 374 genes were called as DE between genotypes.

Interestingly, a very large proportion (88%) of the 374 combined genes DE between the genotypes appeared to be DE at ZT22 (Table 5.2). Visualisation of the circadian expression patterns of these genes for both genotypes revealed that many possessed a twin-peaking pattern in WT (Figure 5.13).

Gene ID	Gene Name	log2 FC	p.adj
Mobp	myelin-associated oligodendrocytic basic protein	1.24	7.55E-09
Rbfox1	RNA binding protein, fox-1 homolog (C. elegans) 1	1.31	5.33E-08
Snca	synuclein, alpha	1.26	1.64E-06
Baiap2	brain-specific angiogenesis inhibitor 1-associated protein 2	1.15	2.47E-05
4921507P07Rik	RIKEN cDNA 4921507P07 gene	-5.84	0.00011
Asb11	ankyrin repeat and SOCS box-containing 11	4.26	0.00011
Chn1	chimerin (chimaerin) 1	0.90	0.00011
Mog	myelin oligodendrocyte glycoprotein	1.30	0.00013
Plp1	proteolipid protein (myelin) 1	0.97	0.00056
Thy1	thymus cell antigen 1, theta	1.05	0.00056

Gene ID	Gene Name	log2 FC	p.adj
Col8a2	collagen, type VIII, alpha 2	2.78	0.00311
Abca4	ATP-binding cassette, sub-family A (ABC1), member 4	2.10	0.00441
Serpinb1b	serine (or cysteine) peptidase inhibitor, clade B, member 1b	2.54	0.00441
Slc47a1	solute carrier family 47, member 1	5.03	0.03402
Egflam	EGF-like, fibronectin type III and laminin G domains	1.52	0.03402

Gene ID	Gene Name	log2 FC	p.adj
Foxp2	forkhead box P2	-2.97	2.64E-06
Mbp	myelin basic protein	1.80	4.05E-06
Slc25a18	solute carrier family 25 (mitochondrial carrier), member 18	1.51	2.93E-05
Rab33b	RAB33B, member of RAS oncogene family	1.61	5.16E-05
Tenm2	teneurin transmembrane protein 2	-1.67	9.39E-05
Fbxo8	F-box protein 8	1.74	0.00010
Kif5a	kinesin family member 5A	1.36	0.00014
Rngtt	RNA guanylyltransferase and 5-phosphatase	1.52	0.00016
Klhdc8b	kelch domain containing 8B	1.51	0.00017
2410002F23Rik	RIKEN cDNA 2410002F23 gene	1.41	0.00020

Table 5.3. Examples of differentially expressed genes.
Top 10 DE genes from **(A)** ZT10, **(B)** ZT14 and **(C)** ZT22. Genes ordered by DESeq padj value. Note: only 5 genes were called as DE at ZT14.

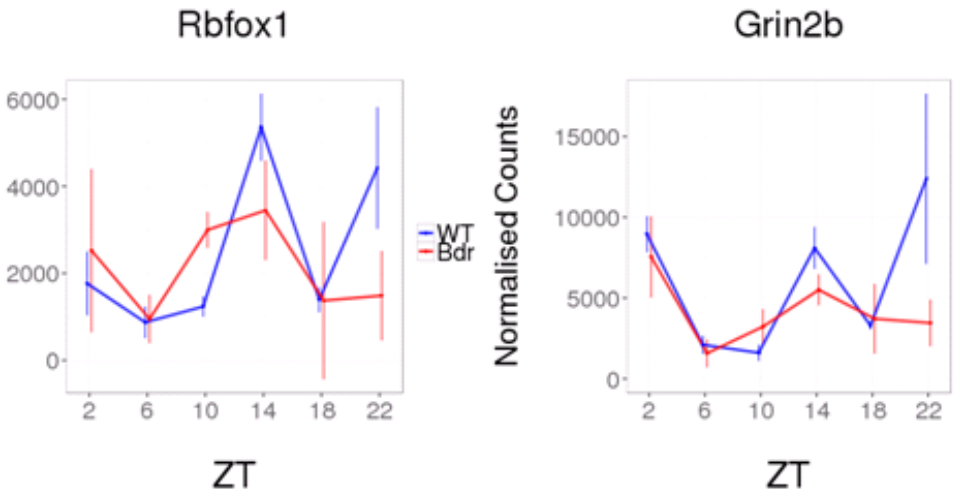


Figure 5.13. Examples of twin-peaking genes differentially expressed at ZT22.
The mean expression levels (± SD) of two genes (*Rbfox1* and *Grin2b*) for both WT (blue) and *Bdr* (red) mice over 24 hours. These genes displayed significant differential expression between the genotypes at ZT22. These genes display a twin peaking pattern in WT (peaks at ZT14 and ZT22) and appear to show systematic expression changes in *Bdr* (i.e. downregulated at ZT14 & ZT22, upregulated at ZT10).

5.3.3 Systematic dysregulation of the twin peaking transcriptional profile

To determine whether the twin-peaking, or other, co-expression modules were systematically disrupted in *Bdr*, the WGCNA co-expression modules (as identified in chapter 3) were tested for enrichment of the 374 DE genes. There was significant enrichment in four co-expression modules. The most enriched co-expression module was the lightsteelblue1 (twin-peaking) module followed by darkgreen, lightcyan and saddlebrown (Table 5.4).

As described in chapter 3, the twin-peaking co-expression module displayed peaks at ZT14 and ZT22 in wild-type mice. However, 57 genes present in this twin-peaking module were called as differentially expressed in at least one timepoint in *Bdr*. Visualisation of the expression pattern of these 57 genes in both the WT and *Bdr* mice over circadian time revealed that the peaks in expression in *Bdr* occur at around ZT2 and ZT12, instead of at ZT14 and ZT22 as observed for the wild-type (Figure 5.14A).

Comparing the mean expression values at each timepoint between genotypes for these 57 genes, it was clear that the largest transcriptomic differences between the genotypes were at ZT22 (where expression was higher in WT) and ZT10 (where expression was higher in *Bdr*) (Figure 5.14A). To assess whether this pattern of dysregulation was conserved across the entire module of 766 genes, the mean expression for genes in this module was displayed for both WT and *Bdr* over circadian time. Although only 57 twin-peaking genes were called as differentially expressed between the genotypes, it appears as though the entire module of 766 genes displayed this pattern of dysregulation, where expression was downregulated in *Bdr* at ZT22 but upregulated at ZT10 (Figure 5.14B). The top 20 twin-peaking genes based upon module membership can be viewed in Table 3.2. Importantly, measuring the circadian expression pattern of 17 twin-peaking genes through

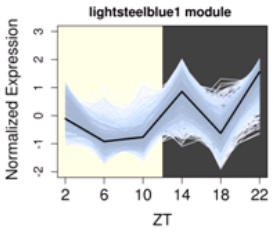
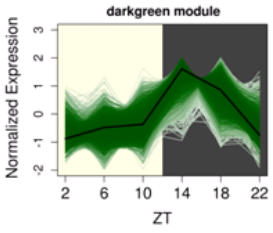
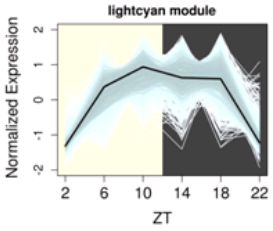
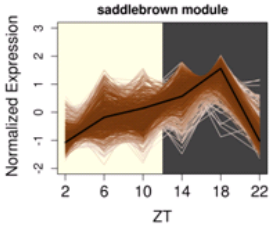
Module	Expression Pattern	Module size	# of DE genes	Enrichment	qval
lightsteelblue1		766	57	3.8	2.07E-17
darkgreen		2,031	77	1.9	2.14E-8
lightcyan		1,131	46	2.1	2.88E-6
saddlebrown		513	22	2.2	6.31E-7

Table 5.4. Co-expression modules enriched for differentially expressed genes.
Four of the 23 co-expression modules identified by WGCNA (Chapter 3) displayed significant enrichment ($q < 0.05$) for DE genes. Note: lightsteelblue1 represents the twin-peaking module.

qPCR in both WT and *Bdr* SCN samples confirmed the same pattern of dysregulation observed in the RNA-seq expression data (Figure 5.14C). These results show that differential expression analysis combined with co-expression network analysis could successfully identify specific transcriptional networks which are disrupted in *Bdr* and may underlie the circadian and behavioural phenotypes observed in this mouse model.

The three additional modules (lightcyan, darkgreen and saddlebrown) significantly enriched for DE genes generally peaked in expression between ZT6-ZT18 in WT, with lowest expression at ZT22/ZT2 (Table 5.4, Figure 5.15). The circadian expression profile in *Bdr* and the FC differences between genotypes for these genes revealed a tendency for expression to be greater at ZT22 for *Bdr* mice (Figure 5.15).

For these three specific modules, when considering the FC expression differences between genotypes for the whole co-expression module, these changes were less pronounced than when considering only those genes called as DE (Figure 5.15). This was a strong contrast to the dysregulation in the twin-peaking module which showed strong dysregulation across the entire co-expression module (Figure 5.15). Furthermore, these transcriptomic differences between the genotypes for these co-expression modules could not be confirmed by qPCR – the genes part of these co-expression modules which were measured by qPCR did not display systematically higher expression in *Bdr* at ZT22, as was described by the RNA-seq data (Figure 5.15). Together, these results suggest that these additional three co-expression modules were not systematically dysregulated and that, for genes in these modules, any transcriptomic differences at ZT22 were likely due to technical biases.

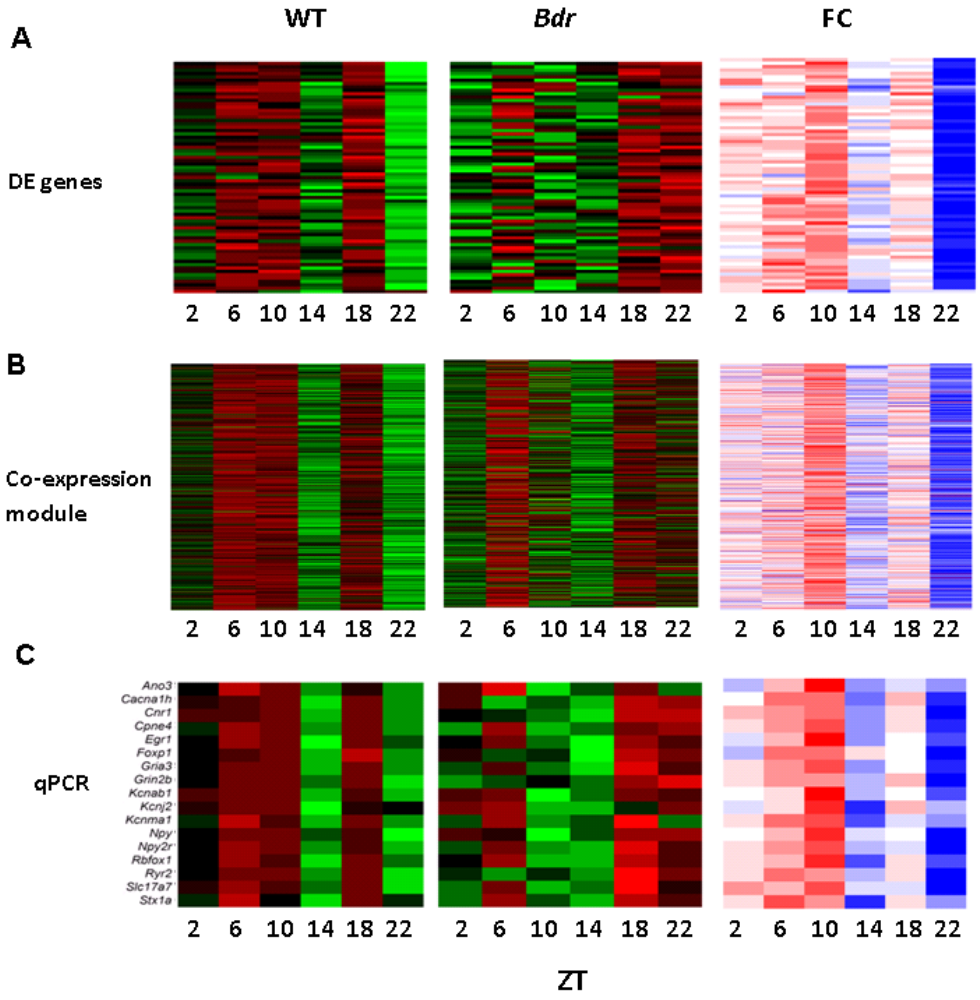


Figure 5.14. Dysregulation of the twin-peaking co-expression module.
First two columns consist of heatmaps which represent the relative mean expression for a particular gene across circadian time for WT and *Bdr* mice (green — high expression; red — low expression). The third column represents the expression difference between the genotypes at each timepoint for the same genes. If coloured blue, then expression is greater in WT, if coloured red, then expression is greater in *Bdr*. Displayed is the relative expression for: **(A)** the 57 DE genes part of the twin-peaking module according to the sequencing data, **(B)** all 766 genes part of the twin-peaking co-expression module, **(C)** 17 twin-peaking genes as determined by qPCR.

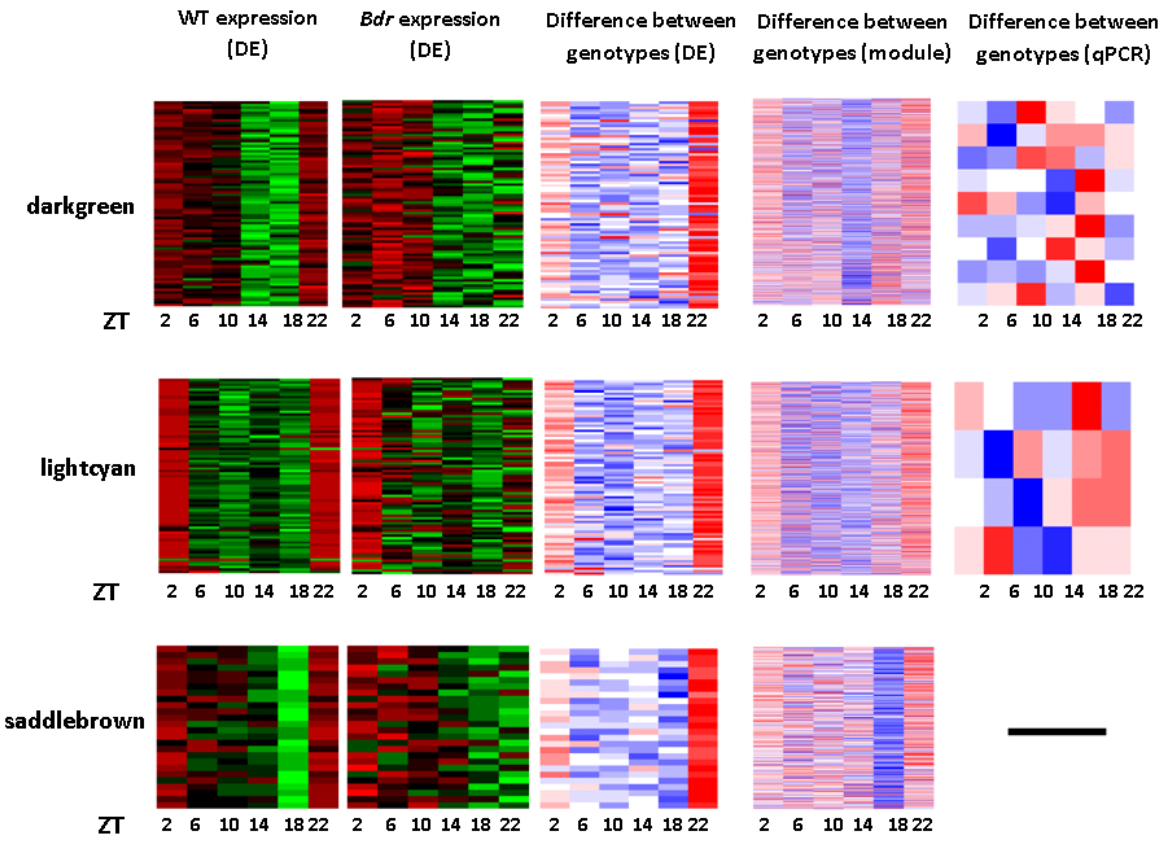

WT expression (DE)
Bdr expression (DE)
Difference between genotypes (DE)
Difference between genotypes (module)
Difference between genotypes (qPCR)
darkgreen
lightcyan
saddlebrown
ZT
2 6 10 14 18 22

Figure 5.15. Dysregulation of additional co-expression modules.
Heatmaps of genes from the three additional co-expression modules enriched with DE genes, where each row represents a different co-expression module (as identified in chapter 3). The first two columns consist of heatmaps which represent the relative mean expression for a particular gene across circadian time for WT (first column) and *Bdr* (second column) mice (green — high expression; red — low expression). The third to fifth columns represent the expression difference between the genotypes at each timepoint. If the expression is greater in WT than *Bdr* at a particular timepoint, the panel is coloured blue. If the expression is greater in *Bdr* than WT, the panel is coloured red. The third column displays the genes which have been called as differentially expressed in that co-expression module – all three modules show greater expression at ZT22 in *Bdr* mice. The fourth column displays *all* genes in that co-expression module. Although these modules tend to show greater expression at ZT22 in *Bdr*, the difference was much more subtle than those DE genes (Column 3). The fifth column represents the difference in expression between the genotypes as determined by qPCR. These genes surveyed did not show higher expression in *Bdr* at ZT22 suggesting that reported transcriptomic differences at ZT22 were likely due to technical biases. Note: no genes from the saddlebrown module were assessed by qPCR as the circadian expression pattern for this module was considered to be driven by an outlier expression (Figure 3.4) and did not possess any circadian relevant genes.

5.3.4 Comparison to previous transcriptomic analysis of *Bdr*

Previous transcriptomic analysis of the SCN in WT and *Bdr* mice determined that the SCN output neuropeptide AVP was phase-advanced by ∼4 hours. Although *Avp* was not called as DE by RNA-seq in this chapter, visualisation of *Avp* expression suggests that it may be phase-advanced (by ∼ 2 hours) but not to the extent described previously (Figure 5.16) (Oliver et al., 2012). In addition, a ∼2-hour phase-advance in gene expression, according to RNA-seq, was observed for PROK2, an SCN output molecule that regulates circadian locomotor behaviour with a potential role in sleep-wake regulation (Zhou and Cheng, 2005, Cheng et al., 2002).

In the previous study investigating the *Bdr* mutant, the clock genes surveyed using qPCR appeared to display normal expression patterns in *Bdr* (Oliver et al., 2012). However, the clock gene *Rora*, which was not assessed in previous studies, appeared to be phase-advanced according to the RNA-seq expression data in a similar manner to both *Avp* and *Prok2*. Therefore, this gene may also possess a role in the SCN output pathway. Although these three genes were not called as statistically differentially expressed, their general shifting of expression was validated by qPCR (Figure 5.16).

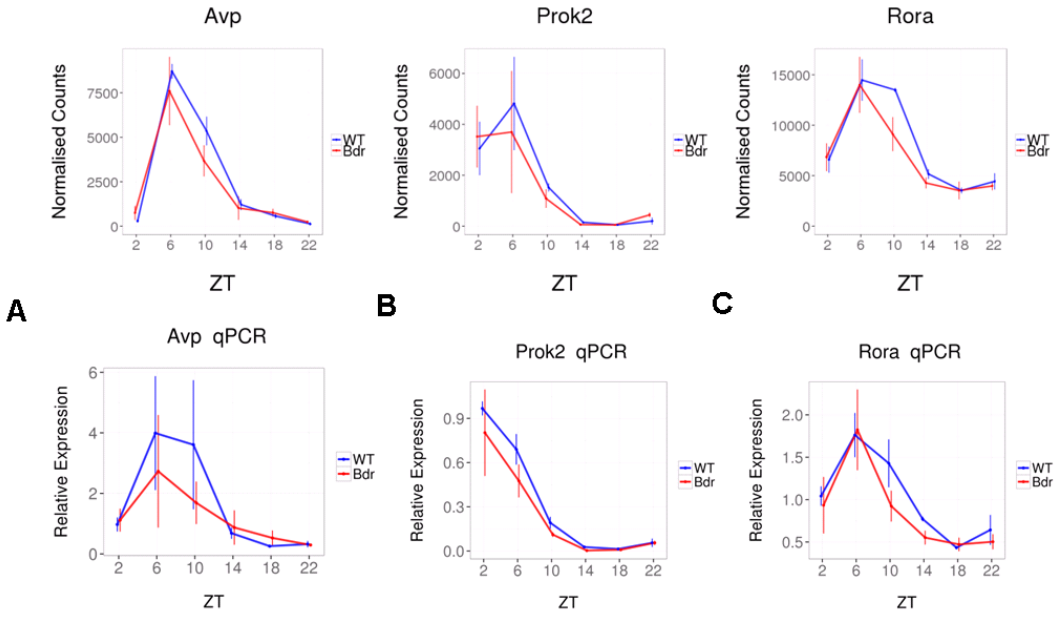


Figure 5.16. Potential phase-advance of SCN-output and circadian relevant gene expression. The expression patterns determined by RNA-seq (top row) and qPCR (bottom row) showed a slight phase-advance in *Bdr* of the SCN output genes **(A)** *Avp*, **(B)** *Prok2* and for the core clock gene **(C)** *Rora*.

5.3.5 Module preservation

In addition to identifying differentially expressed genes, properties of network structure (co-expression modules) can be assessed between genotypes. To detect changes in network structure, the preservation of all modules identified in chapter 3 was assessed between genotypes (Figure 5.17). Despite the systematic dysregulation of the twin-peaking module, the module appeared to be well preserved between the genotypes. Some modules did appear to display significant (Z<2) intramodular connectivity structure. Although this low intramodular connectivity may allude to *Bdr*-specific coexpression modules, these modules may not be reproducible in the initial WT data therefore explaining the poor preservation. For instance, all the modules Z<2 (the dashed blue line) displayed circadian expression patterns which appeared to be driven by outlier expression (Figure 3.3,3.4). Therefore it is unlikely that these expression patterns represent true circadian co-expression modules.

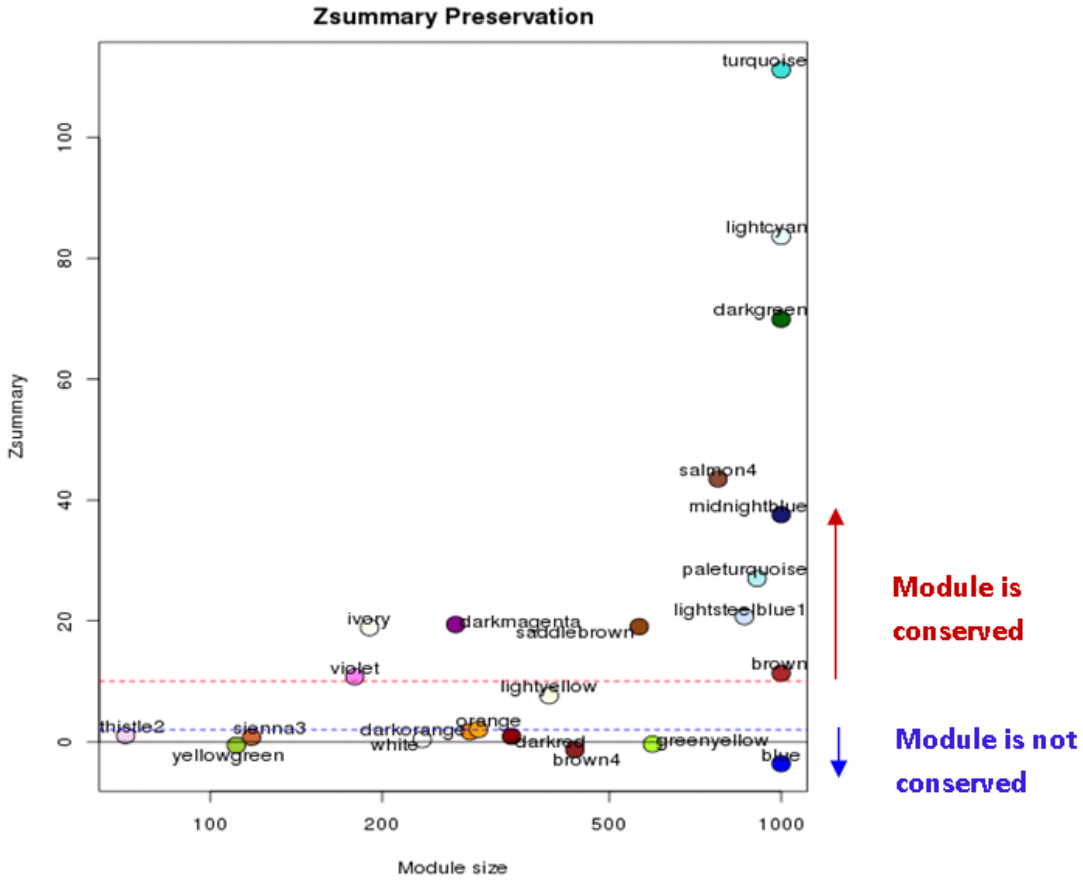


Figure 5.17. Preservation of WT co-expression modules in *Bdr* mice.
Plot shows the module size (number of genes) and Zsummary score for all 23 co-expression modules identified through WGCNA using data from the WT mice in chapter 3. To determine whether a module was preserved in the *Bdr* mouse, a Zsummary preservation score was generated. If the score was <2 (dashed blue line), then there was no evidence that the module is conserved. Modules with a Zsummary score >10 (dashed red line) show strong evidence of preservation. The twin-peaking (lightsteelblue1) module was preserved in *Bdr*.

5.3.6 Disease enrichment of schizophrenia implicated genes

To determine whether specific co-expression modules may have implications in particular neurological diseases, an enrichment test was used to determine whether any co-expression modules identified in chapter 3 were enriched for genes implicated in schizophrenia.

When testing for genes which are implicated to schizophrenia through GWAS, no co-expression module possessed a significant enrichment (Table 5.5). However, when considering genes implicated in schizophrenia through a convergence functional genomics (CFG) approach, a significant enrichment was found in the twin-peaking (lightsteelblue1) module (Table 5.6). Overall this data suggests that the dysregulated twin peaking co-expression module is implicated in schizophrenia, although there may be reasons why schizophrenia implicated genes identified through GWAS and CFG approaches show differing levels of enrichment in the twin-peaking co-expression module as discussed later in this chapter.

InputCategories	NumOverlap	Pvalues	CorrectedPvalues
lightcyan	22	0.056	1
lightsteelblue1	17	0.057	1
sienna3	4	0.068	1
darkgreen	37	0.070	1
midnightblue	60	0.077	1

Table 5.5. Module enrichment for schizophrenia GWAS linked genes.
Table displays the top 5 enrichments.

InputCategories	NumOverlap	Pvalues	CorrectedPvalues
lightsteelblue1	19	7.8 x 10-06	0.00018

Table 5.6. Module enrichment for schizophrenia risk genes determined by the CFG approach. Only the modules with a significant (< 0.05) corrected p value score were displayed

5.3.7 Circadian behavioural wheel-running screens

As discussed in chapter 3, the twin-peaking co-expression module may represent a transcriptional gating mechanism important in clock resetting mechanisms. At the times when these twin-peaking genes are lowly expressed, the clock is resistant to phase-shifts, whereas at times these genes display high levels of gene expression, light can activate this phase-shifting pathway (Ding et al., 1994, Gillette and Mitchell, 2002, Iyer et al., 2014). As the twin-peaking expression pattern is dysregulated in the *Bdr* mouse model, it may underlie the circadian defects observed in *Bdr*. To further dissect the circadian deficits in the *Bdr* mouse model, and determine whether they are a result of the dysregulated twin-peaking co-expression module, a number of circadian paradigms were applied to the *Bdr* mutants and their WT littermates (Figure 5.8,5.9).

If the twin peaking expression pattern is responsible for generating susceptibility windows which allow the clock to shift in response to light impulses, then dysregulation of this expression pattern in *Bdr* may affect the time of day these windows are present. If the susceptibility windows are present during subjective day, then daytime light is therefore likely to send signals to the SCN inducing both a phase-advance (as light impulse at ZT10) and a phase-delay (as a light impulse at ZT2) (Figure 3.19). This may result in

the fragmented rest/activity profiles observed in many *Bdr* mice, as the SCN is receiving contradictory signals and is unable to provide a robust output. To assess whether light during the subjective day was causing specific circadian deficits in *Bdr*, WT and *Bdr* mice were subjected to a skeleton photoperiod (SPP). The reduced light during subjective night may prevent the fragmentation and earlier phase of entrainment reported in *Bdr* mice.

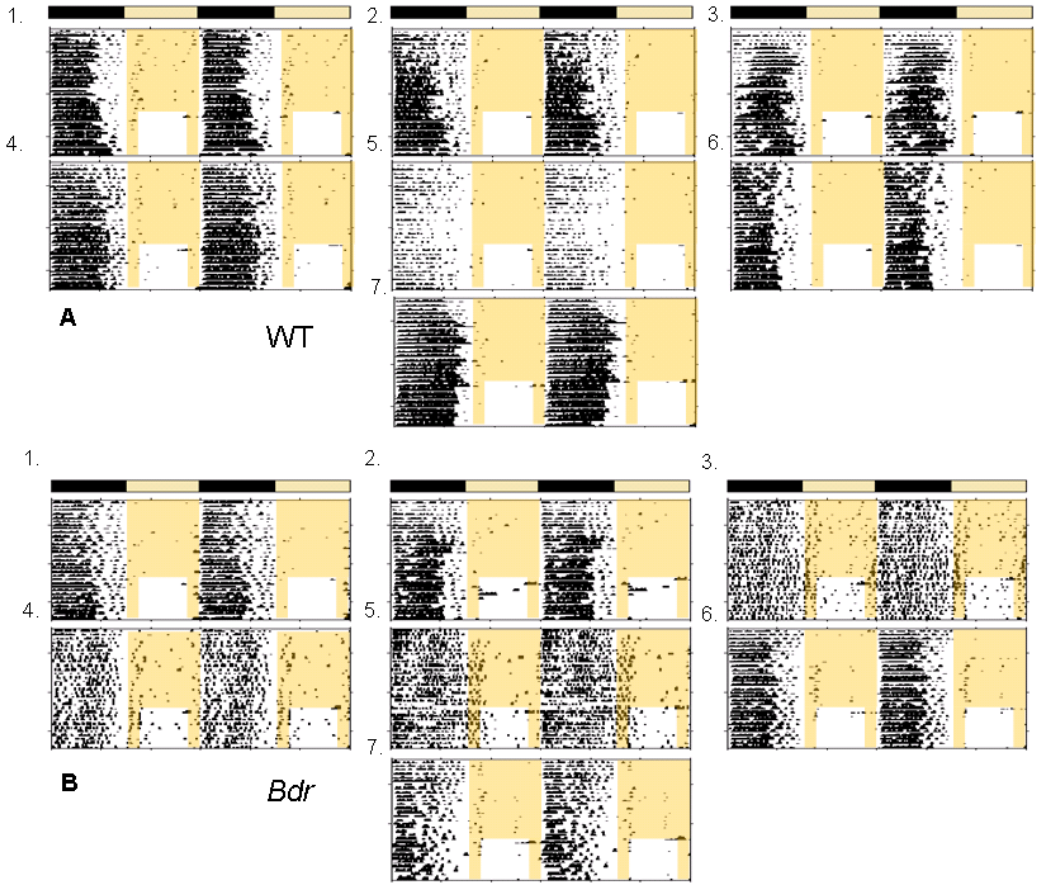


Figure 5.18. Actograms of WT and *Bdr* mice entrained to differing light schedules. Actograms of the **(A)** WT (n=7) and **(B)** *Bdr* (n=7) mice under both LD and a skeleton photoperiod. Mice were entrained to a 12:12 light dark cycle for 10 days, before subject to a skeleton photoperiod (SPP), which represents a 1-hour light pulse at ZT0 and ZT11, signifying the start and end of the light phase.

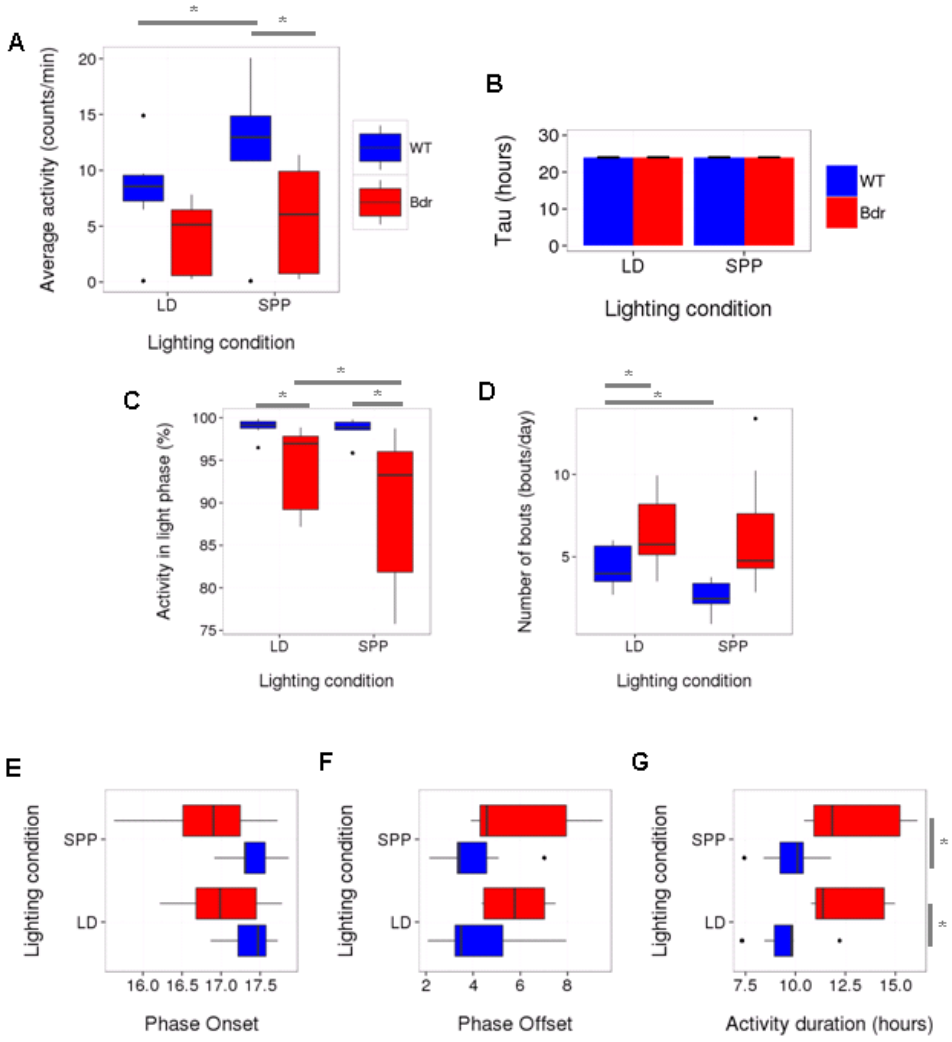


Figure 5.19. *Bdr* mice display a number of circadian deficits in LD which were not altered under a SPP.
A number of circadian behaviours were assessed between WT (blue) and *Bdr* (red) mice housed under both a 12:12 LD cycle and a skeleton photoperiod (SPP). The specific behavioural parameters assessed were **(A)** average activity, **(B)** tau (length of circadian period), **(C)** activity in light phase, **(D)** number of bouts, **(E)** phase onset, **(F)** phase offset and **(G)** activity duration. Significant (*pval < 0.05) differences indicated by grey line.

Both WT and *Bdr* mice were able to entrain to the SPP as they possessed similar values for tau under both LD and SPP lighting conditions and did not display free-running circadian behaviour under SPP conditions as visualised by the actograms (Figure 5.18, Figure 5.19B). It was also determined that *Bdr* mice displayed reduced activity compared to their littermate counterparts, in line with previous findings (Figure 5.19A). As previously described, *Bdr* mice displayed a number of circadian deficits under LD, such as possessing both a fragmented and phase-advanced rest/activity profiles (Oliver et al., 2012). Although not all comparisons were significant (*pval < 0.05), the *Bdr* mice assessed in this study displayed trends for earlier phase onset, later phase onset, greater length of time of activity, increased activity in the light phase and an increased number of bouts, which together suggest increased activity fragmentation, which can also be visualised through the actograms (Figure 5.18,5.19). However, these circadian deficits did not appear to be exacerbated or alleviated under a skeleton photoperiod (Figure 5.18,5.19). These results suggest that light from ZT1 – ZT11 was unlikely to be causing the circadian deficits observed in the *Bdr* mice. This hypothesis was proposed because the dysregulated twin-peaking transcriptional profile in *Bdr* may have generated phase-shifting susceptibility windows during the subjective day, where light could provoke phase-shifting cascades in the SCN.

The dysregulated twin-peaking expression profile in *Bdr* may affect the timing of the phase-shifting susceptibility windows important for clock-resetting and therefore affect the ability of *Bdr* mice to phase-shift in response to light. For instance, the decrease in expression for twin-peaking genes at ZT22 may prevent light-induced phase-advances in *Bdr*. To assess the ability of *Bdr* mice to respond to light at particular times of day, 1-

hour light pulses were applied at ZT14 and ZT22 and locomotor activity was subsequently measured in DD for the following 5 days (Figure 5.20).

As expected, on the application of a light pulse at ZT14, both WT and *Bdr* mice displayed a phase-delay in activity onset (Figure 5.21). The extent of the phase-delay did not significantly differ between genotypes. However, a light impulse at ZT22 was unable to induce a phase-advance in either WT or *Bdr* mice (Figure 5.21). This suggests that while the *Bdr* mice show a similar ability to respond to a phase-delaying light pulse to their littermate counterparts, the light pulse at ZT22 was unable to induce a phase-shift in either genotype, preventing the identification of *Bdr* specific circadian deficits.

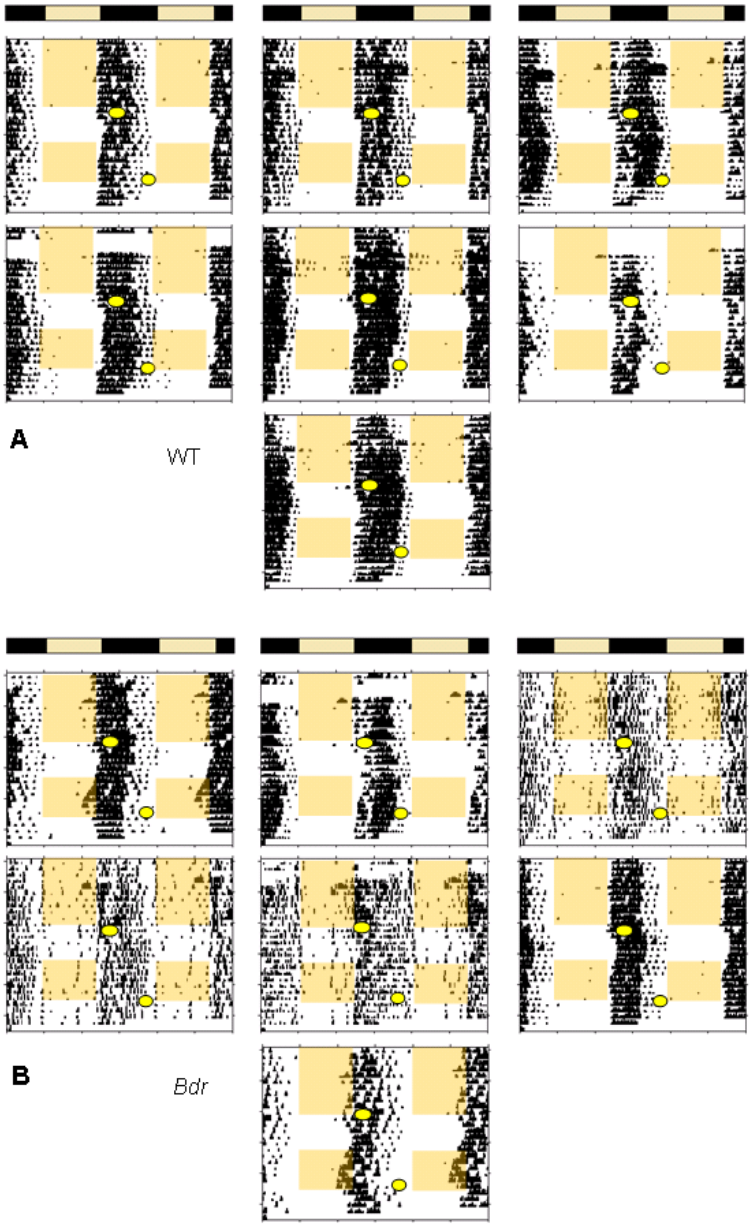


Figure 5.20. Actograms of mice subject to phase-shifting light pulses.
Actograms for **(A)** WT (n=7) and **(B)** *Bdr* (n=7) mice subject to phase-shifting light pulses. Mice were entrained to a 12:12 light dark cycle and then subject to a 1-hour light pulse (yellow oval) at either ZT14 or ZT22. The locomotor activity in DD was then measured to assess the strength of the resultant phase-shift.

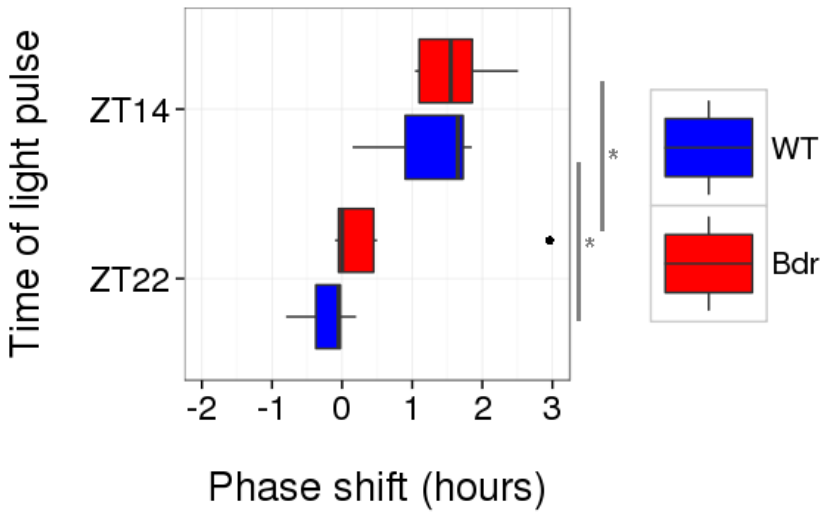


Figure 5.21. *Bdr* mice did not show evidence of altered phase-advance.
WT and *Bdr* mice displayed a ∼1.5-hour phase-shift (phase-delay) when subject to a light pulse at ZT14. A light pulse at ZT22 was not able to induce a phase-shift (phase-advance) in either WT or *Bdr* mice. Significant (*pval < 0.05) differences indicated by grey line.

5.4 Discussion

5.4.1 Identification of differentially expressed genes in *Bdr*

To investigate the specific processes which may underpin the circadian and neurological disturbances observed in the *Bdr* mouse model, SCN samples were extracted using LCM from *Bdr* in parallel with WT littermates (at ZT2, 6, 10, 14, 18 and 22). RNA from these samples was sequenced and used to generate expression values which were subsequently compared against the WT transcriptome (Chapter 3) to identify differentially expressed genes and processes.

Previous analysis of the *Bdr* mouse model utilised microarray technology to measure the transcriptome of SCN punches extracted from WT and *Bdr* mice at 3 timepoints (ZT6, ZT11 and ZT13) (Oliver et al., 2012). However, this analysis revealed only a very small number of protein coding genes to be differentially expressed in *Bdr*. Using a combinatorial DESeq2 /DESeq pairwise comparison approach a total of 374 genes were identified to be differentially expressed (DE) between genotypes (Table 5.2).

Interestingly, the majority of these transcriptomic differences occurred at ZT22, a time which was not surveyed in the previous microarray study (Oliver et al., 2012). A number of these dysregulated genes displayed a twin-peaking pattern in the WT mouse SCN.

DE genes were also enriched in other co-expression modules (such as darkgreen or lightcyan) but these expression differences could not be validated using qPCR. This suggests that there must be some sequencing specific bias between the genotypes, particularly at ZT22/ZT2, because these timepoints incorrectly show the upregulation of genes in *Bdr*.

This systematic bias between the genotypes occurring in the ZT22/ZT2 batch may be introduced during library preparation before sequencing. In addition, an assumption of DESeq2, used to normalise expression levels and determine differentially expressed genes, is that most genes do not significantly change their expression levels over different conditions (Love et al., 2014). As at the entire twin-peaking module was dysregulated, with downregulated expression at ZT22 in *Bdr*, normalisation may have caused a subtle increase in the other genes at this timepoint.

Nevertheless, there were still many other genes which did display transcriptomic differences at ZT22 which were validated by qPCR. As *Bdr* mice have been reported to be phase-advanced relative to their WT counterparts, they may be more likely to be asleep at ZT22, unlike WT mice which may sleep at ∼ZT23 confirmed by video tracking (Oliver et al., 2012). This difference in sleep state may explain the large transcriptomic differences between the genotypes at this time (Cirelli, 2005).

In addition to identifying a large number of DE genes between the genotypes, an increased temporal resolution, compared to the previous microarray study conducted at three timepoints, was important to help characterise the dynamics of circadian disturbances observed in *Bdr*. Even though statistically significant transcriptomic differences were observed at ZT22 for the twin-peaking genes, by considering the expression levels at the other 5 timepoints, it can be speculated that these peaks in expression were shifting to other timepoints and perhaps were not just simply downregulated at ZT22.

Previous transcriptomic analysis of the SCN in WT and *Bdr* mice determined that expression of the transcripts encoding the SCN output neuropeptide AVP was phase-advanced by ∼4 hours. Although *Avp* was not called as DE by RNA-seq in this chapter,

its expression along with that of the SCN output molecule PROK2 appears to be phase-advanced (by $\sim$ 2 hours) but not to the extent described previously (Figure 5.16) (Oliver et al., 2012). AVP has been shown to increase PROK2 expression through the AVPR1A receptor, which implies that the phase-advance of *Prok2* may be due to AVP signalling (Li et al., 2009). A possible reason for the differing degrees of phase-advance of *Avp* expression in *Bdr* may be due to housing conditions. In the previous microarray experiment, mice were group housed, meaning when a mouse wakes early, it may also wake up the other *Bdr* mice, causing all mice to possess this phase-advance profile (Oliver et al., 2012). As there is a great degree of heterogeneity in phase of entrainment for *Bdr* mice, these effects may be dampened when pooling SCN samples from singly housed mice.

In summary, RNA-seq of LCM SCN samples extracted from WT and *Bdr* mice at six timepoints allowed the identification of 374 genes DE between the genotypes, considerably more than the number identified through microarray profiling. As these transcriptomic differences were more prevalent at late subjective night, increased temporal resolution was essential in identifying these differences. Furthermore, increased resolution allowed a detailed description of the changes occurring for specific co-expression patterns. Instead of identifying a set of genes which were downregulated in *Bdr* at ZT22, expression data from other timepoints allowed us to identify a temporal shift of the peaks in expression. In addition, although not significant, the sequencing study revealed a small phase-advance of SCN output molecules, such as *Avp* and *Prok2*, which may induce the advance in locomotor activity as suggested in the previous array study (Oliver et al., 2012).

5.4.2 Disruption of the twin-peaking co-expression module may affect circadian function

Despite sinusoidally expressed genes appearing largely concordant between genotypes, the twin-peaking genes displayed systematic dysregulation in *Bdr* mice. Instead of expression peaks at ZT14 and ZT22 as in the WT SCN, the expression of these genes peaked at ZT2 and ZT10 in mutants. The expression of these genes may have shifted in phase by 12 hours or alternatively, the peak at ZT14 has shifted 2 hours earlier to ZT12 in *Bdr* and the peak at ZT22 has shifted 4 hours later to ZT2 (Figure 5.22).

As the peaks of the twin-peaking transcriptional module in WT mice coincided with the times in which light has been shown to induce a phase-shift, this module was implicated in the resetting of the circadian clock (Chapter 3). Therefore, the shifting of these peaks in expression may underlie the circadian defects observed in the *Bdr* mouse model (Figure 5.22). If the normal twin-peaking expression pattern is responsible for generating the susceptibility windows which allow the clock to shift to light impulses, then in *Bdr* mice, these windows will exist at times during the subjective day (Figure 5.22). Daytime light may therefore send signals to the SCN inducing both a phase-advance (light impulse at ZT10) and a phase-delay (light impulse at ZT2). This may result in the fragmented rest/activity profiles observed in many *Bdr* mice, as the SCN is receiving contradictory signals and is unable to provide a robust output.

The shifting of these peaks in expression may also explain why some *Bdr* mice possess a phase-advanced rest/activity profile (Figure 5.1). WT mice start their activity at ZT12 because the phase -advance/-delay signals received by the SCN are both negligible and opposing. The altered PRC in *Bdr* may cause an earlier phase of entrainment as this

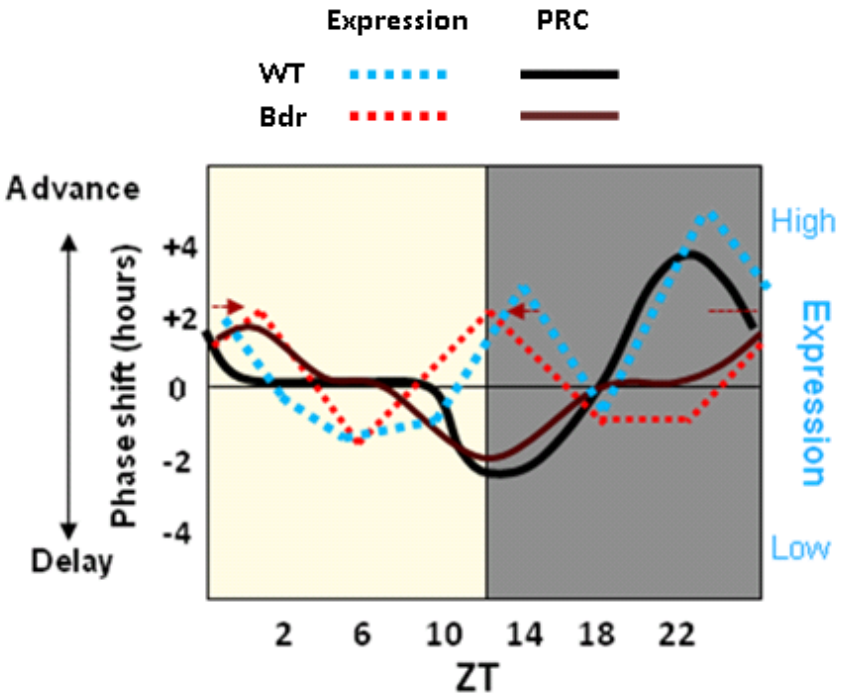


Figure 5.22. Twin peaking expression is shifted in the *Bdr* mouse model.
In WT mice, expression of genes from the twin-peaking module (blue dashed line) peaks at ZT14 and ZT22. When this expression lies above a threshold level (black horizontal line), the phase of the organism is able to shift in response to light, as represented by the Phase Response Curve (PRC; black). In the *Bdr* mouse model, this expression profile (red dashed line) is dysregulated, and the peaks in expression now occur at ZT2 and ZT12. The largest change in expression is at ZT22, which is the timepoint in which most differentially expressed genes were identified. If the twin-peaking expression pattern represents a transcriptional gating mechanism determining when light can induce a phase-shift, then the resultant PRC (dark red) will likely be affected.

would represent the time when the SCN receives a counteracting response to light. This may explain why *Bdr* mice possess normal circadian phenotypes under DD — there is no light which can induce differential responses due to the differing PRC.

If the PRC is disrupted in *Bdr* mice, then these mice should respond differently to light impulses at night. At ZT14, the expression levels appear to be similar, therefore the phase-shifting response would be expected to be similar between genotypes (Figure 5.22). Previous analysis revealed that *Bdr* mice were able to shift to a light pulse at ZT14 similar to WT littermates (Figure 5.21). However, a light pulse at ZT22 should induce a phase-advance in WT, but not in *Bdr* (Figure 5.22). Although a phase-advance inducing light pulse at ZT22 had not been previously performed on *Bdr* mice, a 6-hour jet-lag paradigm was applied to test for phase-advance shifting potential (Oliver et al., 2012). Although no significant differences were observed between the genotypes, the *Bdr* mice appeared to entrain at a slower rate (Figure 5.3). This would be expected as the twin peaking genes display low expression levels at ZT22 preventing strong light-induced phase-advances. A significant difference may not have been detected as the 6-hour jetlag paradigm may not be specific enough to dissect the PRC. As light is present from ZT18-ZT6 in the jet lag paradigm, light at ZT2 will still induce a phase-advance in *Bdr*, albeit to a lesser extent than a light pulse at ZT22 (Figure 5.22). To test the hypothesis that the PRC differs between genotypes, one-hour light pulses were applied at specific circadian times and the phase-shifting potential for both WT and *Bdr* mice was assessed (Figure 5.20). For example, the PACAP$^{-/-}$ mice were shown to display deficits in their phase-shifting response when one-hour light pulses were applied at CT21 (Kawaguchi et al., 2003)

A one-hour light pulse at ZT14 was applied to WT and *Bdr* mice to confirm that the phase-shifting capabilities for *Bdr* were the same as their littermate controls at this timepoint. Measuring the phase-delay from these mice confirmed that there were no significant differences between the genotypes in ability to phase-shift to light (Figure 5.21), which suggests that the PRC is unaltered at this timepoint. This observation was expected as the expression of the twin-peaking module did not greatly differ at ZT14 (Figure 5.22).

A one-hour light pulse was also applied at ZT21 and phase-advances were measured from both genotypes. As twin-peaking genes displayed reduced expression in *Bdr* at ZT22, the *Bdr* mice may be unable to shift their activity on the receipt of a light pulse (Figure 5.22). However, a light-pulse at ZT21 was unable to induce a strong phase-advance in either WT or *Bdr* mice (Figure 5.21). There was perhaps a tendency for WT mice to show small phase-advances (negative phase-shift), whereas the *Bdr* mice did not, however this difference was not significant. The inability for mice to shift to a light pulse at ZT21 may be because mice were generally asleep at ZT21-22 and were not able to receive photic input necessary for a phase-shift. Therefore, to test the hypothesis that the twin-peaking expression pattern is important for resetting the clock, this light-shifting paradigm needs to be optimised so that it can induce a robust phase-advance in WT mice.

5.4.3 SNAP25 and synaptic function affect the timing of circadian expression

As the synaptic twin-peaking co-expression module was disrupted in the *Bdr* mouse model, SNAP25 may have a key role in establishing the timing of this transcriptional pattern. SNAP25 is a multifunctional t-SNARE protein which is anchored to the cytoso-

lic face of the membrane, possessing both a role in neurotransmitter release and neuronal depolarisation (Sollner et al., 1993, Jurado et al., 2013). The disruption of one or more of these processes may therefore affect the transcriptional timing of the twin-peaking module.

As discussed in chapter 3, this twin-peaking module could potentially be regulated by two circadian TFs (or other factors important for inducing transcription) which require different strengths of activation to induce an effect. If the expression level or transcriptional effect of one of the two circadian transcription factors were to increase or decrease, these downstream transcriptional peaks will start to separate from each other as is potentially observed in *Bdr* (Figure 3.17,3.18). Therefore, the I67T *Snap25* mutation in *Bdr* may affect the expression level for one of these regulatory factors important for generating the twin-peaking transcriptional profile and will therefore affect the timing of these transcriptional peaks (Figure 5.23).

The I67T dominant *Snap25* mutation in *Bdr* increases the binding affinity within the core SNARE complex impairing action potential evoked glutamate release, likely due to the inability to replenish the readily releasable pool of synaptic vesicles between stimuli (Jeans et al., 2007). As the spontaneous firing rate is highest during subjective day (ZT6), this defect in neurotransmission may reduce the number of action potentials at this time (Colwell, 2011). If the degree of neuronal firing can affect the expression level of a circadian TF, then this may affect the timing of the peaks in expression as previously described (Figure 5.23). Neuronal activity may regulate gene transcription by affecting the localization, expression or function of particular transcription factors (West et al., 2002, West and Greenberg, 2011). Neuronal transmission may induce a calcium

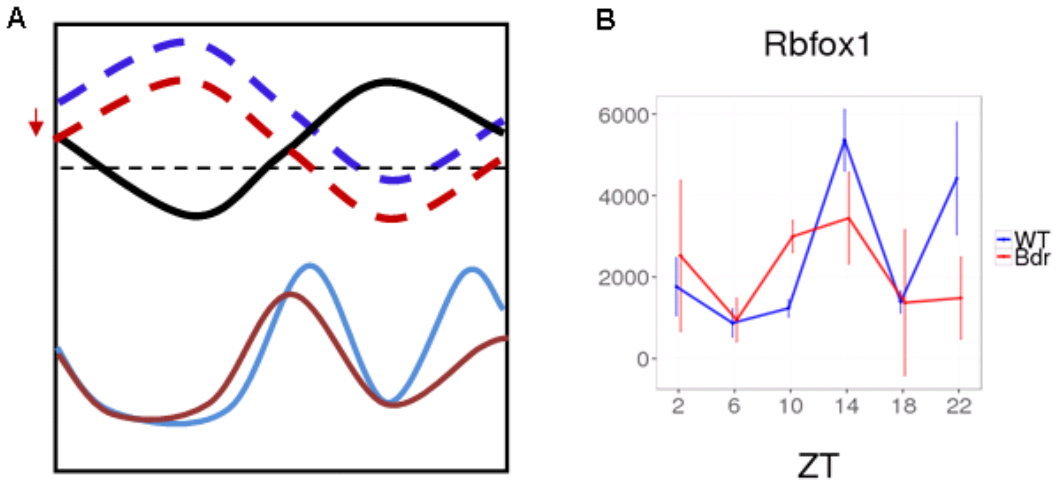


Figure 5.23. Reduction in a circadian TF can affect timing of twin-peaking transcription. **(A)** As described in chapter 3, two circadian transcription factors with opposing phases are able to produce a transcriptional response which approximates the twin-peaking transcriptional profile. As illustrated, the expression of TF1 (black line) and TF2 (dashed blue line) are able to induce a transcriptional response (blue line) when both are above a threshold level (dashed grey line). However, if this TF is to reduce in expression (dashed red line; perhaps due to the SNAP25 mutation) this can effect the timing of the downstream transcriptional profile (red line). The resulting hypothetical transcriptional profile displays a similar circadian expression pattern to those observed for the twin-peaking co-expression module in WT and *Bdr*. **(B)** For example, the mean expression levels (± SD) of the twin-peaking gene *Rbfox1* show an expected twin-peaking profile in WT but altered expression in *Bdr* similar to that hypothesised in **(A)**.

influx, which can bind to a TF altering its function, or activate calcium signalling cascades inducing gene transcription (Carrion et al., 1999, Burgoyne, 2007).

SNAP25 negatively regulates native voltage-gated calcium channels (VGCCs) in glutamatergic neurons by interacting with a synaptic protein interaction (synprint) site decreasing the VGCC's neuronal response to calcium (Catterall and Few, 2008, Pozzi et al., 2008, Antonucci et al., 2013). Reduction of endogenous SNAP25 expression was shown to increase VGCC function, a pattern mirrored in SNAP25 heterozygous mice (Condliffe et al., 2010). VGCCs were shown to play crucial roles in the glutamate-induced phase-shifts of the rat suprachiasmatic circadian clock (Kim et al., 2005). VGCC antagonists applied to rat SCN slices were reported to prevent glutamate-induced phase-shifts. Therefore, this mutation in *Snap25* may increase the binding affinity within the core SNARE complex, preventing SNAP25 from interacting with VGCCs and therefore affecting neuronal response to calcium. If neuronal response to calcium can affect the expression level of the circadian TF responsible for generating the twin-peaking transcriptional profile then this may explain the shifting of the twin-peaking expression.

Additionally, if *Snap25* regulated processes were to occur preferentially at particular times of the day, this may explain the time-dependent change in expression observed for this twin peaking module in *Bdr*. Application of T-type VGCC antagonists to rat SCN slices was reported to prevent glutamate-induced phase-delays, but not advances, whereas application of L-type VGCC antagonists prevented glutamate-induced phase-advances, but only partially prevented delays (Kim et al., 2005). L-type VGCCs are reported to localise in lipid rafts as they have been shown to colocalise with the lipid raft marker THY1 (Sunitha and Joshi, 2010). Cholesterol depletion was reported to enhance

L-type VGCC depolarisation which may be due to the reduced compartmentalisation of SNAP25 in the low density membrane regions of the lipid rafts (Sunitha and Joshi, 2010). Therefore, due to the association of SNAP25 with a specific subtype of VGCC which induce currents at particular times of day, the SNAP25 mutation may affect neuronal function and transcriptional response in a time-dependent manner.

In summary, the I67T dominant *Snap25* mutation in *Bdr* may affect the expression timing of the twin-peaking expression profile in a number of ways due to the multifaceted nature of SNAP25. The mutation increases the binding affinity within the core SNARE complex, which may reduce the degree of synaptic transmission due to the reduction in the reserve of readily releasable vesicles (Jeans et al., 2007). Alternatively, as SNAP25 is restricted to these SNARE complexes in *Bdr* mice, SNAP25 is unable to be partitioned to other regions of the membrane where it is required to interact with membrane proteins such as VGCC which can affect neuronal response to calcium. Interestingly, this *Snap25* mutation affects the expression of the twin-peaking module in a time-dependent manner. This suggests that the mutation in SNAP25 either affects function in a time-specific manner or it affects the expression level of a circadian transcription factor which causes the peaks of the transcriptional module to diverge and peak at differing times (Figure 5.23).

5.4.4 Associations with schizophrenia

Enrichment testing determined that the dysregulated twin-peaking expression module was not enriched for genes overlapping loci discovered through schizophrenia genome-wide association studies (Table 5.5) but was highly enriched for genes implicated in schizophrenia

via a convergent functional genomics (CFG) approach (Table 5.6). There may be a few reasons for this observation.

Although the CFG approach largely integrates non-hypothesis driven data-sets, such as gene expression, GWAS, CNV, linkage and quantitative traits loci, it also considers some hypothesis-driven experimental work which may bias some of the scoring. Genes which have been studied extensively for historical reasons will therefore be likely to show some form of experimental validation even if it does not induce a significant effect in the etiology of schizophrenia. Conversely, genes which have only been recently implicated in schizophrenia are unlikely to have follow-up studies and will therefore score lower in this CFG approach. It is possible that genes in this twin-peaking module have been subject to increased experimentation (perhaps due to their role in synaptic function), and therefore be more likely to be implicated to schizophrenia through this CFG approach. Although the authors attempt to overcome this effect by capping the score received from experimental evidence (Ayalew et al., 2012), this bias may have a small effect on the enrichment of these genes in this module.

Alternatively, the resultant GWAS gene list may not be truly representative of causative agents contributing to schizophrenia, explaining why it was not enriched in this twin-peaking module. Although the 108 variants associated with schizophrenia are likely to induce some effect on schizophrenia pathology, it is difficult to implicate these variants to the genes in which they induce their effect (2014, Sanyal et al., 2012). In this enrichment analysis, the genes overlapping the variant were called to be the disease causing region. The ENCODE Consortium showed that only 27% of distal regulatory elements interacted with the nearest promoter, suggesting that the closest genes to variants

identified through GWAS association studies may not be relevant to disease (Sanyal et al., 2012). For example, the likely causal variant (rs12740374) for low-density lipoprotein cholesterol sits in an enhancer element for the fourth-closest gene, SORT1 (Musunuru et al., 2010). In addition, each distal regulatory element targets, on average, 2.5 local target genes. This suggests that variants identified through GWAS may affect the transcription of numerous genes (Sanyal et al., 2012). This makes it difficult to determine the mechanisms which contribute to disease.

Lastly, these results may imply that the genes of the twin-peaking module, when disrupted, do not contribute to schizophrenia, but instead represent markers or downstream dysfunction of the main causative agent. GWAS is able to identify the genes, or regulatory regions, which likely cause the disease phenotype. On the other hand, gene expression data or experimentation on humans or animals may identify genes or processes which are highly correlated with disease dysfunction, but do not necessarily represent the causative agent. This twin-peaking module, which is dysregulated in *Bdr* could represent a marker of schizophrenia, whereas the upstream causative agent (SNAP25 or other proteins with similar functions) may underlie genetic cause of schizophrenia.

It may be of interest to further investigate some of these dysregulated twin-peaking genes which have been implicated in schizophrenia. For example, TCF4 (transcription factor 4) encodes a transcription factor and is required for neuronal differentiation. Of all the TFs in the twin-peaking transcriptional module, TCF4 displayed the highest module membership (mm = 0.93; Table 3.3) suggesting that it may represent a modular hub gene responsible for driving this transcriptional profile. *Egr1* is also found to be implicated in schizophrenia which also possessed a high module membership (mm = 0.87; Table 3.3).

Therefore, disruption of the genes responsible for driving this transcriptional pattern may contribute to a schizophrenia endophenotype.

5.4.5 Conclusion

In this chapter, it was determined that the twin-peaking transcriptional module identified in chapter 3 appeared to be dysregulated in *Bdr*. This transcriptional profile was hypothesised to play a role in the light-induced phase-shifting pathway and therefore its temporal dysregulation may underlie the circadian deficits observed in *Bdr*. To further investigate the hypothesis that the dysregulated twin-peaking transcriptional profile altered the ability of *Bdr* mice to respond to light at particular times, additional circadian behavioural parameters were assessed for WT and *Bdr* mice. Unfortunately, a 1-hour light pulse was unable to induce a phase-advance in either genotype. A phase-advance in WT, but not *Bdr*, mice would have strengthened the hypothesis that this transcriptional signature is important for circadian rhythmicity. Therefore, this paradigm should be optimised so it is able to elicit a significant shift in WT mice. Interestingly, this transcriptional module also revealed associations with neurological diseases such as schizophrenia, which may provide clues to the molecular link between circadian disruption and psychiatric disease.

6 Conclusions and future perspectives

6.1 Summary

6.1.1 Comprehensive examination into the mouse SCN transcriptome

A comprehensive transcriptomic investigation of the SCN using RNA-seq has been presented in this thesis, offering novel insight into SCN and circadian biology. Prior to this research, transcriptomic information from the SCN was relatively limited. Although some key transcriptional processes were identified to be cycling in the SCN from these previous analyses, the findings were limited by the use of microarray technology, and the tissue sampling and lighting conditions mice were subject to; together, these may have prevented the identification of functionally important transcriptional signatures of the SCN (Panda et al., 2002, Zhang et al., 2014, Hatori et al., 2014). Here, for the first time, an anatomically precise and comprehensive examination of the SCN transcriptional output over 24 hours was assessed using RNA-seq. A web-app interface was developed to allow exploration of various aspects of the RNA-seq dataset and to allow easy visualization of gene expression patterns by the scientific community.

In chapter 3, 5,017 genes were identified that displayed significant fluctuations in the SCN over circadian time. 4,569 of these displayed sinusoidal expression and were enriched for some key processes offering insight into SCN circadian transcriptional regulation. Unlike previous circadian transcriptomic studies, co-expression analysis was also applied to the expression data to identify a novel twin-peaking transcriptional pattern. These results suggest that circadian studies should transition away from sinusoidal fitting algorithms and towards co-expression analysis or other unbiased statistical methods, as

previous analysis may have overlooked functionally important transcriptional profiles and missed much of the richness in such large transcriptomic datasets, including gene-gene relationships and the identification of modular hub genes.

Although the temporal gating of the light-induced resetting pathway in the SCN is well established, the exact mechanistic underpinnings remain unclear (Iyer et al., 2014). The twin-peaking transcriptional profile suggests that a transcriptional gating mechanism may contribute to the SCN's differential response to light depending on the time of day. In addition, genes responsible for actin assembly, a process essential in inducing phase-delays in the early night, were found to peak in expression at ZT14, which may offer additional insight into the molecular pathways regulating light-induced phase-shifts. The identification of this twin-peaking module also allows us to re-interpret existing hypotheses in the circadian field. Interestingly, twin peaking expression, redox state and membrane potential appear to be closely interlinked in the SCN, offering clues into the coupling mechanisms between the transcriptional and non-transcriptional circadian oscillators.

In chapter 4, RNA-seq of anatomically precise SCN samples allowed the identification of thousands of SCN-expressed novel lincRNA and isoforms of annotated genes, many of which may be SCN-specific. Although prior to these analyses, some lncRNA were known to possess roles in circadian biology, this number was limited, especially in mammalian species. For instance, there were no reported lncRNA with importance in the mammalian SCN. Therefore, in addition to identifying a number of lincRNA expressed in the SCN, it was important to further prioritise them based upon expression and sequence conservation to facilitate future functional studies in the circadian system (Lewejohann et al., 2004).

In addition, over a thousand novel exons of previously annotated genes were identified, including one from the clock gene *Cry1*. Interestingly, the novel *Cry1* isoform appeared to be non-coding and to be expressed in antiphase to the canonical protein-coding isoform. This may suggest an isoform switching mechanism evolved to allow the rapid generation of CRY1 protein oscillations. Future work is required to determine the importance of this mechanism and whether it represents a widespread feature of circadian regulation (section 6.2.1).

6.1.2 Elucidation of the molecular links between psychiatric disease and circadian disruption

In this thesis, transcriptional disturbances potentially contributing to circadian disruption and psychiatric disease were identified and discussed. Circadian rhythm disruption is highly associated with a number of neurological and psychiatric diseases such as schizophrenia. The co-morbidity of circadian rhythm disruption and neuropsychiatric disease is proposed to be due to common brain mechanisms underlying the two. However, these molecular mechanisms are largely unknown. To establish these molecular pathways, which when disrupted, contribute to circadian disruption and psychiatric disease, an aim of this thesis was to further investigate the blind-drunk (*Bdr*) *Snap25* mouse mutant. Although microarray transcriptomic analysis was previously implemented on *Bdr* mice, very few expression differences were identified, limiting any novel insight into the molecular pathways linking abnormal brain function with disrupted circadian rhythmicity. To overcome these limitations, RNA-seq was performed on SCN samples extracted by laser capture microdissection (LCM) from *Bdr* at six timepoints (ZT2, 6, 10, 14, 18 and 22) in

parallel with the wild-type mice. In chapter 5, the twin-peaking transcriptional module was reported to be dysregulated in *Bdr*. Given the associations of this transcriptional module with both circadian biology and schizophrenia, this transcriptional module may provide clues to the molecular link between circadian disruption and neuropsychiatric disease (section 6.2.7).

6.2 Discussion and future perspectives

6.2.1 Novel *Cry1* transcript represents novel mechanisms of circadian control

To further assess the potential function of the antiphase non-coding *Cry1* isoform identified in chapter 4, a number of experiments can be performed. This novel *Cry1* isoform may represent an SCN-specific transcript as it had not been previously identified, despite its large size and high expression. The SCN-specificity of this transcript should be confirmed though RT-PCR in other tissues or brain regions. Alternatively, *in situ* hybridisation could be performed for both the canonical and novel isoform in mouse brains, perhaps over multiple timepoints across the day. This will confirm transcript localisation and help determine whether this antiphase expression pattern is occurring in other brain regions.

The core circadian negative feedback loop (Figure 1.3), for which CRY1 is a key component, is present in every cell, therefore it is interesting to speculate why the novel *Cry1* isoform may be SCN specific. As the SCN is the master pacemaker, it may need to produce stronger oscillations in protein expression to drive the circadian rhythms in the whole organism, whereas peripheral tissues may not require such strong protein oscillations but operate more under the influence of systemic cues (Mohawk et al., 2012).

To elucidate the function of this transcript, a genetic complementation assay can be applied, generating and transfecting plasmid constructs to *Cry1*$^{-/-}$*:Cry2*$^{-/-}$ cells (Figure 6.1A) (Ukai-Tadenuma et al., 2011). Firstly, to determine whether the novel *Cry1* transcript has an affect on circadian rhythmicity, a reporter construct, in which only the canonical *Cry1* isoform is expressed, can be added to arrhythmic *Cry1*$^{-/-}$*:Cry2*$^{-/-}$ cells.

The circadian rhythmicity of these *Cry1*$^{-/-}$*:Cry2*$^{-/-}$ cells can be assessed through the addition of a *Per2*-Luc construct, in which luciferase is under the control of a *Per2* promoter. If the *Cry1* construct is unable to regain rhythmicity of these cells, it suggests that the novel transcript is required for circadian rhythmicity. To determine whether the novel *Cry1* isoform has a protein coding function, the *Cry1* construct can be mutated so the resultant transcript has no protein coding potential, if circadian rhythmicity is unable to be recovered, it suggests the coding potential of the novel transcript is important. However, if the novel isoform is part of an isoform switching mechanism important for producing strong CRY1 protein oscillations, the novel transcript should *not* encode protein. If the 5' UTR of the novel *Cry1* exon is mutated in the construct to remove the untranslated open reading frames (uORFs), then the novel transcript may start to produce a functional CRY1 protein, which can be confirmed through western blotting. If the rhythmicity *Cry1*$^{-/-}$*:Cry2*$^{-/-}$ cells cannot be fully recovered, it suggests that it is important for this isoform to be non-coding – as previously suggested (section 4.3.1).

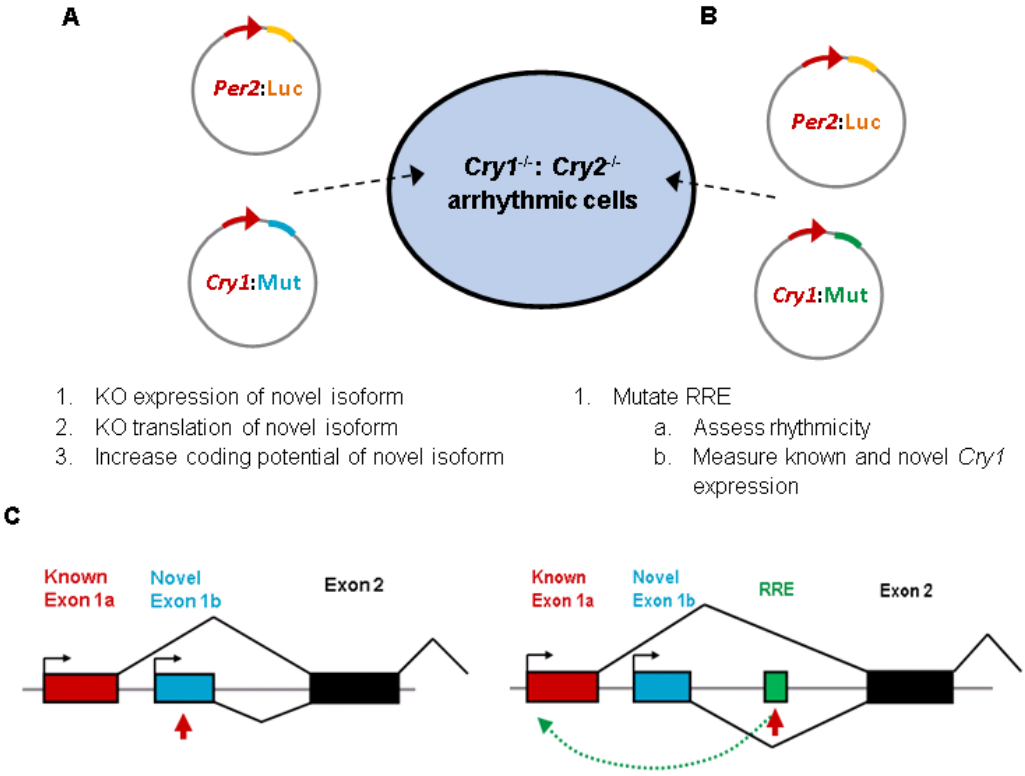


Figure 6.1. Genetic complementation assay in *Cry1*$^{-/-}$*:Cry2*$^{-/-}$ cells.
An array of *Cry1* constructs can be transfected to *Cry1*$^{-/-}$*:Cry2*$^{-/-}$ cells to assess the function and regulation of the novel *Cry1* isoform. **(A)** A variety of *Cry1* constructs can be added to determine how the novel isoform affects circadian rhythmicity (as assessed by the *Per2*:Luc reporter). **(B)** Constructs can be added to assess the regulation of the known and novel *Cry1* isoforms. **(C)** Schematic diagrams of mutant *Cry1* constructs. Red arrow indicates region to be mutated.

Regulation of the timing of clock gene expression depends on the presence of transcription factor binding sites. *Cry1* transcription depends on promoter E'/E-boxes and D-box elements but also a RRE site, found in the first *Cry1* intron (Ukai-Tadenuma et al., 2011). Interestingly, this RRE acts as an enhancer to delay the phase conferred by the canonical *Cry1* promoter. Therefore, the antiphase expression pattern of the novel *Cry1* isoform may be because the RRE is situated too close to the novel exon (also present within the first intron) to be able to regulate, and delay, its transcription, perhaps through a DNA looping mechanism (Figure 4.13).

Adding a construct with a mutated RRE and assessing cellular rhythmicity and the expression levels of both *Cry1* isoforms, through qPCR or reporter constructs, may offer additional insight into the regulation of this gene (Figure 6.1B). If both the novel and canonical *Cry1* transcripts show similar phases of expression after RRE mutation, it suggests that there is no longer a circadian isoform switching mechanism present in this RRE-mutated construct. These results would confirm the differential regulation of the canonical isoform through the RRE which may manifest itself as an isoform-switching mechanism.

Isoform switching mechanisms that regulate protein levels may also be regulating other circadian-relevant genes. uORFs have been proposed to play a role in the circadian system, as loss of function of *Denr*, a regulator of ribosome recycling and reinitiation after uORF usage, induced a shortening of circadian period in fibroblast cell culture (Janich et al., 2015a). In addition, alternative splicing of core clock genes to produce transcripts with premature stop codons had been previously described in *Arabidopsis* and were found to be upregulated in cold conditions, suggesting a role in temperature compensation

mechanisms (Filichkin et al., 2010, Filichkin and Mockler, 2012b). A computational approach could be applied to identify transcripts which are antiphase to the canonical transcript and are either non-coding or possess a large number of uORFs, representing genome-wide circadian isoform switching mechanisms (Carninci et al., 2006, Janich et al., 2015b). The proximal gene regions can be scanned to identify common binding sites which may affect the differential expression over circadian time for each isoform.

6.2.2 Analysing circadian time course expression data

Previous studies examining circadian transcriptomic datasets only considered circadian gene expression that followed a sinusoidal pattern. However, as discussed, this method fails to identify transcriptional patterns which do not match a sinusoidal curve and does not consider gene-gene relationships which can offer much biological insight. In this thesis, co-expression analysis was applied, allowing the identification of a number of transcriptional modules and prioritisation of genes by importance within these modules based on their co-expression. An issue that results from using WGCNA co-expression and similar clustering methods is that they are not designed to consider the ordering of the data points, but just to assess correlations in expression. Thus, these co-expression analyses do not capture time-shifted patterns. Some genes may be regulated by the same transcription factors, but may differ in their lag of expression (time taken from TF binding to discernible levels of mRNA transcripts). Therefore, genes which show similar transcriptional patterns but are slightly offset in circadian phase may be partitioned into different co-expression modules, despite being highly functionally related (Elo and Schwikowski, 2013). This effect is likely to be exacerbated when sampling at higher temporal resolution

as the lags between similar genes will be more noticeable. This lag in transcription between different mRNA transcripts makes hub gene identification complicated. Over 40% of proteins in liver display a lag of >6 hours relative to their transcript levels, including many TFs (Robles et al., 2014). Therefore the TF driving the expression of a transcriptional module is unlikely to also be part of the same module, unless there are mechanisms allowing immediate translation of the TF mRNA. Additional information such as TFBS enrichment may instead be used to determine the hub drivers of a particular transcriptional module. Exploration of other algorithms for identifying transcriptional patterns or analysing circadian data is important, especially when circadian transcriptomes are assessed at a high temporal resolution. For example, dynamic time warp (DTW) has been shown to match transcriptional patterns whilst considering lags in expression (Yuan et al., 2011, Wexler et al., 2011, Remondini et al., 2005, Butte et al., 2001). Briefly, DTW calculates the minimal distance (optimal alignment) between two temporal expression patterns and can provide a measure of both circadian expression similarity and *how* circadian expression differs in the time dimension. Considering lag in expression is also pertinent if integrating different levels of circadian data (such as proteomic or epigenetic) as there is often a delay between each level (Fuhr et al., 2015).

Comparing the circadian expression between *Bdr* and WT mice revealed a small phase-advance of SCN output molecules, such as *Avp* and *Prok2*, which was confirmed by qPCR; however, these differences were not called as significant using the methods applied within this thesis. Perhaps by pooling information between genes or considering the expression data as time course data (as in DTW), the analyses applied will increase in power, allowing detection of subtle changes in circadian expression (Wexler et al., 2011).

Alternatively, the study may be underpowered and require either a greater number of biological replicates per timepoint or a greater sampling resolution to dissect these small phase-shifts. In addition, transcriptional dysregulation in *Bdr* may only occur in particular regions or cellular subtypes of the SCN, therefore greater anatomical resolution of the SCN may help identify robust phase-shifts or other circadian expression differences in the *Bdr* SCN (section 6.2.3). Furthermore, although the dysregulation of the twin-peaking module in *Bdr* mice was confirmed by using a combination of differential expression and co-expression analysis, this method was only able to predict the dysregulation of 57 genes in this twin-peaking module, which was then extrapolated to confirm the dysregulation of the entire module. Therefore, it is desirable to use an alternate approach which can identify systematic changes for particular transcriptional profiles in order to help determine the biological processes which are dysregulated (Tesson et al., 2010). To conclude, although co-expression analysis represents significant improvement into the current sinusoidal based methods applied in circadian datasets, there are also some areas which require optimisation, especially if samples are profiled at high temporal resolution. These methods should also help identify lags in expression between genotypes. The analysis used in this thesis was not sufficiently powerful to identify those genes with small phase-advances. Further considerations may be necessary if attempting to integrate other circadian –omics datasets, an important feat in attempting to understand the true nature of circadian regulation.

6.2.3 High resolution sequencing of the SCN

Anatomically precise dissections of the SCN were important for the identification of SCN-specific transcriptional profiles. However, these transcriptional profiles represent the pooled transcriptional signature of all the cell types within the SCN. It is plausible that different regions of the SCN (such as the SCN core or shell; Figure 1.1) display different transcriptional signatures over the circadian day, which cannot be identified when dissecting and sequencing the whole SCN together. Although, the lag between core and shell has been reported to differ by $\sim$ 1-hour for some genes, it would be interesting to determine whether there are more obvious transcriptional lags (Hamada et al., 2004, Yan et al., 2007). Therefore, to gain a greater understanding of the interplay between SCN core and shell, LCM can be used to dissect these regions separately before sequencing. LCM of SCN core and shell was previously used to extract RNA for high-throughput qPCR (Zhu et al., 2012). However, this was only performed at a single timepoint (ZT15) and could only assess up to 90 genes. The SCN consists of a number of cell types - there are up to 20,000 neurons with similar levels of astroglia. In addition, SCN neurons can be further divided into subpopulations based upon their immunoreactivity to key SCN neuropeptides such as AVP, VIP or GRP (Romijn et al., 1997). Each of these cell populations will also likely possess different transcriptional signatures over circadian time. Fluorescence-activated cell sorting (FACS) can be used to extract the various cell-types from the SCN extracts before sequencing (Cruz et al., 2013, Basu et al., 2010). This will allow the circadian transcriptional profiles for each SCN cell population to be assessed which can help to determine the molecular interplay occurring between these cellular subtypes. The most anatomically precise method of assessing the transcriptome would be

through single-cell RNA-seq. Drop-Seq is a method which can be used to quickly profile the transcriptome of thousands of single cells by separating them into nanoliter-sized aqueous droplets (Macosko et al., 2015). The RNA in each droplet, which corresponds to the transcriptome of a single cell, is then barcoded. The RNA from all cells are subsequently pooled together for sequencing, with the barcoding denoting the cell of origin for each transcript. At a single timepoint (ZT), the population of single cells sequenced can then be clustered based upon their expression to identify distinct cell populations within the SCN. For example, a particular cluster may represent a specific cell type (e.g. oligodendrocyte) or SCN region (e.g. core vs shell). Performing Drop-seq over circadian time allows the circadian expression pattern to be assessed for each cell population. This deluge of transcriptomic data can be used to help determine the processes which work in synchrony with each other, both within or between different SCN cell type populations. This expression data can also be used to help decipher which cell types contribute to the transcriptional patterns observed for the LCM SCN samples within this thesis. Increased anatomical resolution may also help further decipher the transcriptomic disturbances occurring in the SCN of *Bdr* mice. It is possible that SCN-output molecules are phase-advanced in the shell - which sends output signals to the rest of the organism, but possesses a normal transcriptional signature in the core. Increased anatomical resolution will help determine the specific processes and regions which are disrupted. Obtaining SCN sequencing data from individual mice, versus pooled samples, also comes with additional benefits as it allows the variation in gene expression to be more accurately assessed. For example, it can be determined if some genes are more 'noisy' at particular times but show highly regulated levels of transcription at other times. Furthermore, by not pooling SCN

samples, the circadian and behavioural characteristics of each mouse can be correlated to the transcriptomic profile of their SCN at a particular timepoint which provides greater power for analysis.

6.2.4 Circadian isoform quantification

Although not discussed in this thesis, the LCM SCN RNA-seq data could be used to determine the circadian splicing changes which were occurring in the SCN over circadian time. In addition, the RNA-seq data can be used to identify splicing differences in the *Bdr* mice. However, isoform quantification and accurately measuring splicing changes is often problematic in RNA-seq data analysis. One reason for this is because, although, next-generation sequencing has revolutionised the field of genomics, their short read lengths, make them poorly suited for some particular biological problems, such as gene isoform detection and quantification (Au et al., 2013). Recently, Pacific BioSciences (PacBio) had developed an alternative approach known as single-molecule real-time (SMRT) sequencing (Rhoads and Au, 2015). "PacBio sequencing" allows longer read lengths to be sequenced (10 Kb vs 250bp) but is more expensive per base sequenced and often has a higher error rate. These features of PacBio sequencing allow it to suitably complement short read sequencing to answer particular biological questions, such as isoform detection and quantification. The integration of both sequencing approaches is known as hybrid sequencing (Rhoads and Au, 2015). There exist a number of tools which allow the identification and quantification of full-length gene isoforms by using both short and long sequencing reads, such as the Isoform Detection and Prediction tool (IDP) (Au et al., 2013). Compared to Cufflinks, IDP had a much higher sensitivity for isoform identification, for instance

Cufflinks only possessed a 20% true positive rate whereas IDP had 62% (Au et al., 2013, Trapnell et al., 2010). It may therefore be of interest to perform PacBio sequencing on the SCN samples over time, and use both short and long reads to determine the splicing changes and isoform fluctuations over circadian time. For example, the splicing factor *Rbfox1* appears to show twin-peaking expression in the SCN. It would be interesting to determine whether splicing associated with this gene also changes in a twin-peaking pattern. However, as described in chapter 3, this transcriptional pattern may not represent the functional protein levels, unless the SCN receives a light pulse and translates the *Rbfox1* RNA. Therefore, *Rbfox1* related splicing changes in the SCN at certain circadian times (ZT14/ZT22) may only be expected on the receipt of a light pulse. As this expression pattern is disrupted in *Bdr*, there may be disruption in *Rbfox1* regulated splicing which may affect neuronal excitation (Gehman et al., 2011). Once sequencing technology and corresponding algorithms become sufficiently sophisticated to accurately determine isoform expression levels, then isoform co-expression networks can be generated providing increased resolution to biological pathways (Li et al., 2014). This will allows the identification of common splicing changes across circadian time and the biological processes associated with them.

6.2.5 Transcriptional regulation of the twin-peaking transcriptional profile

As the twin-peaking transcriptional profile had not previously been characterised, the mechanisms of its transcriptional regulation are unknown. 12-hour harmonics of circadian expression, however, have been suggested to arise through the combination of two circadian TFs which are out of phase with each other (Westermark and Herzel, 2013). In

chapter 3, it was suggested that the twin peaking transcriptional profile may be modelled by dual circadian activators with differing expression levels or strengths. To investigate the transcriptional regulation of this twin-peaking module, the promoters of all 766 genes could be tested for enrichment of a particular combination of circadian TF binding sites as previously described (Westermark and Herzel, 2013). First, the motifs which correspond to circadian TFs should be identified and determined whether they are annotated as activators or repressors (according to GO annotations). Then these TFs can be modelled to determine how they may interact to produce asymmetrical harmonics of circadian expression. It is possible that only a small proportion of twin-peaking genes are under the transcriptional control of two circadian TFs. Instead, these circadian TFs may regulate the expression for only a small set of TFs which are then responsible for driving the expression of all other genes in this twin-peaking module. Promoter analysis of the genes in the twin-peaking module can determine whether they are enriched for twin peaking TFs. It will be of interest to determine whether TFs such as *Tcf4, Mef2c* and *Egr1*, which display twin peaking gene expression, also display similar patterns of protein expression, and whether their binding is enriched in the promoters of the genes in this module. ChIP-seq could be employed to confirm the time-dependent binding of these TFs to the promoters of genes within this co-expression module (Takahashi et al., 2015). ChIP-seq has been used to identify the DNA binding sites for the clock genes BMAL1, CLOCK, NPAS2, PER1, PER2, CRY1 and CRY2 and to assess their occupancy in the mouse liver over circadian time (Koike et al., 2012). However, as transcription factor occupancy is very often tissue specific, to help determine the transcriptional regulation of this twin-peaking module, circadian ChIP-seq data in the SCN should be collected for the TFs proposed

to regulate the twin-peaking transcriptional module. To gain additional information on the transcriptional regulation in the SCN, ChIP-seq could be performed for the aforementioned core clock proteins and histone marks representing epigenetic modifications (Koike et al., 2012). As the twin-peaking transcriptional profile was dysregulated in *Bdr*, identification of the regulatory TFs may provide clues into the specific factors which are dysregulated in *Bdr*. Identifying the TF which causes disruption of this module will help link the dysregulated transcriptional profile to the synaptic dysfunction caused by the *Snap25* mutation. The TF could subsequently be knocked down in the SCN using *in vivo* RNA interference (Jagannath et al., 2013). Transcriptional signatures and circadian behavioural parameters can then be assessed to elucidate the regulation and functional importance of the twin-peaking transcriptional profile.

6.2.6 Circadian-regulated translation

Interestingly, the genes in the twin-peaking transcriptional module were enriched for FMRP translational targets. This suggests that FMRP may affect the translation of the genes within the twin-peaking transcriptional module. Fragile X mental retardation 1 (FMR1) encodes the RNA-binding protein FMRP (Ascano et al., 2012). FMRP binds to mRNA of a number of genes, many of which have been implicated in autism spectrum disorders (ASD) (Darnell et al., 2011). Interestingly, variable levels of FMRP has been reported to affect protein levels of their RNA targets in human cell culture, mouse ovaries and human brain (Ascano et al., 2012). FMRP has been proposed to represses translation by binding to the translation factor eIF4E preventing translation of the transcript it is bound to (De Rubeis and Bagni, 2011). Interestingly, synaptic stimulation causes the

dissociation of FMRP and eIF4E, allowing the translation of the FMRP targets (twin-peaking genes) (Napoli et al., 2008). Therefore, in addition to the phosphorylation of eIF4E at early and late night which can allow a 3-fold increase in translation in the SCN (section 3.4.4), a light impulse at these times may also allow the translation of the genes in the twin-peaking module due to the dissociation of FMRP and eIF4E.

For example, synaptic plasticity has been proposed to be partially regulated by the TF EGR1 as its inactivation disrupts memory and learning (Jones et al., 2001). Interestingly, synaptic input can induce EGR1 protein expression in only a few minutes (Wei et al., 2000, Simon et al., 2004). As *Egr1* is part of the twin-peaking transcriptional module and also an FMRP target, it suggests that *Egr1* may be controlled by FMRP to allow rapid upregulation of EGR1 on the receipt of a light pulse (Darnell et al., 2011). FMRP and FMRP related genes have been highly implicated in autism and other neurological disease, many of which possess disruptions in sleep and circadian rhythms, therefore the dysregulation of these genes at the translational or transcriptional level may induce abnormal brain function and circadian disruption. Mouse knock outs of *Fmr1* gene family members cause both circadian disruption and altered hippocampal synaptic plasticity (Zhang et al., 2008, Zhang et al., 2009).

6.2.7 Pathways of plasticity in disease and circadian biology

In addition to having a role in synaptic transmission and circadian function, the twin-peaking co-expression module was enriched for genes involved in long-term potentiation (LTP) (Figure 3.14). LTP is the process by which synapses are strengthened based upon the activity of that synapse. The increase in synaptic strength induces increased transmis-

sion across that synapse which is stable long term (Lynch, 2004). Together with long term depression (LTD), which reduces the synaptic strength after synaptic activity, LTP modulates brain plasticity which allows the brain to be "plastic" and to be able to form new connections. For example, plasticity allows an individual to learn or to have a memory by storing this novel information through synaptic connections.

The mechanisms underlying LTP share a remarkable overlap with the genes and signalling cascades involved in the light-induced phase-shifting pathway (Iyer et al., 2014). All the genes labelled in Figure 3.20, plus others such as MAPK, have been shown to possess key roles in both the light-induced phase resetting pathway and LTP. Interestingly, the light-induced resetting of the SCN can be viewed as a form of plasticity. As the brain encodes the time of day information, a change in the brain's circadian phase requires a change in its synaptic connections. Metaplasticity involves mechanisms which can affect the state of the neuron so that they are susceptible or resistant to plasticity-inducing stimuli (Iyer et al., 2014). As described, the SCN is able to respond differently to light depending on the time it receives the signal; during the day and at midnight, a light impulse has no effect on the phase of the organism, however a light impulse at early or late night can induce a phase-delay or a phase-advance respectively (Golombek and Rosenstein, 2010). This gating of the light-induced phase-shifting pathway can therefore be considered a mechanism of metaplasticity.

Currently, there is much evidence that dysregulated neural plasticity underlie schizophrenia and numerous other neurological disease (Crabtree and Gogos, 2014, Stephan et al., 2009, Keshavan et al., 2015). Therefore, the co-morbidity of circadian disruption and psychiatric disease may be due to a dysregulation of synaptic metaplasticity. If

this hypothesis was true, then it would be expected that in *Bdr*, synaptic metaplasticity and LTP pathways would be disrupted in brain regions linked to schizophrenia, such as the hippocampus. The *Snap25* mutation may therefore disrupt the regulation of plasticity in the SCN, causing circadian disruption, and disrupt the regulation of plasticity in schizophrenia-implicated brain regions, such as the hippocampus, resulting in schizophrenia-like endophenotypes.

LTP and brain plasticity are extremely important in the hippocampus given its role in creating new memories. Understanding the light-induced phase-shifting pathway may therefore contribute to our understanding to memory formation. For example, this may mean that memory training may function as a zeitgeber in brain regions important for memory, just as light functions as a zeitgeber in the SCN, and therefore disruption of the same pathways can produce a number of phenotypes (Eckel-Mahan and Storm, 2009). Circadian disruption has also been shown to affect memory formation as SCN lesions affect hippocampal-dependent memory formation in rats and light-induced phase-shifting can induce retrograde amnesia (Stephan and Kovacevic, 1978, Tapp and Holloway, 1981). It is therefore possible that the hippocampal circadian clock may modulate the potential for LTP and other forms of synaptic plasticity by iterative metaplasticity in a similar mechanism to the SCN (Iyer et al., 2014).

This common molecular pathway may also explain the high co-morbidity circadian rhythm disruption and decline in episodic memory in aging adults. Associative memory in elderly human subjects was positively correlated with both consistent circadian activity rhythms and hippocampal activity as determined by fMRI (Sherman et al., 2015). This

suggests that as humans get older, a decrease in synaptic plasticity may affect the SCN and hippocampus, disrupting circadian rhythmicity and memory formation respectively.

Circadian disruption has been described in a number of other models of neuropsychiatric disease (Pritchett et al., 2012a, Li et al., 2015). It would be therefore be interesting to perform circadian screens of these mouse models and to perform transcriptomic profiling of certain brain regions (such as the SCN and hippocampus) over circadian time in order to identify the dysregulated processes and determine whether the same molecular pathways are disturbed. For example, the heterozygous reeler mouse ($rl^{+/-}$) possesses many features of schizophrenia including altered synaptic plasticity, pre-pulse inhibition and behavioural deficits in associative learning and memory, which were improved with reelin (Reln) supplementation (Rogers et al., 2013). It would be interesting to determine whether $rl^{+/-}$ mice possessed circadian deficits and whether Reelin supplementation improved rhythmicity. As *Reln* displays twin-peaking expression in the LCM SCN dataset, it may be part of the same pathway that is disrupted in the *Bdr* mouse model.

6.3 Conclusion

A comprehensive transcriptomic investigation of the SCN using RNA-seq has been presented in this thesis. A number of previously unannotated transcripts were identified in the SCN, suggesting that a large portion of the SCN transcriptome remains undefined. Further study is required to assess the relative importance of these transcripts and uncover the additional layers of transcriptional regulation occurring in the SCN. In addition, co-expression analysis identified an unexpected twin-peaking transcriptional profile which was associated with both synaptic transmission and clock resetting pathways. This indicates that previous analyses may have been overlooking functionally important transcriptional signatures. Interestingly, this module appeared to be highly dysregulated in *Bdr*, which provides initial clues to the molecular link between circadian disruption and psychiatric disease. Further work is required to decipher the precise mechanisms which result in circadian disruption and the schizophrenia endophenotypes in *Bdr* mice, as well as the complex interplay between other neurological systems. For instance, human patients of psychiatric disease often show sleep deficits in addition to, or instead of, circadian disruption, it will therefore be of interest to determine the shared molecular mechanisms regulating sleep and circadian disruption (Sutton, 2014, Goldstein and Walker, 2014). In addition, as a number of other neurological disorders are co-morbid with sleep and circadian disruption, it will be of utmost interest to determine whether these disorders share similar molecular pathways of disruption and how these neurological defects can be treated (Karatsoreos, 2014, Tordjman et al., 2015).

7 References

2014. Biological insights from 108 schizophrenia-associated genetic loci. Nature, 511, 421-7.

ABRAHAMSON, E. E. & MOORE, R. Y. 2001. Suprachiasmatic nucleus in the mouse: retinal innervation, intrinsic organization and efferent projections. Brain Res, 916, 172-91.

ADAMS, M. D., KELLEY, J. M., GOCAYNE, J. D., DUBNICK, M., POLYMEROPOULOS, M. H., XIAO, H., MERRIL, C. R., WU, A., OLDE, B., MORENO, R. F. & ET AL. 1991. Complementary DNA sequencing: expressed sequence tags and human genome project. Science, 252, 1651-6.

ADICONIS, X., BORGES-RIVERA, D., SATIJA, R., DELUCA, D. S., BUSBY, M. A., BERLIN, A. M., SIVACHENKO, A., THOMPSON, D. A., WYSOKER, A., FENNELL, T., GNIRKE, A., POCHET, N., REGEV, A. & LEVIN, J. Z. 2013. Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods, 10, 623-9.

ADLER, M., SHERIDAN, R. E., DESHPANDE, S. S. & OYLER, G. A. 2001. Neuromuscular transmission and muscle contractility in SNAP-25-deficient coloboma mice. Neurotoxicology, 22, 775-86.

AGARWAL, A., ZHANG, M., TREMBAK-DUFF, I., UNTERBARNSCHEIDT, T., RADYUSHKIN, K., DIBAJ, P., MARTINS DE SOUZA, D., BORETIUS, S., BRZOZKA, M. M., STEFFENS, H., BERNING, S., TENG, Z., GUMMERT, M. N., TANTRA, M., GUEST, P. C., WILLIG, K. I., FRAHM, J., HELL, S. W., BAHN, S., ROSSNER, M. J.,

NAVE, K. A., EHRENREICH, H., ZHANG, W. & SCHWAB, M. H. 2014. Dysregulated expression of neuregulin-1 by cortical pyramidal neurons disrupts synaptic plasticity. Cell Rep, 8, 1130-45.

AGOSTINO, P. V., FERREYRA, G. A., MURAD, A. D., WATANABE, Y. & GOLOMBEK, D. A. 2004. Diurnal, circadian and photic regulation of calcium/calmodulin-dependent kinase II and neuronal nitric oxide synthase in the hamster suprachiasmatic nuclei. Neurochem Int, 44, 617-25.

AHITUV, N., ZHU, Y., VISEL, A., HOLT, A., AFZAL, V., PENNACCHIO, L. A. & RUBIN, E. M. 2007. Deletion of ultraconserved elements yields viable mice. PLoS Biol, 5, e234.

ALBERS, H. E. & FERRIS, C. F. 1984. Neuropeptide Y: role in light-dark cycle entrainment of hamster circadian rhythms. Neurosci Lett, 50, 163-8.

ALBUS, H., BONNEFONT, X., CHAVES, I., YASUI, A., DOCZY, J., VAN DER HORST, G. T. & MEIJER, J. H. 2002. Cryptochrome-deficient mice lack circadian electrical activity in the suprachiasmatic nuclei. Curr Biol, 12, 1130-3.

ALEXANDER, R. P., FANG, G., ROZOWSKY, J., SNYDER, M. & GERSTEIN, M. B. 2010. Annotating non-coding regions of the genome. Nat Rev Genet, 11, 559-71.

ALHOLA, P. & POLO-KANTOLA, P. 2007. Sleep deprivation: Impact on cognitive performance. Neuropsychiatr Dis Treat, 3, 553-67.

AMARAL, P. P., DINGER, M. E., MERCER, T. R. & MATTICK, J. S. 2008. The eukaryotic genome as an RNA machine. Science, 319, 1787-9.

AMBROS, V. 2004. The functions of animal microRNAs. Nature, 431, 350-5.

ANDERS, S., PYL, P. T. & HUBER, W. 2015. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics, 31, 166-9.

ANTONUCCI, F., CORRADINI, I., MORINI, R., FOSSATI, G., MENNA, E., POZZI, D., PACIONI, S., VERDERIO, C., BACCI, A. & MATTEOLI, M. 2013. Reduced SNAP-25 alters short-term plasticity at developing glutamatergic synapses. EMBO Rep, 14, 645-51.

ARNOLD, J. 2013. Actin‘s role in signaling light-induced phase shifts in the mammalian circadian clock. University of Illinois at Urbana-Champaign.

ASCANO, M., JR., MUKHERJEE, N., BANDARU, P., MILLER, J. B., NUSBAUM, J. D., CORCORAN, D. L., LANGLOIS, C., MUNSCHAUER, M., DEWELL, S., HAFNER, M., WILLIAMS, Z., OHLER, U. & TUSCHL, T. 2012. FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature, 492, 382-6.

ATON, S. J., COLWELL, C. S., HARMAR, A. J., WASCHEK, J. & HERZOG, E. D. 2005. Vasoactive intestinal polypeptide mediates circadian rhythmicity and synchrony in mammalian clock neurons. Nat Neurosci, 8, 476-83.

AU, K. F., SEBASTIANO, V., AFSHAR, P. T., DURRUTHY, J. D., LEE, L., WILLIAMS, B. A., VAN BAKEL, H., SCHADT, E. E., REIJO-PERA, R. A., UNDERWOOD, J. G. & WONG, W. H. 2013. Characterization of the human ESC transcriptome by hybrid sequencing. Proc Natl Acad Sci U S A, 110, E4821-30.

AUSLANDER, L. A., PERRY, W. & JESTE, D. V. 2002. Assessing disturbed thinking and cognition using the Ego Impairment Index in older schizophrenia patients: paranoid vs. nonparanoid distinction. Schizophr Res, 53, 199-207.

AYALEW, M., LE-NICULESCU, H., LEVEY, D. F., JAIN, N., CHANGALA, B., PATEL, S. D., WINIGER, E., BREIER, A., SHEKHAR, A., AMDUR, R., KOLLER, D., NURNBERGER, J. I., CORVIN, A., GEYER, M., TSUANG, M. T., SALOMON, D., SCHORK, N. J., FANOUS, A. H., O'DONOVAN, M. C. & NICULESCU, A. B. 2012. Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. Mol Psychiatry, 17, 887-905.

AZUAJE, F. J. 2014. Selecting biologically informative genes in co-expression networks with a centrality score. Biol Direct, 9, 12.

AZZI, A., DALLMANN, R., CASSERLY, A., REHRAUER, H., PATRIGNANI, A., MAIER, B., KRAMER, A. & BROWN, S. A. 2014. Circadian behavior is light-reprogrammed by plastic DNA methylation. Nat Neurosci, 17, 377-82.

BACHUS, S. E., HYDE, T. M., HERMAN, M. M., EGAN, M. F. & KLEINMAN, J. E. 1997. Abnormal cholecystokinin mRNA levels in entorhinal cortex of schizophrenics. J Psychiatr Res, 31, 233-56.

BAI, F. & WITZMANN, F. A. 2007. Synaptosome proteomics. Subcell Biochem, 43, 77-98.

BAINBRIDGE, M. N., WARREN, R. L., HIRST, M., ROMANUIK, T., ZENG, T., GO, A., DELANEY, A., GRIFFITH, M., HICKENBOTHAM, M., MAGRINI, V., MARDIS, E. R., SADAR, M. D., SIDDIQUI, A. S., MARRA, M. A. & JONES, S. J. 2006. Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC Genomics, 7, 246.

BARANAUSKAS, G., TKATCH, T., NAGATA, K., YEH, J. Z. & SURMEIER, D. J. 2003. Kv3.4 subunits enhance the repolarizing efficiency of Kv3.1 channels in fast-spiking neurons. Nat Neurosci, 6, 258-66.

BARBOSA-MORAIS, N. L., IRIMIA, M., PAN, Q., XIONG, H. Y., GUEROUSSOV, S., LEE, L. J., SLOBODENIUC, V., KUTTER, C., WATT, S., COLAK, R., KIM, T., MISQUITTA-ALI, C. M., WILSON, M. D., KIM, P. M., ODOM, D. T., FREY, B. J. & BLENCOWE, B. J. 2012. The evolutionary landscape of alternative splicing in vertebrate species. Science, 338, 1587-93.

BARK, C., BELLINGER, F. P., KAUSHAL, A., MATHEWS, J. R., PARTRIDGE, L. D. & WILSON, M. C. 2004. Developmentally regulated switch in alternatively spliced SNAP-25 isoforms alters facilitation of synaptic transmission. J Neurosci, 24, 8796-805.

BARK, I. C., HAHN, K. M., RYABININ, A. E. & WILSON, M. C. 1995. Differential expression of SNAP-25 protein isoforms during divergent vesicle fusion events of neural development. Proc Natl Acad Sci U S A, 92, 1510-4.

BARNARD, A. R. & NOLAN, P. M. 2008. When clocks go bad: neurobehavioural consequences of disrupted circadian timing. PLoS Genet, 4, e1000040.

BASS, J. 2012. Circadian topology of metabolism. Nature, 491, 348-56.

BASU, S., CAMPBELL, H. M., DITTEL, B. N. & RAY, A. 2010. Purification of specific cell population by fluorescence activated cell sorting (FACS). J Vis Exp.

BAYLISS, D. A. & BARRETT, P. Q. 2008. Emerging roles for two-pore-domain potassium channels and their potential therapeutic impact. Trends Pharmacol Sci, 29, 566-75.

BECHTOLD, D. A., GIBBS, J. E. & LOUDON, A. S. 2010. Circadian dysfunction in disease. Trends Pharmacol Sci, 31, 191-8.

BELGARD, T. G., MARQUES, A. C., OLIVER, P. L., ABAAN, H. O., SIREY, T. M., HOERDER-SUABEDISSEN, A., GARCIA-MORENO, F., MOLNAR, Z., MARGULIES, E. H. & PONTING, C. P. 2011. A transcriptomic atlas of mouse neocortical layers. Neuron, 71, 605-16.

BENJAMINI, Y. & SPEED, T. P. 2012. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res, 40, e72.

BENLOUCIF, S. & DUBOCOVICH, M. L. 1996. Melatonin and light induce phase shifts of circadian activity rhythms in the C3H/HeN mouse. J Biol Rhythms, 11, 113-25.

BIELLO, S. M., GOLOMBEK, D. A., SCHAK, K. M. & HARRINGTON, M. E. 1997. Circadian phase shifts to neuropeptide Y In vitro: cellular communication and signal transduction. J Neurosci, 17, 8468-75.

BIELLO, S. M., JANIK, D. & MROSOVSKY, N. 1994. Neuropeptide Y and behaviorally induced phase shifts. Neuroscience, 62, 273-9.

BLACK, D. L. 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem, 72, 291-336.

BOIVIN, D. B. 2000. Influence of sleep-wake and circadian rhythm disturbances in psychiatric disorders. J Psychiatry Neurosci, 25, 446-58.

BRADBURY, M. J., DEMENT, W. C. & EDGAR, D. M. 1997. Serotonin-containing fibers in the suprachiasmatic hypothalamus attenuate light-induced phase delays in mice. Brain Res, 768, 125-34.

BRAWAND, D., SOUMILLON, M., NECSULEA, A., JULIEN, P., CSARDI, G., HARRIGAN, P., WEIER, M., LIECHTI, A., AXIMU-PETRI, A., KIRCHER, M., ALBERT, F. W., ZELLER, U., KHAITOVICH, P., GRUTZNER, F., BERGMANN, S., NIELSEN, R., PAABO, S. & KAESSMANN, H. 2011. The evolution of gene expression levels in mammalian organs. Nature, 478, 343-8.

BRAY, M. S., SHAW, C. A., MOORE, M. W., GARCIA, R. A., ZANQUETTA, M. M., DURGAN, D. J., JEONG, W. J., TSAI, J. Y., BUGGER, H., ZHANG, D., ROHRWASSER, A., RENNISON, J. H., DYCK, J. R., LITWIN, S. E., HARDIN, P. E., CHOW, C. W., CHANDLER, M. P., ABEL, E. D. & YOUNG, M. E. 2008. Disruption of the circadian clock within the cardiomyocyte influences myocardial contractile function, metabolism, and gene expression. Am J Physiol Heart Circ Physiol, 294, H1036-47.

BROSIUS, J. 2005. Waste not, want not–transcript excess in multicellular eukaryotes. Trends Genet, 21, 287-8.

BROWN, T. M. & PIGGINS, H. D. 2007. Electrophysiology of the suprachiasmatic circadian clock. Prog Neurobiol, 82, 229-55.

BURGOYNE, R. D. 2007. Neuronal calcium sensor proteins: generating diversity in neuronal Ca2+ signalling. Nat Rev Neurosci, 8, 182-93.

BUTTE, A. J., BAO, L., REIS, B. Y., WATKINS, T. W. & KOHANE, I. S. 2001. Comparing the similarity of time-series gene expression using signal processing metrics. J Biomed Inform, 34, 396-405.

CABEJ, N. R. 2013. Building the Most Complex Structure on Earth: An Epigenetic Narrative of Development and Evolution of Animals. Building the Most Complex

Structure on Earth: An Epigenetic Narrative of Development and Evolution of Animals, 1-298.

CABILI, M. N., TRAPNELL, C., GOFF, L., KOZIOL, M., TAZON-VEGA, B., REGEV, A. & RINN, J. L. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev, 25, 1915-27.

CAGAMPANG, F. R., RATTRAY, M., CAMPBELL, I. C., POWELL, J. F. & COEN, C. W. 1998a. Variation in the expression of the mRNA for protein kinase C isoforms in the rat suprachiasmatic nuclei, caudate putamen and cerebral cortex. Brain Res Mol Brain Res, 53, 277-84.

CAGAMPANG, F. R., SHEWARD, W. J., HARMAR, A. J., PIGGINS, H. D. & COEN, C. W. 1998b. Circadian changes in the expression of vasoactive intestinal peptide 2 receptor mRNA in the rat suprachiasmatic nuclei. Brain Res Mol Brain Res, 54, 108-12.

CAHOY, J. D., EMERY, B., KAUSHAL, A., FOO, L. C., ZAMANIAN, J. L., CHRISTOPHERSON, K. S., XING, Y., LUBISCHER, J. L., KRIEG, P. A., KRUPENKO, S. A., THOMPSON, W. J. & BARRES, B. A. 2008. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci, 28, 264-78.

CAMPOS COSTA, I., NOGUEIRA CARVALHO, H. & FERNANDES, L. 2013. Aging, circadian rhythms and depressive disorders: a review. Am J Neurodegener Dis, 2, 228-46.

CAO, R., GKOGKAS, C. G., DE ZAVALIA, N., BLUM, I. D., YANAGIYA, A., TSUKUMO, Y., XU, H., LEE, C., STORCH, K. F., LIU, A. C., AMIR, S. & SONEN-

BERG, N. 2015. Light-regulated translational control of circadian behavior by eIF4E phosphorylation. Nat Neurosci, 18, 855-62.

CARNINCI, P., KASUKAWA, T., KATAYAMA, S., GOUGH, J., FRITH, M. C., MAEDA, N., OYAMA, R., RAVASI, T., LENHARD, B., WELLS, C., KODZIUS, R., SHIMOKAWA, K., BAJIC, V. B., BRENNER, S. E., BATALOV, S., FORREST, A. R., ZAVOLAN, M., DAVIS, M. J., WILMING, L. G., AIDINIS, V., ALLEN, J. E., AMBESI-IMPIOMBATO, A., APWEILER, R., ATURALIYA, R. N., BAILEY, T. L., BANSAL, M., BAXTER, L., BEISEL, K. W., BERSANO, T., BONO, H., CHALK, A. M., CHIU, K. P., CHOUDHARY, V., CHRISTOFFELS, A., CLUTTERBUCK, D. R., CROWE, M. L., DALLA, E., DALRYMPLE, B. P., DE BONO, B., DELLA GATTA, G., DI BERNARDO, D., DOWN, T., ENGSTROM, P., FAGIOLINI, M., FAULKNER, G., FLETCHER, C. F., FUKUSHIMA, T., FURUNO, M., FUTAKI, S., GARIBOLDI, M., GEORGII-HEMMING, P., GINGERAS, T. R., GOJOBORI, T., GREEN, R. E., GUSTINCICH, S., HARBERS, M., HAYASHI, Y., HENSCH, T. K., HIROKAWA, N., HILL, D., HUMINIECKI, L., IACONO, M., IKEO, K., IWAMA, A., ISHIKAWA, T., JAKT, M., KANAPIN, A., KATOH, M., KAWASAWA, Y., KELSO, J., KITAMURA, H., KITANO, H., KOLLIAS, G., KRISHNAN, S. P., KRUGER, A., KUMMERFELD, S. K., KUROCHKIN, I. V., LAREAU, L. F., LAZAREVIC, D., LIPOVICH, L., LIU, J., LIUNI, S., MCWILLIAM, S., MADAN BABU, M., MADERA, M., MARCHIONNI, L., MATSUDA, H., MATSUZAWA, S., MIKI, H., MIGNONE, F., MIYAKE, S., MORRIS, K., MOTTAGUI-TABAR, S., MULDER, N., NAKANO, N., NAKAUCHI, H., NG, P., NILSSON, R., NISHIGUCHI, S., NISHIKAWA, S., et al. 2005. The transcriptional landscape of the mammalian genome. Science, 309, 1559-63.

CARNINCI, P., SANDELIN, A., LENHARD, B., KATAYAMA, S., SHIMOKAWA, K., PONJAVIC, J., SEMPLE, C. A., TAYLOR, M. S., ENGSTROM, P. G., FRITH, M. C., FORREST, A. R., ALKEMA, W. B., TAN, S. L., PLESSY, C., KODZIUS, R., RAVASI, T., KASUKAWA, T., FUKUDA, S., KANAMORI-KATAYAMA, M., KITAZUME, Y., KAWAJI, H., KAI, C., NAKAMURA, M., KONNO, H., NAKANO, K., MOTTAGUI-TABAR, S., ARNER, P., CHESI, A., GUSTINCICH, S., PERSICHETTI, F., SUZUKI, H., GRIMMOND, S. M., WELLS, C. A., ORLANDO, V., WAHLESTEDT, C., LIU, E. T., HARBERS, M., KAWAI, J., BAJIC, V. B., HUME, D. A. & HAYASHIZAKI, Y. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet, 38, 626-35.

CARRION, A. M., LINK, W. A., LEDO, F., MELLSTROM, B. & NARANJO, J. R. 1999. DREAM is a Ca2+-regulated transcriptional repressor. Nature, 398, 80-4.

CATTERALL, W. A. & FEW, A. P. 2008. Calcium channel regulation and presynaptic plasticity. Neuron, 59, 882-901.

CAVENER, D. R. & RAY, S. C. 1991. Eukaryotic start and stop translation sites. Nucleic Acids Res, 19, 3185-92.

CHALEI, V., SANSOM, S. N., KONG, L., LEE, S., MONTIEL, J. F., VANCE, K. W. & PONTING, C. P. 2014. The long non-coding RNA Dali is an epigenetic regulator of neural differentiation. Elife, 3, e04530.

CHALLET, E., SCARBROUGH, K., PENEV, P. D. & TUREK, F. W. 1998. Roles of suprachiasmatic nuclei and intergeniculate leaflets in mediating the phase-shifting effects of a serotonergic agonist and their photic modulation during subjective day. J Biol Rhythms, 13, 410-21.

CHENG, M. Y., BULLOCK, C. M., LI, C., LEE, A. G., BERMAK, J. C., BELLUZZI, J., WEAVER, D. R., LESLIE, F. M. & ZHOU, Q. Y. 2002. Prokineticin 2 transmits the behavioural circadian rhythm of the suprachiasmatic nucleus. Nature, 417, 405-10.

CHIANG, C. K., MEHTA, N., PATEL, A., ZHANG, P., NING, Z., MAYNE, J., SUN, W. Y., CHENG, H. Y. & FIGEYS, D. 2014. The proteomic landscape of the suprachiasmatic nucleus clock reveals large-scale coordination of key biological processes. PLoS Genet, 10, e1004695.

CHO, W. K., LIAN, S., KIM, S. M., SEO, B. Y., JUNG, J. K. & KIM, K. H. 2015. Time-Course RNA-Seq Analysis Reveals Transcriptional Changes in Rice Plants Triggered by Rice stripe virus Infection. PLoS One, 10, e0136736.

CHODROFF, R. A., GOODSTADT, L., SIREY, T. M., OLIVER, P. L., DAVIES, K. E., GREEN, E. D., MOLNAR, Z. & PONTING, C. P. 2010. Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. Genome Biol, 11, R72.

CHONG, V. Z., THOMPSON, M., BELTAIFA, S., WEBSTER, M. J., LAW, A. J. & WEICKERT, C. S. 2008. Elevated neuregulin-1 and ErbB4 protein in the prefrontal cortex of schizophrenic patients. Schizophr Res, 100, 270-80.

CHOU, W. C., CHENG, A. L., BROTTO, M. & CHUANG, C. Y. 2014. Visual gene-network analysis reveals the cancer gene co-expression in human endometrial cancer. BMC Genomics, 15, 300.

CHU, A., CALDWELL, J. S. & CHEN, Y. A. 2010. Identification and characterization of a small molecule antagonist of human VPAC(2) receptor. Mol Pharmacol, 77, 95-101.

CIRELLI, C. 2005. A molecular window on sleep: changes in gene expression between sleep and wakefulness. Neuroscientist, 11, 63-74.

COGILL, S. B. & WANG, L. 2014. Co-expression Network Analysis of Human lncRNAs and Cancer Genes. Cancer Inform, 13, 49-59.

COHRS, S. 2008. Sleep disturbances in patients with schizophrenia : impact and effect of antipsychotics. CNS Drugs, 22, 939-62.

COLWELL, C. S. 2001. NMDA-evoked calcium transients and currents in the suprachiasmatic nucleus: gating by the circadian system. Eur J Neurosci, 13, 1420-8.

COLWELL, C. S. 2011. Linking neural activity and molecular oscillations in the SCN. Nat Rev Neurosci, 12, 553-69.

COLWELL, C. S. & MENAKER, M. 1992. NMDA as well as non-NMDA receptor antagonists can prevent the phase-shifting effects of light on the circadian system of the golden hamster. J Biol Rhythms, 7, 125-36.

CONDLIFFE, S. B., CORRADINI, I., POZZI, D., VERDERIO, C. & MATTEOLI, M. 2010. Endogenous SNAP-25 regulates native voltage-gated calcium channels in glutamatergic neurons. J Biol Chem, 285, 24968-76.

COON, S. L., MUNSON, P. J., CHERUKURI, P. F., SUGDEN, D., RATH, M. F., MOLLER, M., CLOKIE, S. J., FU, C., OLANICH, M. E., RANGEL, Z., WERNER, T., MULLIKIN, J. C. & KLEIN, D. C. 2012. Circadian changes in long noncoding RNAs in the pineal gland. Proc Natl Acad Sci U S A, 109, 13319-24.

CORRADINI, I., VERDERIO, C., SALA, M., WILSON, M. C. & MATTEOLI, M. 2009. SNAP-25 in neuropsychiatric disorders. Ann N Y Acad Sci, 1152, 93-9.

CRABTREE, G. W. & GOGOS, J. A. 2014. Synaptic plasticity, neural circuits, and the emerging role of altered short-term information processing in schizophrenia. Front Synaptic Neurosci, 6, 28.

CROWLEY, K. 2011. Sleep and sleep disorders in older adults. Neuropsychol Rev, 21, 41-53.

CRUZ, F. C., KOYA, E., GUEZ-BARBER, D. H., BOSSERT, J. M., LUPICA, C. R., SHAHAM, Y. & HOPE, B. T. 2013. New technologies for examining the role of neuronal ensembles in drug addiction and fear. Nat Rev Neurosci, 14, 743-54.

CUI, L. N., CODERRE, E. & RENAUD, L. P. 2001. Glutamate and GABA mediate suprachiasmatic nucleus inputs to spinal-projecting paraventricular neurons. Am J Physiol Regul Integr Comp Physiol, 281, R1283-9.

CUI, M., ZHENG, M., SUN, B., WANG, Y., YE, L. & ZHANG, X. 2015. A long noncoding RNA perturbs the circadian rhythm of hepatoma cells to facilitate hepatocarcinogenesis. Neoplasia, 17, 79-88.

CURTIS, A. M., CHENG, Y., KAPOOR, S., REILLY, D., PRICE, T. S. & FITZGERALD, G. A. 2007. Circadian variation of blood pressure and the vascular response to asynchronous stress. Proc Natl Acad Sci U S A, 104, 3450-5.

CZARNA, A., BERNDT, A., SINGH, H. R., GRUDZIECKI, A., LADURNER, A. G., TIMINSZKY, G., KRAMER, A. & WOLF, E. 2013. Structures of Drosophila cryptochrome and mouse cryptochrome1 provide insight into circadian function. Cell, 153, 1394-405.

CZEISLER, C. A., MOORE-EDE, M. C. & COLEMAN, R. H. 1982. Rotating shift work schedules that disrupt sleep are improved by applying circadian principles. Science, 217, 460-3.

DAMIOLA, F., LE MINH, 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 Dev, 14, 2950-61.

DARNELL, J. C., VAN DRIESCHE, S. J., ZHANG, C., HUNG, K. Y., MELE, A., FRASER, C. E., STONE, E. F., CHEN, C., FAK, J. J., CHI, S. W., LICATALOSI, D. D., RICHTER, J. D. & DARNELL, R. B. 2011. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell, 146, 247-61.

DE RUBEIS, S. & BAGNI, C. 2011. Regulation of molecular pathways in the Fragile X Syndrome: insights into Autism Spectrum Disorders. J Neurodev Disord, 3, 257-69.

DENZLER, R., AGARWAL, V., STEFANO, J., BARTEL, D. P. & STOFFEL, M. 2014. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. Mol Cell, 54, 766-76.

DERRIEN, T., JOHNSON, R., BUSSOTTI, G., TANZER, A., DJEBALI, S., TILGNER, H., GUERNEC, G., MARTIN, D., MERKEL, A., KNOWLES, D. G., LAGARDE, J., VEERAVALLI, L., RUAN, X., RUAN, Y., LASSMANN, T., CARNINCI, P., BROWN, J. B., LIPOVICH, L., GONZALEZ, J. M., THOMAS, M., DAVIS, C. A., SHIEKHATTAR, R., GINGERAS, T. R., HUBBARD, T. J., NOTREDAME, C., HARROW, J. & GUIGO, R. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res, 22, 1775-89.

DERRIGO, M., CESTELLI, A., SAVETTIERI, G. & DI LIEGRO, I. 2000. RNA-protein interactions in the control of stability and localization of messenger RNA (review). Int J Mol Med, 5, 111-23.

DESAN, P. H., OREN, D. A., MALISON, R., PRICE, L. H., ROSENBAUM, J., SMOLLER, J., CHARNEY, D. S. & GELERNTER, J. 2000. Genetic polymorphism at the CLOCK gene locus and major depression. Am J Med Genet, 96, 418-21.

DIBNER, C., SCHIBLER, U. & ALBRECHT, U. 2010. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. Annu Rev Physiol, 72, 517-49.

DILLIES, M. A., RAU, A., AUBERT, J., HENNEQUET-ANTIER, C., JEANMOUGIN, M., SERVANT, N., KEIME, C., MAROT, G., CASTEL, D., ESTELLE, J., GUERNEC, G., JAGLA, B., JOUNEAU, L., LALOE, D., LE GALL, C., SCHAEFFER, B., LE CROM, S., GUEDJ, M. & JAFFREZIC, F. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Brief Bioinform, 14, 671-83.

DING, J. M., BUCHANAN, G. F., TISCHKAU, S. A., CHEN, D., KURIASHKINA, L., FAIMAN, L. E., ALSTER, J. M., MCPHERSON, P. S., CAMPBELL, K. P. & GILLETTE, M. U. 1998. A neuronal ryanodine receptor mediates light-induced phase delays of the circadian clock. Nature, 394, 381-4.

DING, J. M., CHEN, D., WEBER, E. T., FAIMAN, L. E., REA, M. A. & GILLETTE, M. U. 1994. Resetting the biological clock: mediation of nocturnal circadian shifts by glutamate and NO. Science, 266, 1713-7.

DINGER, M. E., PANG, K. C., MERCER, T. R. & MATTICK, J. S. 2008. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. PLoS Comput Biol, 4, e1000176.

DIOUM, E. M., RUTTER, J., TUCKERMAN, J. R., GONZALEZ, G., GILLES-GONZALEZ, M. A. & MCKNIGHT, S. L. 2002. NPAS2: a gas-responsive transcription factor. Science, 298, 2385-7.

DJEBALI, S., DAVIS, C. A., MERKEL, A., DOBIN, A., LASSMANN, T., MORTAZAVI, A., TANZER, A., LAGARDE, J., LIN, W., SCHLESINGER, F., XUE, C., MARINOV, G. K., KHATUN, J., WILLIAMS, B. A., ZALESKI, C., ROZOWSKY, J., RODER, M., KOKOCINSKI, F., ABDELHAMID, R. F., ALIOTO, T., ANTOSHECHKIN, I., BAER, M. T., BAR, N. S., BATUT, P., BELL, K., BELL, I., CHAKRABORTTY, S., CHEN, X., CHRAST, J., CURADO, J., DERRIEN, T., DRENKOW, J., DUMAIS, E., DUMAIS, J., DUTTAGUPTA, R., FALCONNET, E., FASTUCA, M., FEJES-TOTH, K., FERREIRA, P., FOISSAC, S., FULLWOOD, M. J., GAO, H., GONZALEZ, D., GORDON, A., GUNAWARDENA, H., HOWALD, C., JHA, S., JOHNSON, R., KAPRANOV, P., KING, B., KINGSWOOD, C., LUO, O. J., PARK, E., PERSAUD, K., PREALL, J. B., RIBECA, P., RISK, B., ROBYR, D., SAMMETH, M., SCHAFFER, L., SEE, L. H., SHAHAB, A., SKANCKE, J., SUZUKI, A. M., TAKAHASHI, H., TILGNER, H., TROUT, D., WALTERS, N., WANG, H., WROBEL, J., YU, Y., RUAN, X., HAYASHIZAKI, Y., HARROW, J., GERSTEIN, M., HUBBARD, T., REYMOND, A., ANTONARAKIS, S. E., HANNON, G., GIDDINGS, M. C., RUAN, Y., WOLD, B., CARNINCI, P., GUIGO, R. & GINGERAS, T. R. 2012. Landscape of transcription in human cells. Nature, 489, 101-8.

DOBIN, A., DAVIS, C. A., SCHLESINGER, F., DRENKOW, J., ZALESKI, C., JHA, S., BATUT, P., CHAISSON, M. & GINGERAS, T. R. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29, 15-21.

DOHM, J. C., LOTTAZ, C., BORODINA, T. & HIMMELBAUER, H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res, 36, e105.

DOI, M., ISHIDA, A., MIYAKE, A., SATO, M., KOMATSU, R., YAMAZAKI, F., KIMURA, I., TSUCHIYA, S., KORI, H., SEO, K., YAMAGUCHI, Y., MATSUO, M., FUSTIN, J. M., TANAKA, R., SANTO, Y., YAMADA, H., TAKAHASHI, Y., ARAKI, M., NAKAO, K., AIZAWA, S., KOBAYASHI, M., OBRIETAN, K., TSUJIMOTO, G. & OKAMURA, H. 2011. Circadian regulation of intracellular G-protein signalling mediates intercellular synchrony and rhythmicity in the suprachiasmatic nucleus. Nat Commun, 2, 327.

DRUCKER-COLIN, R., AGUILAR-ROBLERO, R., GARCIA-HERNANDEZ, F., FERNANDEZ-CANCINO, F. & BERMUDEZ RATTONI, F. 1984. Fetal suprachiasmatic nucleus transplants: diurnal rhythm recovery of lesioned rats. Brain Res, 311, 353-7.

DULNEVA, A., LEE, S., OLIVER, P. L., DI GLERIA, K., KESSLER, B. M., DAVIES, K. E. & BECKER, E. B. 2015. The mutant Moonwalker TRPC3 channel links calcium signaling to lipid metabolism in the developing cerebellum. Hum Mol Genet, 24, 4114-25.

DUNLAP, J. C. 1999. Molecular bases for circadian clocks. Cell, 96, 271-90.

EASTON, A., ARBUZOVA, J. & TUREK, F. W. 2003. The circadian Clock mutation increases exploratory activity and escape-seeking behavior. Genes Brain Behav, 2, 11-9.

EBLING, F. J. 1996. The role of glutamate in the photic regulation of the suprachiasmatic nucleus. Prog Neurobiol, 50, 109-32.

EBLING, F. J., MAYWOOD, E. S., STALEY, K., HUMBY, T., HANCOCK, D. C., WATERS, C. M., EVANT, G. I. & HASTINGS, M. H. 1991. The role of N-methyl-d-aspartate-type glutamatergic neurotransmission in the photic induction of immediate-early gene expression in the suprachiasmatic nuclei of the Syrian hamster. J Neuroendocrinol, 3, 641-52.

ECKEL-MAHAN, K. L., PATEL, V. R., MOHNEY, R. P., VIGNOLA, K. S., BALDI, P. & SASSONE-CORSI, P. 2012. Coordination of the transcriptome and metabolome by the circadian clock. Proc Natl Acad Sci U S A, 109, 5541-6.

ECKEL-MAHAN, K. L. & STORM, D. R. 2009. Circadian rhythms and memory: not so simple as cogs and gears. EMBO Rep, 10, 584-91.

EDELSTEIN, K. & AMIR, S. 1999. The role of the intergeniculate leaflet in entrainment of circadian rhythms to a skeleton photoperiod. J Neurosci, 19, 372-80.

EDGAR, D. M., MILLER, J. D., PROSSER, R. A., DEAN, R. R. & DEMENT, W. C. 1993. Serotonin and the mammalian circadian system: II. Phase-shifting rat behavioral rhythms with serotonergic agonists. J Biol Rhythms, 8, 17-31.

ELBASHIR, S. M., HARBORTH, J., LENDECKEL, W., YALCIN, A., WEBER, K. & TUSCHL, T. 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature, 411, 494-8.

ELO, L. L. & SCHWIKOWSKI, B. 2013. Analysis of time-resolved gene expression measurements across individuals. PLoS One, 8, e82340.

EPPIG, J. T., BLAKE, J. A., BULT, C. J., KADIN, J. A. & RICHARDSON, J. E. 2015. The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic Acids Res, 43, D726-36.

EULER, P., FRIEDRICH, B., ZIEGLER, R., KUHN, A., LINDENBERG, K. S., WEILLER, C. & ZUCKER, B. 2012. Gene expression analysis on a single cell level in Purkinje cells of Huntington‘s disease transgenic mice. Neurosci Lett, 517, 7-12.

EWING, B. & GREEN, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res, 8, 186-94.

EWING, B., HILLIER, L., WENDL, M. C. & GREEN, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res, 8, 175-85.

FARAONE, S. V., PERLIS, R. H., DOYLE, A. E., SMOLLER, J. W., GORALNICK, J. J., HOLMGREN, M. A. & SKLAR, P. 2005. Molecular genetics of attention-deficit/hyperactivity disorder. Biol Psychiatry, 57, 1313-23.

FERREIRA, J. J., DESBOEUF, K., GALITZKY, M., THALAMAS, C., BREFEL-COURBON, C., FABRE, N., SENARD, J. M., MONTASTRUC, J. L., SAMPAIO, C. & RASCOL, O. 2006. Sleep disruption, daytime somnolence and ‘sleep attacks‘ in Parkinson‘s disease: a clinical survey in PD patients and age-matched healthy volunteers. Eur J Neurol, 13, 209-14.

FERREYRA, G. A. & GOLOMBEK, D. A. 2001. Rhythmicity of the cGMP-related signal transduction pathway in the mammalian circadian system. Am J Physiol Regul Integr Comp Physiol, 280, R1348-55.

FILICHKIN, S. A. & MOCKLER, T. C. 2012a. Unproductive alternative splicing and nonsense mRNAs: A widespread phenomenon among plant circadian clock genes. Biol Direct, 7.

FILICHKIN, S. A. & MOCKLER, T. C. 2012b. Unproductive alternative splicing and nonsense mRNAs: a widespread phenomenon among plant circadian clock genes. Biol Direct, 7, 20.

FILICHKIN, S. A., PRIEST, H. D., GIVAN, S. A., SHEN, R., BRYANT, D. W., FOX, S. E., WONG, W. K. & MOCKLER, T. C. 2010. Genome-wide mapping of alternative splicing in Arabidopsis thaliana. Genome Res, 20, 45-58.

FORABOSCO, P., RAMASAMY, A., TRABZUNI, D., WALKER, R., SMITH, C., BRAS, J., LEVINE, A. P., HARDY, J., POCOCK, J. M., GUERREIRO, R., WEALE, M. E. & RYTEN, M. 2013. Insights into TREM2 biology by network analysis of human brain gene expression data. Neurobiol Aging, 34, 2699-714.

FOSTER, R. G. & HELFRICH-FORSTER, C. 2001. The regulation of circadian clocks by light in fruitflies and mice. Philos Trans R Soc Lond B Biol Sci, 356, 1779-89.

FU, L. & LEE, C. C. 2003. The circadian clock: pacemaker and tumour suppressor. Nat Rev Cancer, 3, 350-61.

FUHR, L., ABREU, M., PETT, P. & RELOGIO, A. 2015. Circadian systems biology: When time matters. Comput Struct Biotechnol J, 13, 417-26.

FUKUHARA, C., BREWER, J. M., DIRDEN, J. C., BITTMAN, E. L., TOSINI, G. & HARRINGTON, M. E. 2001. Neuropeptide Y rapidly reduces Period 1 and Period 2 mRNA levels in the hamster suprachiasmatic nucleus. Neurosci Lett, 314, 119-22.

GANNON, R. L. 2003. Serotonergic serotonin (1A) mixed agonists/antagonists elicit large-magnitude phase shifts in hamster circadian wheel-running rhythms. Neuroscience, 119, 567-76.

GARFIELD, D. A., RUNCIE, D. E., BABBITT, C. C., HAYGOOD, R., NIELSEN, W. J. & WRAY, G. A. 2013. The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network. PLoS Biol, 11, e1001696.

GAU, D., LEMBERGER, T., VON GALL, C., KRETZ, O., LE MINH, N., GASS, P., SCHMID, W., SCHIBLER, U., KORF, H. W. & SCHUTZ, G. 2002. Phosphorylation of CREB Ser142 regulates light-induced phase shifts of the circadian clock. Neuron, 34, 245-53.

GEHMAN, L. T., STOILOV, P., MAGUIRE, J., DAMIANOV, A., LIN, C. H., SHIUE, L., ARES, M., JR., MODY, I. & BLACK, D. L. 2011. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat Genet, 43, 706-11.

GEHRING, W. & ROSBASH, M. 2003. The coevolution of blue-light photoreception and circadian rhythms. J Mol Evol, 57 Suppl 1, S286-9.

GERBER, A., ESNAULT, C., AUBERT, G., TREISMAN, R., PRALONG, F. & SCHIBLER, U. 2013. Blood-borne circadian signal stimulates daily oscillations in actin dynamics and SRF activity. Cell, 152, 492-503.

GILLETTE, M. U. & MITCHELL, J. W. 2002. Signaling in the suprachiasmatic nucleus: selectively responsive and integrative. Cell Tissue Res, 309, 99-107.

GILLETTE, M. U. & WANG, T. A. 2014. Brain circadian oscillators and redox regulation in mammals. Antioxid Redox Signal, 20, 2955-65.

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, 238-41.

GLASS, J. D., SELIM, M. & REA, M. A. 1994. Modulation of light-induced C-Fos expression in the suprachiasmatic nuclei by 5-HT1A receptor agonists. Brain Res, 638, 235-42.

GOLDSTEIN, A. N. & WALKER, M. P. 2014. The role of sleep in emotional brain function. Annu Rev Clin Psychol, 10, 679-708.

GOLOMBEK, D. A., BIELLO, S. M., RENDON, R. A. & HARRINGTON, M. E. 1996. Neuropeptide Y phase shifts the circadian clock in vitro via a Y2 receptor. Neuroreport, 7, 1315-9.

GOLOMBEK, D. A. & RALPH, M. R. 1994. KN-62, an inhibitor of Ca2+/calmodulin kinase II, attenuates circadian responses to light. Neuroreport, 5, 1638-40.

GOLOMBEK, D. A. & ROSENSTEIN, R. E. 2010. Physiology of circadian entrainment. Physiol Rev, 90, 1063-102.

GOWDA, M., JANTASURIYARAT, C., DEAN, R. A. & WANG, G. L. 2004. Robust-LongSAGE (RL-SAGE): a substantially improved LongSAGE method for gene discovery and transcriptome analysis. Plant Physiol, 134, 890-7.

GRANT, G. R., FARKAS, M. H., PIZARRO, A. D., LAHENS, N. F., SCHUG, J., BRUNK, B. P., STOECKERT, C. J., HOGENESCH, J. B. & PIERCE, E. A. 2011. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinformatics, 27, 2518-28.

GROSSMAN, G. H., FARNBAUCH, L. & GLASS, J. D. 2004. Regulation of serotonin release in the Syrian hamster intergeniculate leaflet region. Neuroreport, 15, 103-6.

GROSSMAN, G. H., MISTLBERGER, R. E., ANTLE, M. C., EHLEN, J. C. & GLASS, J. D. 2000. Sleep deprivation stimulates serotonin release in the suprachiasmatic nucleus. Neuroreport, 11, 1929-32.

GUNN, R. K., KEENAN, M. E. & BROWN, R. E. 2011. Analysis of sensory, motor and cognitive functions of the coloboma (C3Sn.Cg-Cm/J) mutant mouse. Genes Brain Behav, 10, 579-88.

GUO, Q., CHENG, Y., LIANG, T., HE, Y., REN, C., SUN, L. & ZHANG, G. 2015. Comprehensive analysis of lncRNA-mRNA co-expression patterns identifies immune-associated lncRNA biomarkers in ovarian cancer malignant progression. Sci Rep, 5, 17683.

GUTTMAN, M., DONAGHEY, J., CAREY, B. W., GARBER, M., GRENIER, J. K., MUNSON, G., YOUNG, G., LUCAS, A. B., ACH, R., BRUHN, L., YANG, X., AMIT, I., MEISSNER, A., REGEV, A., RINN, J. L., ROOT, D. E. & LANDER, E. S. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature, 477, 295-300.

HAAK, L. L. 1999. Metabotropic glutamate receptor modulation of glutamate responses in the suprachiasmatic nucleus. J Neurophysiol, 81, 1308-17.

HAERTY, W. & PONTING, C. P. 2013. Mutations within lncRNAs are effectively selected against in fruitfly but not in human. Genome Biol, 14, R49.

HAHN, C. G., WANG, H. Y., CHO, D. S., TALBOT, K., GUR, R. E., BERRETTINI, W. H., BAKSHI, K., KAMINS, J., BORGMANN-WINTER, K. E., SIEGEL, S. J., GALLOP, R. J. & ARNOLD, S. E. 2006. Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. Nat Med, 12, 824-8.

HALL, J. R., MESSENGER, Z. J., TAM, H. W., PHILLIPS, S. L., RECIO, L. & SMART, R. C. 2015. Long noncoding RNA lincRNA-p21 is the major mediator of UVB-induced and p53-dependent apoptosis in keratinocytes. Cell Death Dis, 6, e1700.

HAMADA, T., ANTLE, M. C. & SILVER, R. 2004. Temporal and spatial expression patterns of canonical clock genes and clock-controlled genes in the suprachiasmatic nucleus. Eur J Neurosci, 19, 1741-8.

HAN, J., DINCULESCU, A., DAI, X., DU, W., SMITH, W. C. & PANG, J. 2013. Review: the history and role of naturally occurring mouse models with Pde6b mutations. Mol Vis, 19, 2579-89.

HANSEN, K. D., BRENNER, S. E. & DUDOIT, S. 2010. Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res, 38, e131.

HANSEN, K. D., IRIZARRY, R. A. & WU, Z. 2012. Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics, 13, 204-16.

HARA, R., WAN, K., WAKAMATSU, H., AIDA, R., MORIYA, T., AKIYAMA, M. & SHIBATA, S. 2001. Restricted feeding entrains liver clock without participation of the suprachiasmatic nucleus. Genes Cells, 6, 269-78.

HARROW, J., FRANKISH, A., GONZALEZ, J. M., TAPANARI, E., DIEKHANS, M., KOKOCINSKI, F., AKEN, B. L., BARRELL, D., ZADISSA, A., SEARLE, S., BARNES, I., BIGNELL, A., BOYCHENKO, V., HUNT, T., KAY, M., MUKHERJEE, G., RAJAN, J., DESPACIO-REYES, G., SAUNDERS, G., STEWARD, C., HARTE, R., LIN, M., HOWALD, C., TANZER, A., DERRIEN, T., CHRAST, J., WALTERS, N., BALASUBRAMANIAN, S., PEI, B., TRESS, M., RODRIGUEZ, J. M., EZKURDIA, I., VAN BAREN, J., BRENT, M., HAUSSLER, D., KELLIS, M., VALENCIA, A., REYMOND, A., GERSTEIN, M., GUIGO, R. & HUBBARD, T. J. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res, 22, 1760-74.

HARTLEY, S. W. & MULLIKIN, J. C. 2015. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. BMC Bioinformatics, 16, 224.

HASTINGS, M. H., REDDY, A. B. & MAYWOOD, E. S. 2003. A clockwork web: circadian timing in brain and periphery, in health and disease. Nat Rev Neurosci, 4, 649-61.

HATORI, M., GILL, S., MURE, L. S., GOULDING, M., O'LEARY, D. D. & PANDA, S. 2014. Lhx1 maintains synchrony among circadian oscillator neurons of the SCN. Elife, 3, e03357.

HEAD, S. R., KOMORI, H. K., LAMERE, S. A., WHISENANT, T., VAN NIEUWERBURGH, F., SALOMON, D. R. & ORDOUKHANIAN, P. 2014. Library construction for next-generation sequencing: overviews and challenges. Biotechniques, 56, 61-4, 66, 68, passim.

HECKMAN, M. 2009. The Circadian Clock and Nr2f2 (COUP-TF) Regulation of Adipogenesis.

HENDRICKSON, A. E., WAGONER, N. & COWAN, W. M. 1972. An autoradiographic and electron microscopic study of retino-hypothalamic connections. Z Zellforsch Mikrosk Anat, 135, 1-26.

HERZOG, E. D., TAKAHASHI, J. S. & BLOCK, G. D. 1998. Clock controls circadian period in isolated suprachiasmatic nucleus neurons. Nat Neurosci, 1, 708-13.

HINOI, E., OGITA, K., TAKEUCHI, Y., OHASHI, H., MARUYAMA, T. & YONEDA, Y. 2001. Characterization with [3H]quisqualate of group I metabotropic glutamate receptor subtype in rat central and peripheral excitable tissues. Neurochem Int, 38, 277-85.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc, 4, 44-57.

HUGHES, A. T. & PIGGINS, H. D. 2008. Behavioral responses of Vipr2-/- mice to light. J Biol Rhythms, 23, 211-9.

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 Genet, 5, e1000442.

HUGHES, M. E., GRANT, G. R., PAQUIN, C., QIAN, J. & NITABACH, M. N. 2012a. Deep sequencing the circadian and diurnal transcriptome of Drosophila brain. Genome Res, 22, 1266-81.

HUGHES, M. E., HOGENESCH, J. B. & KORNACKER, K. 2010. JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. J Biol Rhythms, 25, 372-80.

HUGHES, M. E., HONG, H. K., CHONG, J. L., INDACOCHEA, A. A., LEE, S. S., HAN, M., TAKAHASHI, J. S. & HOGENESCH, J. B. 2012b. Brain-specific rescue of Clock reveals system-driven transcriptional rhythms in peripheral tissue. PLoS Genet, 8, e1002835.

HUHMAN, K. L. & ALBERS, H. E. 1994. Neuropeptide Y microinjected into the suprachiasmatic region phase shifts circadian rhythms in constant darkness. Peptides, 15, 1475-8.

HURLEY, J. M., DASGUPTA, A., EMERSON, J. M., ZHOU, X., RINGELBERG, C. S., KNABE, N., LIPZEN, A. M., LINDQUIST, E. A., DAUM, C. G., BARRY, K. W., GRIGORIEV, I. V., SMITH, K. M., GALAGAN, J. E., BELL-PEDERSEN, D., FREITAG, M., CHENG, C., LOROS, J. J. & DUNLAP, J. C. 2014. Analysis of clock-regulated genes in Neurospora reveals widespread posttranscriptional control of metabolic potential. Proc Natl Acad Sci U S A, 111, 16995-7002.

INOUYE, S. T. & KAWAMURA, H. 1979. Persistence of circadian rhythmicity in a mammalian hypothalamic "island" containing the suprachiasmatic nucleus. Proc Natl Acad Sci U S A, 76, 5962-6.

IWASAKI, S., KATAOKA, M., SEKIGUCHI, M., SHIMAZAKI, Y., SATO, K. & TAKAHASHI, M. 2000. Two distinct mechanisms underlie the stimulation of neurotransmitter release by phorbol esters in clonal rat pheochromocytoma PC12 cells. J Biochem, 128, 407-14.

IYER, R., WANG, T. A. & GILLETTE, M. U. 2014. Circadian gating of neuronal functionality: a basis for iterative metaplasticity. Front Syst Neurosci, 8, 164.

JAGANNATH, A., BUTLER, R., GODINHO, S. I., COUCH, Y., BROWN, L. A., VASUDEVAN, S. R., FLANAGAN, K. C., ANTHONY, D., CHURCHILL, G. C., WOOD, M. J., STEINER, G., EBELING, M., HOSSBACH, M., WETTSTEIN, J. G., DUFFIELD, G. E., GATTI, S., HANKINS, M. W., FOSTER, R. G. & PEIRSON, S. N. 2013. The CRTC1-SIK1 pathway regulates entrainment of the circadian clock. Cell, 154, 1100-11.

JANICH, P., ARPAT, A. B., CASTELO-SZEKELY, V., LOPES, M. & GATFIELD, D. 2015a. Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. Genome Res.

JANICH, P., ARPAT, A. B., CASTELO-SZEKELY, V., LOPES, M. & GATFIELD, D. 2015b. Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames. Genome Res, 25, 1848-59.

JANIK, D. & MROSOVSKY, N. 1994. Intergeniculate leaflet lesions and behaviorally-induced shifts of circadian rhythms. Brain Res, 651, 174-82.

JEANS, A. F., OLIVER, P. L., JOHNSON, R., CAPOGNA, M., VIKMAN, J., MOLNAR, Z., BABBS, A., PARTRIDGE, C. J., SALEHI, A., BENGTSSON, M., ELIASSON, L., RORSMAN, P. & DAVIES, K. E. 2007. A dominant mutation in Snap25 causes impaired vesicle trafficking, sensorimotor gating, and ataxia in the blind-drunk mouse. Proc Natl Acad Sci U S A, 104, 2431-6.

JEYARAJ, D., HALDAR, S. M., WAN, X., MCCAULEY, M. D., RIPPERGER, J. A., HU, K., LU, Y., EAPEN, B. L., SHARMA, N., FICKER, E., CUTLER, M. J., GULICK, J., SANBE, A., ROBBINS, J., DEMOLOMBE, S., KONDRATOV, R. V.,

SHEA, S. A., ALBRECHT, U., WEHRENS, X. H., ROSENBAUM, D. S. & JAIN, M. K. 2012. Circadian rhythms govern cardiac repolarization and arrhythmogenesis. Nature, 483, 96-9.

JIANG, L., SCHLESINGER, F., DAVIS, C. A., ZHANG, Y., LI, R., SALIT, M., GINGERAS, T. R. & OLIVER, B. 2011. Synthetic spike-in standards for RNA-seq experiments. Genome Res, 21, 1543-51.

JOHANSSON, J. U., ERICSSON, J., JANSON, J., BERAKI, S., STANIC, D., MANDIC, S. A., WIKSTROM, M. A., HOKFELT, T., OGREN, S. O., ROZELL, B., BERGGREN, P. O. & BARK, C. 2008. An ancient duplication of exon 5 in the Snap25 gene is required for complex neuronal development/function. PLoS Genet, 4, e1000278.

JOHNSON, R. F., MORIN, L. P. & MOORE, R. Y. 1988. Retinohypothalamic projections in the hamster and rat demonstrated using cholera toxin. Brain Res, 462, 301-12.

JONES, M. W., ERRINGTON, M. L., FRENCH, P. J., FINE, A., BLISS, T. V., GAREL, S., CHARNAY, P., BOZON, B., LAROCHE, S. & DAVIS, S. 2001. A requirement for the immediate early gene Zif268 in the expression of late LTP and long-term memories. Nat Neurosci, 4, 289-96.

JONES, S. H. 2001. Circadian rhythms, multilevel models of emotion and bipolar disorder–an initial step towards integration? Clin Psychol Rev, 21, 1193-209.

JOUFFE, C., CRETENET, G., SYMUL, L., MARTIN, E., ATGER, F., NAEF, F. & GACHON, F. 2013. The circadian clock coordinates ribosome biogenesis. PLoS Biol, 11, e1001455.

JUD, C., SCHMUTZ, I., HAMPP, G., OSTER, H. & ALBRECHT, U. 2005. A guideline for analyzing circadian wheel-running behavior in rodents under different lighting conditions. Biol Proced Online, 7, 101-16.

JURADO, S., GOSWAMI, D., ZHANG, Y., MOLINA, A. J., SUDHOF, T. C. & MALENKA, R. C. 2013. LTP requires a unique postsynaptic SNARE fusion machinery. Neuron, 77, 542-58.

KAASIK, K. & LEE, C. C. 2004. Reciprocal regulation of haem biosynthesis and the circadian clock in mammals. Nature, 430, 467-71.

KADENER, S., MENET, J. S., SUGINO, K., HORWICH, M. D., WEISSBEIN, U., NAWATHEAN, P., VAGIN, V. V., ZAMORE, P. D., NELSON, S. B. & ROSBASH, M. 2009. A role for microRNAs in the Drosophila circadian clock. Genes Dev, 23, 2179-91.

KAIKKONEN, M. U., SPANN, N. J., HEINZ, S., ROMANOSKI, C. E., ALLISON, K. A., STENDER, J. D., CHUN, H. B., TOUGH, D. F., PRINJHA, R. K., BENNER, C. & GLASS, C. K. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. Mol Cell, 51, 310-25.

KANTROWITZ, J. T., OAKMAN, E., BICKEL, S., CITROME, L., SPIELMAN, A., SILIPO, G., BATTAGLIA, J. & JAVITT, D. C. 2010. The importance of a good night‘s sleep: an open-label trial of the sodium salt of gamma-hydroxybutyric acid in insomnia associated with schizophrenia. Schizophr Res, 120, 225-6.

KAPP, L. D. & LORSCH, J. R. 2004. The molecular mechanics of eukaryotic translation. Annu Rev Biochem, 73, 657-704.

KARATSOREOS, I. N. 2014. Links between Circadian Rhythms and Psychiatric Disease. Front Behav Neurosci, 8, 162.

KARATSOREOS, I. N., YAN, L., LESAUTER, J. & SILVER, R. 2004. Phenotype matters: identification of light-responsive cells in the mouse suprachiasmatic nucleus. J Neurosci, 24, 68-75.

KARL, T., BURNE, T. H. & HERZOG, H. 2006. Effect of Y1 receptor deficiency on motor activity, exploration, and anxiety. Behav Brain Res, 167, 87-93.

KATAOKA, M., YAMAMORI, S., SUZUKI, E., WATANABE, S., SATO, T., MIYAOKA, H., AZUMA, S., IKEGAMI, S., KUWAHARA, R., SUZUKI-MIGISHIMA, R., NAKAHARA, Y., NIHONMATSU, I., INOKUCHI, K., KATOH-FUKUI, Y., YOKOYAMA, M. & TAKAHASHI, M. 2011. A single amino acid mutation in SNAP-25 induces anxiety-related behavior in mouse. PLoS One, 6, e25158.

KAWAGUCHI, C., TANAKA, K., ISOJIMA, Y., SHINTANI, N., HASHIMOTO, H., BABA, A. & NAGAI, K. 2003. Changes in light-induced phase shift of circadian rhythm in mice lacking PACAP. Biochem Biophys Res Commun, 310, 169-75.

KEIFER, J. & ZHENG, Z. 2010. AMPA receptor trafficking and learning. Eur J Neurosci, 32, 269-77.

KESHAVAN, M. S., MEHTA, U. M., PADMANABHAN, J. L. & SHAH, J. L. 2015. Dysplasticity, metaplasticity, and schizophrenia: Implications for risk, illness, and novel interventions. Dev Psychopathol, 27, 615-35.

KESSELS, H. W. & MALINOW, R. 2009. Synaptic AMPA receptor plasticity and behavior. Neuron, 61, 340-50.

KHAN, S. K., XU, H., UKAI-TADENUMA, M., BURTON, B., WANG, Y., UEDA, H. R. & LIU, A. C. 2012. Identification of a novel cryptochrome differentiating domain required for feedback repression in circadian clock function. J Biol Chem, 287, 25917-26.

KHVOROVA, A. & WOLFSON, A. 2012. New competition in RNA regulation. Nat Biotechnol, 30, 58-9.

KIM, D. Y., CHOI, H. J., KIM, J. S., KIM, Y. S., JEONG, D. U., SHIN, H. C., KIM, M. J., HAN, H. C., HONG, S. K. & KIM, Y. I. 2005. Voltage-gated calcium channels play crucial roles in the glutamate-induced phase shifts of the rat suprachiasmatic circadian clock. Eur J Neurosci, 21, 1215-22.

KIM, T. K., HEMBERG, M., GRAY, J. M., COSTA, A. M., BEAR, D. M., WU, J., HARMIN, D. A., LAPTEWICZ, M., BARBARA-HALEY, K., KUERSTEN, S., MARKENSCOFF-PAPADIMITRIOU, E., KUHL, D., BITO, H., WORLEY, P. F., KREIMAN, G. & GREENBERG, M. E. 2010. Widespread transcription at neuronal activity-regulated enhancers. Nature, 465, 182-7.

KISS, T. 2002. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell, 109, 145-8.

KO, C. H. & TAKAHASHI, J. S. 2006. Molecular components of the mammalian circadian clock. Hum Mol Genet, 15 Spec No 2, R271-7.

KOEFOED, P., HANSEN, T. V., WOLDBYE, D. P., WERGE, T., MORS, O., HANSEN, T., JAKOBSEN, K. D., NORDENTOFT, M., WANG, A., BOLWIG, T. G. & REHFELD, J. F. 2009. An intron 1 polymorphism in the cholecystokinin-A receptor gene associated with schizophrenia in males. Acta Psychiatr Scand, 120, 281-7.

KOGEL, U., GONZALEZ SUAREZ, I., XIANG, Y., DOSSIN, E., GUY, P. A., MATHIS, C., MARESCOTTI, D., GOEDERTIER, D., MARTIN, F., PEITSCH, M. C. & HOENG, J. 2015. Biological impact of cigarette smoke compared to an aerosol produced

from a prototypic modified risk tobacco product on normal human bronchial epithelial cells. Toxicol In Vitro.

KOHLI, M., YOUNG, C. Y., TINDALL, D. J., NANDY, D., MCKENZIE, K. M., BEVAN, G. H. & DONKENA, K. V. 2015. Whole blood defensin mRNA expression is a predictive biomarker of docetaxel response in castration-resistant prostate cancer. Onco Targets Ther, 8, 1915-22.

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, 349-54.

KONDRATOVA, A. A. & KONDRATOV, R. V. 2012. The circadian clock and pathology of the ageing brain. Nat Rev Neurosci, 13, 325-35.

KONG, L., ZHANG, Y., YE, Z. Q., LIU, X. Q., ZHAO, S. Q., WEI, L. & GAO, G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res, 35, W345-9.

KONONENKO, N. I., SHAO, L. R. & DUDEK, F. E. 2004. Riluzole-sensitive slowly inactivating sodium current in rat suprachiasmatic nucleus neurons. J Neurophysiol, 91, 710-8.

KORNMANN, B., SCHAAD, O., BUJARD, H., TAKAHASHI, J. S. & SCHIBLER, U. 2007. System-driven and oscillator-dependent circadian transcription in mice with a conditionally active liver clock. PLoS Biol, 5, e34.

KOSIR, R., ACIMOVIC, J., GOLICNIK, M., PERSE, M., MAJDIC, G., FINK, M. & ROZMAN, D. 2010. Determination of reference genes for circadian studies in different tissues and mouse strains. BMC Mol Biol, 11, 60.

KRISHNAN, N., DAVIS, A. J. & GIEBULTOWICZ, J. M. 2008. Circadian regulation of response to oxidative stress in Drosophila melanogaster. Biochem Biophys Res Commun, 374, 299-303.

KRISHNAN, N., KRETZSCHMAR, D., RAKSHIT, K., CHOW, E. & GIEBULTOWICZ, J. M. 2009. The circadian clock gene period extends healthspan in aging Drosophila melanogaster. Aging (Albany NY), 1, 937-48.

KUDO, T., LOH, D. H., KULJIS, D., CONSTANCE, C. & COLWELL, C. S. 2011. Fast delayed rectifier potassium current: critical for input and output of the circadian system. J Neurosci, 31, 2746-55.

KUNG, J. T., COLOGNORI, D. & LEE, J. T. 2013. Long noncoding RNAs: past, present, and future. Genetics, 193, 651-69.

KUNST, M., HUGHES, M. E., RACCUGLIA, D., FELIX, M., LI, M., BARNETT, G., DUAH, J. & NITABACH, M. N. 2014. Calcitonin gene-related peptide neurons mediate sleep-specific circadian output in Drosophila. Curr Biol, 24, 2652-64.

KYRIACOU, C. P. & HASTINGS, M. H. 2010. Circadian clocks: genes, sleep, and cognition. Trends Cogn Sci, 14, 259-67.

LAFONTAINE, I. & DUJON, B. 2010. Origin and fate of pseudogenes in Hemiascomycetes: a comparative analysis. BMC Genomics, 11, 260.

LAM, M. T., CHO, H., LESCH, H. P., GOSSELIN, D., HEINZ, S., TANAKA-OISHI, Y., BENNER, C., KAIKKONEN, M. U., KIM, A. S., KOSAKA, M., LEE, C. Y., WATT, A., GROSSMAN, T. R., ROSENFELD, M. G., EVANS, R. M. & GLASS, C. K. 2013. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature, 498, 511-5.

LAM, M. T., LI, W., ROSENFELD, M. G. & GLASS, C. K. 2014. Enhancer RNAs and regulated transcriptional programs. Trends Biochem Sci, 39, 170-82.

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, 552-6.

LAMIA, K. A., STORCH, K. F. & WEITZ, C. J. 2008. Physiological significance of a peripheral tissue circadian clock. Proc Natl Acad Sci U S A, 105, 15172-7.

LANCASTER, M. A. & GLEESON, J. G. 2009. The primary cilium as a cellular signaling center: lessons from disease. Curr Opin Genet Dev, 19, 220-9.

LANDGRAF, D., MCCARTHY, M. J. & WELSH, D. K. 2014. The role of the circadian clock in animal models of mood disorders. Behav Neurosci, 128, 344-59.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9, 559.

LEE, S., BOOKOUT, A. L., LEE, C. E., GAUTRON, L., HARPER, M. J., ELIAS, C. F., LOWELL, B. B. & ELMQUIST, J. K. 2012. Laser-capture microdissection and transcriptional profiling of the dorsomedial nucleus of the hypothalamus. J Comp Neurol, 520, 3617-32.

LEEK, J. T., JOHNSON, W. E., PARKER, H. S., JAFFE, A. E. & STOREY, J. D. 2012. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics, 28, 882-3.

LEEK, J. T., SCHARPF, R. B., BRAVO, H. C., SIMCHA, D., LANGMEAD, B., JOHNSON, W. E., GEMAN, D., BAGGERLY, K. & IRIZARRY, R. A. 2010. Tackling

the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet, 11, 733-9.

LEMING, M. T., RUND, S. S., BEHURA, S. K., DUFFIELD, G. E. & O'TOUSA, J. E. 2014. A database of circadian and diel rhythmic gene expression in the yellow fever mosquito Aedes aegypti. BMC Genomics, 15, 1128.

LEWEJOHANN, L., SKRYABIN, B. V., SACHSER, N., PREHN, C., HEIDUSCHKA, P., THANOS, S., JORDAN, U., DELL'OMO, G., VYSSOTSKI, A. L., PLESKACHEVA, M. G., LIPP, H. P., TIEDGE, H., BROSIUS, J. & PRIOR, H. 2004. Role of a neuronal small non-messenger RNA: behavioural alterations in BC1 RNA-deleted mice. Behav Brain Res, 154, 273-89.

LEWIS, C. M., LEVINSON, D. F., WISE, L. H., DELISI, L. E., STRAUB, R. E., HOVATTA, I., WILLIAMS, N. M., SCHWAB, S. G., PULVER, A. E., FARAONE, S. V., BRZUSTOWICZ, L. M., KAUFMANN, C. A., GARVER, D. L., GURLING, H. M., LINDHOLM, E., COON, H., MOISES, H. W., BYERLEY, W., SHAW, S. H., MESEN, A., SHERRINGTON, R., O'NEILL, F. A., WALSH, D., KENDLER, K. S., EKELUND, J., PAUNIO, T., LONNQVIST, J., PELTONEN, L., O'DONOVAN, M. C., OWEN, M. J., WILDENAUER, D. B., MAIER, W., NESTADT, G., BLOUIN, J. L., ANTONARAKIS, S. E., MOWRY, B. J., SILVERMAN, J. M., CROWE, R. R., CLONINGER, C. R., TSUANG, M. T., MALASPINA, D., HARKAVY-FRIEDMAN, J. M., SVRAKIC, D. M., BASSETT, A. S., HOLCOMB, J., KALSI, G., MCQUILLIN, A., BRYNJOLFSON, J., SIGMUNDSSON, T., PETURSSON, H., JAZIN, E., ZOEGA, T. & HELGASON, T. 2003. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. Am J Hum Genet, 73, 34-48.

LI, J. D., BURTON, K. J., ZHANG, C., HU, S. B. & ZHOU, Q. Y. 2009. Vasopressin receptor V1a regulates circadian rhythms of locomotor activity and expression of clock-controlled genes in the suprachiasmatic nuclei. Am J Physiol Regul Integr Comp Physiol, 296, R824-30.

LI, J. Z., BUNNEY, B. G., MENG, F., HAGENAUER, M. H., WALSH, D. M., VAWTER, M. P., EVANS, S. J., CHOUDARY, P. V., CARTAGENA, P., BARCHAS, J. D., SCHATZBERG, A. F., JONES, E. G., MYERS, R. M., WATSON, S. J., JR., AKIL, H. & BUNNEY, W. E. 2013. Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. Proc Natl Acad Sci U S A, 110, 9950-5.

LI, Q., LOH, D. H., KUDO, T., TRUONG, D., DERAKHSHESH, M., KASWAN, Z. M., GHIANI, C. A., TSOA, R., CHENG, Y., SUN, Y. E. & COLWELL, C. S. 2015. Circadian rhythm disruption in a mouse model of Rett syndrome circadian disruption in RTT. Neurobiol Dis, 77, 155-64.

LI, W., DAI, C., KANG, S. & ZHOU, X. J. 2014. Integrative analysis of many RNA-seq datasets to study alternative splicing. Methods, 67, 313-24.

LIU, C., WEAVER, D. R., STROGATZ, S. H. & REPPERT, S. M. 1997. Cellular construction of a circadian clock: period determination in the suprachiasmatic nuclei. Cell, 91, 855-60.

LIU, S., LAMAZE, A., LIU, Q., TABUCHI, M., YANG, Y., FOWLER, M., BHARADWAJ, R., ZHANG, J., BEDONT, J., BLACKSHAW, S., LLOYD, T. E., MONTELL, C., SEHGAL, A., KOH, K. & WU, M. N. 2014. WIDE AWAKE mediates the circadian timing of sleep onset. Neuron, 82, 151-66.

LOUDON, A. S. 2012. Circadian biology: a 2.5 billion year old clock. Curr Biol, 22, R570-1.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol, 15, 550.

LOVEN, J., ORLANDO, D. A., SIGOVA, A. A., LIN, C. Y., RAHL, P. B., BURGE, C. B., LEVENS, D. L., LEE, T. I. & YOUNG, R. A. 2012. Revisiting global gene expression analysis. Cell, 151, 476-82.

LUDEWIG, K., GEYER, M. A. & VOLLENWEIDER, F. X. 2003. Deficits in prepulse inhibition and habituation in never-medicated, first-episode schizophrenia. Biol Psychiatry, 54, 121-8.

LYNCH, M. A. 2004. Long-term potentiation and memory. Physiol Rev, 84, 87-136.

MACOSKO, E. Z., BASU, A., SATIJA, R., NEMESH, J., SHEKHAR, K., GOLDMAN, M., TIROSH, I., BIALAS, A. R., KAMITAKI, N., MARTERSTECK, E. M., TROMBETTA, J. J., WEITZ, D. A., SANES, J. R., SHALEK, A. K., REGEV, A. & MCCARROLL, S. A. 2015. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell, 161, 1202-14.

MARCHEVA, B., RAMSEY, K. M., BUHR, E. D., KOBAYASHI, Y., SU, H., KO, C. H., IVANOVA, G., OMURA, C., MO, S., VITATERNA, M. H., LOPEZ, J. P., PHILIPSON, L. H., BRADFIELD, C. A., CROSBY, S. D., JEBAILEY, L., WANG, X., TAKAHASHI, J. S. & BASS, J. 2010. Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. Nature, 466, 627-31.

MATHUR, A., GOLOMBEK, D. A. & RALPH, M. R. 1996. cGMP-dependent protein kinase inhibitors block light-induced phase advances of circadian rhythms in vivo. Am J Physiol, 270, R1031-6.

MATSUMURA, H., REICH, S., ITO, A., SAITOH, H., KAMOUN, S., WINTER, P., KAHL, G., REUTER, M., KRUGER, D. H. & TERAUCHI, R. 2003. Gene expression analysis of plant host-pathogen interactions by SuperSAGE. Proc Natl Acad Sci U S A, 100, 15718-23.

MAURY, E., HONG, H. K. & BASS, J. 2014. Circadian disruption in the pathogenesis of metabolic syndrome. Diabetes Metab, 40, 338-46.

MAYWOOD, E. S. & MROSOVSKY, N. 2001. A molecular explanation of interactions between photic and non-photic circadian clock-resetting stimuli. Brain Res Gene Expr Patterns, 1, 27-31.

MAYWOOD, E. S., REDDY, A. B., WONG, G. K., O'NEILL, J. S., O'BRIEN, J. A., MCMAHON, D. G., HARMAR, A. J., OKAMURA, H. & HASTINGS, M. H. 2006. Synchronization and maintenance of timekeeping in suprachiasmatic circadian clock cells by neuropeptidergic signaling. Curr Biol, 16, 599-605.

MCLEAN, C. Y., BRISTOR, D., HILLER, M., CLARKE, S. L., SCHAAR, B. T., LOWE, C. B., WENGER, A. M. & BEJERANO, G. 2010. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol, 28, 495-501.

MEDANIC, M. & GILLETTE, M. U. 1992. Serotonin regulates the phase of the rat suprachiasmatic circadian pacemaker in vitro only during the subjective day. J Physiol, 450, 629-42.

MEDANIC, M. & GILLETTE, M. U. 1993. Suprachiasmatic circadian pacemaker of rat shows two windows of sensitivity to neuropeptide Y in vitro. Brain Res, 620, 281-6.

MELO, C. A., DROST, J., WIJCHERS, P. J., VAN DE WERKEN, H., DE WIT, E., OUDE VRIELINK, J. A., ELKON, R., MELO, S. A., LEVEILLE, N., KALLURI, R., DE LAAT, W. & AGAMI, R. 2013. eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol Cell, 49, 524-35.

MENDLIN, A., MARTIN, F. J., RUETER, L. E. & JACOBS, B. L. 1996. Neuronal release of serotonin in the cerebellum of behaving rats: an in vivo microdialysis study. J Neurochem, 67, 617-22.

MENET, J. S., RODRIGUEZ, J., ABRUZZI, K. C. & ROSBASH, M. 2012. Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. Elife, 1, e00011.

MERCER, T. R., DINGER, M. E., SUNKIN, S. M., MEHLER, M. F. & MATTICK, J. S. 2008. Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A, 105, 716-21.

MERKIN, J., RUSSELL, C., CHEN, P. & BURGE, C. B. 2012. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. Science, 338, 1593-9.

MEYER-BERNSTEIN, E. L. & MORIN, L. P. 1996. Differential serotonergic innervation of the suprachiasmatic nucleus and the intergeniculate leaflet and its role in circadian rhythm modulation. J Neurosci, 16, 2097-111.

MEYER-BERNSTEIN, E. L. & MORIN, L. P. 1999. Electrical stimulation of the median or dorsal raphe nuclei reduces light-induced FOS protein in the suprachiasmatic nucleus and causes circadian activity rhythm phase shifts. Neuroscience, 92, 267-79.

MILLAR, J. K., PICKARD, B. S., MACKIE, S., JAMES, R., CHRISTIE, S., BUCHANAN, S. R., MALLOY, M. P., CHUBB, J. E., HUSTON, E., BAILLIE, G. S., THOMSON, P. A., HILL, E. V., BRANDON, N. J., RAIN, J. C., CAMARGO, L. M., WHITING, P. J., HOUSLAY, M. D., BLACKWOOD, D. H., MUIR, W. J. & PORTEOUS, D. J. 2005. DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. Science, 310, 1187-91.

MISTLBERGER, R. E. & SKENE, D. J. 2004. Social influences on mammalian circadian rhythms: animal and human studies. Biol Rev Camb Philos Soc, 79, 533-56.

MITTERAUER, B. 2000. Clock genes, feedback loops and their possible role in the etiology of bipolar disorders: an integrative model. Med Hypotheses, 55, 155-9.

MOGHADDAM, B. 2004. Targeting metabotropic glutamate receptors for treatment of the cognitive symptoms of schizophrenia. Psychopharmacology (Berl), 174, 39-44.

MOHAWK, J. A., GREEN, C. B. & TAKAHASHI, J. S. 2012. Central and peripheral circadian clocks in mammals. Annu Rev Neurosci, 35, 445-62.

MOORE, R. Y. 1973. Retinohypothalamic projection in mammals: a comparative study. Brain Res, 49, 403-9.

MOORE, R. Y. & EICHLER, V. B. 1972. Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. Brain Res, 42, 201-6.

MOORE, R. Y. & KLEIN, D. C. 1974. Visual pathways and the central neural control of a circadian rhythm in pineal serotonin N-acetyltransferase activity. Brain Res, 71, 17-33.

MORENO, D., GOURLET, P., DE NEEF, P., CNUDDE, J., WAELBROECK, M. & ROBBERECHT, P. 2000. Development of selective agonists and antagonists for the human vasoactive intestinal polypeptide VPAC(2) receptor. Peptides, 21, 1543-9.

MORIN, L. P. 2013. Neuroanatomy of the extended circadian rhythm system. Exp Neurol, 243, 4-20.

MORRIS, K. V. & MATTICK, J. S. 2014. The rise of regulatory RNA. Nat Rev Genet, 15, 423-37.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods, 5, 621-8.

MORTON, A. J., WOOD, N. I., HASTINGS, M. H., HURELBRINK, C., BARKER, R. A. & MAYWOOD, E. S. 2005. Disintegration of the sleep-wake cycle and circadian timing in Huntington‘s disease. J Neurosci, 25, 157-63.

MOUSAVI, K., ZARE, H., DELL‘ORSO, S., GRONTVED, L., GUTIERREZ-CRUZ, G., DERFOUL, A., HAGER, G. L. & SARTORELLI, V. 2013. eRNAs promote transcription by establishing chromatin accessibility at defined genomic loci. Mol Cell, 51, 606-17.

MROSOVSKY, N. 1995. A non-photic gateway to the circadian clock of hamsters. Ciba Found Symp, 183, 154-67; discussion 167-74.

MUKHOPADHYAY, S. & ROHATGI, R. 2014. G-protein-coupled receptors, Hedgehog signaling and primary cilia. Semin Cell Dev Biol, 33, 63-72.

MURRAY, G. & HARVEY, A. 2010. Circadian rhythms and sleep in bipolar disorder. Bipolar Disord, 12, 459-72.

MUSIEK, E. S., LIM, M. M., YANG, G., BAUER, A. Q., QI, L., LEE, Y., ROH, J. H., ORTIZ-GONZALEZ, X., DEARBORN, J. T., CULVER, J. P., HERZOG, E. D., HOGENESCH, J. B., WOZNIAK, D. F., DIKRANIAN, K., GIASSON, B. I., WEAVER, D. R., HOLTZMAN, D. M. & FITZGERALD, G. A. 2013. Circadian clock proteins regulate neuronal redox homeostasis and neurodegeneration. J Clin Invest, 123, 5389-400.

MUSUNURU, K., STRONG, A., FRANK-KAMENETSKY, M., LEE, N. E., AH-FELDT, T., SACHS, K. V., LI, X., LI, H., KUPERWASSER, N., RUDA, V. M., PIRRUC-CELLO, J. P., MUCHMORE, B., PROKUNINA-OLSSON, L., HALL, J. L., SCHADT, E. E., MORALES, C. R., LUND-KATZ, S., PHILLIPS, M. C., WONG, J., CANTLEY, W., RACIE, T., EJEBE, K. G., ORHO-MELANDER, M., MELANDER, O., KOTELIAN-SKY, V., FITZGERALD, K., KRAUSS, R. M., COWAN, C. A., KATHIRESAN, S. & RADER, D. J. 2010. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature, 466, 714-9.

MYERS, E., STARTUP, H. & FREEMAN, D. 2011. Cognitive behavioural treatment of insomnia in individuals with persistent persecutory delusions: a pilot trial. J Behav Ther Exp Psychiatry, 42, 330-6.

NAGALAKSHMI, U., WANG, Z., WAERN, K., SHOU, C., RAHA, D., GERSTEIN, M. & SNYDER, M. 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science, 320, 1344-9.

NAGOSHI, R. N. & BAKER, B. S. 1990. Regulation of sex-specific RNA splicing at the Drosophila doublesex gene: cis-acting mutations in exon sequences alter sex-specific RNA splicing patterns. Genes Dev, 4, 89-97.

NAISMITH, S. L., HICKIE, I. B., TERPENING, Z., RAJARATNAM, S. M., HODGES, J. R., BOLITHO, S., ROGERS, N. L. & LEWIS, S. J. 2014. Circadian misalignment and sleep disruption in mild cognitive impairment. J Alzheimers Dis, 38, 857-66.

NAKAGAWA, S., NAGANUMA, T., SHIOI, G. & HIROSE, T. 2011. Paraspeckles are subpopulation-specific nuclear bodies that are not essential in mice. J Cell Biol, 193, 31-9.

NAKAMURA, N., RUEBEL, K., JIN, L., QIAN, X., ZHANG, H. & LLOYD, R. V. 2007. Laser capture microdissection for analysis of single cells. Methods Mol Med, 132, 11-8.

NAKAMURA, W., HONMA, S., SHIRAKAWA, T. & HONMA, K. 2002. Clock mutation lengthens the circadian period without damping rhythms in individual SCN neurons. Nat Neurosci, 5, 399-400.

NAKATA, Y., YASUDA, T., FUKAYA, M., YAMAMORI, S., ITAKURA, M., NIHIRA, T., HAYAKAWA, H., KAWANAMI, A., KATAOKA, M., NAGAI, M., SAKAGAMI, H., TAKAHASHI, M., MIZUNO, Y. & MOCHIZUKI, H. 2012. Accumulation of alpha-synuclein triggered by presynaptic dysfunction. J Neurosci, 32, 17186-96.

NAPOLI, I., MERCALDO, V., BOYL, P. P., ELEUTERI, B., ZALFA, F., DE RUBEIS, S., DI MARINO, D., MOHR, E., MASSIMI, M., FALCONI, M., WITKE, W., COSTA-MATTIOLI, M., SONENBERG, N., ACHSEL, T. & BAGNI, C. 2008. The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP. Cell, 134, 1042-54.

NILSSON, T., MANN, M., AEBERSOLD, R., YATES, J. R., 3RD, BAIROCH, A. & BERGERON, J. J. 2010. Mass spectrometry in high-throughput proteomics: ready for the big time. Nat Methods, 7, 681-5.

NOBLE, F., WANK, S. A., CRAWLEY, J. N., BRADWEJN, J., SEROOGY, K. B., HAMON, M. & ROQUES, B. P. 1999. International Union of Pharmacology. XXI. Structure, distribution, and functions of cholecystokinin receptors. Pharmacol Rev, 51, 745-81.

NOLAN, P. M., HUGILL, A. & COX, R. D. 2002. ENU mutagenesis in the mouse: application to human genetic disease. Brief Funct Genomic Proteomic, 1, 278-89.

NOLAN, P. M., PETERS, J., STRIVENS, M., ROGERS, D., HAGAN, J., SPURR, N., GRAY, I. C., VIZOR, L., BROOKER, D., WHITEHILL, E., WASHBOURNE, R., HOUGH, T., GREENAWAY, S., HEWITT, M., LIU, X., MCCORMACK, S., PICKFORD, K., SELLEY, R., WELLS, C., TYMOWSKA-LALANNE, Z., ROBY, P., GLENISTER, P., THORNTON, C., THAUNG, C., STEVENSON, J. A., ARKELL, R., MBURU, P., HARDISTY, R., KIERNAN, A., ERVEN, A., STEEL, K. P., VOEGELING, S., GUENET, J. L., NICKOLS, C., SADRI, R., NASSE, M., ISAACS, A., DAVIES, K., BROWNE, M., FISHER, E. M., MARTIN, J., RASTAN, S., BROWN, S. D. & HUNTER, J. 2000. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. Nat Genet, 25, 440-3.

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, 949-53.

O'NEILL, J. S. & REDDY, A. B. 2011. Circadian clocks in human red blood cells. Nature, 469, 498-503.

OHIRA, K., KOBAYASHI, K., TOYAMA, K., NAKAMURA, H. K., SHOJI, H., TAKAO, K., TAKEUCHI, R., YAMAGUCHI, S., KATAOKA, M., OTSUKA, S., TAKAHASHI, M. & MIYAKAWA, T. 2013. Synaptosomal-associated protein 25 mutation induces immaturity of the dentate granule cells of adult mice. Mol Brain, 6, 12.

OHISHI, H., SHIGEMOTO, R., NAKANISHI, S. & MIZUNO, N. 1993. Distribution of the mRNA for a metabotropic glutamate receptor (mGluR3) in the rat brain: an in situ hybridization study. J Comp Neurol, 335, 252-66.

OLIVER, P. L. & DAVIES, K. E. 2009. Interaction between environmental and genetic factors modulates schizophrenic endophenotypes in the Snap-25 mouse mutant blind-drunk. Hum Mol Genet, 18, 4576-89.

OLIVER, P. L., SOBCZYK, M. V., MAYWOOD, E. S., EDWARDS, B., LEE, S., LIVIERATOS, A., OSTER, H., BUTLER, R., GODINHO, S. I., WULFF, K., PEIRSON, S. N., FISHER, S. P., CHESHAM, J. E., SMITH, J. W., HASTINGS, M. H., DAVIES, K. E. & FOSTER, R. G. 2012. Disrupted circadian rhythms in a mouse model of schizophrenia. Curr Biol, 22, 314-9.

OMORI, Y., CHAYA, T., YOSHIDA, S., IRIE, S., TSUJII, T. & FURUKAWA, T. 2015. Identification of G Protein-Coupled Receptors (GPCRs) in Primary Cilia and Their Possible Involvement in Body Weight Control. PLoS One, 10, e0128422.

PAGEL, J. I. & DEINDL, E. 2012. Disease progression mediated by egr-1 associated signaling in response to oxidative stress. Int J Mol Sci, 13, 13104-17.

PANDA, S., ANTOCH, M. P., MILLER, B. H., SU, A. I., SCHOOK, A. B., STRAUME, M., SCHULTZ, P. G., KAY, S. A., TAKAHASHI, J. S. & HOGENESCH, J. B. 2002. Coordinated transcription of key pathways in the mouse by the circadian clock. Cell, 109, 307-20.

PANG, K. C., FRITH, M. C. & MATTICK, J. S. 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. Trends Genet, 22, 1-5.

PARANJPE, D. A. & SHARMA, V. K. 2005. Evolution of temporal order in living organisms. J Circadian Rhythms, 3, 7.

PARIKSHAK, N. N., GANDAL, M. J. & GESCHWIND, D. H. 2015. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat Rev Genet, 16, 441-58.

PARIKSHAK, N. N., LUO, R., ZHANG, A., WON, H., LOWE, J. K., CHANDRAN, V., HORVATH, S. & GESCHWIND, D. H. 2013. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell, 155, 1008-21.

PARTONEN, T., TREUTLEIN, J., ALPMAN, A., FRANK, J., JOHANSSON, C., DEPNER, M., ARON, L., RIETSCHEL, M., WELLEK, S., SORONEN, P., PAUNIO, T., KOCH, A., CHEN, P., LATHROP, M., ADOLFSSON, R., PERSSON, M. L., KASPER, S., SCHALLING, M., PELTONEN, L. & SCHUMANN, G. 2007. Three circadian clock genes Per2, Arntl, and Npas2 contribute to winter depression. Ann Med, 39, 229-38.

PASCHOS, G. K. 2015. Circadian clocks, feeding time, and metabolic homeostasis. Front Pharmacol, 6, 112.

PEMBROKE, W. G., BABBS, A., DAVIES, K., PONTING, C. P. & OLIVER, P. L. 2015. Temporal transcriptomics suggest that twin-peaking genes reset the clock. Elife, 4.

PENNY, G. D., KAY, G. F., SHEARDOWN, S. A., RASTAN, S. & BROCKDORFF, N. 1996. Requirement for Xist in X chromosome inactivation. Nature, 379, 131-7.

PEPKE, S., WOLD, B. & MORTAZAVI, A. 2009. Computation for ChIP-seq and RNA-seq studies. Nat Methods, 6, S22-32.

PERILHOU, A., TOURREL-CUZIN, C., KHARROUBI, I., HENIQUE, C., FAUVEAU, V., KITAMURA, T., MAGNAN, C., POSTIC, C., PRIP-BUUS, C. & VASSEUR-COGNET, M. 2008. The transcription factor COUP-TFII is negatively regulated by insulin and glucose via Foxo1- and ChREBP-controlled pathways. Mol Cell Biol, 28, 6568-79.

PICKARD, G. E., RALPH, M. R. & MENAKER, M. 1987. The intergeniculate leaflet partially mediates effects of light on circadian rhythms. J Biol Rhythms, 2, 35-56.

PONJAVIC, J., PONTING, C. P. & LUNTER, G. 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. Genome Res, 17, 556-65.

PONTING, C. P., OLIVER, P. L. & REIK, W. 2009. Evolution and functions of long noncoding RNAs. Cell, 136, 629-41.

PORTERFIELD, V. M., PIONTKIVSKA, H. & MINTZ, E. M. 2007. Identification of novel light-induced genes in the suprachiasmatic nucleus. BMC Neurosci, 8, 98.

POZZI, D., CONDLIFFE, S., BOZZI, Y., CHIKHLADZE, M., GRUMELLI, C., PROUX-GILLARDEAUX, V., TAKAHASHI, M., FRANCESCHETTI, S., VERDERIO, C. & MATTEOLI, M. 2008. Activity-dependent phosphorylation of Ser187 is required for SNAP-25-negative modulation of neuronal voltage-gated calcium channels. Proc Natl Acad Sci U S A, 105, 323-8.

PRITCHETT, D., JAGANNATH, A., BROWN, L. A., TAM, S. K., HASAN, S., GATTI, S., HARRISON, P. J., BANNERMAN, D. M., FOSTER, R. G. & PEIRSON, S. N. 2015. Deletion of Metabotropic Glutamate Receptors 2 and 3 (mGlu2 & mGlu3) in Mice Disrupts Sleep and Wheel-Running Activity, and Increases the Sensitivity of the Circadian System to Light. PLoS One, 10, e0125523.

PRITCHETT, D., WULFF, K., OLIVER, P. L., BANNERMAN, D. M., DAVIES, K. E., HARRISON, P. J., PEIRSON, S. N. & FOSTER, R. G. 2012a. Evaluating the links between schizophrenia and sleep and circadian rhythm disruption. J Neural Transm (Vienna), 119, 1061-75.

PRITCHETT, D., WULFF, K., OLIVER, P. L., BANNERMAN, D. M., DAVIES, K. E., HARRISON, P. J., PEIRSON, S. N. & FOSTER, R. G. 2012b. Evaluating the links between schizophrenia and sleep and circadian rhythm disruption. J Neural Transm, 119, 1061-75.

PROSSER, R. A., DEAN, R. R., EDGAR, D. M., HELLER, H. C. & MILLER, J. D. 1993. Serotonin and the mammalian circadian system: I. In vitro phase shifts by serotonergic agonists and antagonists. J Biol Rhythms, 8, 1-16.

PROSSER, R. A., HELLER, H. C. & MILLER, J. D. 1994. Serotonergic phase advances of the mammalian circadian clock involve protein kinase A and K+ channel opening. Brain Res, 644, 67-73.

PROSSER, R. A., MCARTHUR, A. J. & GILLETTE, M. U. 1989. cGMP induces phase shifts of a mammalian circadian pacemaker at night, in antiphase to cAMP effects. Proc Natl Acad Sci U S A, 86, 6812-5.

PROSSER, R. A., MILLER, J. D. & HELLER, H. C. 1990. A serotonin agonist phase-shifts the circadian clock in the suprachiasmatic nuclei in vitro. Brain Res, 534, 336-9.

RAJ, A., PESKIN, C. S., TRANCHINA, D., VARGAS, D. Y. & TYAGI, S. 2006. Stochastic mRNA synthesis in mammalian cells. PLoS Biol, 4, e309.

RALPH, M. R., FOSTER, R. G., DAVIS, F. C. & MENAKER, M. 1990. Transplanted suprachiasmatic nucleus determines circadian period. Science, 247, 975-8.

RAMOS, A. D., DIAZ, A., NELLORE, A., DELGADO, R. N., PARK, K. Y., GONZALES-ROYBAL, G., OLDHAM, M. C., SONG, J. S. & LIM, D. A. 2013. Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. Cell Stem Cell, 12, 616-28.

RANDS, C. M., MEADER, S., PONTING, C. P. & LUNTER, G. 2014. 8.2

RAVASI, T., SUZUKI, H., PANG, K. C., KATAYAMA, S., FURUNO, M., OKUNISHI, R., FUKUDA, S., RU, K., FRITH, M. C., GONGORA, M. M., GRIMMOND, S. M., HUME, D. A., HAYASHIZAKI, Y. & MATTICK, J. S. 2006. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res, 16, 11-9.

REDDY, A. B., MAYWOOD, E. S., KARP, N. A., KING, V. M., INOUE, Y., GONZALEZ, F. J., LILLEY, K. S., KYRIACOU, C. P. & HASTINGS, M. H. 2007. Glucocorticoid signaling synchronizes the liver circadian transcriptome. Hepatology, 45, 1478-88.

REGHUNANDANAN, V. & REGHUNANDANAN, R. 2006. Neurotransmitters of the suprachiasmatic nuclei. J Circadian Rhythms, 4, 2.

REMONDINI, D., O'CONNELL, B., INTRATOR, N., SEDIVY, J. M., NERETTI, N., CASTELLANI, G. C. & COOPER, L. N. 2005. Targeting c-Myc-activated genes with a correlation method: detection of global changes in large gene expression network dynamics. Proc Natl Acad Sci U S A, 102, 6902-6.

RHOADS, A. & AU, K. F. 2015. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics, 13, 278-89.

RICHTER, C. P. 1960. Biological Clocks in Medicine and Psychiatry: Shock-Phase Hypothesis. Proc Natl Acad Sci U S A, 46, 1506-30.

RISSO, D., NGAI, J., SPEED, T. P. & DUDOIT, S. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol, 32, 896-902.

ROBERTS, T. C., MORRIS, K. V. & WOOD, M. J. 2014. The role of long non-coding RNAs in neurodevelopment, brain function and neurological disease. Philos Trans R Soc Lond B Biol Sci, 369.

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, 139-40.

ROBLES, M. S., COX, J. & MANN, M. 2014. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. PLoS Genet, 10, e1004047.

RODRIGUEZ, J., TANG, C. H., KHODOR, Y. L., VODALA, S., MENET, J. S. & ROSBASH, M. 2013. Nascent-Seq analysis of Drosophila cycling gene expression. Proc Natl Acad Sci U S A, 110, E275-84.

ROGERS, J. T., ZHAO, L., TROTTER, J. H., RUSIANA, I., PETERS, M. M., LI, Q., DONALDSON, E., BANKO, J. L., KEENOY, K. E., REBECK, G. W., HOE, H. S., D'ARCANGELO, G. & WEEBER, E. J. 2013. Reelin supplementation recovers sensorimotor gating, synaptic plasticity and associative learning deficits in the heterozygous reeler mouse. J Psychopharmacol, 27, 386-95.

ROMIJN, H. J., SLUITER, A. A., POOL, C. W., WORTEL, J. & BUIJS, R. M. 1997. Evidence from confocal fluorescence microscopy for a dense, reciprocal innervation between AVP-, somatostatin-, VIP/PHI-, GRP-, and VIP/PHI/GRP-immunoreactive neurons in the rat suprachiasmatic nucleus. Eur J Neurosci, 9, 2613-23.

ROYBAL, K., THEOBOLD, D., GRAHAM, A., DINIERI, J. A., RUSSO, S. J., KRISHNAN, V., CHAKRAVARTY, S., PEEVEY, J., OEHRLEIN, N., BIRNBAUM, S., VITATERNA, M. H., ORSULAK, P., TAKAHASHI, J. S., NESTLER, E. J., CARLEZON, W. A., JR. & MCCLUNG, C. A. 2007. Mania-like behavior induced by disruption of CLOCK. Proc Natl Acad Sci U S A, 104, 6406-11.

RUDIC, R. D., MCNAMARA, P., CURTIS, A. M., BOSTON, R. C., PANDA, S., HOGENESCH, J. B. & FITZGERALD, G. A. 2004. BMAL1 and CLOCK, two essential components of the circadian clock, are involved in glucose homeostasis. PLoS Biol, 2, e377.

RUDY, B. & MCBAIN, C. J. 2001. Kv3 channels: voltage-gated K+ channels designed for high-frequency repetitive firing. Trends Neurosci, 24, 517-26.

RUSAK, B., MEIJER, J. H. & HARRINGTON, M. E. 1989. Hamster circadian rhythms are phase-shifted by electrical stimulation of the geniculo-hypothalamic tract. Brain Res, 493, 283-91.

RUSTICI, G., MATA, J., KIVINEN, K., LIO, P., PENKETT, C. J., BURNS, G., HAYLES, J., BRAZMA, A., NURSE, P. & BAHLER, J. 2004. Periodic gene expression program of the fission yeast cell cycle. Nat Genet, 36, 809-17.

RUTTER, J., REICK, M. & MCKNIGHT, S. L. 2002. Metabolism and the control of circadian rhythms. Annu Rev Biochem, 71, 307-31.

RUTTER, J., REICK, M., WU, L. C. & MCKNIGHT, S. L. 2001. Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. Science, 293, 510-4.

SAKAI, T., TAMURA, T., KITAMOTO, T. & KIDOKORO, Y. 2004. A clock gene, period, plays a key role in long-term memory formation in Drosophila. Proc Natl Acad Sci U S A, 101, 16058-63.

SAKHARKAR, M. K., PERUMAL, B. S., SAKHARKAR, K. R. & KANGUEANE, P. 2005. An analysis on gene architecture in human and mouse genomes. In Silico Biol, 5, 347-65.

SALGADO-DELGADO, R., ANGELES-CASTELLANOS, M., SADERI, N., BUIJS, R. M. & ESCOBAR, C. 2010. Food intake during the normal activity phase prevents obesity and circadian desynchrony in a rat model of night work. Endocrinology, 151, 1019-29.

SANYAL, A., LAJOIE, B. R., JAIN, G. & DEKKER, J. 2012. The long-range interaction landscape of gene promoters. Nature, 489, 109-13.

SATO, P. Y., CHUPRUN, J. K., SCHWARTZ, M. & KOCH, W. J. 2015. The evolving impact of g protein-coupled receptor kinases in cardiac health and disease. Physiol Rev, 95, 377-404.

SAUVAGEAU, M., GOFF, L. A., LODATO, S., BONEV, B., GROFF, A. F., GERHARDINGER, C., SANCHEZ-GOMEZ, D. B., HACISULEYMAN, E., LI, E., SPENCE, M., LIAPIS, S. C., MALLARD, W., MORSE, M., SWERDEL, M. R., D'ECCLESSIS, M. F., MOORE, J. C., LAI, V., GONG, G., YANCOPOULOS, G. D., FRENDEWEY, D., KELLIS, M., HART, R. P., VALENZUELA, D. M., ARLOTTA, P. & RINN, J. L. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife, 2, e01749.

SCHENA, M., SHALON, D., DAVIS, R. W. & BROWN, P. O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 270, 467-70.

SCHOENHARD, J. A., SMITH, L. H., PAINTER, C. A., EREN, M., JOHNSON, C. H. & VAUGHAN, D. E. 2003. Regulation of the PAI-1 promoter by circadian clock components: differential activation by BMAL1 and BMAL2. J Mol Cell Cardiol, 35, 473-81.

SCHUMAN, E. M., DYNES, J. L. & STEWARD, O. 2006. Synaptic regulation of translation of dendritic mRNAs. J Neurosci, 26, 7143-6.

SHARMA, V. K. 2003. Adaptive significance of circadian clocks. Chronobiol Int, 20, 901-19.

SHENDE, V. R., GOLDRICK, M. M., RAMANI, S. & EARNEST, D. J. 2011. Expression and rhythmic modulation of circulating microRNAs targeting the clock gene Bmal1 in mice. PLoS One, 6, e22586.

SHENDURE, J. 2008. The beginning of the end for microarrays? Nat Methods, 5, 585-7.

SHERMAN, S. M., MUMFORD, J. A. & SCHNYER, D. M. 2015. Hippocampal activity mediates the relationship between circadian activity rhythms and memory in older adults. Neuropsychologia, 75, 617-25.

SHIMAZAKI, Y., NISHIKI, T., OMORI, A., SEKIGUCHI, M., KAMATA, Y., KOZAKI, S. & TAKAHASHI, M. 1996. Phosphorylation of 25-kDa synaptosome-associated protein. Possible involvement in protein kinase C-mediated regulation of neurotransmitter release. J Biol Chem, 271, 14548-53.

SHIMAZOE, T., MORITA, M., OGIWARA, S., KOJIYA, T., GOTO, J., KAMAKURA, M., MORIYA, T., SHINOHARA, K., TAKIGUCHI, S., KONO, A., MIYASAKA, K., FUNAKOSHI, A. & IKEDA, M. 2008. Cholecystokinin-A receptors regulate photic input pathways to the circadian clock. FASEB J, 22, 1479-90.

SHINOHARA, K., TOMINAGA, K., FUKUHARA, C., OTORI, Y. & INOUYE, S. I. 1993a. Processing of photic information within the intergeniculate leaflet of the lateral geniculate body: assessed by neuropeptide Y immunoreactivity in the suprachiasmatic nucleus of rats. Neuroscience, 56, 813-22.

SHINOHARA, K., TOMINAGA, K., ISOBE, Y. & INOUYE, S. T. 1993b. Photic regulation of peptides located in the ventrolateral subdivision of the suprachiasmatic nucleus of the rat: daily variations of vasoactive intestinal polypeptide, gastrin-releasing peptide, and neuropeptide Y. J Neurosci, 13, 793-800.

SIMON, P., SCHOTT, K., WILLIAMS, R. W. & SCHAEFFEL, F. 2004. Posttranscriptional regulation of the immediate-early gene EGR1 by light in the mouse retina. Eur J Neurosci, 20, 3371-7.

SIMS, D., ILOTT, N. E., SANSOM, S. N., SUDBERY, I. M., JOHNSON, J. S., FAWCETT, K. A., BERLANGA-TAYLOR, A. J., LUNA-VALERO, S., PONTING, C. P. & HEGER, A. 2014a. CGAT: computational genomics analysis toolkit. Bioinformatics, 30, 1290-1.

SIMS, D., SUDBERY, I., ILOTT, N. E., HEGER, A. & PONTING, C. P. 2014b. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet, 15, 121-32.

SIOMI, M. C., SATO, K., PEZIC, D. & ARAVIN, A. A. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol, 12, 246-58.

SNODGRASS-BELT, P., GILBERT, J. L. & DAVIS, F. C. 2005. Central administration of transforming growth factor-alpha and neuregulin-1 suppress active behaviors and cause weight loss in hamsters. Brain Res, 1038, 171-82.

SOETEDJO, L., GLOVER, D. A. & JIN, H. 2013. Targeting of vasoactive intestinal peptide receptor 2, VPAC2, a secretin family G-protein coupled receptor, to primary cilia. Biol Open, 2, 686-94.

SOLLNER, T., BENNETT, M. K., WHITEHEART, S. W., SCHELLER, R. H. & ROTHMAN, J. E. 1993. A protein assembly-disassembly pathway in vitro that may correspond to sequential steps of synaptic vesicle docking, activation, and fusion. Cell, 75, 409-18.

SOSHNEV, A. A., ISHIMOTO, H., MCALLISTER, B. F., LI, X., WEHLING, M. D., KITAMOTO, T. & GEYER, P. K. 2011. A conserved long noncoding RNA affects sleep behavior in Drosophila. Genetics, 189, 455-68.

SOWALSKY, A. G., XIA, Z., WANG, L., ZHAO, H., CHEN, S., BUBLEY, G. J., BALK, S. P. & LI, W. 2015. Whole transcriptome sequencing reveals extensive unspliced mRNA in metastatic castration-resistant prostate cancer. Mol Cancer Res, 13, 98-106.

STEGLE, O., PARTS, L., PIIPARI, M., WINN, J. & DURBIN, R. 2012. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc, 7, 500-7.

STEPHAN, F. K. & KOVACEVIC, N. S. 1978. Multiple retention deficit in passive avoidance in rats is eliminated by suprachiasmatic lesions. Behav Biol, 22, 456-62.

STEPHAN, F. K. & ZUCKER, I. 1972. Circadian rhythms in drinking behavior and locomotor activity of rats are eliminated by hypothalamic lesions. Proc Natl Acad Sci U S A, 69, 1583-6.

STEPHAN, K. E., FRISTON, K. J. & FRITH, C. D. 2009. Dysconnection in schizophrenia: from abnormal synaptic plasticity to failures of self-monitoring. Schizophr Bull, 35, 509-27.

STOPA, E. G., VOLICER, L., KUO-LEBLANC, V., HARPER, D., LATHI, D., TATE, B. & SATLIN, A. 1999. Pathologic evaluation of the human suprachiasmatic nucleus in severe dementia. J Neuropathol Exp Neurol, 58, 29-39.

STRAUME, M. 2004. DNA microarray time series analysis: automated statistical assessment of circadian rhythms in gene expression patterning. Methods Enzymol, 383, 149-66.

SUNITHA, S. N. & JOSHI, P. G. 2010. Differential Regulation of P/Q and L-Type Voltage Gated Calcium Channels by Lipid Rafts in PC12 Cells. In: LIM, C. T. & GOH, J. C. H. (eds.) 6th World Congress of Biomechanics (WCB 2010). August

1-6, 2010 Singapore: In Conjunction with 14th International Conference on Biomedical Engineering (ICBME) and 5th Asia Pacific Conference on Biomechanics (APBiomech). Berlin, Heidelberg: Springer Berlin Heidelberg.

SUPEK, F., BOSNJAK, M., SKUNCA, N. & SMUC, T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One, 6, e21800.

SUTTON, E. L. 2014. Psychiatric disorders and sleep issues. Med Clin North Am, 98, 1123-43.

SZETO, C. Y., LIN, C. H., CHOI, S. C., YIP, T. T., NGAN, R. K., TSAO, G. S. & LI LUNG, M. 2014. Integrated mRNA and microRNA transcriptome sequencing characterizes sequence variants and mRNA-microRNA regulatory network in nasopharyngeal carcinoma model systems. FEBS Open Bio, 4, 128-40.

TACHIKAWA, H., HARADA, S., KAWANISHI, Y., OKUBO, T. & SUZUKI, T. 2001. Linked polymorphisms (-333G¿T and -286A¿G) in the promoter region of the CCK-A receptor gene may be associated with schizophrenia. Psychiatry Res, 103, 147-55.

TAKAHASHI, J. S., KUMAR, V., NAKASHE, P., KOIKE, N., HUANG, H. C., GREEN, C. B. & KIM, T. K. 2015. ChIP-seq and RNA-seq methods to study circadian control of transcription in mammals. Methods Enzymol, 551, 285-321.

TAPP, W. N. & HOLLOWAY, F. A. 1981. Phase shifting circadian rhythms produces retrograde amnesia. Science, 211, 1056-8.

TARAZONA, S., GARCIA-ALCALDE, F., DOPAZO, J., FERRER, A. & CONESA, A. 2011. Differential expression in RNA-seq: a matter of depth. Genome Res, 21, 2213-23.

TARIQ, M. A., KIM, H. J., JEJELOWO, O. & POURMAND, N. 2011. Whole-transcriptome RNAseq analysis from minute amount of total RNA. Nucleic Acids Res, 39, e120.

TEICHER, M. H., GLOD, C. A., MAGNUS, E., HARPER, D., BENSON, G., KRUEGER, K. & MCGREENERY, C. E. 1997. Circadian rest-activity disturbances in seasonal affective disorder. Arch Gen Psychiatry, 54, 124-30.

TESSON, B. M., BREITLING, R. & JANSEN, R. C. 2010. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinformatics, 11, 497.

THASTRUP, O., CULLEN, P. J., DROBAK, B. K., HANLEY, M. R. & DAWSON, A. P. 1990. Thapsigargin, a tumor promoter, discharges intracellular Ca2+ stores by specific inhibition of the endoplasmic reticulum Ca2(+)-ATPase. Proc Natl Acad Sci U S A, 87, 2466-70.

THOMPSON, P. M., EGBUFOAMA, S. & VAWTER, M. P. 2003. SNAP-25 reduction in the hippocampus of patients with schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry, 27, 411-7.

TIWARI, H. K. & SCHORK, N. J. 2011. Grand challenges in statistical genetics/genomics methodology. Front Genet, 2, 5.

TORDJMAN, S., DAVLANTIS, K. S., GEORGIEFF, N., GEOFFRAY, M. M., SPERANZA, M., ANDERSON, G. M., XAVIER, J., BOTBOL, M., ORIOL, C., BELLISSANT, E., VERNAY-LECONTE, J., FOUGEROU, C., HESPEL, A., TAVENARD, A., COHEN, D., KERMARREC, S., COULON, N., BONNOT, O. & DAWSON, G. 2015.

Autism as a disorder of biological and behavioral rhythms: toward new therapeutic perspectives. Front Pediatr, 3, 1.

TRAPNELL, C., PACHTER, L. & SALZBERG, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics, 25, 1105-11.

TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc, 7, 562-78.

TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol, 28, 511-5.

TRIPATHI, V., ELLIS, J. D., SHEN, Z., SONG, D. Y., PAN, Q., WATT, A. T., FREIER, S. M., BENNETT, C. F., SHARMA, A., BUBULYA, P. A., BLENCOWE, B. J., PRASANTH, S. G. & PRASANTH, K. V. 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell, 39, 925-38.

TSAI, G. & COYLE, J. T. 2002. Glutamatergic mechanisms in schizophrenia. Annu Rev Pharmacol Toxicol, 42, 165-79.

TSAI, G., PASSANI, L. A., SLUSHER, B. S., CARTER, R., BAER, L., KLEINMAN, J. E. & COYLE, J. T. 1995. Abnormal excitatory neurotransmitter metabolism in schizophrenic brains. Arch Gen Psychiatry, 52, 829-36.

TUREK, F. W., JOSHU, C., KOHSAKA, A., LIN, E., IVANOVA, G., MCDEARMON, E., LAPOSKY, A., LOSEE-OLSON, S., EASTON, A., JENSEN, D. R., ECKEL, R. H., TAKAHASHI, J. S. & BASS, J. 2005. Obesity and metabolic syndrome in circadian Clock mutant mice. Science, 308, 1043-5.

TURETSKY, B. I. & MOBERG, P. J. 2009. An odor-specific threshold deficit implicates abnormal intracellular cyclic AMP signaling in schizophrenia. Am J Psychiatry, 166, 226-33.

TURNER, F. S. 2014. Assessment of insert sizes and adapter content in fastq data from NexteraXT libraries. Front Genet, 5, 5.

UEDA, H. R., CHEN, W., ADACHI, A., WAKAMATSU, H., HAYASHI, S., TAKASUGI, T., NAGANO, M., NAKAHAMA, K., SUZUKI, Y., SUGANO, S., IINO, M., SHIGEYOSHI, Y. & HASHIMOTO, S. 2002a. A transcription factor response element for gene expression during circadian night. Nature, 418, 534-9.

UEDA, H. R., MATSUMOTO, A., KAWAMURA, M., IINO, M., TANIMURA, T. & HASHIMOTO, S. 2002b. Genome-wide transcriptional orchestration of circadian rhythms in Drosophila. J Biol Chem, 277, 14048-52.

UKAI-TADENUMA, M., YAMADA, R. G., XU, H., RIPPERGER, J. A., LIU, A. C. & UEDA, H. R. 2011. Delay in feedback repression by cryptochrome 1 is required for circadian clock function. Cell, 144, 268-81.

VACIC, V., MCCARTHY, S., MALHOTRA, D., MURRAY, F., CHOU, H. H., PEOPLES, A., MAKAROV, V., YOON, S., BHANDARI, A., COROMINAS, R., IAKOUCHEVA, L. M., KRASTOSHEVSKY, O., KRAUSE, V., LARACH-WALTERS, V., WELSH, D. K., CRAIG, D., KELSOE, J. R., GERSHON, E. S., LEAL, S. M., DELL

AQUILA, M., MORRIS, D. W., GILL, M., CORVIN, A., INSEL, P. A., MCCLELLAN, J., KING, M. C., KARAYIORGOU, M., LEVY, D. L., DELISI, L. E. & SEBAT, J. 2011. Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. Nature, 471, 499-503.

VELCULESCU, V. E., ZHANG, L., VOGELSTEIN, B. & KINZLER, K. W. 1995. Serial analysis of gene expression. Science, 270, 484-7.

VINDLACHERUVU, R. R., EBLING, F. J., MAYWOOD, E. S. & HASTINGS, M. H. 1992. Blockade of Glutamatergic Neurotransmission in the Suprachiasmatic Nucleus Prevents Cellular and Behavioural Responses of the Circadian System to Light. Eur J Neurosci, 4, 673-679.

VOLICER, L., HARPER, D. G., MANNING, B. C., GOLDSTEIN, R. & SATLIN, A. 2001. Sundowning and circadian rhythms in Alzheimer‘s disease. Am J Psychiatry, 158, 704-11.

VOLLMERS, C., GILL, S., DITACCHIO, L., PULIVARTHY, S. R., LE, H. D. & PANDA, S. 2009. Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. Proc Natl Acad Sci U S A, 106, 21453-8.

WANG, C., GONG, B., BUSHEL, P. R., THIERRY-MIEG, J., THIERRY-MIEG, D., XU, J., FANG, H., HONG, H., SHEN, J., SU, Z., MEEHAN, J., LI, X., YANG, L., LI, H., LABAJ, P. P., KREIL, D. P., MEGHERBI, D., GAJ, S., CAIMENT, F., VAN DELFT, J., KLEINJANS, J., SCHERER, A., DEVANARAYAN, V., WANG, J., YANG, Y., QIAN, H. R., LANCASHIRE, L. J., BESSARABOVA, M., NIKOLSKY, Y., FURLANELLO, C., CHIERICI, M., ALBANESE, D., JURMAN, G., RICCADONNA, S., FILOSI, M., VISINTAINER, R., ZHANG, K. K., LI, J., HSIEH, J. H., SVOBODA, D.

L., FUSCOE, J. C., DENG, Y., SHI, L., PAULES, R. S., AUERBACH, S. S. & TONG, W. 2014. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. Nat Biotechnol, 32, 926-32.

WANG, D., LIANG, X., CHEN, X. & GUO, J. 2013. Ribonucleoprotein complexes that control circadian clocks. Int J Mol Sci, 14, 9018-36.

WANG, J., SI, Y. M., LIU, Z. L. & YU, L. 2003. Cholecystokinin, cholecystokinin-A receptor and cholecystokinin-B receptor gene polymorphisms in Parkinson's disease. Pharmacogenetics, 13, 365-9.

WANG, J., YIN, L. & LAZAR, M. A. 2006. The orphan nuclear receptor Rev-erb alpha regulates circadian expression of plasminogen activator inhibitor type 1. J Biol Chem, 281, 33842-8.

WANG, K., SINGH, D., ZENG, Z., COLEMAN, S. J., HUANG, Y., SAVICH, G. L., HE, X., MIECZKOWSKI, P., GRIMM, S. A., PEROU, C. M., MACLEOD, J. N., CHIANG, D. Y., PRINS, J. F. & LIU, J. 2010. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res, 38, e178.

WANG, T. A., YU, Y. V., GOVINDAIAH, G., YE, X., ARTINIAN, L., COLEMAN, T. P., SWEEDLER, J. V., COX, C. L. & GILLETTE, M. U. 2012. Circadian rhythm of redox state regulates excitability in suprachiasmatic nucleus neurons. Science, 337, 839-42.

WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet, 10, 57-63.

WASHBOURNE, P., THOMPSON, P. M., CARTA, M., COSTA, E. T., MATHEWS, J. R., LOPEZ-BENDITO, G., MOLNAR, Z., BECHER, M. W., VALENZUELA, C. F.,

PARTRIDGE, L. D. & WILSON, M. C. 2002. Genetic ablation of the t-SNARE SNAP-25 distinguishes mechanisms of neuroexocytosis. Nat Neurosci, 5, 19-26.

WATANABE, K. A., RINGLER, P., GU, L. & SHEN, Q. J. 2014. RNA-sequencing reveals previously unannotated protein- and microRNA-coding genes expressed in aleurone cells of rice seeds. Genomics, 103, 122-34.

WATANABE, S., YAMAMORI, S., OTSUKA, S., SAITO, M., SUZUKI, E., KATAOKA, M., MIYAOKA, H. & TAKAHASHI, M. 2015. Epileptogenesis and epileptic maturation in phosphorylation site-specific SNAP-25 mutant mice. Epilepsy Res, 115, 30-44.

WEBER, E. T., GANNON, R. L. & REA, M. A. 1995. cGMP-dependent protein kinase inhibitor blocks light-induced phase advances of circadian rhythms in vivo. Neurosci Lett, 197, 227-30.

WEBER, E. T. & REA, M. A. 1997. Neuropeptide Y blocks light-induced phase advances but not delays of the circadian activity rhythm in hamsters. Neurosci Lett, 231, 159-62.

WEI, C. L., NG, P., CHIU, K. P., WONG, C. H., ANG, C. C., LIPOVICH, L., LIU, E. T. & RUAN, Y. 2004. 5‘ Long serial analysis of gene expression (LongSAGE) and 3‘ LongSAGE for transcriptome characterization and genome annotation. Proc Natl Acad Sci U S A, 101, 11701-6.

WEI, F., XU, Z. C., QU, Z., MILBRANDT, J. & ZHUO, M. 2000. Role of EGR1 in hippocampal synaptic enhancement induced by tetanic stimulation and amputation. J Cell Biol, 149, 1325-34.

WELSH, D. K., LOGOTHETIS, D. E., MEISTER, M. & REPPERT, S. M. 1995. Individual neurons dissociated from rat suprachiasmatic nucleus express independently phased circadian firing rhythms. Neuron, 14, 697-706.

WELSH, D. K., TAKAHASHI, J. S. & KAY, S. A. 2010. Suprachiasmatic nucleus: cell autonomy and network properties. Annu Rev Physiol, 72, 551-77.

WERSINGER, S. R., CALDWELL, H. K., CHRISTIANSEN, M. & YOUNG, W. S., 3RD 2007. Disruption of the vasopressin 1b receptor gene impairs the attack component of aggressive behavior in mice. Genes Brain Behav, 6, 653-60.

WEST, A. E. & GREENBERG, M. E. 2011. Neuronal activity-regulated gene transcription in synapse development and cognitive function. Cold Spring Harb Perspect Biol, 3.

WEST, A. E., GRIFFITH, E. C. & GREENBERG, M. E. 2002. Regulation of transcription factors by neuronal activity. Nat Rev Neurosci, 3, 921-31.

WESTERMARK, P. O. & HERZEL, H. 2013. Mechanism for 12 hr rhythm generation by the circadian clock. Cell Rep, 3, 1228-38.

WETHMAR, K., BARBOSA-SILVA, A., ANDRADE-NAVARRO, M. A. & LEUTZ, A. 2014. uORFdb–a comprehensive literature database on eukaryotic uORF biology. Nucleic Acids Res, 42, D60-7.

WEXLER, E. M., ROSEN, E., LU, D., OSBORN, G. E., MARTIN, E., RAYBOULD, H. & GESCHWIND, D. H. 2011. Genome-wide analysis of a Wnt1-regulated transcriptional network implicates neurodegenerative pathways. Sci Signal, 4, ra65.

WICHERT, S., FOKIANOS, K. & STRIMMER, K. 2004. Identifying periodically expressed transcripts in microarray time series data. Bioinformatics, 20, 5-20.

WILSON, M. C. 2000. Coloboma mouse mutant as an animal model of hyperkinesis and attention deficit hyperactivity disorder. Neurosci Biobehav Rev, 24, 51-7.

WU, T. D. & NACU, S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics, 26, 873-81.

WU, T. D. & WATANABE, C. K. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics, 21, 1859-75.

WULFF, K., DIJK, D. J., MIDDLETON, B., FOSTER, R. G. & JOYCE, E. M. 2012. Sleep and circadian rhythm disruption in schizophrenia. Br J Psychiatry, 200, 308-16.

WULFF, K., GATTI, S., WETTSTEIN, J. G. & FOSTER, R. G. 2010. Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. Nat Rev Neurosci, 11, 589-99.

XUE, Z., YE, Q., ANSON, S. R., YANG, J., XIAO, G., KOWBEL, D., GLASS, N. L., CROSTHWAITE, S. K. & LIU, Y. 2014. Transcriptional interference by antisense RNA is required for circadian clock function. Nature, 514, 650-3.

YAN, L., KARATSOREOS, I., LESAUTER, J., WELSH, D. K., KAY, S., FOLEY, D. & SILVER, R. 2007. Exploring spatiotemporal organization of SCN circuits. Cold Spring Harb Symp Quant Biol, 72, 527-41.

YAN, L., MIYAKE, S. & OKAMURA, H. 2000. Distribution and circadian expression of dbp in SCN and extra-SCN areas in the mouse brain. J Neurosci Res, 59, 291-5.

YEAN, S. L., WUENSCHELL, G., TERMINI, J. & LIN, R. J. 2000. Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. Nature, 408, 881-4.

YOKOTA, S., YAMAMOTO, M., MORIYA, T., AKIYAMA, M., FUKUNAGA, K., MIYAMOTO, E. & SHIBATA, S. 2001. Involvement of calcium-calmodulin protein kinase but not mitogen-activated protein kinase in light-induced phase delays and Per gene expression in the suprachiasmatic nucleus of the hamster. J Neurochem, 77, 618-27.

YOO, S. H., YAMAZAKI, S., LOWREY, P. L., SHIMOMURA, K., KO, C. H., BUHR, E. D., SIEPKA, S. M., HONG, H. K., OH, W. J., YOO, O. J., MENAKER, M. & TAKAHASHI, J. S. 2004. PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. Proc Natl Acad Sci U S A, 101, 5339-46.

YU, Y., FUSCOE, J. C., ZHAO, C., GUO, C., JIA, M., QING, T., BANNON, D. I., LANCASHIRE, L., BAO, W., DU, T., LUO, H., SU, Z., JONES, W. D., MOLAND, C. L., BRANHAM, W. S., QIAN, F., NING, B., LI, Y., HONG, H., GUO, L., MEI, N., SHI, T., WANG, K. Y., WOLFINGER, R. D., NIKOLSKY, Y., WALKER, S. J., DUERKSEN-HUGHES, P., MASON, C. E., TONG, W., THIERRY-MIEG, J., THIERRY-MIEG, D., SHI, L. & WANG, C. 2014. A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. Nat Commun, 5, 3230.

YUAN, Y., CHEN, Y. P., NI, S., XU, A. G., TANG, L., VINGRON, M., SOMEL, M. & KHAITOVICH, P. 2011. Development and application of a modified dynamic time warping algorithm (DTW-S) to analyses of primate brain expression time series. BMC Bioinformatics, 12, 347.

ZACHRISSON, O., DE BELLEROCHE, J., WENDT, K. R., HIRSCH, S. & LINDEFORS, N. 1999. Cholecystokinin CCK(B) receptor mRNA isoforms: expression in schizophrenic brains. Neuroreport, 10, 3265-8.

ZAK, D. E., HAO, H., VADIGEPALLI, R., MILLER, G. M., OGUNNAIKE, B. A. & SCHWABER, J. S. 2006. Systems analysis of circadian time-dependent neuronal epidermal growth factor receptor signaling. Genome Biol, 7, R48.

ZHANG, B. & HORVATH, S. 2005. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol, 4, Article17.

ZHANG, E. E., LIU, Y., DENTIN, R., PONGSAWAKUL, P. Y., LIU, A. C., HIROTA, T., NUSINOW, D. A., SUN, X., LANDAIS, S., KODAMA, Y., BRENNER, D. A., MONTMINY, M. & KAY, S. A. 2010. Cryptochrome mediates circadian regulation of cAMP signaling and hepatic gluconeogenesis. Nat Med, 16, 1152-6.

ZHANG, J., FANG, Z., JUD, C., VANSTEENSEL, M. J., KAASIK, K., LEE, C. C., ALBRECHT, U., TAMANINI, F., MEIJER, J. H., OOSTRA, B. A. & NELSON, D. L. 2008. Fragile X-related proteins regulate mammalian circadian behavioral rhythms. Am J Hum Genet, 83, 43-52.

ZHANG, J., HOU, L., KLANN, E. & NELSON, D. L. 2009. Altered hippocampal synaptic plasticity in the FMR1 gene family knockout mouse models. J Neurophysiol, 101, 2572-80.

ZHANG, L., KASIF, S., CANTOR, C. R. & BROUDE, N. E. 2004a. GC/AT-content spikes as genomic punctuation marks. Proc Natl Acad Sci U S A, 101, 16855-60.

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. Proc Natl Acad Sci U S A, 111, 16219-24.

ZHANG, S., ROSS, K. D., SEIDNER, G. A., GORMAN, M. R., POON, T. H., WANG, X., KEITHLEY, E. M., LEE, P. N., MARTINDALE, M. Q., JOINER, W. J.

& HAMILTON, B. A. 2015. Nmf9 Encodes a Highly Conserved Protein Important to Neurological Function in Mice and Flies. PLoS Genet, 11, e1005344.

ZHANG, Y., VILAYTHONG, A. P., YOSHOR, D. & NOEBELS, J. L. 2004b. Elevated thalamic low-voltage-activated currents precede the onset of absence epilepsy in the SNAP25-deficient mouse mutant coloboma. J Neurosci, 24, 5239-48.

ZHOU, Q. Y. & CHENG, M. Y. 2005. Prokineticin 2 and circadian clock output. FEBS J, 272, 5703-9.

ZHU, H., VADIGEPALLI, R., RAFFERTY, R., GONYE, G. E., WEAVER, D. R. & SCHWABER, J. S. 2012. Integrative gene regulatory network analysis reveals light-induced regional gene expression phase shift programs in the mouse suprachiasmatic nucleus. PLoS One, 7, e37833.

ZUBER, A. M., CENTENO, G., PRADERVAND, S., NIKOLAEVA, S., MAQUELIN, L., CARDINAUX, L., BONNY, O. & FIRSOV, D. 2009. Molecular clock is involved in predictive circadian adjustment of renal function. Proc Natl Acad Sci U S A, 106, 16523-8.

ZUEGER, M., URANI, A., CHOURBAJI, S., ZACHER, C., LIPP, H. P., ALBRECHT, U., SPANAGEL, R., WOLFER, D. P. & GASS, P. 2006. mPer1 and mPer2 mutant mice show regular spatial and contextual learning in standardized tests for hippocampus-dependent learning. J Neural Transm, 113, 347-56.