Genetic and bioinformatic screening for behavioural mutations in *Drosophila melanogaster*

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Edward Wilhelm Green MA (Cantab)

Department of Genetics

University of Leicester

July 2009

Abstract

Genetic and bioinformatic screening for behavioural mutations in *Drosophila melanogaster*

Edward Wilhelm Green

In the post-genomic era, the question of how genes give rise to the observable diversity of morphology, physiology, behaviour and disease susceptibility is becoming of one central importance. Even in a model system as well studied as the fruit fly *Drosophila melanogaster*, the function of the vast majority of genes, and the mechanisms by which they give rise to such diversity, remains unknown.

Drosophila behaviour represents a sensitive system in which to evaluate novel methods of determining gene function. Traditionally the analysis of behavioural phenotypes has represented a time consuming, highly subjective process. I have developed a suite of automated analysis tools (the BeFly! package) that has not only made such analyses both quicker and more objective, but has also allowed data to be examined in greater depth by making complex algorithms more accessible to users.

The BeFly! package was initially used to characterise a serendipitously identified circadian mutant strain provisionally named *Party on.* As the *Party on* gene could not be conclusively mapped, and a meta analysis of existing circadian microarray data suggested that many circadian genes remained to be identified, BeFly!'s high throughput tools were employed in a novel systems biology screen in which phenotypic analysis was combined with gene expression data to identify likely gene function.

This approach generated a number of novel candidate clock genes, the roles of which were further analysed using RNAi knockdown, confirming that the neuropeptide gene *Adipokinetic hormone-like* played a role in the clock mechanism. Given the success of our new strategy, it was widened to identify genes controlling sleep in *Drosophila*, leading to the identification of several genes associated with distinct aspects of sleep.

In conclusion, the tools and methods developed in this thesis represent a novel, sensitive method for determining gene function applicable beyond the *Drosophila* model.

Acknowledgements

This work could not have been done without the support of my supervisor Bambos, whose squash playing prowess and encyclopaedic memory must surely epitomise the ideal work/life balance of the modern working professor. Without Bambos I can safely say I would never have embarked upon (or finished) this PhD – though that is not to say that the production of this weighty tome is entirely his fault!

Indeed, my passion for so many aspects of the study of Genetics was engendered by my undergraduate supervisor Dr Mike Majerus, whose untimely passing early this year has robbed the scientific community not only of one of its most brilliant entomologists, but the wider community of a true character in the best tradition of the Victorian gentleman scientist. He is sorely missed.

It would be remiss of me not to thank colleagues who taught me important techniques and helped in generating some of the data in this work, including Karen Garner, Katie Beaumont, Helen Roe, Rozi Andretic, David Evans, Jenée Wagner, Bruno van Swinderen and Ralph Greenspan.

Notable amongst the ever changing kaleidoscope of friends and colleagues supporting me over the course of this work are my teammates at RED, the Leicester Ultimate Frisbee team, who have accepted me as one of their own through thick and thin, both on the field and off.

Last but not least, this work could not have been completed without my parents Anna and Mark, whose support picked up where that of the MRC left off. I should like to take this opportunity to ask that they remember not to make it look too easy when they come to write their own theses in the coming years.

Table of Contents

Abbreviations used

Abbreviation	Definition
LD 10:14	Environmental conditions set to a cycle of 10 hours light and 14 hours darkness.
LL	Constant light conditions
DD	Constant dark conditions
GCA	General Combining Ability
SCA	Specific Combining Ability
SD	Sleep deprivation
ZT	Zeitgeber Time
CT	Clock Time (or Constant Time)
RNAi	Ribonuclei acid interference (Fire et al., 1998)
Userform	Excel 2007 form that allows user to enter control code operation
Dialog box	Excel 2007 notification box
RibbonUI	The bar at the top of the screen in Excel 2007 containing a number of tabs, each tab containing numerous buttons which control macros.
tG4(A3) M	*tim>GAL4(line A3); MKRS* – a genetic lineage may be indicated within parentheses
cry[b]	Allele names may be indicated within square brackets or as a superscript
τ=24.28 ±0.07, n=22	Period of 24.28 hours, standard error of 0.07, estimated using 22 individuals
♂	Male
♀	Female

1 Introduction – the problems of identifying gene function

Drosophila, with its rapid life cycle, large progeny size, tractable genetics and large research community, has been perhaps the most successful model organism in genetics. Sequencing of the *Drosophila* genome provided the proof of concept for shotgun sequencing (Adams et al., 2000), which itself has become the basis of modern massively parallel sequencing technologies. Uniquely this has allowed the rapid sequencing of 12 closely related *Drosophilid* species, identifying both conserved and rapidly diverging regions of the *Drosophilid* genome (Drosophila 12 Genomes Consortium et al., 2007).

Despite the massive increases in computational power over the past decade that have allowed increasingly sophisticated algorithmic annotation of genomes, annotating and interpreting the wealth of sequencing information has proven difficult, as advances in functional genomics and proteomics have lagged behind those in sequencing technology. The function of the vast majority of genes, even in a system as well characterised as *Drosophila,* remains unknown.

Determining gene function therefore remains largely the preserve of experimentalists using either traditional 'top down' quantitative genetics, or 'bottom up' mutagenesis. This thesis presents the results of the development and use of novel, high throughput techniques to quantify phenotypes and uncover gene function using a new 'systems biology' synthesis of both approaches (Toma et al., 2002).

1.1 Quantitative genetics

As recently as a decade ago there was no clear consensus as to the number of genes a simple model organism might posses, nor the number of genes contributing to the generation of a single phenotype. Whilst T.H. Morgan's work on *Drosophila* eye colour mutants

suggested that linear pathways with clear patterns of epistasis and complementation underlay phenotypes (Fisher, 1918), more recent experimental results showing extensive genetic pleiotropy (Hall, 1994) suggest that pathways contributing to phenotypes must be branched or overlapping.

In an effort to address such questions, early experimentalists began bidirectional selection experiments in which the quantitative change in each generation's mean phenotypic score was recorded (Tully, 1996); analysis of the patterns of phenotypic inheritance then provided an indication of the complexity of the genetic architecture underlying a given trait:

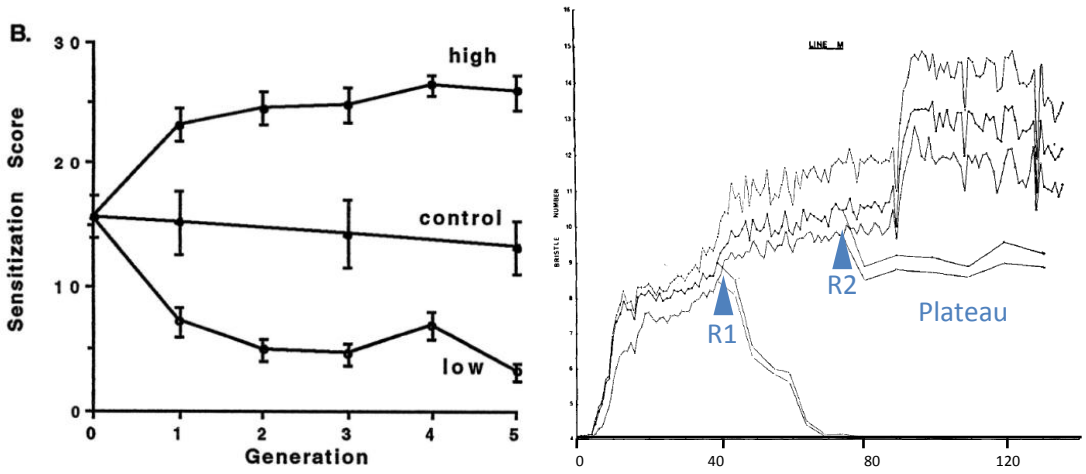


Figure 1-1 Quantitative genetics examples. Left panel: bidirectional selection for learning in blowflies shows the response to selection is almost complete after one generation. Data from McGuire (McGuire, 1981). **Right panel**: bidirectional selection for sensory bristle number in *Drosophila* over 140 generations. Relaxing selection following 80 generations of selection (**R2**) causes the population to decline to a higher plateau than relaxing selection after only 40 generations (**R1**), suggesting a number of alleles had become fixed within the population. Figure from Sheldon (Sheldon and Milton, 1972).

Figure 1-1 shows that whilst variation in some phenotypes may be controlled by a single, biallelic gene (left panel), other traits respond to selection for much longer periods (right panel); bidirectional selection for geotaxic behaviour in *Drosophila* only reached its zenith after almost 1,000 generations of intermittent selection over a period of over 30 years (Ricker and Hirsch, 1985). In general the more complex the output phenotype being selected,

the longer the period of selective response, indicative of the broad genetic architecture underlying that trait.

Using quantitative genetics it has been possible to calculate the contribution of any one gene to a phenotype (Falconer and Mackay, 1995). As is evident in the right panel of Figure 1-1, in the case of complex traits a small number of genes contribute much of the variance, and therefore bidirectional selection initially produces a very marked response. However, there are many more genes which have minor effects on the trait (Dilda and Mackay, 2002), and selection for these alleles takes long periods of time (as predicted by the Kimura's neutral theory), therefore complex traits show an extended period of response to selection. However, over evolutionary timescales, polymorphisms are fixed by selection, and as quantitative genetics studies polymorphisms present in populations, fixed loci are 'invisible' to this approach, limiting the number of loci that can be associated with a given trait.

Furthermore meiotic mapping the loci identified by quantitative genetics is difficult, as the large phenotypic effect caused by a major locus may be opposed by the net contribution of many other loci with small opposing effects on the phenotype. Such 'genetic modifiers' obfuscate the mapping of major loci (Greenspan, 1997), though recent developments of high resolution SNP maps has improved mapping techniques greatly (Chen et al., 2008).

As a result of the difficulties inherent in mapping genes using quantitative genetics, Seymour Benzer developed a novel approach based on the idea that the function of a gene can be determined by careful examination and characterisation of mutant alleles; a hypothesis that forms the basis what is now known as 'forward 'genetics.

1.2 Benzerian forward genetics and the circadian clock

Benzer hypothesised that by a process of unbiased chemical mutagenesis (later refined using *P* elements), a saturating screen should identify all genes contributing to a phenotype by

creating alleles much more severe (and therefore more easily mapped) than the polymorphisms present in natural populations. Using these techniques, Benzer was able to show for the first time that a single locus could control a complex behaviour such as the rhythmic outputs of the circadian clock (Konopka and Benzer, 1971).

Humans have been fascinated with the concept of time; indeed efforts to measure the passage of time using clocks (as opposed to calendars) date back at least 5000 years to the earliest Sumerian and Egyptian obelisks and sundials, whilst accurate measurement of time underlies modern digital communications (Allan et al., 1997). Although colloquially people are often referred to as 'night owls' or 'larks', the first formal demonstration that organisms might have an endogenous clock mechanism was not made until 1729 when deMarain showed that heliotrope plant leaf opening rhythms persisted in total darkness, exhibiting a period of approximately 24 hours, *i.e.* a *circadian* rhythm (Pittendrigh, 1965). Indeed, biological clocks may be extremely accurate, the clock of the flying squirrel, *Glaucomys volans,* has been calculated to be accurate to ±190 seconds per 24 hour cycle (Dowse, 2008), a level of accuracy not surpassed by human designed clocks until Huygen's introduction of the pendulum clock in 1656 (Bennett et al., 2002).

Aschoff proposed that a true circadian clock must exhibit several properties (Aschoff, 1965):

1. It must anticipate environmental change.
2. It must be entrainable to environmental conditions, and be sufficiently plastic to accommodate seasonal changes.
3. It must be temperature compensated.
4. It must persist under free running conditions, i.e. be self sustaining.

Circadian clocks can be found even in organisms as simple as cyanobacteria, suggesting that clocks emerged early in evolution (Rosbash, 2009), perhaps as an 'from light'

to prevent UV mediated DNA damage (Pittendrigh, 1965). Indeed, the ability to anticipate environmental change has been shown in the laboratory to be a highly adaptive strategy (Ouyang et al., 1998; Woelfle et al., 2004).

Perhaps as a result of its early evolution, the circadian clock has deep roots in the body's physiological processes; recent studies indicate that up to 20% of hepatic proteins (and presumably those in other systems) may show daily cycles in abundance (Reddy et al., 2006). As a possible consequence of such cycles, the presentation of symptoms in medical emergencies follows a clear temporal profile (Wang et al., 1995). As might therefore be expected, disruption of the phasic relationship between the environment and the body clock, such as that following rapid transition across time zones, results in a series of symptoms including general fatigue, loss of appetite and altered bowel function collectively termed "Jetlag" (Waterhouse et al., 2002). Even subtle phasic desynchrony, such as that following the biannual transition between summer and winter time, can disrupt the body's seasonal adaptation mechanisms (Kantermann et al., 2007), failure of which may lead to Seasonal Affective Disorder (Murray et al., 2003).

More chronic disruption of the phasic relationship, such as that following long term shift work or as a result of genetic perturbation of the clock mechanism (such as the inherited Familial Advanced Sleep Phase Syndrome), has been associated with more serious metabolic and endocrine impairment, immunity decline and an increased incidence of cancers (Spaggiari, 2008).

1.3 The circadian clock mechanism in *Drosophila*

Following Benzer's pioneering demonstration that alleles of the gene *period* (*per*) could not only change the periodicity of the circadian clock in constant conditions, but also render it arrhythmic, considerable work has been done to reveal the molecular composition of

the circadian clock. In many ways the molecular clock recapitulates the design principals of modern mechanical clocks, composing of three parts:

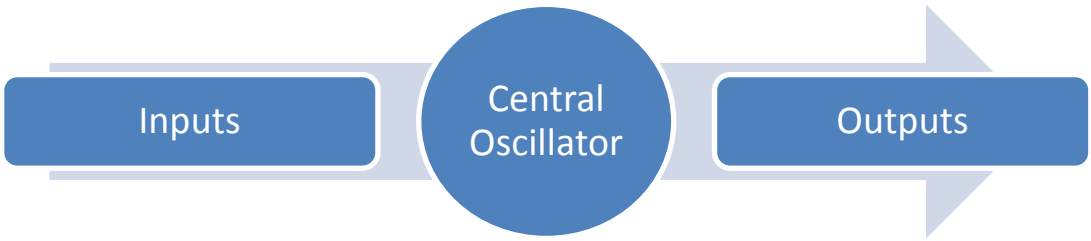


Figure 1-2 The classical circadian clock model.

The clock has a number of inputs, including light and temperature, and myriad independent physiological and behavioural outputs. The central oscillator mechanism itself may require as few as three proteins and an ATP energy source in unicellular cyanobacteria (Kageyama et al., 2006), and even complex multicellular organisms such as the marine snail *Aplysia* may retain such cell autonomous clocks (reviewed in Block et al., 1996). However, in mammals the vast majority of cells in the body do not possess a functioning body clock, relying on neuroendocrine and metabolic cues from central clock cells for time cues (Hastings et al., 2003). Whilst these cells may themselves retain an intact intracellular clock when dissociated in cell culture, the central oscillator system in higher eukaryotes is inherently multicellular.

1.3.1 The negative feedback loop of the intracellular oscillator

The intracellular clock, as currently understood, is a complex machinery consisting of both transcriptional activators and repressors, as well as genes affecting the post translational stability and subcellular localisation of proteins. It is therefore often referred to as a Transcription-Translation-Oscillator (TTO) model, although evidence is accumulating that such a model does not fully encompass the circadian clock's complexity in multicellular organisms (S. Kay, Presidential Symposium SRBR 2008).

The *per* gene identified by Konopka and Benzer is a central component of the molecular clockwork in *Drosophila* (as illustrated in Figure 1-4), but is not itself a transcription

factor capable of driving a transcriptional cycle. The transcription of *per* is promoted by a heterodimer of CLOCK (CLK) (Allada et al., 1998; Darlington et al., 1998) and CYCLE (CYC) (Rutila et al., 1998b). These basic helix-loop-helix transcription factors recognise both E-box *(CACGTG)* motifs in the *per* enhancer region (Darlington et al., 2000; Yu et al., 2006), and non-essential motifs which regulate spatial aspects of *per*'s expression (Lyons et al., 2000). New evidence suggests that *per*'s transcription is also regulated in part by a process of histone remodelling (Taylor and Hardin, 2008).

per and *timless* (TIM, the PAS domain binding partner of PER: Vosshall et al., 1994; Myers et al., 1995) show rhythmic cycles of transcription, mRNA levels peaking early during the night, with their corresponding protein levels peaking several hours afterwards as a result of post-translational mechanisms modifying protein accumulation, stability and activity (illustrated in Figure 1-3). This "intracellular interval timer" (Meyer et al., 2006) is essential for maintaining the proper period and phase of the circadian clock.

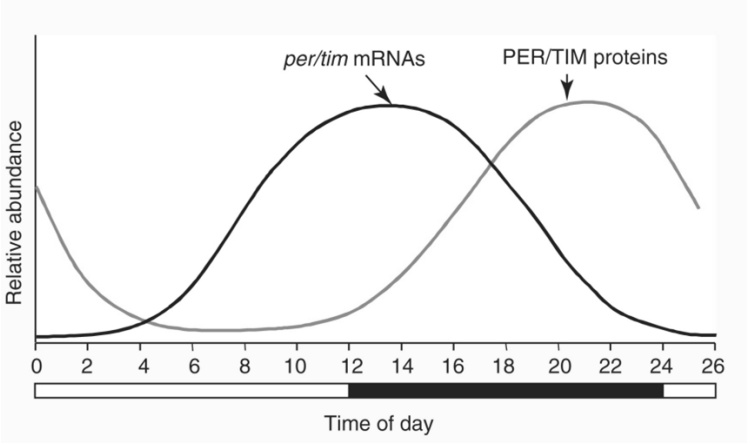


Figure 1-3 mRNA and protein cycles of *per* and *tim*. Figure from Nitabach and Taghert (Nitabach and Taghert, 2008).

Cytoplasmic PER is progressively phosphorylated by the mammalian casein kinase 1ε homologue DOUBLE-TIME (DBT) (Kloss et al., 1998), though this may be potentiated by prior phosphorylation of a key DBT target serine residue mediated by casein kinase 2 (CKII, also identified as *Andante*) (Lin et al., 2002; Akten et al., 2003; Konopka et al., 1991). The

progressive phosphorylation of PER is temporally gated by the occupancy of the key N terminal S47 site, the phosphorylation of which is antagonised by protein phophatase 2A (PP2A) (Sathyanarayanan et al., 2004), as shown in Figure 1-4. Once phosphorylated, PER may be bound by Supernumerary Limbs (SLIMB), an F-box/WD40-repeat E3 ligase (Grima et al., 2002; Ko et al., 2002), that targets PER for 26S proteasomal degradation (Price et al., 1998; Edery et al., 1994). The binding of TIM to phosphorylated PER (possibly within a PER-DBT complex) (Zeng et al., 1996) protects PER from such degradation.

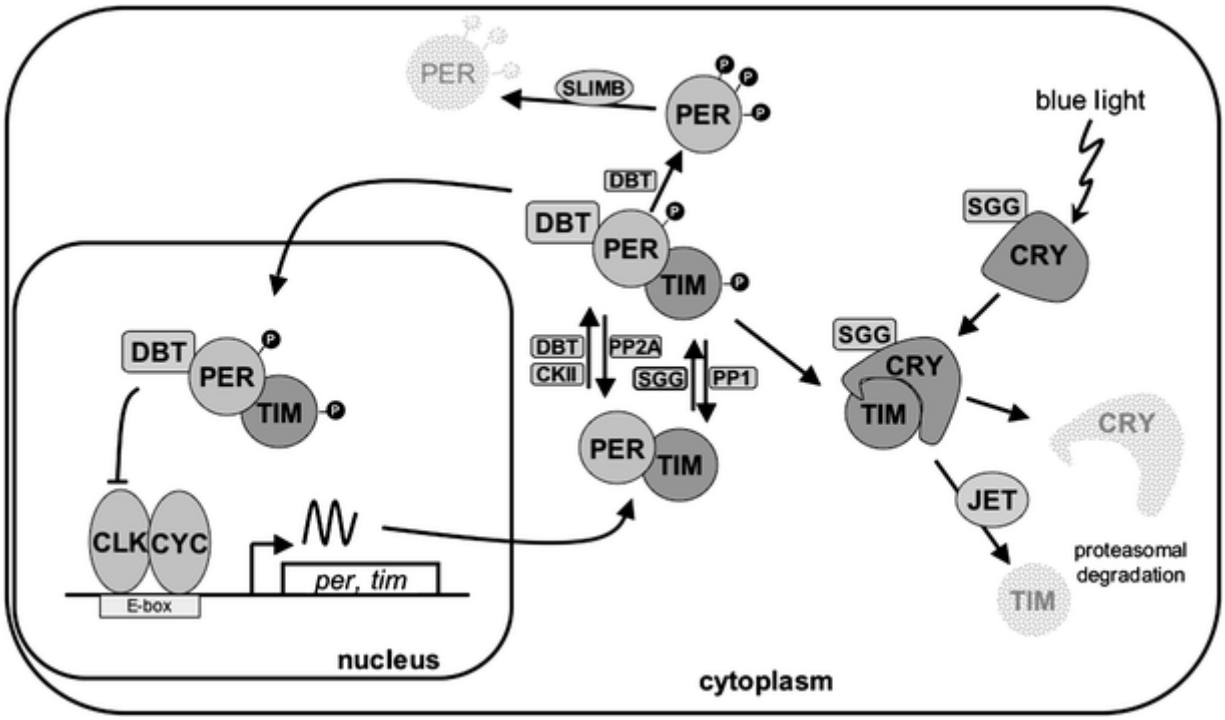


Figure 1-4 The *Drosophila* molecular clock and the CRY light input pathway. Speckled shapes indicate proteins undergoing proteasomal degradation. ***P*** = phosphate groups. See main text for details. Figure and legend from Dubruille and Emery (Dubruille and Emery, 2008).

TIM-PER-DBT complexes are phosphorylated at TIM residues by SHAGGY (SGG) in a process that appears to be localised to discrete cytoplasmic foci (Meyer et al., 2006), possibly mediated by the cellular scaffold protein dAxin (Karen Garner, PhD Thesis). Phosphorylation of TIM is opposed by stabilising PP1 phosphatase action, although phenotypic data suggests that SGG and PP1 target different TIM residues (Fang et al., 2007).

The mechanisms controlling nuclear translocation of the TIM-PER-DBT complex are subject to considerable debate; it has been argued that TIM (and possibly PER) phosphorylation is a prerequisite for nuclear translocation of the TIM-PER-DBT complex (Martinek et al., 2001), whilst other data suggests that DBT and CKII phosphorylation of PER increases PER's repressor activity, increasing its nuclear retention rather than altering PER nuclear import or export dynamics (Nawathean and Rosbash, 2004). In addition, PER and TIM have been shown to move independently into the nucleus, as shown by antisera (Shafer et al., 2002) and fluorescence resonance energy transfer (FRET) measurements (Meyer et al., 2006), and supported by mathematical modelling suggesting that dissociation of PER-TIM complexes prior to nuclear entry is a means of adjusting the period and phase of molecular oscillations (Leise and Moin, 2007). It appears therefore that the interaction of PER with TIM stimulates, but is not obligatory for, nuclear localization and retention of both proteins.

There is also debate as to whether PER and DBT translocate together, as DBT is found in the nucleus in per^{01} mutants, suggesting PER is not essential for its translocation. Indeed, PER appears to enter the nucleus as a homodimer (Huang et al., 1995; Landskron et al., 2009), as disruption of PER-PER homodimer formation affects PER's subcellular distribution, *per* transcription and lengthens locomotor period (Landskron et al., 2009).

Once in the nucleus PER can inhibit CLK/CYC mediated transcription by binding CLK to form a complex (Lee et al., 1999) in a process that is somehow accelerated by the TIM's earlier drop in nuclear abundance (Zeng et al., 1996). During this process there is a coincident increase in levels of hyperphosphorylated PER and hyperphosphorylated CLK, both of which are dependent on DBT (Yu et al., 2006). Hyperphosphorylated CLK and PER are targeted for degradation in the early morning, allowing another cycle of transcription to begin.

Intriguingly the phosphorylation of PER is not required for its repressor function (Yu et al., 2009), however the phosphorylation of CLK is required for transcriptional repression,

presumably by reducing CLK's affinity for E-boxes. Though phosphorylation requires DBT, DBT itself does not directly phosphorylate CLK (Yu et al., 2009), suggesting that DBT recruits other – as yet unknown – kinases to the complex, as shown in Figure 1-5:

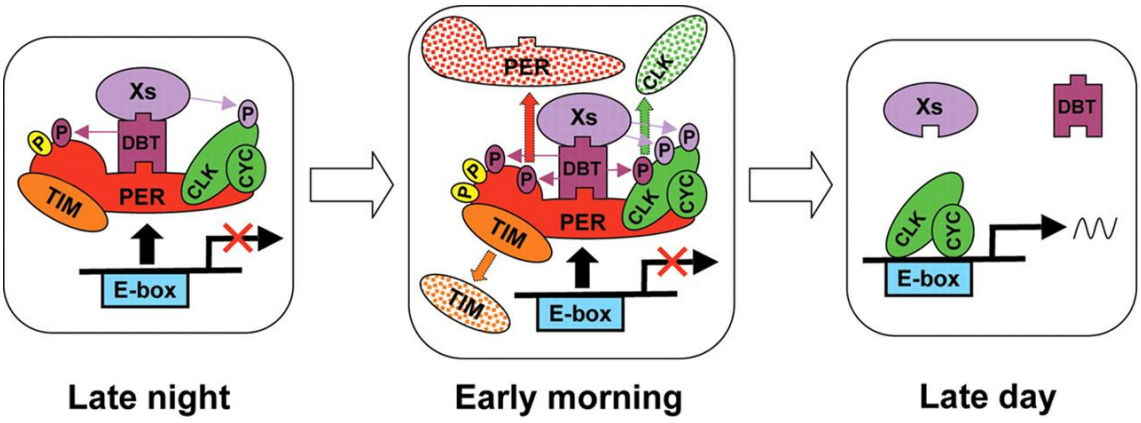


Figure 1-5 Model of non-catalytic DBT function during the circadian cycle. Speckled shapes represent degraded proteins. **XS** represents the DBT recruited kinases that phosphorylate CLK. Figure from Yu *et al.* (Yu et al., 2009).

Together the negative feedback loop leads to the rhythmic transcription of *clk*/*cyc* and *per*/*tim* that has long been thought to represent the core mechanism by which the molecular clock generates rhythmic outputs. However, blocking the rhythmic expression of *Clk* (Kim et al., 2002), *per* or *tim* independently (Cheng and Hardin, 1998) or together (Yang and Sehgal, 2001) is sufficient to abrogate rhythmic behaviour. Indeed, whilst CLK levels have been observed to cycle in phase with *Clk* (Lee et al., 1998), this result is dependent on the CLK antisera used, other groups reporting that CLK levels may actually be constant throughout the diurnal cycle (Houl et al., 2006) – as has been demonstrated for the expression of CLK's binding partner *cyc* (Bae et al., 2000).

Together these data suggest that additional transcriptional and post translational rhythms exist which act to increase both the amplitude and robustness of the molecular clock.

1.3.2 The accessory negative feedback loop

clockwork orange (cwo) is a transcriptional repressor of the basic helix-loop-helix-O (bHLH-O) superfamily which is transcribed in a cyclical fashion under CLK/CYC control in a similar manner to *per* and *tim*, though showing a lower amplitude of expression and an earlier phase peak. CWO competes with CLK/CYC for access to E-boxes (Matsumoto et al., 2007); this probably serves to terminate CLK/CYC mediated transcription, as *cwo* mutants have a long period and delayed phase under DD conditions. bHLH-O proteins are linked to histone deacetylase recruitment which epigenetically marks chromatin to adopt its closed conformation (Davis and Turner, 2001), which may antagonise the CLK mediated histone acetylase activity. Supporting this, the expression levels of genes under control of CLK/CYC do not fall to wild type levels in *cwo* mutants, suggesting that chromatin is retained in a more open configuration (Kadener et al., 2007).

1.3.3 The positive feedback loop

Microarray studies have shown that CLK/CYC activates the transcription of many other genes in addition to *per*, *tim* and *cwo* (Ueda et al., 2002; Etter and Ramaswami, 2002; McDonald and Rosbash, 2001; Ceriani et al., 2002; Lin et al., 2002; Claridge-Chang et al., 2001), including the transcription factors *vrille* (*vri*) (Blau and Young, 1999; George and Terracol, 1997) and *Par domain protein-1ϵ* (*Pdp1*) (Lin et al., 1997; Cyran et al., 2003) which compete for access to VRI/PDP1ϵ (V/P) boxes.

VRI is a transcriptional repressor that binds V/P boxes in the *Clk* enhancer during the early night (~ZT15). PDP1 levels peak approximately three hours after VRI, at which time PDP1ϵ competes for access to V/P boxes, displaying VRI and acting as a transcriptional activator of *Clk* and thereby creating a positive feedback loop controlling *Clk* expression. Transgenically altering the expression of *vri* can be shown to affect period length and may result in behavioural arrhythmicity (Blau and Young, 1999; Cyran et al., 2003), whilst disruption

of the PDP1 ϵ-isoform alone is sufficient to disrupt rhythmic phenotypes (Zheng et al., 2009 contrary to results in; Benito et al., 2007).

1.3.4 Starting the clockwork

The interlocked feedback loops of the circadian clock, like mechanical clockwork, can create a self sustaining oscillator, however how and when is the first cycle of transcription initiated? It is clear that the circadian timekeeping system of *Drosophila* is functioning as early as the first larval instar (Kaneko et al., 1997), and given that oviposition itself follows a circadian rhythm (Paranjpe et al., 2004), there may be some degree of maternal effect contribution.

It has been shown that the ectopic mis-expression of *Clk* is sufficient to generate ectopic peripheral clocks (Zhao et al., 2003), suggesting that *Clk* expression is the key stimulus. It appears that transcription factors such as *eyeless* may stimulate the transcription of *Clk* (N. Glossop, pers. comm.), and a recent study has demonstrated that the *Fer2* transcription factor also plays a role (Nagoshi et al., 2010). The existence of as-yet unidentified transcription factors promoting *Clk* expression is further supported by the observation that *Clk* levels are high in both Clk^{Jrk} and cyc^{0} mutants, in which one would expect their levels to be low (Glossop et al., 1999; Glossop et al., 2003).

1.3.5 Entraining the oscillator

Under natural conditions, temporal cues are provided by daily cycles in light intensity and temperature, both of which are capable of independently entraining the circadian clock (Zimmerman et al., 1968), and are therefore referred to as 'Zeitgebers' (time-giving stimuli).

Light is the major Zeitgeber in *Drosophila*, detected both by light sensitive organs in the head, and by the blue light photoreceptor CRYPTOCHROME in a cell autonomous fashion.

Careful dissection of the photoentrainment system has revealed that the different photoreceptors have subtly different roles in entrainment (Rieger et al., 2003).

Whilst the principals of entrainment are examined in considerable detail in chapter 4, for the purposes of this introduction it is sufficient to say that CRY, on activation by light, binds to TIM (Ceriani et al., 1999), causing the phosphorylation of key tyrosine residues on TIM. This phosphorylation triggers the sequential degradation of TIM and then CRY in the proteasome (Naidoo et al., 1999), thereby resetting the intracellular oscillator mechanism as shown in Figure 1-4. The extent to which light 'resets' the clock is a function of the state of the clock; during the early night high levels of *tim* mRNA ensure that degraded TIM is replaced, albeit after a phase delay. In contrast, the administration of light during the late night speeds the degradation of the remaining TIM, causing a phase advance.

1.4 The *Drosophila* circadian clock as an intercellular network

Although the intracellular clock mechanism is now well understood, the relationship between the various timekeeping centres of the *Drosophila* brain has only recently become amenable to experimental dissection. The concept of a multi-oscillator clock was first proposed by Pittendrigh and Daan in 1976 as a model to explain how a circadian clock might adapt to seasonal changes in the environment (Pittendrigh and Daan, 1976). Furthermore, evidence is beginning to emerge that communication between oscillators may serve to stabilise - and possibly synchronise - intracellular rhythms (Nitabach et al., 2005), in much the same way that accurate measurements of time now use the consensus result of several independent atomic clocks (Allan et al., 1997).

Immunostaining of PER and TIM expression identifies ~200 cells in the fly brain, as shown in Figure 1-6. These cells can be subdivided into – arguably – 10 bilaterally symmetrical groups according to their positions and transcriptional profiles (Kaneko and Hall, 2000;

Helfrich-Forster et al., 2007; Shafer et al., 2008). Careful genetic dissection has revealed that these groups play distinct, hierarchical roles within the circadian clock, both as a result of the distinct transcriptional identity of cell groups, and of the pattern of synaptic connectivity linking groups to form a network:

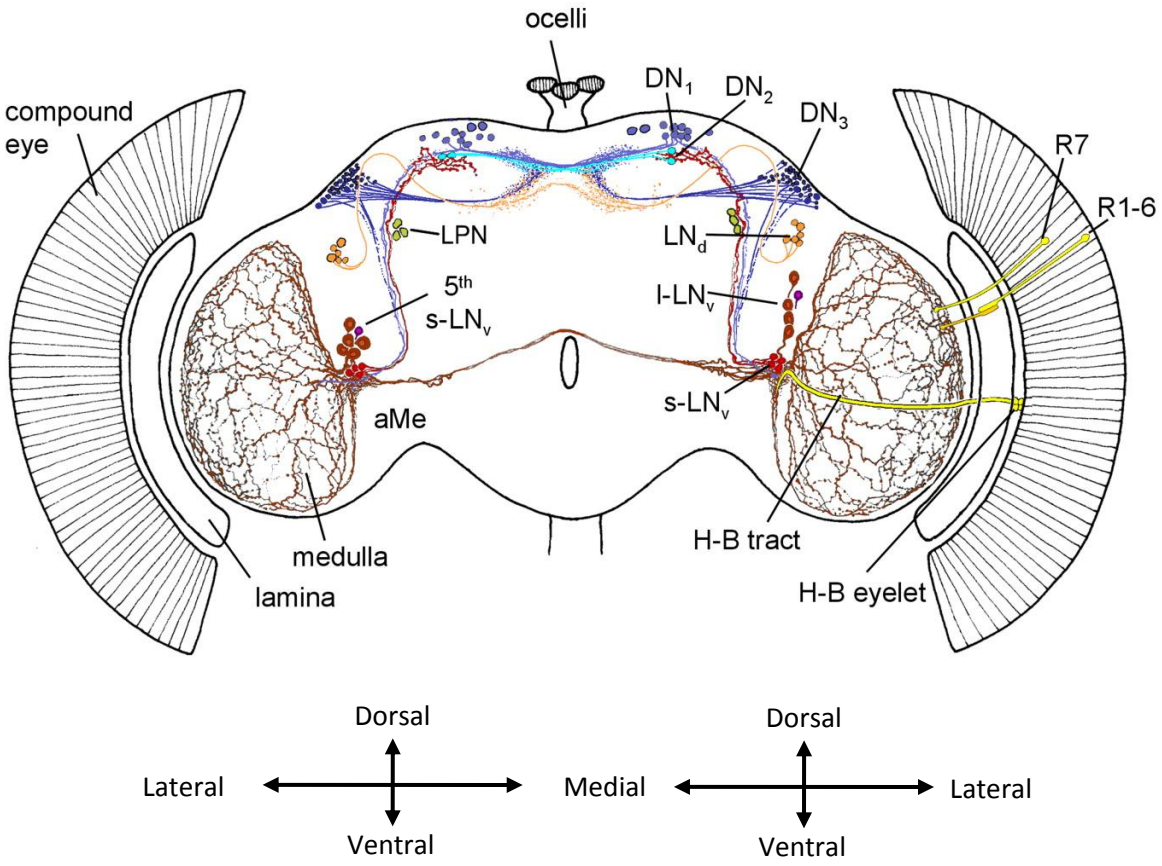


Figure 1-6 Photoreceptors and clock cells within the *Drosophila* head. Clock cells are identified by the presence of the core clock proteins PER and TIM, and are named in relation to their location within the head. Some clock cells also contain the circadian neuropeptide PDF, as detailed in the body text. The lateral neurons consist of the LN_d cells (**orange**), the PDF-positive l-LN_v and s-LN_v cells (**red**), and the PDF-negative fifth s-LNv cell (**violet**). The dorsal neurons comprise the DN_1, DN_2, and DN_3 cells (all in **blue**). Furthermore, three PER/TIM positive neurons are located in the posterior lateral brain (lateral posterior neurons; LPN, **green**). The known projections of the circadian neurons are illustrated; almost all clock neurons whose projections have been mapped synapse within the dorsal protocerebrum. Clock cells receive light information from the photoreceptor cells (R1-8) of the compound eye and the four Hofbauer-Bucher (**H-B**) eyelet cells as shown. Further details concerning the connectivity of clock cells, and the original figure, can be found in Helfrich-Forster *et al.* (Helfrich-Forster et al., 2007).

1.4.1 Functions of the clock cell clusters

The various clock neurons of the brain have been shown by a number of groups to play very specific roles within the clock (Grima et al., 2004; Stoleru et al., 2004); although all clusters express the *per/tim* intracellular clock machinery, each controls a distinct aspect of rhythmic outputs, the ensemble of their output creating *Drosophila's* crepuscular pattern of locomotor activity in LD conditions. Furthermore, chronic manipulation of various clock cell clusters (using both genetic ablation and transgenic rescue of the per^0 clock mutant) has shown that whilst some clusters act as essential 'pacemakers', others play accessory roles in generating rhythmic outputs.

The LN_vs appear to be the key pacemakers, as they alone are both necessary and sufficient to generate crepuscular locomotor activity patterns under light-dark conditions (Renn et al., 1999), and robust locomotor rhythms under free running conditions (Ewer et al., 1992; Frisch et al., 1994). The LN_vs appear to be especially important in controlling 'morning' activity, as transgenically blocking their output reduces flies' anticipation of dawn (Kaneko and Hall, 2000; Blanchardon et al., 2001), whilst rescue of clock function in all cells but the LN_vs is sufficient for evening anticipation (Stoleru et al., 2004).

Whilst the morning peak is dependent solely on the LN_vs, the evening peak can itself be split under dim light conditions (Yoshii et al., 2004; Rieger et al., 2006), suggesting that it represents a conjugate peak. The morning and evening peaks are clearly dependent upon each other; selectively speeding the clock in the 'morning' LN_v cells by over-expressing SGG (Martinek et al., 2001) not only speeds mRNA oscillations in the sLN_v cells, but also in the 'evening' LN_d, DN_1 and DN_3 cells (Stoleru et al., 2005). Surprisingly however, this manipulation did not alter the phase of the SGG overexpressing lLN_vs, nor the DN_2s. Indeed, performing the reciprocal experiment reinforced the conclusion that the DN_2s and lLN_vs clusters form a distinct clock network (Stoleru et al., 2005).

Such a result is not wholly unexpected; the DN_2s are known to be molecularly antiphasic to other clock cells, both in larval stages (Kaneko et al., 1997; Kaneko and Hall, 2000) and in the adult (Costa, R. unpublished observations). Indeed, the correlation between these cells' antiphasic molecular cycling, and the almost perfect 12 hour antiphasic peaks of morning and evening activity logically argues that these cells might play a role in driving the evening peak of activity. Though these cells are not capable of driving self sustaining oscillations in DD conditions (Klarsfeld et al., 2004), they appear particularly sensitive to light (Veleri et al., 2003), and hyper-excitation of the DN cells has been shown to significantly reduce daytime activity and induce some degree of LN_v desynchrony (J. Blau pers. comm., S. Dissel pers. comm.).

Whilst the function of the lLN_v cells within this distinct network remains unclear, a number of recent publications have shown that light-arousal and circadian photoreception circuits intersect at the lLN_v cells, suggesting important roles in entrainment and sleep (Shang et al., 2008; Sheeba et al., 2008a).

It is less clear whether the distinct clusters of clock cells constitute a bipartite oscillator as predicted by Pittendrigh and Daan's model. One feature of a multiple oscillator model is that the oscillators are differently responsive to entraining stimuli, therefore under certain conditions may become decoupled or 'split'. Indeed, such decoupling of behavioural components can be observed in *Drosophila* under particular environmental conditions or as a result of certain mutations (Yoshii et al., 2004; Rieger et al., 2006; Helfrich, 1986; Helfrich-Forster, 2000; Nitabach et al., 2006). However, categorising cells as belonging to 'morning' or 'evening' oscillators is likely to represent an over-simplification; whilst the relative contribution of each cluster to clock output is likely to be gated by light, it is not yet clear that they represent truly independent, though coupled, oscillators.

1.4.2 Communication between clock cell clusters

Intracellular communication is a key component of the clock mechanism in *Drosophila*; indeed inhibiting electrochemical communication between clock cells stops the free running clock (Nitabach et al., 2005; Nitabach et al., 2002), whilst a number of ion channel mutants have compromised circadian clocks (Lear et al., 2005a; Elkins et al., 1986). Advances in the resolution of patch clamping technology have made it clear that the neuro-excitation profile of individual clock neurons changes significantly throughout the day (Sheeba et al., 2008b).

In addition to electrochemical signalling, the circadian clock employs a diverse range of neuroendocrine signals to convey time information. These include *neuropeptide F* (Lee et al., 2006), *neuropeptide-like precursor 1* (Baggerman et al., 2002), GABA (Parisky et al., 2008), glutamate (Hamasaka et al., 2007), aminergic signals (Hamasaka and Nassel, 2006), serotonin (Nichols, 2007) and CCAP (Park et al., 2003). For an in depth review consult Chang (Chang, 2006).

However, the most important neuropeptide for conveying temporal cues appears to be *Pigment dispersing factor (Pdf),* the transcript most enriched in the LN_v cells (J. Blau pers. comm.). First identified in the eyestalks of the fiddler crab *Uca pugilatorwas* (Rao et al., 1985) where it was shown to have a role in synchronising oscillators (Petri and Stengl, 1997), the *Drosophila* homologue was cloned and shown to be expressed in the LN_v neurons (Helfrich-Forster, 1995; Park and Hall, 1998). Whilst both Pdf^0 mutants and transgenic flies lacking *Pdf* expressing neurons can entrain to light, the majority of such flies rapidly become arrhythmic under DD conditions (Renn et al., 1999). There is some evidence that PDF acts locally to synchronise the intracellular clocks of adjacent pacemaker cells (Lin et al., 2004).

PDF reactivity in the termini of s-LN_vs shows a strong circadian oscillation (which may represent cyclical release of PDF), reaching its zenith at ZT2-3 and nadir around ZT13-15, and this rhythm is dependent both on the molecular clockwork and electrochemical signalling (Park

et al., 2000; Wu et al., 2008b), though this cyclical accumulation appears dispensable for its function (Wu et al., 2008b; Kula et al., 2006).

The PDF receptor (PDFR) was independently identified by three groups, who showed that PDFR stimulates downstream signalling mediated by cAMP (Hyun et al., 2005; Lear et al., 2005b; Mertens et al., 2005). Using a novel FRET reporter, Shafer *et al.* were able to show that PDFR is present in all the clock neurons of the brain, with the exception of the $DN1_p$s and LN_ds, which themselves lay in close proximity to non-clock cells in which PDFR is present, and may therefore receive some processed PDF input (Shafer et al., 2008). This strongly suggests that PDF is the key factor controlling communication between cells in the intracellular clock network (for discussion see Chang, 2006)

1.5 Clock output and peripheral oscillators

As described, the circadian clock is deeply rooted in body physiology, and as such many bodily outputs show a clear diurnal rhythm (described in detail in chapter 5). The most commonly studied clock output is that of locomotor activity, which can easily be assessed using automated activity monitoring equipment as described in the methods section and chapter 3.

The connection between the molecular feedback loops, the intercellular timekeeping network, and the peripheral oscillators and rhythmic outputs is largely unknown. A number of mutants have been identified that specifically abrogate a single rhythmic output, suggesting that a number of distinct output signalling pathways are employed. The most widely known such mutation is *ebony*, which solely affects eclosion rhythms, as does the neuropeptide *Ccap* (Park et al., 2003; Clark et al., 2004). Other mutations known to affect circadian output but not the oscillation of pacemaker cells include another RNA binding protein *lark* (Newby and Jackson, 1993), its binding partner *dfmr*, and *takeout* (So et al., 2000; Sarov-Blat et al., 2000).

An alternate explanation for the independence of output rhythms is that outputs may depend not on cues from the central oscillator, but on information generated by peripheral clocks distributed throughout the body (Bell-Pedersen et al., 2005). The mechanism of peripheral clocks may differ from that of the central oscillators in that CRY may play a role in the transcriptional feedback loop in a manner similar to its role in the mammalian clock system (Collins et al., 2006; Krishnan et al., 2001; Ivanchenko et al., 2001) however this is not the case in all peripheral oscillators (Ito et al., 2008). This mechanistic degeneracy may be the underlying cause for the dissociation of different rhythmic outputs.

1.6 Project aims

Identifying the function of genes is the key to unlocking the promise of the genomic era, however both quantitative genetics and Benzerian forward genetics have failed to determine the function of the majority of genes. This failure is not the result of a lack of screening effort; many screens in *Drosophila* have reached saturation (Greenspan, 1997); therefore further screening using these techniques has little potential to generate further information.

This is especially the case in well characterised systems such as the circadian clock, which due to its therapeutic potential has been subject to intense scrutiny for almost 40 years. Although the intracellular clock machinery is now understood in great detail, there remain a number of unresolved issues. Why are *Clk* levels elevated in Clk^{JRK} mutants in which the positive feedback loop is not functional? What stimuli prime the circadian machinery, and how do central and peripheral oscillators differ? If PER is targeted for degradation using the E3 ligase SLIMB, might this not also require E1 and E2 ligases? How does an inherently biphasic ON/OFF TTO system give rise to rhythmic outputs that peak throughout the day and night? These and other questions make it clear that the circadian clock model as it stands is incomplete (Blau et al., 2007).

This thesis addresses the problem of identifying gene function using a number of forward and reverse genetics techniques using the circadian clock as its primary model.

2 General Materials and Methods

2.1 Fly Keeping

Fly stocks were kept in glass vials (10cm x 2.2cm) filled with ~2cm of food medium. Three mediums were used;

- Sugar medium (4.63g sucrose, 4.63g live yeast, 0.71g agar and 0.2g Nipagin in 100ml water).
- Oatmeal medium (13g rolled oats, 4ml black treacle, 0.7g agar and 0.6ml 20% Nipagin solution in 100ml water).
- Maize medium (7.20g maize meal, 7.93g glucose, 5g brewers' yeast, 0.85g agar, 0.3ml propionic acid and 1.35ml 20% Nipagin solution in 100ml water).

Vials were tightly bunged with cotton wool balls and kept in temperature controlled rooms (either at 18°C or at 25°C) subject to a 24 hour fluorescent lighting cycle of 12 hours light followed by 12 hours of dark.

Flies were handled in accordance with the principals laid out by Ralph Greenspan (Greenspan, 1997) using CO_2 anaesthesia. Crosses were performed using appropriate balancers as detailed by Ashburner (Ashburner et al., 2005).

2.2 Locomotor activity experiments

Clean, cylindrical glass tubes (10cm x 0.5cm) were prepared containing ~1cm of sugar/agar medium. Tubes containing food were left open for ~4 hours to allow the food to dry, before being closed with a rubber cap (TRIKINETICS) to prevent further desiccation. Activity experiments were carried out using anaesthetised flies loaded singly into the glass

tubes, sealed with a coloured cotton wool bung. Caps, bungs and glass tubes were cleaned in IMS and reused between experiments.

Tubes filled with flies were secured within 32 channel *Drosophila* activity monitors (DAM5, TRIKINETICS) using a rubber band such that the infra-red emitter/detector pair was located centrally on each tube.

Up to four DAM5 units were loaded inside fan cooled metal 'light boxes' containing LED lighting. The lighting inside each box was controlled using either digital timing units (sure time Stu27), or a Trikinetics experiment controller. Unless otherwise stated, flies were subjected to 3 full days of entraining LD conditions.

Up to 6 light boxes were loaded into incubators (LMS 303A) running at a constant temperature of 25°C unless otherwise stated. Humidity within incubators was not controlled.

The Trikinetics DAM software (v3.03, http://www.trikinetics.com/) is designed such that only a single count is recorded every time a fly crosses the infra-red beam. Data records were truncated so as to begin at 0200 (ZT18), and beam crossing events were summed to form bins of either 5 minutes (for sleep studies), or 30 minutes (for circadian studies).

2.3 Eclosion experiments

For each genotype tested, 20 vials of sugar food were seeded with an equal number of flies and placed in light boxes within a temperature controlled incubator. Vials were kept in LD 12:12 conditions such that half the vials were antiphasic to the other half. After a week the adults were removed and vials were transferred to DD conditions. 11 days after the start of the experiment the vials were once again emptied and newly emerging flies were collected and counted every two hours for 12 hours a day (the first collection removing flies that had emerged overnight).

Eclosion results were normalised using the number of vials producing progeny (as some vials produced no flies due to bacterial infection), and this normalised number was expressed as a ratio relative to the total number of flies emerging over the course of the five day experiment (including those flies emerging overnight). Eclosion ratios for vials kept in opposite phases were then spliced together to form a continuous five day eclosion record.

2.4 Polymerase chain reactions

PCR reactions were carried out on a DYAD PELTIER THERMOCYCLER from Bio-Rad using the following protocols:

2.4.1 DNA Preparation Protocol

Unless otherwise stated, DNA for PCR reactions was prepared using Georg Dietzl's protocol:

1. The squishing buffer (SB) is 10mM Tris-Cl pH 8.2, 1mM EDTA, 25mM NaCl, and 200µg/ml Proteinase K, with the enzyme diluted fresh from a frozen stock each day.
2. Place one fly in a 0.5ml tube and mash the fly for 5 - 10 seconds with a pipette tip containing 50µl of SB, without expelling any liquid (sufficient liquid escapes from the tip). Then expel the remaining SB.
3. Incubate at 25-37 °C (or room temp.) for 20-30 minutes.
4. Inactivate the Proteinase K by heating to 95 °C for 1-2 minutes.

2.4.2 Wing PCR

In some cases DNA was amplified from *Drosophila* wings rather than entire individuals. This protocol differed from that above such that in step 2, fly wings rather than entire bodies were placed in 200µl tubes and mashed with 10µl of squishing buffer, whilst in step 3 the tubes were left to incubate overnight.

2.4.3 Polymerases

Regular PCR was carried out using DNA polymerase from either KAPA (http://www.kapabiosystems.com/products/standard/kapataq-dna-polymerase) or NEB (http://www.neb.com/nebecomm/products/productM0273.asp). High fidelity PCR was performed using Finnzymes' 'Phusion' high-fidelity polymerase (http://www.finnzymes.fi/pcr/phusion_flash_high_fidelity_PCR_Master_Mix.html).

2.4.4 Primers

All primers were designed against genomic sequences retrieved from FlyBase (www.FlyBase.org) using the Primer3 primer design program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Primers were manufactured by either Sigma or Invitrogen.

2.4.5 Reagent mix

All PCR reactions were carried out following the appropriate manufacturer's instructions and using appropriate buffers provided by the manufacturer. A typical reagent mix consisted of:

Reagent	Volume (µl)
DNA	1.0
Forward Primer	1.0
Reverse Primer	1.0
Polymerase	0.2
Buffer (5x)	4.0
H_20	12.8
Total	**20.0**

2.4.6 Reaction conditions

Appropriate reaction conditions for PCR products of different length were used following manufacturer's instructions. An example PCR reaction using KAPA taq to amplify a 400bp amplicon from genomic DNA might be:

Step	Temp (°C)	Time (s)
1	92	120 (melting)
2	92	30
3	64	30 (annealing)
4	72	120 (extension)
5	-	Cycle to step 2 x34
6	72	600 (final extension)
7	10	∞ (hold)

All PCR products were run on a 0.8% agarose/TBE gel using an appropriate loading buffer and DNA size markers supplied by NEB.

2.4.7 Touchdown PCR

In some cases regular PCR cycling was preceded by 5 rounds of 'touchdown' PCR, in which a higher annealing temperature is used in an effort to reduce non-specific amplification of DNA during the PCR reaction. For such reactions the following protocol was used:

Step	Temp (°C)	Time (s)
1	92	120
2	92	30
3	70(-cycle#)	30
4	72	120
5	-	Cycle to step 2 x4
6	92	30
7	64	30
8	72	120
9	-	Cycle to step 6 x34
10	72	600 (
11	10	∞ (hold)

2.4.8 Statistics

All statistics were performed using Statsoft's STATISTICA program.

3 Interpretation and Analysis of Circadian Datasets

3.1 Introduction

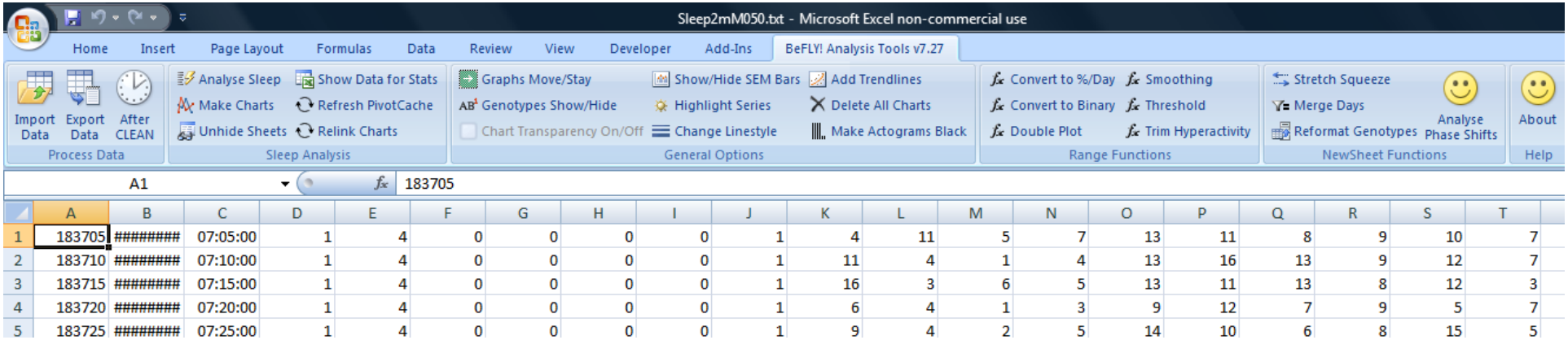
This chapter will introduce the concepts and methods required to understand and interpret the often complex patterns of behaviour which form the large body of the results within this thesis. The reader is reassured that quantitative comparison of various algorithms is well beyond the scope of this work (and likely this author). The emphasis will be on qualitative examples; for a more comprehensive, step based guide to generating such results using the BeFly! package developed by the author, the reader is directed to the manuals within appendix D1.

3.2 Discussion

3.2.1 Activity recording systems

Numerous hardware systems capable of recording Drosophila activity are available, the two systems in use in Leicester being the custom 'Drosophix' system and the newer, commercially available Trikinetics system (http://www.trikinetics.com/, illustrated in Figure 3-1). Although higher resolution infrared (Shaw et al., 2000) or video based (Zimmerman et al., 2008; Heward et al., 2005) activity recording solutions exist, these are largely restricted by cost to small scale studies considering very detailed activity metrics, notably aggression and courtship interactions (Heward et al., 2005).

Figure 3-1 Trikinetics DAM5 apparatus loaded with a single activity tube.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	183705	########	07:05:00	1	4	0	0	0	0	1	4	11	5	7	13	11	8	9	10	7
2	183710	########	07:10:00	1	4	0	0	0	0	1	11	4	1	4	13	16	13	9	12	7
3	183715	########	07:15:00	1	4	0	0	0	0	1	16	3	6	5	13	11	13	8	12	3
4	183720	########	07:20:00	1	4	0	0	0	0	1	6	4	1	3	9	12	7	9	5	7
5	183725	########	07:25:00	1	4	0	0	0	0	1	9	4	2	5	14	10	6	8	15	5

Figure 3-2 A DAM5 results file opened in Microsoft Excel 2007 using the author's 'BeFly!' behavioural analysis tools package.

3.2.1 Basic visualisation of activity records

Due to the file limitations of legacy software, activity recording systems generally present results as columns of numbers, each column representing the activity of an individual fly, and each row representing the total activity occurring during a time interval or 'bin' (as shown in Figure 3-2). For most circadian analyses, a bin size of 30 minutes is the most appropriate interval that balances temporal resolution against the loss of amplitudinal range between peak and trough activity values inherent in the use of smaller bins.

3.2.1.1 Actograms

Typically the crepuscular activity rhythm shown by *Drosophila* under laboratory conditions is visualised and described in terms of a waveform by plotting beam crossing events ('activity') against 'bins' (discrete units of time):

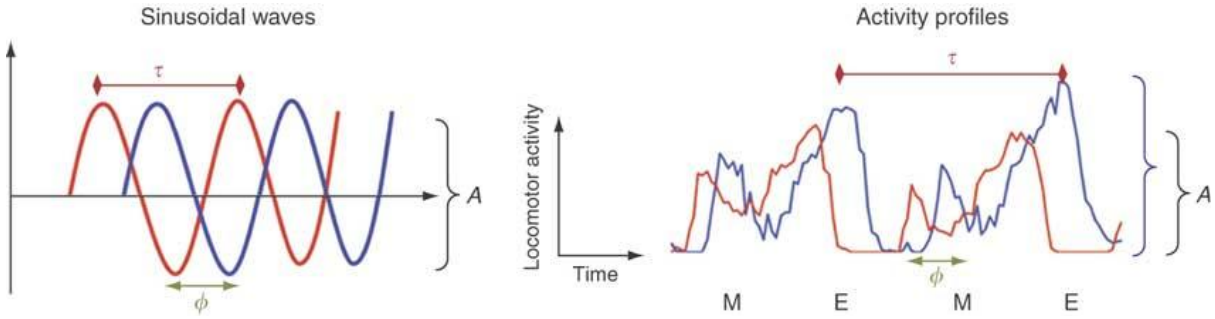


Figure 3-3 Properties of waveforms. Left panel: waveforms are characterized by their period (τ), amplitude (*A*) and phase (ϕ). This example shows two waveforms having the same τ and *A,* but different ϕ. Correspondingly, the same parameters can be used to describe the activity profiles in the right panel, which plots *Drosophila* activity against time, in which the two traces have the same τ but different *A* and ϕ. M and E indicate the crepuscular 'morning' (M) and 'evening' (E) peaks of activity, respectively. Adapted from Rosato and Kyriacou (Rosato and Kyriacou, 2006).

The convention within the field is to plot the mean activity of a genotype against time, as shown in Figure 3-3. This representation is generally referred to as an *actogram*, and formally should be plotted as a bar chart as bins are discrete units of time, though is commonly plotted as a line chart to facilitate comparisons between genotypes. There are however cases

when it is more appropriate to plot the genotype median, as shown in Figure 3-4 and Figure 3-5:

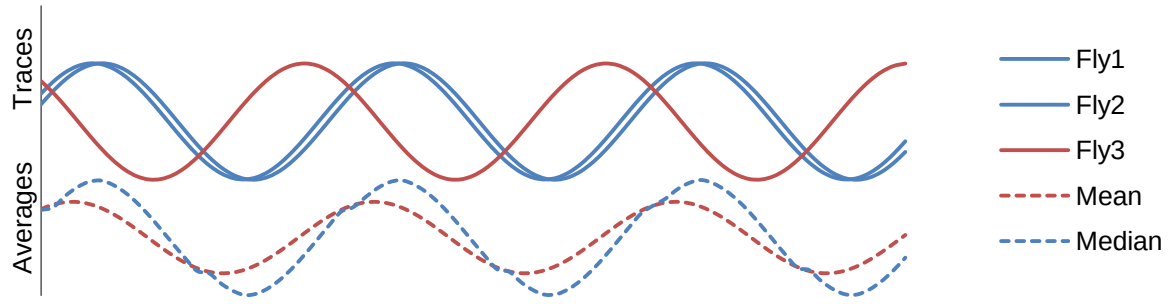


Figure 3-4 Comparison of mean and median values when a minority of flies are out of phase. Top panel, '**Traces'**, represents a simplified representation of three flies of the same genotype displaying locomotor activity of same amplitude and period, but different phase. The lower panel, '**Averages**', represents both the mean and median plots of all three flies.

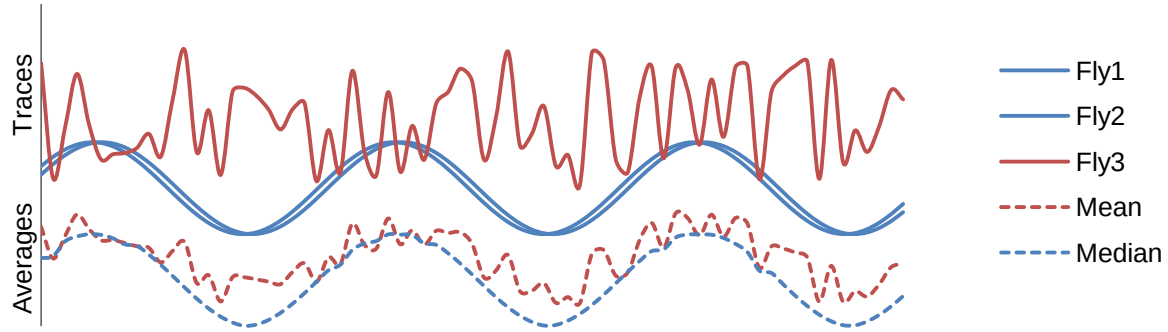


Figure 3-5 Comparison of mean and median values when a minority of flies are arrhythmic. Panels same as in Figure 3-4**,** but this example shows the response of group averages when Fly3 is arrhythmic.

In these examples, where the sample size is small, the median might be argued to better reflect the true 'average' of a fly genotype due to its lesser sensitivity to outlying flies, (which may or may not be experimentally significant). By reflecting the majority of the flies' activity, it is less susceptible to outliers which can cause amplitude or phase changes. These examples make it clear that sample sizes should be a large as possible to minimize the effect of outliers on analysis.

In cases where mutations show variable penetrance, such as the *Party on* mutant described in Chapter4, flies may fall into discrete categories which should not be averaged.

Simple actograms are best generated using the Excel line or bar chart tools, together with a custom styleset that ensures that charts look consistent between experiments.

3.2.1.2 Double plotted actograms

Although an actogram is informative when comparing two or more samples, it is difficult to visualise changes in phase or period during the course of an experiment using an actogram; instead a double plot can be created. Double plots are a form of stacked chart, in which each row contains the data for two units of time (most commonly 2 days, i.e. 48 hours or 96 x 30 minute bins), and the data plotted in each row overlaps one unit of time with the previous row:

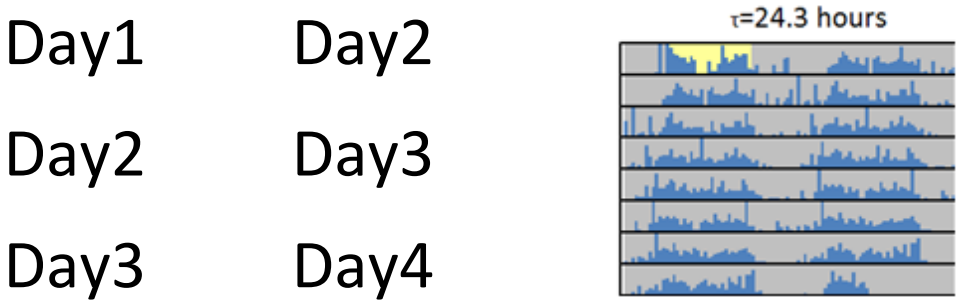


Figure 3-6 Double plot actogram. Left panel shows the format of an actogram, an example of which is shown in the right panel.

This design is particularly useful for visualising slow changes in activity and complex rhythmicities in which several rhythmic components may combine, generating complex harmonic rhythms not easily detected using an actogram (Yoshii et al., 2004):

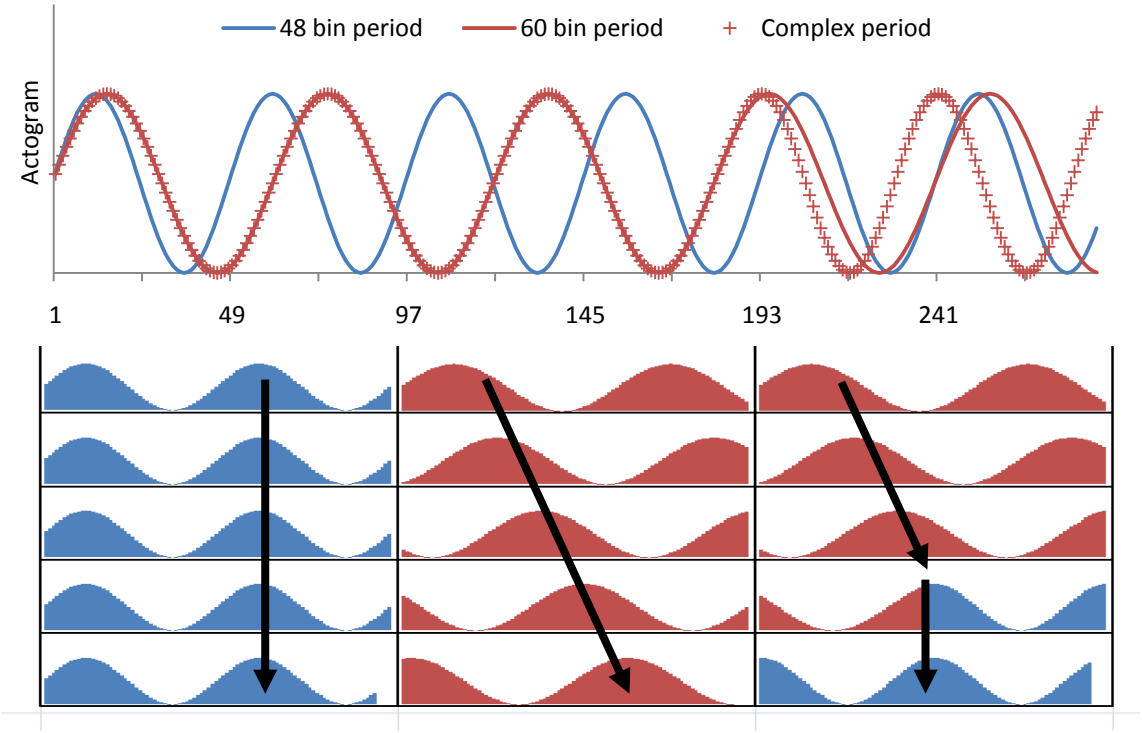


Figure 3-7 The use of double plots to identify changing rhythmicities. Top panel shows a normal actogram of three waveforms representing idealised traces for rhythms with a 24 hour (48 bin) period (blue trace), a long 60 bin period (red trace), and a complex rhythm which switches between 48 and 60 bin periodicities (red hatched trace). Although differences between genotypes are evident, the actogram is difficult to interpret. Lower panel shows a double plot of the data in top panel, which allows facile identification of long periods and period changes (indicated by black arrows).

Although not a common technique, double plots can be used to provide an estimate of periodicity; for an example implementation concerning complex ultradian rhythmicities see Figure 5-7.

3.2.2 Data processing

When analysing circadian data, it is often required to filter or 'smooth' raw data to facilitate later circadian analyses and comparisons (Helfrich-Forster, 2000; Dowse, 2007; Levine et al., 2002b). A number of filter functions are available in the BeFly!, package, and each function has a different purpose, and must be used in an appropriate context.

3.2.2.1 % activity per unit time

When recording activity for long periods of time, flies show both linear and non-linear trends in activity as a result of both factors such as ageing and changes in the environment within the activity tube (including build-up of faecal matter and desiccation of the food source). These trends can be removed from data using Fourier based methods to exclude low frequency noise - i.e. long term trends (Levine et al., 2002b).

Trends can also be removed by expressing raw activity data as a ratio relative to a unit of time (i.e. % activity per day). This filter should be used either when comparing flies which have activity profiles of dramatically different amplitude as a result of their genetic background (notably mutants in genes such as *couch potato* which affect activity levels (Bellen et al., 1992)), or flies of different ages:

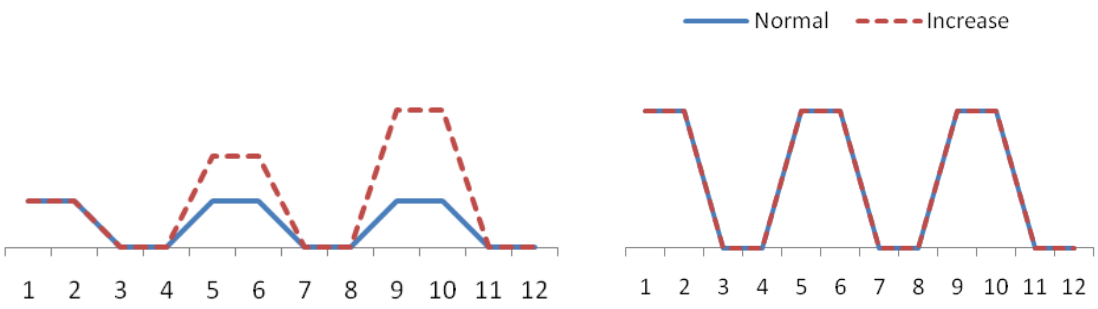


Figure 3-8 Eliminating long term trends by converting raw activity data to % activity per day. Left panel shows two in phase waveforms, the dotted red line showing a waveform which increases in amplitude over time. By converting the raw data for each waveform into a relative value, namely the % activity per cycle (in this case 3 complete cycles are shown for each waveform), the trend is eliminated and the waveforms overlap (**right panel**).

3.2.2.2 Convert to binary

By convention, double plot actograms (specialised time activity plots described later in this chapter) may be plotted in a binary fashion, particularly when data is recorded at a high temporal resolution, as shown in Figure 3-9. Data can easily be converted to a binary form in BeFly! using a range filter.

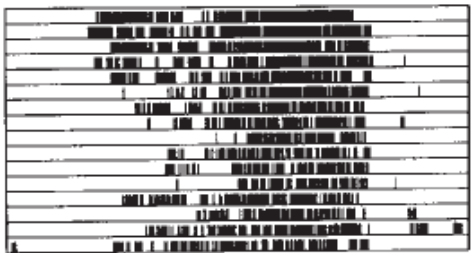


Figure 3-9 Double plot of binary activity data. Figure from Helfrich-Forster (Helfrich-Forster, 2000).

3.2.2.3 Threshold Activity

This filter converts all activity values below a given threshold value to 0, therefore removing low amplitude counts (which are likely to represent stochastic noise) from an activity record:

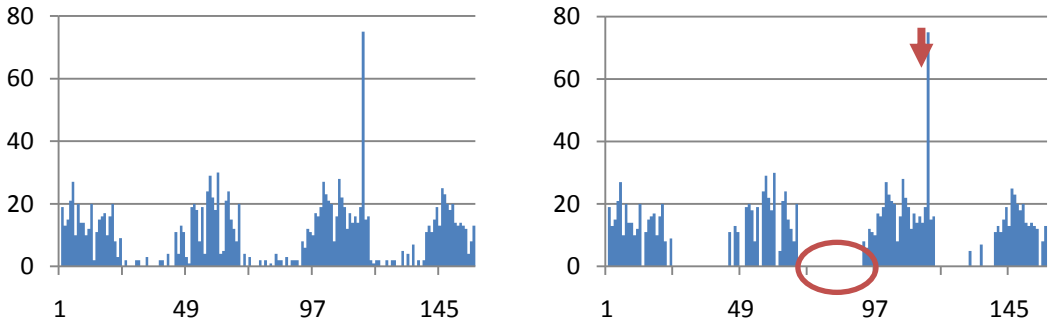


Figure 3-10 The action of a threshold activity filter. Figure shows same activity trace before (**left**) and after (**right**) a minimum threshold filter set to 5 was passed over the dataset. This creates better definition of activity onsets and offsets (**red circle**). Note that this filter has no effect on the aberrantly high data point at bin 116 (**red arrow**) as it uses only a minimum activity threshold.

3.2.2.4 Smoothing

BeFly! implements two smoothing filters which do not change the phase of the original data; an implementation of the two pass, finite response Butterworth filter (adapted from the Brandeis Rhythms Package), and a weighted moving average filter (here referred to as a 'Triangular' filter due to the usual weighting of points).

Simply put, the triangular filter 'smooths' a data point (x) by calculating the mean of (x) and both its adjacent points (x-1 and x+1), and the adjacent but one points (x-2, x+2), and

finally substituting the new value for (x). The extent to which the adjacent points contribute to the smoothed point is determined by coefficients *a, b* and *c* as in the equation:

$$\boldsymbol{smoothed\ x} = \frac{[\mathbf{a} \times (\mathbf{x}-\mathbf{2})] + [\mathbf{b} \times (\mathbf{x}-\mathbf{1})] + [\mathbf{c} \times (\mathbf{x})] + [\mathbf{b} \times (\mathbf{x}+\mathbf{1})] + [\mathbf{a} \times (\mathbf{x}+\mathbf{2})]}{[\mathbf{a} \times \mathbf{2}] + [\mathbf{b} \times \mathbf{2}] + [\mathbf{c}]}$$

The filter is passed once over the data from the start to the end. The filter is termed 'triangular' as the coefficients are generally valued such that c > b > a, forming a triangular shape:

a b **C** b a

The Butterworth filter is also a smoothing filter, but as it smoothes in only one direction (similar to Excel's moving average filter), it is passed over the data once in each direction to prevent phase shifting effects. For further consideration of the Butterworth filter, see Dowse and references therein (Dowse, 2007). An example is presented below:

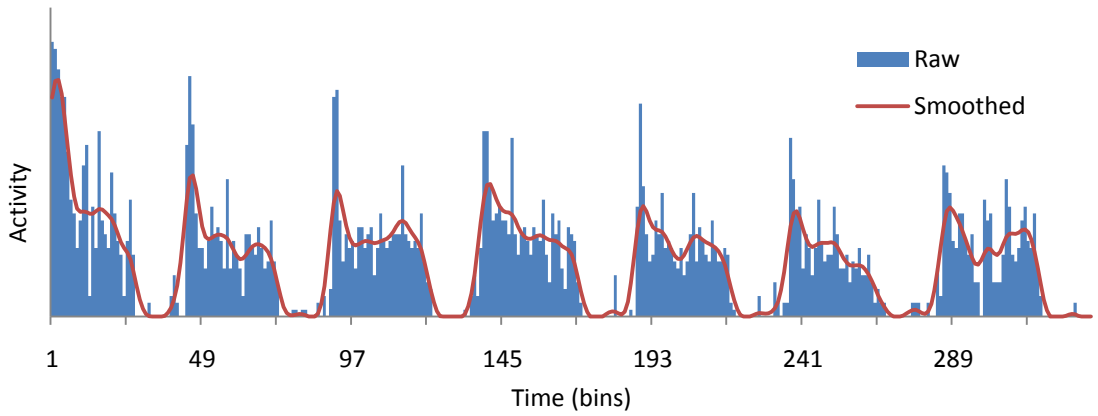


Figure 3-11 The action of a smoothing filter on activity data. The original raw activity data recorded at 30minute resolution (**blue trace**) is passed through the Butterworth smoothing filter to attenuate high frequency noise using the default parameters as used in the Brandeis Rhythms Package to create a smoothed result (**red trace**).

3.2.2.5 *Trim Hyperactivity*

When placed in locomotor activity tubes, typically flies become hyperactive for a short period before death. This may be the result of desiccation of the food source within the activity tube, causing flies to become hyperactivity as a response to starvation (Lee and Park, 2004). This period of hyperactivity distorts the results of activity analysis, and should be removed prior to further analysis. A typical activity trace for a single fly is shown below:

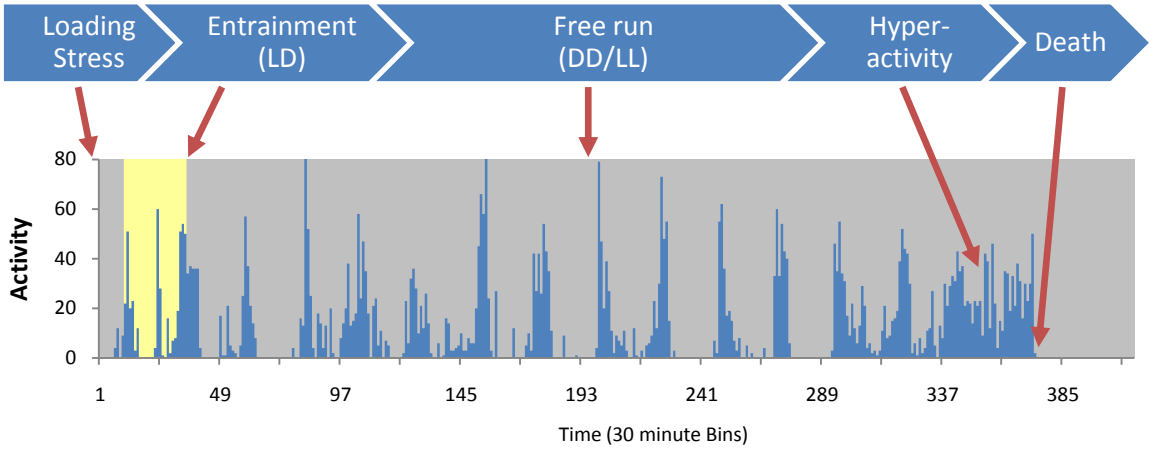


Figure 3-12 Example activity recording for a single fly. Example shows entrainment to light (yellow background), then free running activity in DD conditions (grey background). Note the period of hyperactivity (bins 337-370) preceding death (bin 370).

Behavioural records can be trimmed using the 'Trim Hyperactivity' tool. Dead flies are identified by long series of values falling below a threshold (generally set to 0). The filter identifies the last bin of activity preceding death, and then optionally trims a set period of time preceding that point (which represents hyperactive activity), leaving the 'healthy' activity record.

3.2.3 Quantitative analysis of activity records

It is often required to analyse circadian data in a quantitative fashion to allow objective, statistical analysis of data. The most commonly assessed metric is that of circadian period. Estimates of period are limited in resolution to twice the sampling rate of the data (termed the Nyquist frequency), however the convention within the circadian field is that by

taking the average period of many different individual flies, a measurement of period of greater accuracy than that determined by the Nyquist limit can be achieved (Levine et al., 2002b). A number of quantitative analyses are commonly used to analyse circadian datasets:

3.2.3.1 Autocorrelation analysis

Cross-correlation of two rhythmic datasets creates rhythmic peaks in the correlogram which reflect the periodicity of the data (Figure 3-22). In a similar manner, if a dataset is lagged against itself (autocorrelated), an estimate of that dataset's periodicity can be obtained from the peaks in the correlogram. By convention only half the data series is analysed due to the problems inherent in correlating small numbers of datapoints.

Confidence limits can easily be assigned to peaks in an autocorrelogram, as the 95% confidence limit can be calculated as 2/square root of the total number of time points. The periodicity of the dataset is generally taken as being the highest peak above the confidence threshold, though for small time series the significance threshold may be unrealistically high (Levine et al., 2002b).

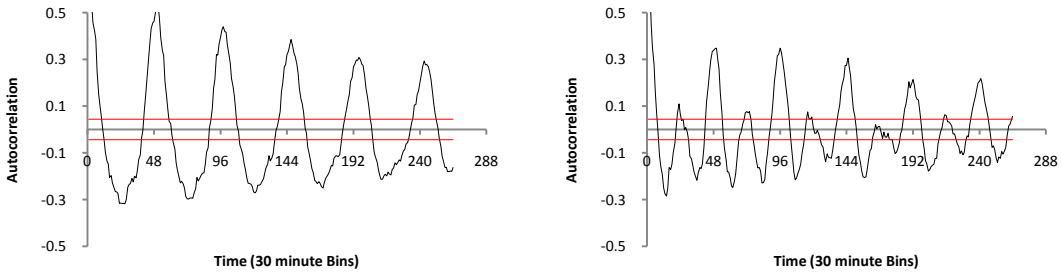


Figure 3-13 Examples of autocorrelograms for unimodal and biomodal activity traces. Correlation in black, 95% confidence limits in red. **Left panel** shows an autocorrelation trace for a fly showing unimodal activity in DD conditions, **right panel** shows the autocorrelation trace for a fly showing considerable bimodality, showing twice as many peaks as the trace on the left. Both flies have the same τ of 24 hours (48 bins).

Autocorrelation analysis is particularly useful in providing an estimate of the quality of a rhythmic dataset (as shown in Figure 3-14); this can be quantified by measuring the height of the third peak in the autocorrelation plot (see Levine et al., 2002b).

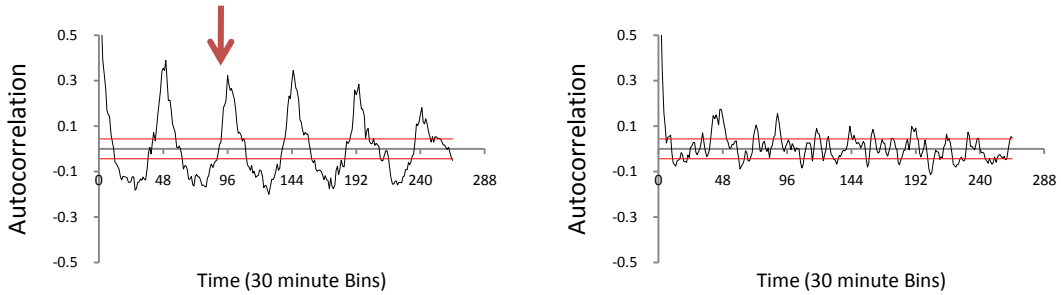


Figure 3-14 Examples of high quality and low quality autocorrelograms. Left panel shows a high quality autocorrelogram, in which a rhythmicity index can be calculated using the height of the third peak (indicated by **red arrow**). Right panel shows an example of a fly showing relatively poor rhythmicity, reflected in the noisy autocorrelogram.

3.2.3.2 Discrete Fourier Transform/CLEAN spectral analysis

Some circadian datasets, particularly those generated in cell culture using luciferase reporters to measure rhythmicity, are particularly complex and require considerable signal conditioning before they can be analysed, and for such datasets autocorrelation analysis is inappropriate. Instead a number of algorithms designed to detect the period of datasets in which considerable noise is superimposed on the underlying rhythmic signal can be employed for circadian analysis.

The CLEAN algorithm is a sophisticated and robust spectral analysis tool developed by radio-astronomers (Högbom, 1974). The CLEAN algorithm relies on the Fourier transformation, a method by which a function (in this case a data series) can be deconvoluted into a series of harmonic sine and cosine terms with coefficients determined by the goodness-of-fit to the data, limited by the Nyquist frequency. The vector sum of the coefficients for the sine and cosine terms at a given frequency (i.e. period) represents the power in the signal attributable to that frequency:

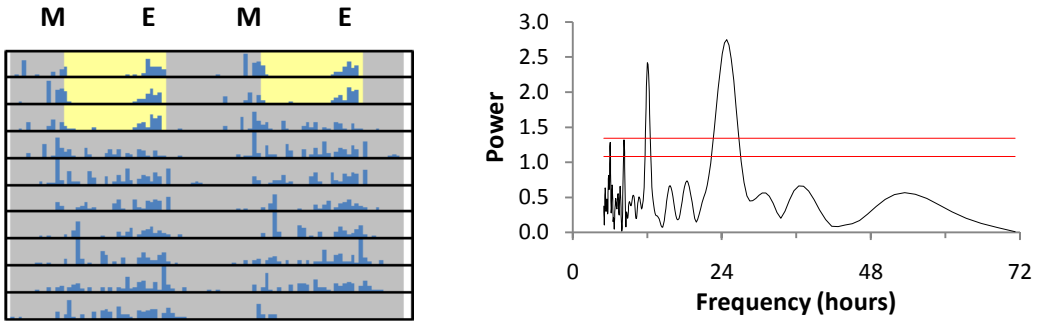


Figure 3-15 Example output of CLEAN analysis of a wild type fly's DD behaviour. Left panel: double plot of a fly showing bimodal activity in both LD and DD conditions. Morning (**M**) and evening (E**)** peaks are indicated. **Right panel**: diagram of spectral power under DD conditions shows both an ultradian, 12 hour peak (reflecting the M-E and E-M periodicity arising as a result of bimodal behaviour) and a circadian, 24 hour peak (reflecting the true M-M and E-E relationship).

Roberts *et al.'s* implementation of CLEAN (Roberts et al., 1987) is an iterative procedure able to cope with common problems in the analysis of experimental time series, such as suboptimal length of sampling and missing data, and therefore this implementation has been adapted for the analysis of both *Drosophila* male courtship songs (Kyriacou and Hall, 1989) and locomotor activity (Sawyer et al., 1997). For additional detail regarding the mathematics of CLEAN see Crane *et al.* or Levine *et al.* (Levine et al., 2002b; Crane, 2001); a more accessible review is found in Rosato and Kyriacou (Rosato and Kyriacou, 2006).

The most recent implementation of CLEAN is part of the MAZ package developed by Zordan and colleagues written in Python (Zordan et al., 2007). In their implementation the significance of the different frequency components identified by CLEAN is assessed relative to 95% and 99% confidence limits calculated by a Monte Carlo data randomisation strategy (Sawyer et al., 1997).

3.2.3.3 Maximum Entropy Spectral Analysis (MESA)

MESA is a variation of spectral analysis developed by Burg (PhD thesis 1978), and like CLEAN is derived from the Fourier transform. MESA is thought to be particularly robust, and has been shown to perform particularly well in determining periodicity in artificial datasets

containing large amounts of noise (Dowse, 2007). MESA is implemented in both Matlab and TurboBasic (Dowse et al., 1989), however unlike the MAZ implementation of CLEAN, MESA implementations do not assess the statistical significance of periodicities, relying instead on autocorrelation metrics. MESA results appear in much the same way as the CLEAN results presented in the right panel of Figure 3-15; an example result is shown in Figure 6-16.

3.2.4 Comparing activity records

3.2.4.1 Waveform regression

If quantitative analysis of two datasets suggests that there is a period difference between them, an important question is whether this is the result of generalised speeding or slowing of the clock mechanism, or whether only a particular portion (or 'phase') of the circadian cycle is affected. This requires comparison of the waveforms of each sample.

For such a comparison to be meaningful, the period of the two waveforms must be normalised, allowing comparison of phase and amplitude information. This is colloquially termed 'stretch and squeeze':

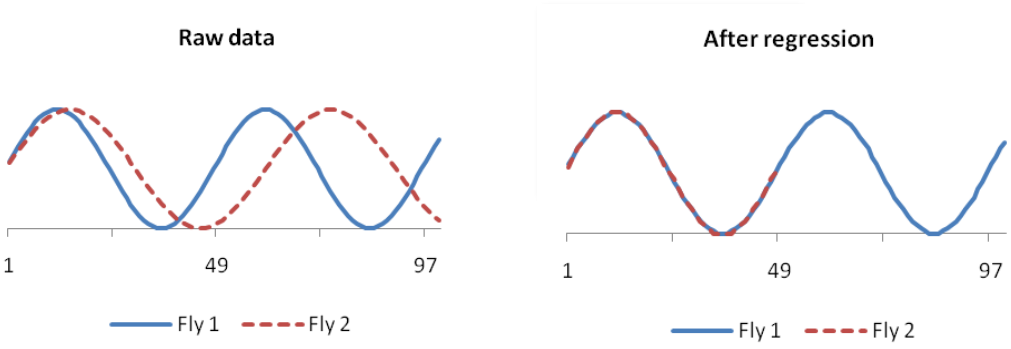


Figure 3-16 Example of stretch and squeeze. Idealised waveforms of tow flies, one of 24 hour periodicity (**blue trace**), the other of 30 hour periodicity (**red trace**). After regression, both waveforms have a normalised period of 24 hours, therefore the profiles can be superimposed to show that they are identical in phase and amplitude.

Once the period has been normalised, the amplitude can be normalised using the '% activity per day' filter described in section 3.2.2.1. The paradigm example of such an analysis is Meng *et al.*'s dissection of the precise nature of short period CK1ε^{tau} mutant hamsters, in

which normalised activity profiles for wild type and mutant hamsters were compared to reveal CK1ε^{tau} to be a phase specific mutation (Meng et al., 2008):

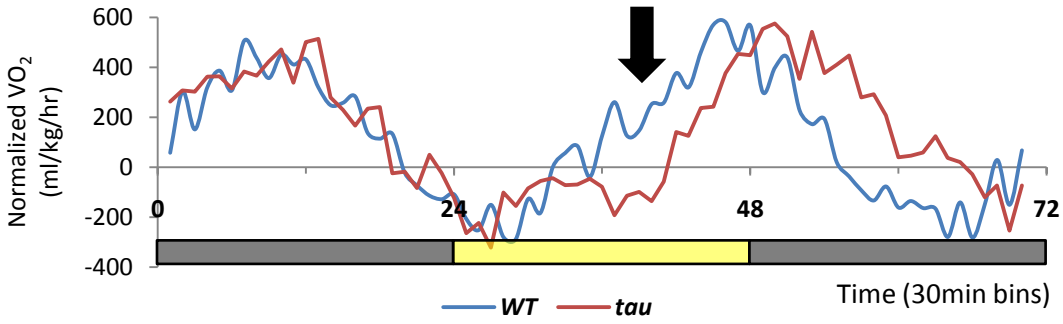


Figure 3-17 Using synchronised activity profiles to dissect the nature of the clock mutation CK1ε^{tau}. Oxygen intake for wild type (***WT***) and mutant (***tau***) hamsters, normalised to the midpoint of the trough-to-peak amplitude. Genotypes show initial phase synchrony during the night when they are active, but desynchony develops during the day (**black arrow**), revealing CK1ε^{tau} to have a phase specific effect on circadian period. Figure recreated from Meng *et al.* (Meng et al., 2008).

3.2.4.2 Merge Days

Locomotor activity output is the result of brain motor centres integrating input from many sources, including circadian time, metabolic state, short term arousal, recent activity and social interactions. In general the effects of such competing factors are removed by taking the average activity of many flies of the same genotype to form an activity profile reflecting circadian effects alone, and this average can then be subjected to regression as described above.

However, if individuals are to be compared, this averaging approach is not possible. Instead each individual's activity can be averaged over time, as the stochastic factors acting at any given time on day 1 will not be the same as those acting at the same time on day 2, whilst the underlying circadian oscillation will be the same. By averaging, or 'merging', several days of activity recorded under identical environmental conditions, an activity profile can be built up for each fly. Before performing such a merge, data should be detrended using the % activity per day filter to compensate for any amplitude changes occurring between days.

Calculating a merged day is useful when calculating sensitive phase metrics, such as those described by Helfrich-Forster revealing subtly sexually dimorphic aspects to *Drosophila* activity patterns (Helfrich-Forster, 2000):

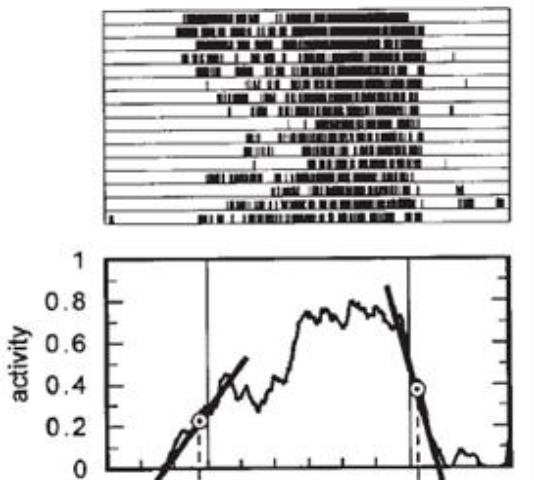


Figure 3-18 Data merging. **Top panel**: single fly binary double plot recorded over 16 days in DD conditions**.** **Bottom panel**: results of merging 16 days of activity to form an average day actogram. Figure adapted from Helfrich-Forster (Helfrich-Forster, 2000).

3.2.4.3 Phase shifts and the phase response curve

The techniques above (including autocorrelation and spectral approaches) assume that individuals' free running behaviour remains constant under free running conditions. However, this is not always the case, and certain genotypes may show changes in amplitude, period and phase over time. Fortunately however, the state of an organism's clock can be directly assayed by administering light pulses at different points during the night, eliciting a phase shift which reveals the phase of the clock without recourse to complex regression or data merging techniques (for a description of the effects of light on the clock, see section 1.3.5).

In order that flies of different periods can be compared, experiments generating phase response curves are generally 'anchored' such that the light pulse is administered whilst the system is still under entrainment conditions, and therefore the phase of the organism is equal to that of the Zeitgeber (Rosato and Kyriacou, 2006).

3.2.4.3.1 Subjective phase shift determination

The simplest measure of phase is the position of a reference marker (commonly the onset or offset of activity; see Figure 3-18) relative to a given time point (generally lights off, ZT 12), which can easily be measured using an actogram or double plot. Often the position of the reference marker is assessed over several cycles of activity, both to increase the accuracy of measurement and to assess whether the phase change is instantaneous or transient in nature. To facilitate the identification of reference markers, data is often grouped and/or smoothed:

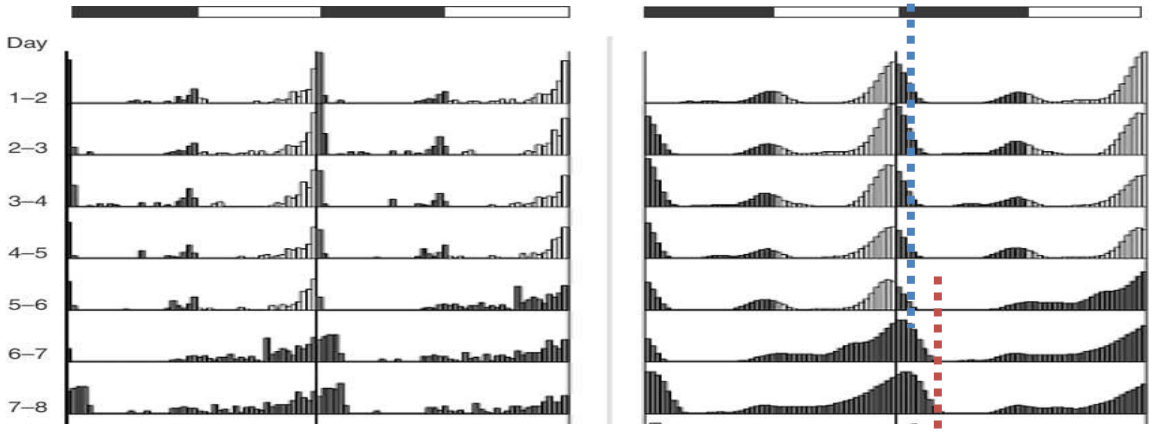


Figure 3-19 Example of phase analysis using the MAZ package's subjective peak detection. Left panel shows 5 days LD and 5 days DD activity, **right panel** shows a Butterworth Filter smoothed version of the raw data. The graph clearly shows the activity offset point is fixed under LD condition (days 1-5, **blue line**). Following a phase shifting stimulus at the end of day 5, the activity offset point has clearly moved to the right (**red line**). In this example the phase shift is clearly instantaneous rather than transient. Figure taken from Rosato and Kyriacou (Rosato and Kyriacou, 2006).

By plotting the phase change of the endogenous clock elicited by the light pulse as a function of the time of the delivery of the pulse to generate a phase response curve (PRC), the clocks of different genotypes can be compared:

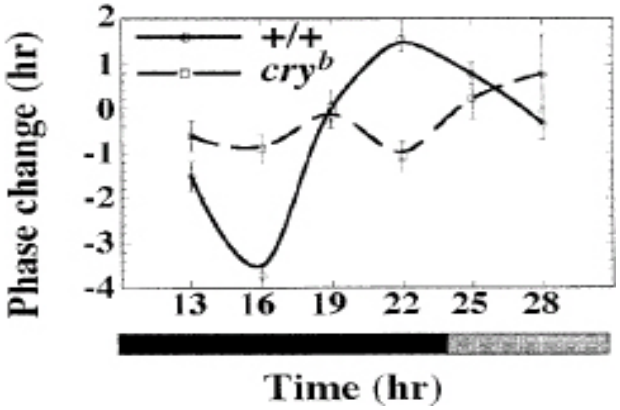


Figure 3-20 Example of phase response curves for wild type and cry mutant flies. *cry*^{b} mutants show reduced phase shifts, showing that cell autonomous photoreception mediated by CRY is essential for proper phase shifting. Figure from Stanewsky *et al.* (Stanewsky et al., 1998).

3.2.4.3.2 Cross-correlation – an objective measure of phase shifts

Performing phase analysis using reference markers is a highly subjective and time consuming process when considering individual flies. The BeFly! package implements an alternative, automated cross-correlation routine for phase analysis, which is more objective as a result of analysing the entirety of a dataset rather than arbitrary landmark features susceptible to stochastic disruption. Phase shift assessments using BeFly! have been validated by comparison with subjective results obtained by several independent observers.

Phase shift experiments can be run following a serial or parallel design. In a serial design, flies are entrained to LD conditions, phase shifted, then re-entrained and allowed to free run, whilst in a parallel experiment different flies of the same genotype are entrained together, but only a subset of the flies are phase shifted:

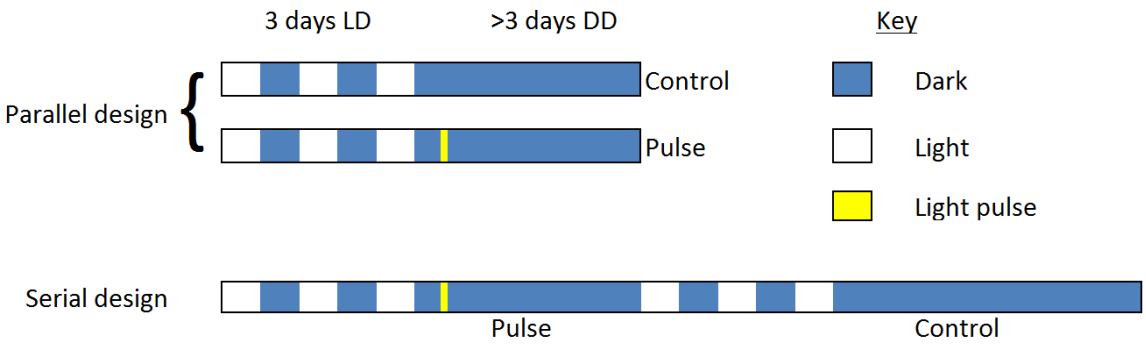


Figure 3-21 Serial and parallel phase shift experiment designs.

The advantage of the serial experimental design is that flies act as their own controls, but such a design cannot account for ageing related activity changes that occur during the experiment. The BeFly! implementation of cross-correlation analysis supports both serial and parallel experimental designs using cross correlation.

In cross-correlation, a probing series is correlated against the experimental dataset, and the result recorded. The probing dataset is then iteratively moved or 'lagged' across the experimental dataset by one bin and the correlation measurement repeated. By plotting correlation against lag to form a correlogram, the position of the highest correlation value on the lag axis represents the phase difference between the data series. Rhythmic datasets will therefore show repeating peaks in the correlogram every time the datasets come into phase:

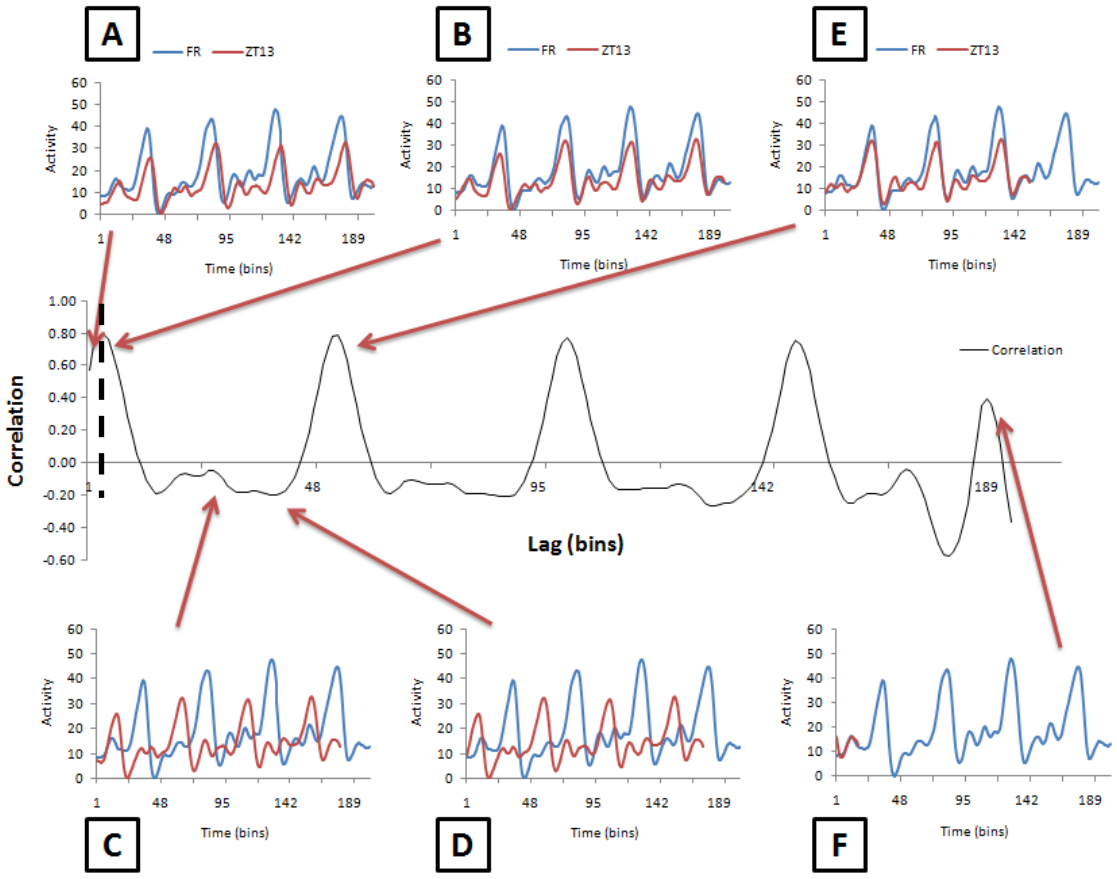


Figure 3-22 Example of cross-correlation phase shift analysis. The central panel shows the correlogram formed when two median activity traces are lagged over each other (cross-correlated). One trace represents the smoothed median free running (i.e. non phase shifted) activity of 24 flies (**blue**), the other represents congenic flies phase shifted by a 15 minute light pulse administered at ZT13 (eliciting a phase advance). Initially there is relatively good correlation between the two traces (**A**), as light pulses at ZT 13 elicit only small phase shifts. The size of the phase shift is equal to the lag that creates the maximum correlation between the two traces (**B, hashed line**). Further lagging of the two datasets reduces the correlation between them, reaching its nadir when the traces are entirely out of phase with respect to each other (**D**). A small increase in correlation occurs when the morning peak of one trace overlies the evening peak of the other (**C**). As the traces are rhythmic, further lagging the traces increases the correlation as the peaks come back into phase (**E**), although the correlation is generally not as good. Further lagging results can lead to artifactual correlations as the extent of overlap between the two traces diminishes (**F**).

3.2.4.3.3 Transient phase shifts

An additional complication inherent to phase shift experiments is that of transient shifts, in which the phase shift is only complete after several cycles of activity. In some cases

the phase shifting stimulus may also induce a period change, which further complicates the calculation of the true phase shift.

BeFly! includes tools for analysing such complex shifts, however an understanding of these methods is not required for this thesis; an interested reader is directed to the digital appendix for further details.

3.3 Conclusion

With some practise, a competent BeFly! user can learn to rapidly analyse activity data using the different visualisations and analyses presented together to allow a holistic interpretation of circadian rhythmicity. Not only is activity analysis more comprehensive than using existing tools, but analysis can be completed much more rapidly than has hitherto been possible; period analysis can be completed in less than 25% of the time it might otherwise take using the MAZ package alone, whilst even greater time savings can be achieved in the calculation of phase shifts, a particularly laborious task which BeFly! completes almost instantaneously.

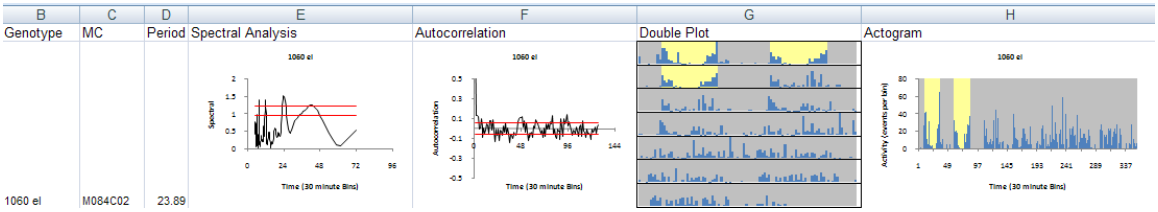


Figure 3-23 An example of BeFly! presenting the results of analyses run using the MAZ package. Figure shows the autocorrelation, CLEAN, double plot and actogram for a single fly's activity. For further details, see appendix.

BeFly! is therefore a significant advance in allowing the high throughput analysis of behavioural mutations, and these tools were instrumental in identifying and analysing the complex rhythmicities found in a new strain of *Drosophila* which was serendipitously discovered to have unusual behavioural rhythms, as described in the next chapter.

4 Identification of the *Party on* strain

4.1 Introduction

In temperate regions, organisms can adapt to seasonal changes in day length by using environmental cues to entrain the clock. Although biological clocks continue to cycle in the absence of Zeitgeber signals ('free run'), in the wild such conditions only occur at extreme latitudes, therefore free running behaviour may be regarded as an evolutionary by-product of selection for a robust central oscillator, rather than as a core feature of the oscillator itself (Roenneberg et al., 2003).

Whilst clocks free run in the absence of a Zeitgeber, the general consensus is that constant Zeitgeber stimulation of the clock (i.e. constant light, LL) results in severe disruption of rhythmicity as a result of the clock being constantly 'reset' (Stanewsky et al., 1998). The intensity of the constant light stimulus determines its effect; whilst low intensity illumination (such as that created by moonlight in natural conditions) suppresses circadian rhythmicity (Winfree, 1974; Dissel et al., 2004) high intensity light induces behavioural arrhythmicity (Stanewsky et al., 1998). Furthermore, light may induce changes in activity independent of the clock mechanism (Kempinger et al., 2009).

In *Drosophila* the 'resetting' of the clock by light is mediated by the photolyase CRY, which when activated by light mediates the degradation of clock protein TIM, preventing its accumulation (Stanewsky et al., 1998), as detailed in the introduction. A recent elaboration of this mechanism has shown that both TIM and CRY are targeted for proteasomal degradation by the F-box protein JETLAG (JET), as shown in Figure 4-1 (Koh et al., 2006; Peschel et al., 2006). Mutations in either *cry* or *jet* compromise flies' ability to reset their clocks, and as a consequence show rhythmic behaviour in LL as well as DD conditions.

Cell culture studies have shown that TIM only becomes an appropriate substrate for JET after association with CRY, therefore fly strains carrying the S-TIM isoform, which interacts more strongly with CRY than the L-TIM isoform, show increased responsiveness to light (Peschel et al., 2009). This variation in light sensitivity appears to be a highly adaptive phenotype (Tauber et al., 2007; Sandrelli et al., 2007; Kyriacou et al., 2008).

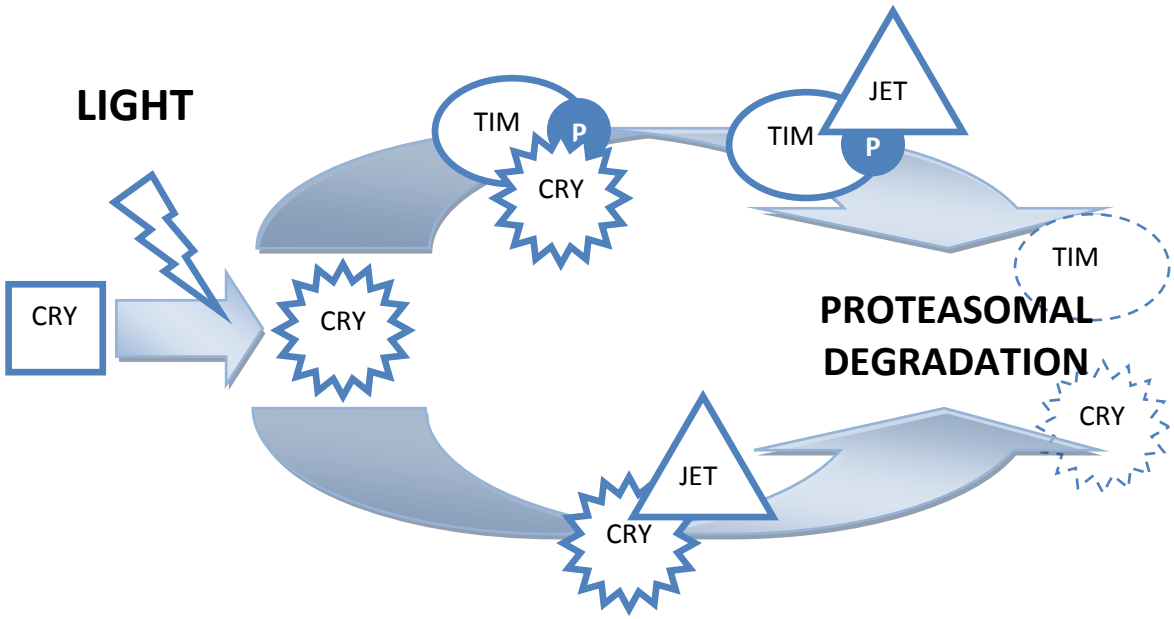


Figure 4-1 The cell autonomous molecular pathway by which light entrains the molecular clock through light-dependent interactions between JET, CRY, and L-TIM/S-TIM. See text for details.

It is likely that the exact sequence of steps involved in proteasomal degradation has yet to be fully elucidated; it has recently been shown to require the COP9 signalosome (Knowles et al., 2009), whilst an RNAi screen identifying genes affecting CRY degradation implicated a number of additional genes including the E2/E3 ubiquitin ligase *Bruce* (which presumably acts before the E3 ligase JET), as well as a dual specific phosphatase *ssh* and the HECT domain-containing E3 ligase *CG17735* (Sathyanarayanan et al., 2008). Over-expression of *morgue*, an F box/ubiquitin conjugase domain protein in clock cells has also been shown to induce LL rhythmicity (Murad et al., 2007), suggesting it too plays a role in light mediated entrainment of the clock.

Whilst the response of the intracellular clock to constant light is well established, comparatively little work has been done to further dissect the behavioural arrhythmicity seen in LL. Of the few studies in this area, perhaps the most interesting is the observation by Yoshii *et al.* (Yoshii et al., 2005), that the addition of low amplitude temperature cycles to LL conditions is sufficient for wild type flies to regain both behavioural and molecular rhythms. The authors hypothesise that temperature induced transcription of *per* and *tim* might somehow circumvent the constant resetting of the central oscillator mechanism mediated by CRY (Yoshii et al., 2002), and have recently elaborated their theory by showing that raising or lowering the ambient temperature in LL conditions induces different transcriptional changes (Yoshii et al., 2007). Another possible explanation is that the temperature Zeitgeber acts in a tissue autonomous fashion via *norpA* and *nocte*, and therefore entrains motor centres downstream of the central clock neurons (Glaser and Stanewsky, 2005; Sehadova et al., 2009). The exact relationship between the light and temperature Zeitgebers, as well a possible role for social entrainment by chemosensory cues (Levine et al., 2002a), remains a subject of considerable research interest.

The canonical phototransduction pathway in *Drosophila* consists of a number of light sensitive organs including the compound eyes, the Hofbauer-Buchner (HB) eyelets, and ocelli (Figure 4-2) which synapse with neurons of the circadian clock in the brain (Figure 1-6). Although the canonical pathway does not appear to play a major role in re-entraining the clock (as assessed by phase shifting experiments, Suri et al., 1998), mutants lacking some or all these structures have noticeably different activity profiles compared with wild type flies (Helfrich, 1986; Wheeler et al., 1993). Elimination of all known autonomous and non-autonomous photoreceptors in $glass^{60j}$; cry^{b} double mutants has been shown to make flies 'blind' to light (Helfrich-Forster et al., 2001), though more recent experiments suggest that there may be an additional light sensing pathway, possibly mediated by a novel rhodopsin in R8 cells

(Papatsenko et al., 1997), that is not ablated in such mutants (Breda, C. pers comm., Rouyer, F. unpublished observations).

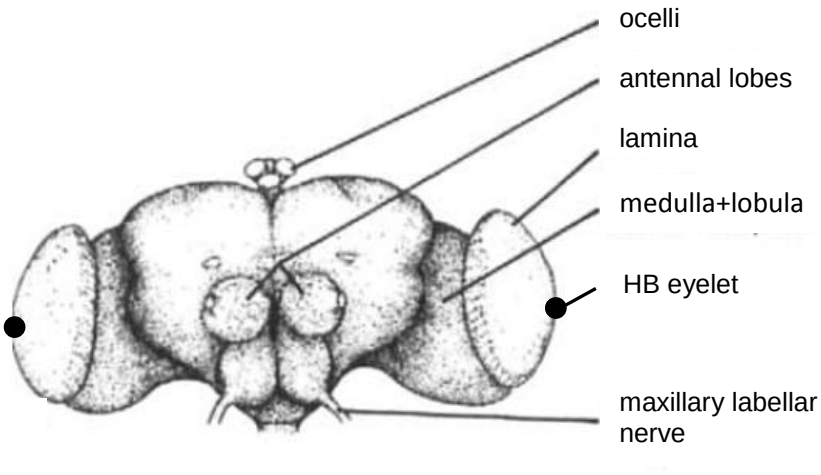


Figure 4-2 Dissection of *Drosophila* brain showing light sensitive ocelli, the lamina immediately beneath the rhabdomeres of the compound eye, and the optic lobes (labelled medulla+lobula complex) in which projections from the compound eye and HB eyelet may synapse with the l-LN$_v$ and s-LN$_v$ cells of the circadian clock. Figure adapted from Helfrich (Helfrich, 1986).

These complexities aside, it is clear that each of these structures plays a subtly different role in the entrainment and masking pathways of *Drosophila* activity (Rieger et al., 2003; Mealey-Ferrara et al., 2003), presumably as a result of their dissimilar connectivity to clock neurons. The exact nature of these interactions remains to be elucidated, leaving open the intriguing possibility that different photoreceptive organs show some degree of specificity in terms of light intensity or wavelength (Hanai et al., 2008).

Flies in which the cell autonomous light sensing capacity has been ablated using the extreme hypomorphic mutation *cry*^{b} show rhythmic behaviour in constant light, suggesting that the cell autonomous Zeitgeber information processed through CRY is more important to the circadian clock than that imparted by the other light receptors. However, when the intensity of constant light is increased, the behavioural rhythms of *cry*^{b} can be induced to split into distinct components (as shown in Figure 4-3). Interestingly this rhythm splitting effect is not observed when the *cry*^{b} mutation is combined with the mutations in the photoreceptor

genes *norpA* or *sine oculis* (*so*) (Yoshii et al., 2004), suggesting that splitting requires the canonical pathway. Confusingly genetic ablation of the optic lobes using combinations of *so* mutants can itself induce complex, split rhythmicity (albeit in DD conditions), therefore the exact nature of this splitting effect remains unclear.

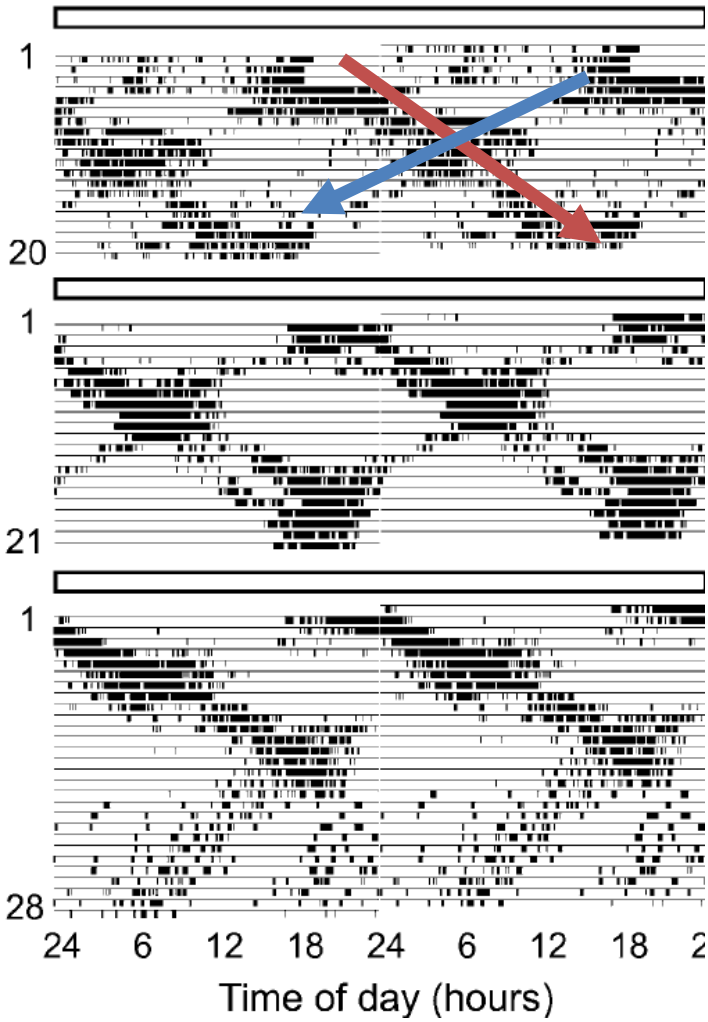


Figure 4-3 Example of split oscillators in *cry^b* flies under high intensity LL conditions. Top and middle panels: 2 examples of flies showing activity controlled by split oscillators, one oscillator showing a period <24 hours (**blue arrow**) and the other >24 hours (**red arrow**), forming a herringbone pattern. **Bottom panel**: example of fly showing single long period rhythm component spontaneously switching to a short period rhythm. Figure and caption adapted from Yoshii *et al.* (Yoshii et al., 2004).

As can be seen from Figure 4-3, when split oscillators interact, activity levels increase to form dense bands of activity, suggesting that the decoupled oscillators mutually suppress activity output when out of phase, as they do in the cricket *Gryllus bimaculatus* (Tomioka et al., 1991). Immunohistological examination of flies showing split oscillators showed that different

clock cell populations had different molecular periods matching those of the split behavioural rhythm components; this was one of the earliest indications that *Drosophila* locomotor rhythms were driven by two separate *per*-dependent oscillators, which respond differentially to constant light. This theory has recently come to prominence with the re-emergence of the distributed two oscillator model (Grima et al., 2004; Stoleru et al., 2004 - see chapter 1 and discussion for details).

In the context of this work, it was serendipitously noticed that, contrary to expectation, a number of supposedly wild type Canton-S flies remained rhythmic throughout a nine day LL experiment in a manner similar to cry^b mutant flies. Given the incomplete understanding of LL rhythms, it was decided to further investigate and attempt to map the source of this atypical behavioural phenotype, which was provisionally named the '*Party on'* (*Po*) strain to reflect its sustained rhythmicity.

4.2 Methods

4.2.1 Linkage analysis

4.2.1.1 Excluding the X chromosome

The X chromosome's contribution to a phenotype can be easily assessed using a reciprocal crossing scheme and scoring only the phenotype of male F1 flies which carry a single X chromosome derived from the maternal line:

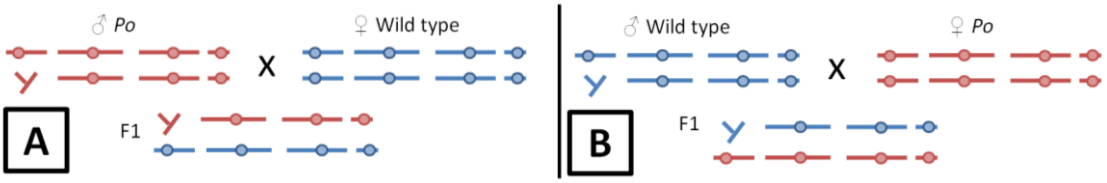


Figure 4-4 Reciprical crossing scheme to generate male F1 progeny. (A) flies carrying an X chromosome derived from the wild type strain are compared with flies carrying an X from the *Po* strain (**B**).

4.2.1.2 Autosomal linkage

Autosomes can be manipulated by maintaining them over balancer chromosomes to prevent recombination. To determine whether the *Po* phenotype mapped to an autosome, strongly rhythmic males were selected from the *Po* mutant stock and crossed to the balancer stocks w^{1118}; noc^{Sco}/CyO; MKRS/TM6B, Tb^{1} or w^{1118}; noc^{Sco}/CyO; MKRS/TM6B, Dp^{1} (neither of which exhibited significant rhythmicity under LL conditions). Using the dominant markers on the balancer chromosomes the inheritance of the *Po* autosomes could be tracked and excluded in a pairwise fashion to determine whether they played a major role in determining LL rhythmicity, as shown in Figure 4-5.

Due to a chronic problem of bacterial infection of our fly food, as well as the difficulties of single pair matings when using double balancers, these crosses were carried out *en masse* to generate F1 and F2 flies.

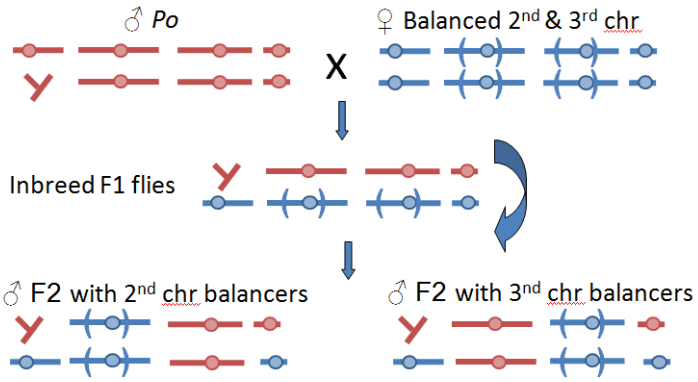


Figure 4-5 Linkage analysis crossing scheme used to generate F2 lines in which one autosome from the *Po* line had been been excluded.

4.2.2 Meiotic mapping

Meiotic mapping was performed at 29°C using a multiply marked 2nd chromosome carrying the recessive alleles *al*[1], *dp*[ov1], *b*[1], *pr*[1], *cn*[1], *c*[1], *px*[1], *sp*[1], (Bloomington stock 4187) following guidelines set out by Greenspan (Greenspan, 1997) using the following crossing scheme:

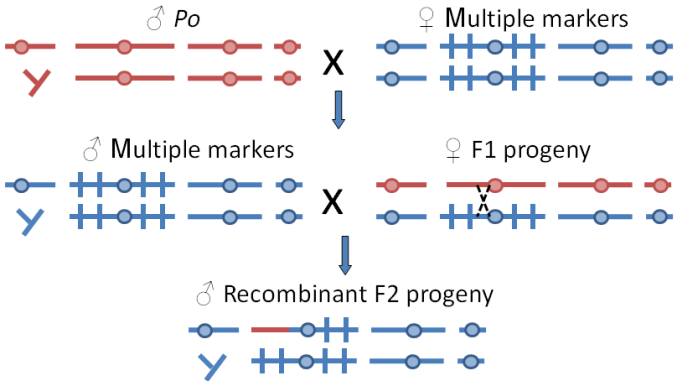


Figure 4-6 Crossing scheme used to generated recombinant 2nd chromosomes carrying multiple recessive markers used to meiotically map the *Po* mutation.

Male F2 progeny of were examined, and recombinants' LL behaviour was tested.

4.3 Results

4.3.1 Discovery of a rhythmic strain of *Drosophila* maintained in LL

During an LL experiment a Canton-S control line was observed to show extended rhythmicity under LL conditions:

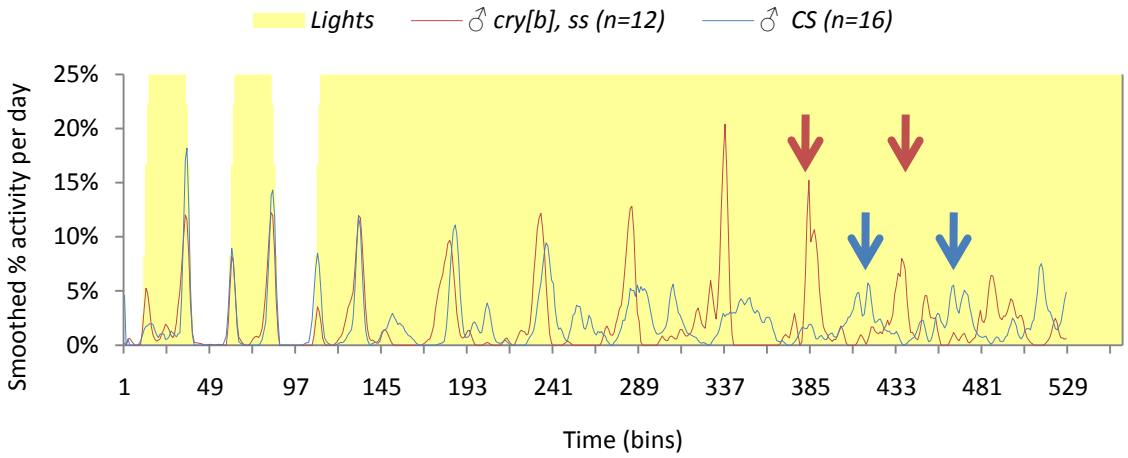


Figure 4-7 Actogram showing LL activity of *cry*^{b}, *ss* and a rhythmic strain of Canton-S. Note that although the Canton-S flies retain a rhythmic signal, it is of a longer period than the *cry*^{b}, *ss* flies and of a lower quality. This is particularly evident after a week in LL conditions (**red** and **blue** arrows denote respective activity peaks). Data smoothed and normalised to daily peak values for clarity.

To preclude experimental error, the experiment was repeated using a number of different Canton-S isolates held in the lab:

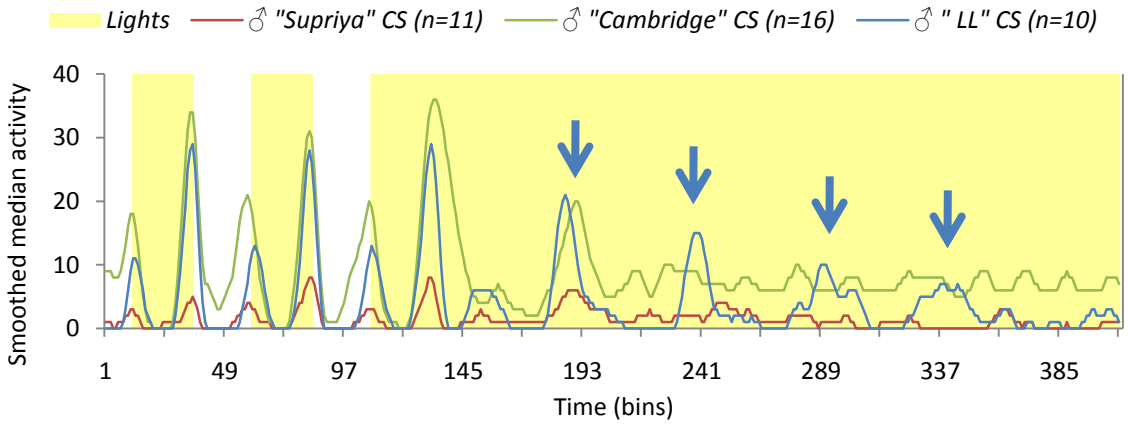


Figure 4-8 Activity of different Canton-S isolates under LL conditions. Only the "LL" isolate, later named *Po*, shows sustained - though dampening – LL rhythmicity (**blue arrows**). "Supriya" and "Cambridge" stocks a gift of S. Bhutani.

The new isolate of the Canton-S strain was provisionally named the '*Party On*' (*Po*) strain to reflect its sustained rhythmicity in constant light. In male flies the period of this rhythmicity was estimated to be 27.60 ±0.30 hours (n=50), with individuals having a large spread of periods ranging from 23.6 to 33.1 hours.

Although when considered as a group average the *Po* strain maintains rhythmic behaviour for about a week in LL before becoming arrhythmic (Figure 4-7), individual *Po* flies show robust rhythmicity throughout a 23 day LL experiment in a manner similar to cry^b, *ss* mutants (shown in Figure 4-9). This suggested that *Po* flies were not as phase synchronised as cry^b flies; indeed examination of individual fly records revealed a number of distinct behavioural patterns (Figure 4-9) which, when averaged, cause the apparent dampening of rhythmicity under sustained LL conditions. This implied that the *Po* strain was either genetically heterogeneous, or that the *Po* mutation showed a degree of variable penetrance causing some stochastic variation in period and phase. The relative proportion of flies showing strong, complex or arrhythmic behaviours is shown in Figure 4-13 and Figure 4-14.

At the time of discovery the only other known LL mutants were cry^b and *jetlag* (also known as *Veela*), both kept in the Leicester fly laboratories but phenotypically distinct from *Po* (Figure 4-9), therefore contamination by known stocks could be ruled out. Although sustained LL rhythmicity precluded gross circadian clock disruptions, the behaviour of the *Po* strain was examined under DD conditions to determine whether the long period phenotype under LL conditions was also seen in DD:

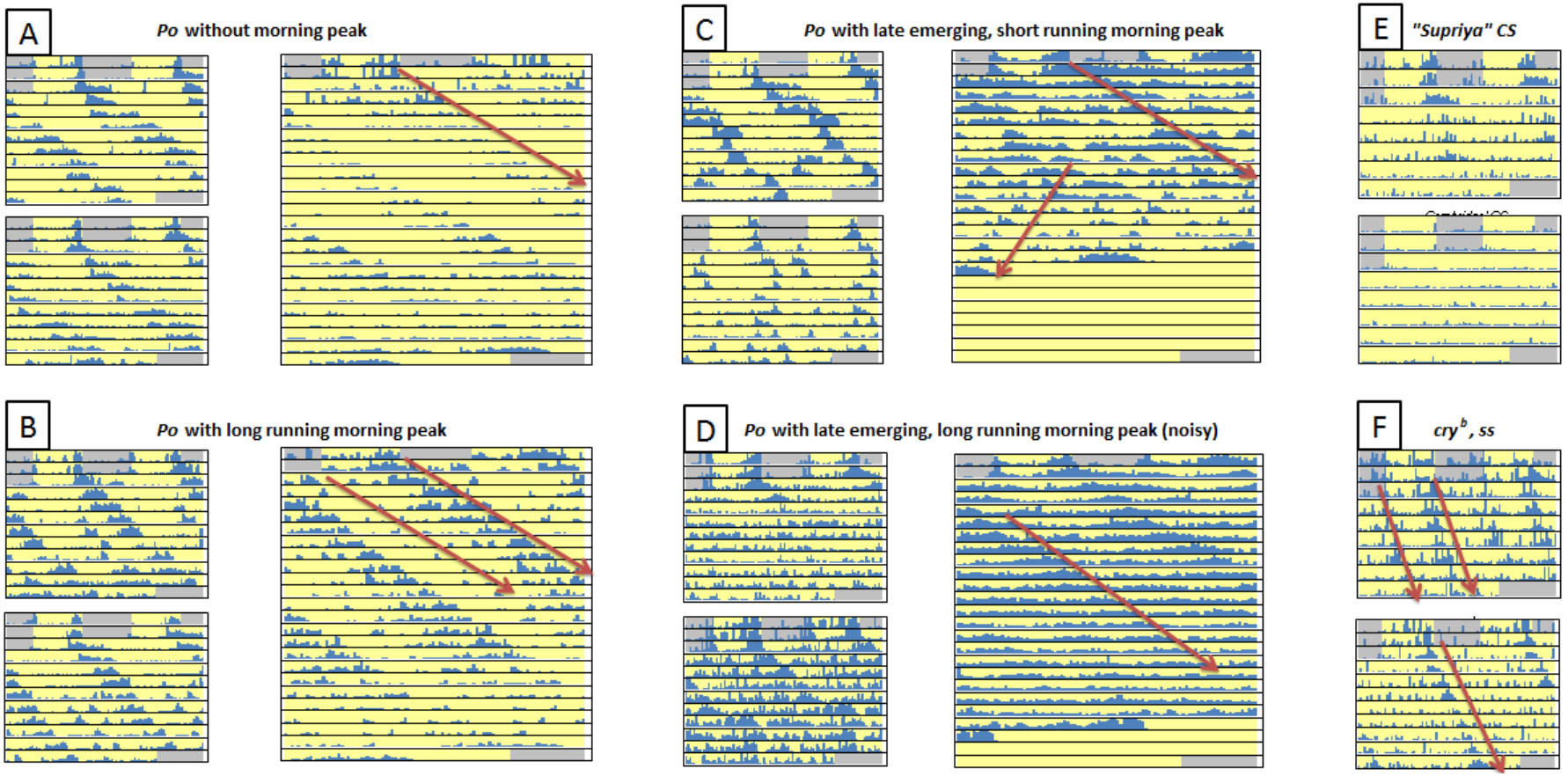


Figure 4-9 Examples of distinct *Po* behavioural patterns in LL. Small panels show data recorded over 9 days of LL using male flies, long panels show results of 23 day LL experiment using a mixture of male and virgin female flies (activity patterns are not correlated to sex). **Red arrows** show pattern of rhythmicity in long panels. *Po* flies show distinct behavioural patterns in LL, including **A**: long period LL activity derived solely from the evening peak. **B**:long period activity contains elements of both morning and evening activity peaks. **C**: evening peak becomes long running, morning peak re-emerges after several days as a short running split component. **D**: significant long period activity, but noisy background activity evident throughout. **E**: "Supriya"Canton-S flies showing no significant rhythmicity in LL. **F**: cry^{b}, *ss* flies showing long running LL activity. All *Po* flies in this figure were subject to CLEAN period analysis and shown to have statistically significant long periods.

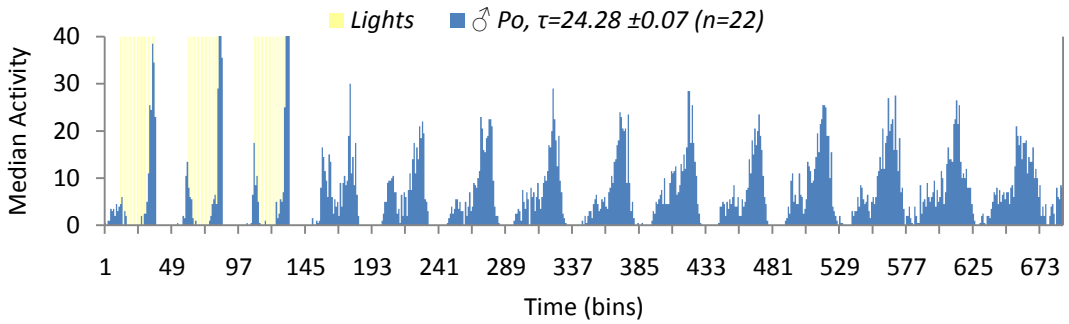


Figure 4-10 DD actogram for *Po* flies, showing first bimodal then unimodal behaviour.

Although a τ of 24.28 is relatively long for a Canton-S strain in DD conditions, free running period is sensitive to genetic background (the Oregon R strain having a τ of 24.51 ±0.10, n=22), therefore it is difficult to determine whether the free running DD period of *Po* is truly 'long' or not; it is clearly significantly shorter than the period of LL rhythmicity. Visual inspection of double plots for individual flies suggested that 18% of *Po* flies exhibited 'weak' circadian rhythms in which a circadian pattern could not easily be observed by eye; however CLEAN analysis showed that these 'weak' rhythms were nevertheless significant at the 99% level, and therefore cannot be classed as arrhythmic flies (data not shown).

A further experiment was performed to determine whether eclosion rhythms under DD conditions were comparable to wild type:

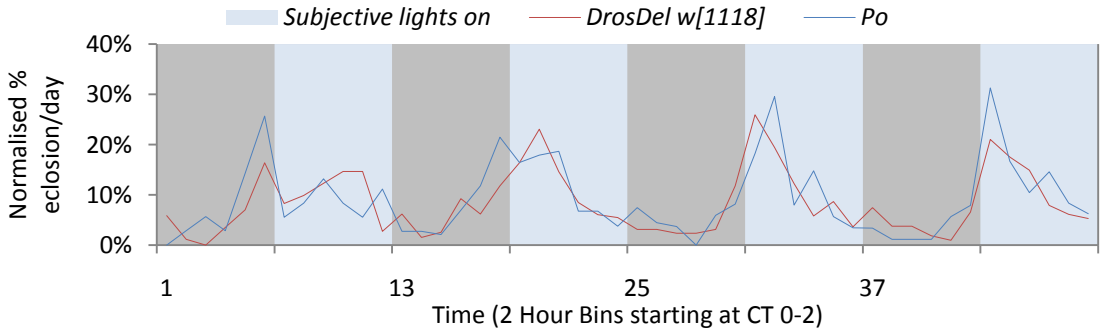


Figure 4-11 Eclosion rhythms for *Po* and white eyed controls from the DrosDel collection.

Eclosion data is generally recorded at low resolution, and is therefore not amenable to period determination using spectral techniques, however the eclosion rhythms observed in *Po* flies and controls appeared superficially similar, peaking around subjective dawn. Given the nature of the *Po* strain, an effort was made to map the *Po* gene using linkage analysis.

4.3.2 Linkage analysis – excluding the X chromosome

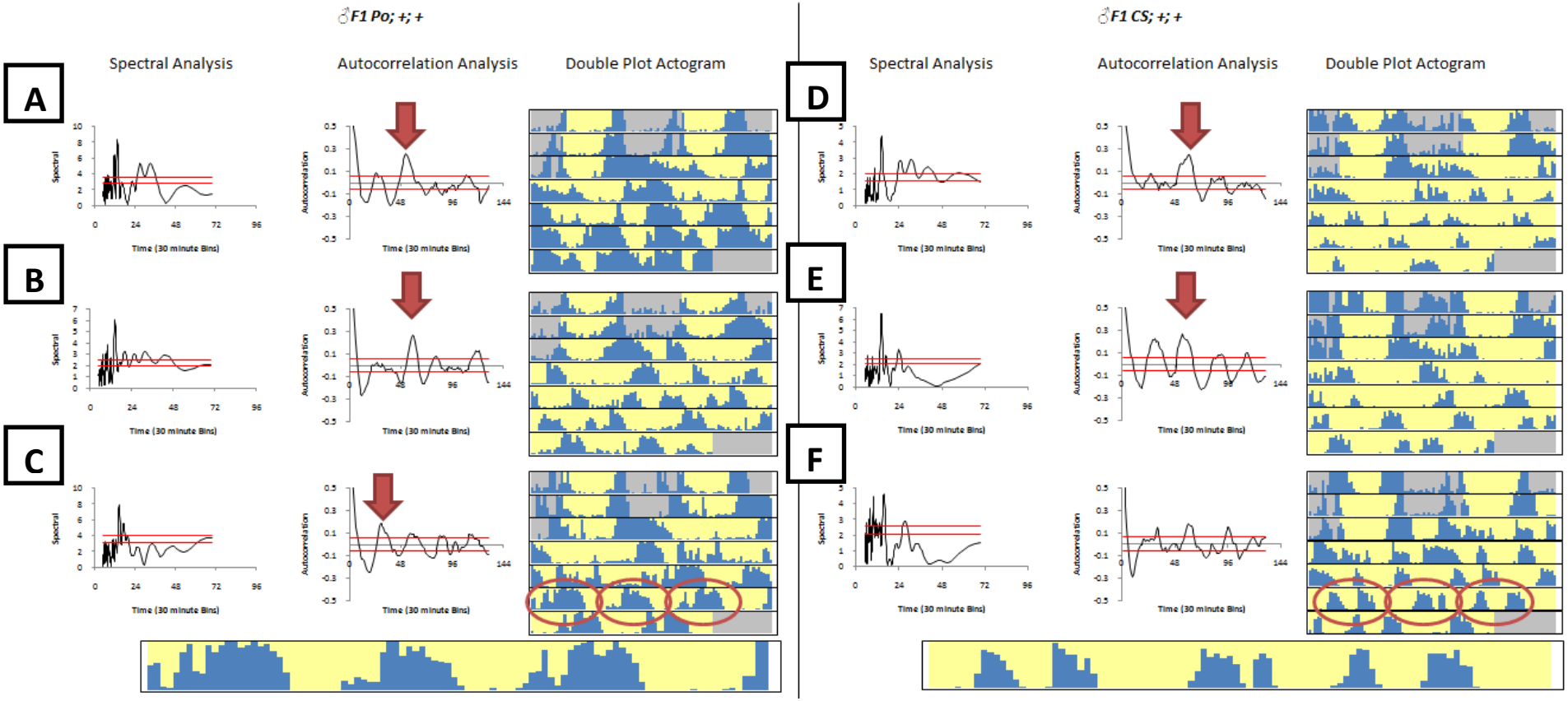


Figure 4-12 Examples of individual reciprocally outcrossed *Po* flies in LL conditions. In this figure individual double plot actograms are presented with their accompanying spectral and autocorrelation analysis panels. *CS* = X chromosome from Canton-S line, *Po* = *Po* line. For full details see text.

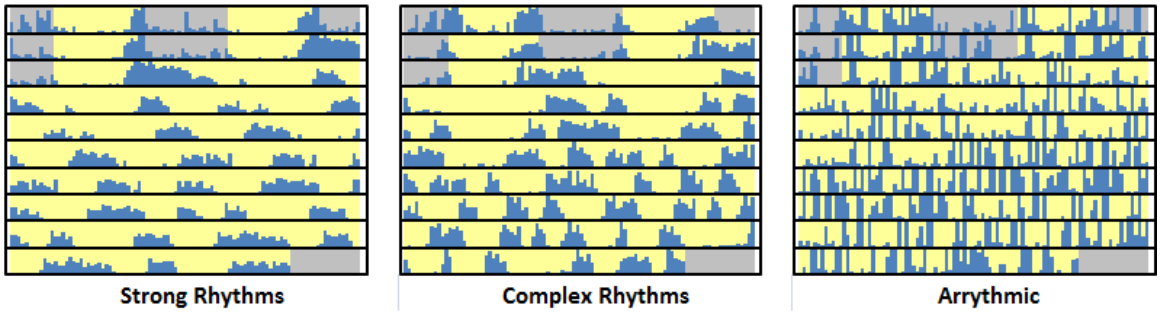
It is clear from Figure 4-12 that *Po* shows a dominant mode of inheritance, but this dominance is not fully penetrant; comparison of the actograms of heterozygous F1 flies in Figure 4-12 with those of the (presumably homozygous) *Po* parents shown in Figure 4-9 shows that the behaviour of outcrossed heterozygotes was of a generally more complex character, containing many more discrete bouts of activity than the parents.

Comparing the left and right panels of Figure 4-12 it is evident that both F1 lines from the reciprocal cross show a roughly equal mixture of long periods of various quality (**A**, **B**, **D**, **E**) and more complex ultradian rhythms (**C**, **F**), showing that the *Po* mutation does not segregate with the X chromosome. Genetic background clearly plays an important role in mediating rhythmicity; flies carrying an X chromosome from the Canton-S (CS) parents show more discrete bouts of activity (**D**,**E**,**F**), whilst those carrying a *Po* derived X generally show broader, less defined bouts (**A**,**B**,**C**).

Though the complex rhythms (panels **C** and **F**) may appear superficially to resemble long period rhythms, closer consideration of a 48 hour period of activity (zoomed panels) shows that these flies exhibit three clear bouts of activity (delineated by **red circles**). Due to the strong bimodal pattern of activity in all the flies illustrated, spectral analysis generally detects the ultradian components of the LL rhythms (i.e. spectral peaks with a period of <24 hours arising due to the presence of both morning and evening peaks). Due to the different shapes of morning and evening activity bout, autocorrelation analysis can be used to determine the most appropriate periodicity of such lines. In panels **A**, **B**, **D**, and **E** the highest autocorrelation peak (**red arrow**) suggests that the LL rhythmicity observed in the double plot has a long period (i.e. a peak >48 hours on the x axis). In contrast, the autocorrelation data for the fly in panel **C** confirms that the dominant rhythmicity in this fly is ultradian, whilst in panel **F** the circadian and ultradian peaks are of similar magnitude.

4.3.3 Linkage analysis – pairwise exclusion of autosomes

As the X chromosome did not harbour the *Po* mutation(s), it was determined to test whether the *Po* mutation lay on either the large 2nd or 3rd autosomes using standard balancer crosses as described in the methods. As previous experiments had suggested that *Po* inheritance was only partially dominant, resulting in a mixture of long period and complex behaviours under LL conditions, the behaviour of F2 flies was classified as falling into one of three qualitative categories, and the proportion of each was recorded:



Genotype	n	Strong Rhythms	Complex Rhythms	Arrhythmic
w(*); Cyo/noc[sco]; Po/Po	29	17%	31%	52%
w(*); Po/Po; MKRS/TM6B, Tb[1]	22	77%	23%	0%
Supriya CS	11	0%	9%	91%

Figure 4-13 Autosomal linkage analysis using discrete categories of LL behaviours patterns. Top panel: examples of individual fly double plots showing strongly rhythmic, complex and arrhythmic behaviours. **Bottom panel**: Autosomal linkage analysis suggested that flies carrying the 2nd chromosome from the Po strain (i.e. *w(*); Po/Po; MKRS/TM6B, Tb*^{1} flies) were much more rhythmic in LL conditions that those carrying the 3rd chromosome or Canton-S (**CS**) controls from the 'Supriya' collection which had previously been shown not to harbour the *Po* mutation (**Figure 4-8**). *w** represents an equal mix of flies carrying w^{+} and w^{1118} alleles on the X chromosome.

The results suggested that the *Po* mutation maps to the 2nd chromosome, but due to the small number of flies recovered it was not possible to exclude a contribution from chromosome 3 (or the small 4th autosome). It was determined to inbreed the *Po* line in an effort to increase the ratio of flies showing strong LL rhythms rather than more complex patterns which might obfuscate the mapping.

Every generation *Po* flies' LL rhythmicity was therefore assayed, and only the most rhythmic individuals allowed to breed. Although an attempt was made to perform single pair matings of both strongly rhythmic and weakly rhythmic flies (in order to determine Mendelian segregation ratios in the progeny), this proved impossible; partly as a result of the reduced vigour flies displayed following the long LL activity experiments required to assay their rhythmicity prior to breeding, and partly due to chronic bacterial infection of food supplies compromising the viability of single pair crosses.

Instead crosses had to be performed such that a single male was crossed to several virgin females of similar rhythmicity. The results of inbreeding are shown below:

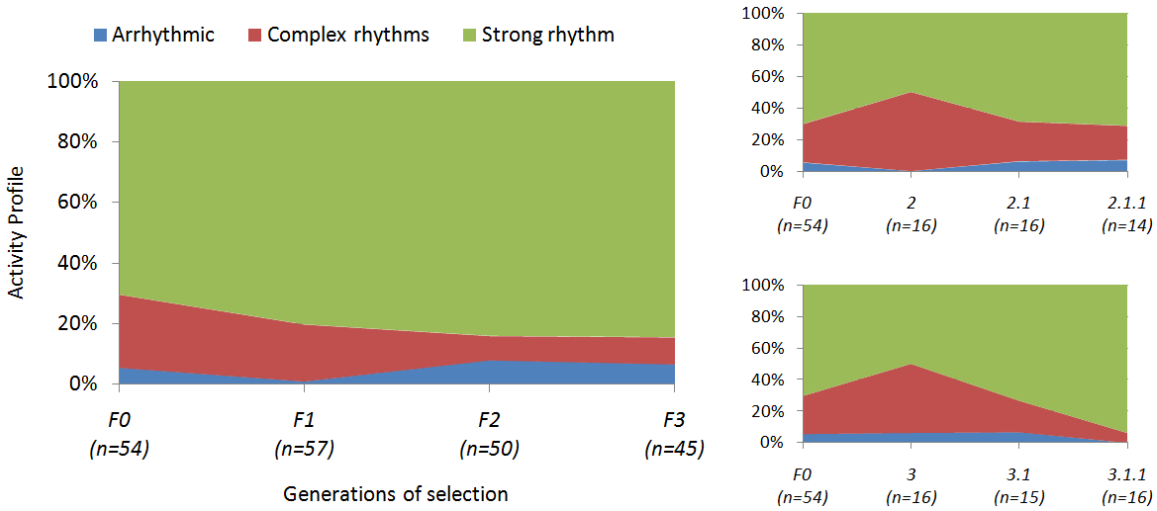


Figure 4-14 Results of inbreeding the *Po* strain. Left panel shows the average response of inbreeding the *Po* line; **right panel** shows the results of inbreeding in two different lineages.

Figure 4-14 shows the inbreeding the *Po* line produces only a small selective response, as the percentage of complex rhythms appears to decrease then stabilise. However, this may be an artefact; examination of separate inbreeding lineages (right panels) suggests that the rhythmicity profile for a given line is subject to some fluctuation each generation, therefore the decrease in complex rhythmicities may not have a genetic basis.

The limited response to selection was evident not only in terms of the absolute proportion of flies showing evidence of some LL rhythmicity (Figure 4-14), but also in terms of the quality of the rhythmic behaviour displayed by individual flies:

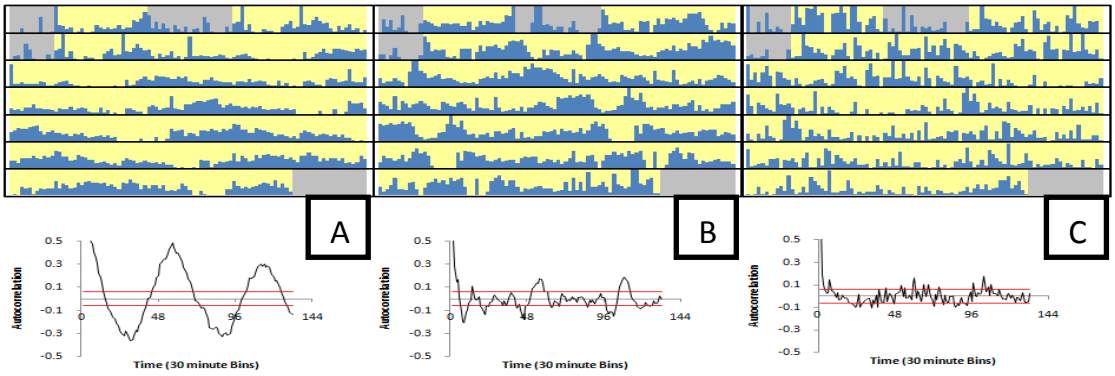


Figure 4-15 Examples of individual fly double plots (top) and autocorrelation plots (bottom) illustrating the variation in the quality of rhythmic behaviour within a *Po* line selectively inbred for 4 generations. Inbreeding did not significantly improve the quality of rhythmicity, as individual inbred flies showed considerable phenotypic variation in terms of rhythm quality (assessed using autocorrelograms), from strong rhythms (**A**) to weakly rhythmic flies (**B** and **C**).

The data in Figure 4-14 imply that the mutation underlying the *Po* phenotype is either fixed or present at a high gene frequency within the *Po* population, itself a surprising result given its serendipitous discovery in a vial of supposedly Canton-S flies. However, the continued segregation of rhythms of various qualities after four generations of selection suggests a more complex mode of inheritance. As previous outcrossing experiments suggested that the *Po* mutation was partially dominant, the most parsimonious explanation is that the *Po* phenotype is subject to modifiers, a not unexpected result given that behaviour is the result of many interacting genes and systems (Hall, 1994; Mackay, 2008).

As assaying LL activity every generation was not only time consuming and highly subjective, but also that the results suggested that the *Po* mutation might already be fixed within the population, it was decided to abandon further inbreeding experiments as it was deemed unlikely that a robust further improvement in LL rhythmicity could be engendered.

Further experiments were performed using the inbred lineage *Po*$^{2.1.1}$; although this lineage did not show the highest proportion of LL rhythmicity of the inbred lines, it appeared to be fitter than other lines, generating more progeny in crosses, and therefore most appropriate for further work.

The previous linkage analysis was therefore repeated using the inbred *Po*$^{2.1.1}$ line. As an additional step to minimise the influence of genetic modifiers, the rhythmicity of balanced F1 flies was assessed, and only the most rhythmic individuals crossed to generate the F2 generation. Although F2 phenotypes in such situations will not reflect the underlying Mendelian inheritance of the trait and its modifiers, due to the complexity of the phenotype, the length of LL experiments and the limited number of activity monitors this was a necessary compromise.

As a first step, F2 data were grouped and examined for evidence that the presence of eye pigment (w^+) or sex might interact with the LL rhythmicity phenotype:

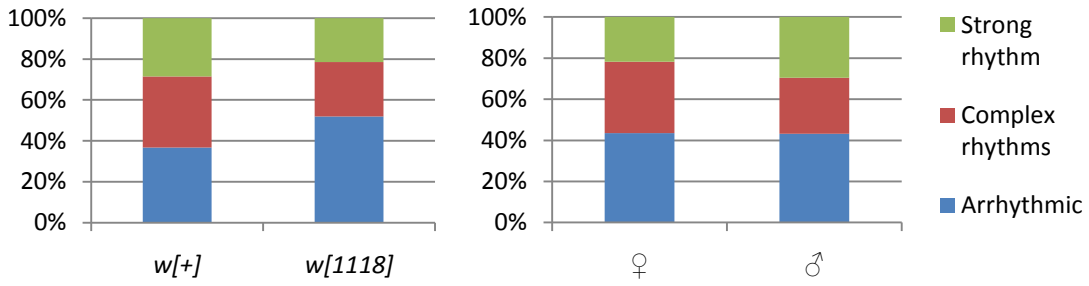


Figure 4-16 *Po* phenotype does not interact with sex nor the presence of eye pigment. The linkage data provides no evidence for an interaction with sex or eye colour (flies carrying the w^{1118} allele are blind). n~100 per group.

As neither factor correlated with the rhythmic phenotype, data were grouped for linkage analysis irrespective of these two variables:

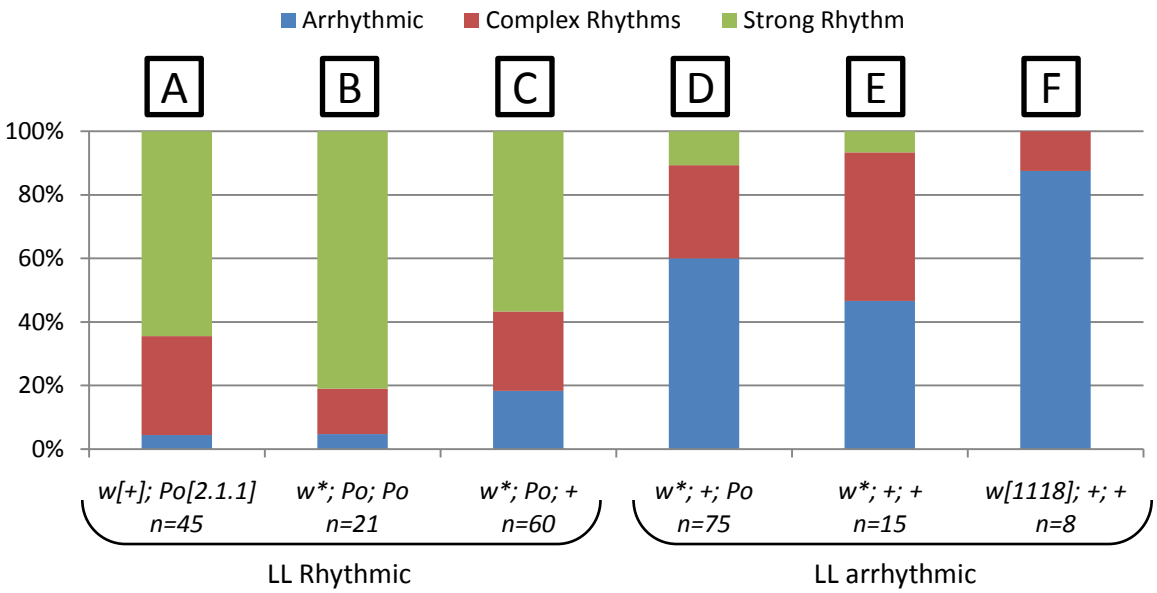


Figure 4-17 Linkage analysis for the *Po* mutation implicates the 2nd chromosome. $Po^{2.1.1}$ flies (**A**) were crossed to a line carrying 2nd and 3rd chromosome balancers which showed no LL rhythmicity (**F** - in this figure balancer chromosomes are indicated by the symbol '+' to indicate that they show no LL rhythmicity and therefore represent the wild type state. Balancer genotypes are detailed more fully in the methods section). F2 flies carrying chromosome 2 from the $Po^{2.1.1}$ line (**B** and **C**) had a behavioural profile that closely matched the $Po^{2.1.1}$ line (**A**), whilst those F2 flies not carrying this chromosome (**D** and **E**) resembled the balancer line (**F**).

Although not entirely unequivocal, the repeated linkage analysis in Figure 4-17 was in accordance with the preliminary analysis shown in Figure 4-13, strongly suggesting that the *Po* mutation lay on the 2nd chromosome, as F2 flies not carrying this chromosome had a rhythmicity profile which more closely resembled the balancer controls (albeit with an increase in the number of complex rhythmicities found). This data might indicate that the *Po* background is particularly permissive for LL rhythmicity.

A further mapping experiment was attempted to rule out the 4th chromosome, but the results were unclear due to the small number of flies that could be run (data not shown).

4.3.4 Meiotic mapping

Despite the considerable degree of variation in rhythmicity even within inbred *Po* lines, an effort was made to map the position of the *Po* mutation to a region of the 2nd chromosome using the $Po^{2.1.1}$ strain and a multiply recessively marked 2nd chromosome strain (see methods section 4.2.2). Several different mapping strains' LL rhythmicity was tested, many of which showed complex ultradian rhythms to some degree in LL. This posed a complication for mapping, as complex ultradian rhythmicities - though distinct from the clear long period LL rhythms of the homozygous $Po^{2.1.1}$ strain - in some cases resembled the more complex rhythms seen in *Po* heterozygotes, as shown in Figure 4-12. Nevertheless, by breeding only the most rhythmic F1 flies in an effort to increase the behavioural signal to noise ratio in the F2 generation, it was possible to obtain the results shown in Figure 4-18.

The meiotic mapping results suggested that the *Po* mutation lies close to the gene *curved* on chromosome arm 2R. All recombinant flies assayed which survived the LL experiment were killed and their DNA stored with the intent of facilitating later DNA marker based mapping. Although some effort had been expended in sequencing candidate clock genes using DNA samples from the *Po* strain, this effort was abandoned when the meiotic mapping implicated a region devoid of known clock genes (both when this work was performed and after the more recent discovery of additional genes such as *bruce*).

Given the difficulties inherent in mapping the *Po* mutation due to its susceptibility to genetic modifiers, it was decided instead to characterise the *Po* phenotype further in a circadian context.

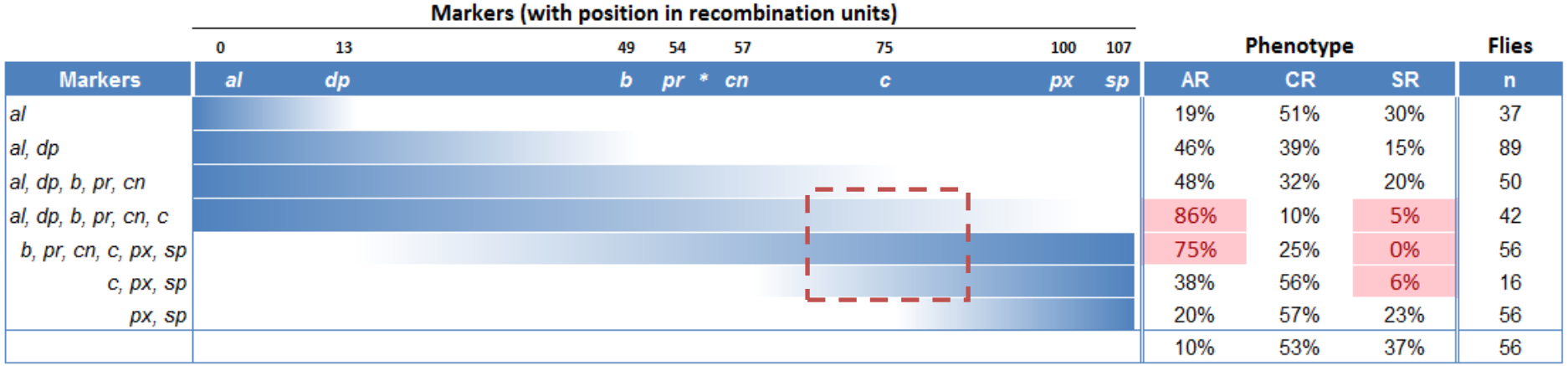


Figure 4-18 Meiotic mapping the *Po* mutation onto the 2nd chromosome. Marker positions are indicated to a relative scale. * indicates the position of the centromere, a region of low recombination. Only recombinant genotypes for which >10 males were recovered are shown. AR = arrhythmic, CR = complex rhythms (including ultradian and weak circadian), SR = strong, long period rhythm. **Left panel** shows the composition of the recombinant chromosomes recovered, such that blue regions are derived from the mapping strain, and white regions derived from the $Po^{2.1.1}$ line. The graduated transition between blue and white reflects the uncertainty in positioning of the recombination breakpoint between two markers. The first line therefore represents a recombinant chromosome carrying the marker al^{1} but not dp^{ov1}, and the second represents a chromosome with a breakpoint between the markers dp^{ov1} and b^{1}. The bottom row represents virgins heterozygous for $Po^{2.1.1}$ and all recessive markers, showing the penetrance of the LL rhythmicity phenotype when a single copy of *Po* is present in this genetic background. **Right panel** the behaviour of the recombinant flies in LL was assessed, and notable results highlighted in red. Lines carrying 2 copies of the c^{1} allele are noticeably less rhythmic than other lines, suggesting that the *Po* mutation lies near this marker.

4.3.5 Low temperature activity

It has been established for some time that although the circadian clock's period is temperature compensated, the daily behaviour patterns of flies are temperature sensitive, and reflect underlying temperature dependent changes in molecular cycles (Majercak et al., 1997). At temperatures lower than the 25°C commonly used for activity experiments, flies' activity under DD conditions becomes unimodal, the single activity peak lying between the morning and evening peaks seen at higher temperatures (Majercak et al., 1999).

It has been established that *per* mutants that cause long *(per^L)* or short *(per^S)* rhythms in DD conditions are poorly temperature compensated, such that at 18°C both run with a period much closer to 24 hours than they do at 25°C (Konopka and Benzer, 1971; Konopka et al., 1989). It was therefore decided to test the activity of $Po^{2.1.1}$ flies at 18°C to ascertain whether they might also exhibit a change in period, and whether the complex rhythms seen in this genotype could be eliminated by suppressing the morning peak of activity, facilitating further mapping studies. All flies were raised at 25°C before being transferred to the experimental temperature regime.

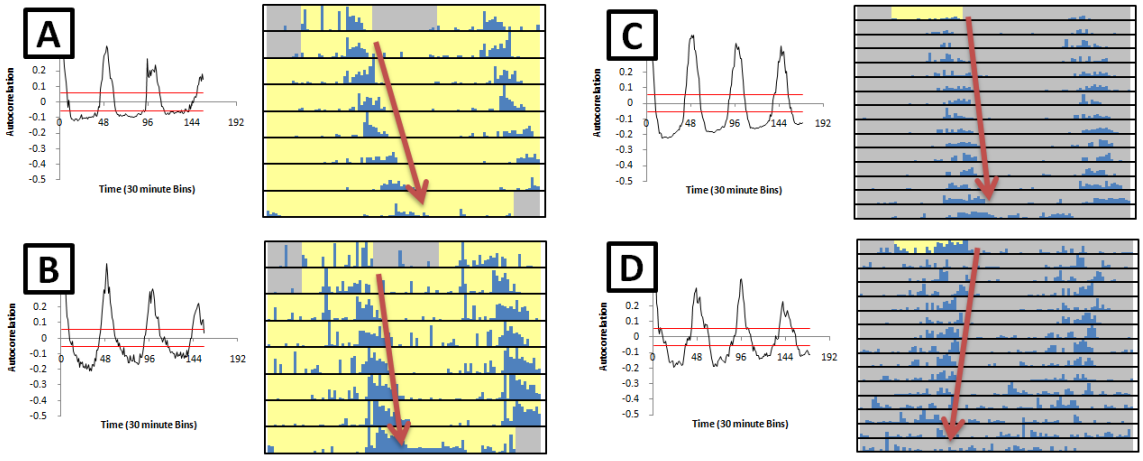


Figure 4-19 Autocorrelograms and double plot examples of male $Po^{2.1.1}$ flies' activity at 18°C. Under LL conditions, $Po^{2.1.1}$ flies may show a long period rhythm (**A**) or an almost circadian rhythm (**B**). Under DD conditions, flies showed periods either slightly longer (**C**) or shorter (**D**) than 24 hours.

The results in Figure 4-19 were analysed using CLEAN to generate an estimate of the periodicity of the flies. The LL period for *Po* flies at 18°C was found to be 25.1 ±0.15 hours, (n=32): clearly different from the value of 27.60 ±0.30 hours (n=50) at 25°C. Under DD conditions Po flies exhibited a 23.77 ±0.08 hour period (n=30), which was found to be significantly different (ANOVA P=0.00002) from the 24.27 ±0.77 (n=22) hour period at 25°C.

Not only was the *Po* phenotype found to be temperature sensitive in terms of period, but also comparison of the double plots in Figure 4-19 and Figure 4-9 (as well as the autocorrelograms for these flies) showed a qualitative improvement in rhythmicity at 18°C due to the more unimodal pattern of activity.

4.3.6 Clock dependency

As the *Po* mutation affects circadian rhythms, it was decided to combine the *Po* mutation with known clock mutants in an effort to elucidate *Po*'s method of action.

4.3.6.1 s-tim/l-tim *isoform*

It has been shown that the S-TIM isoform of TIM interacts more strongly with CRY than the L-TIM isoform, therefore an effort was made to sequence a number of lines carrying the *Po* mutation. All *Po* flies tested carried the *ls-tim* allele whilst there was no clear association between *tim* isoform and LL rhythmicity patterns in other heterozygous genotypes tested (data not shown).

4.3.6.2 per

As a preliminary experiment, a reciprocal cross was performed between the *Po* strain and the arrhythmic circadian clock mutant per^{01} to determine whether the LL rhythms seen in the *Po* strain were dependent on a functioning core clock mechanism. Due to the sensitivity of many *per* alleles to temperature, all experiments were run at 25°C:

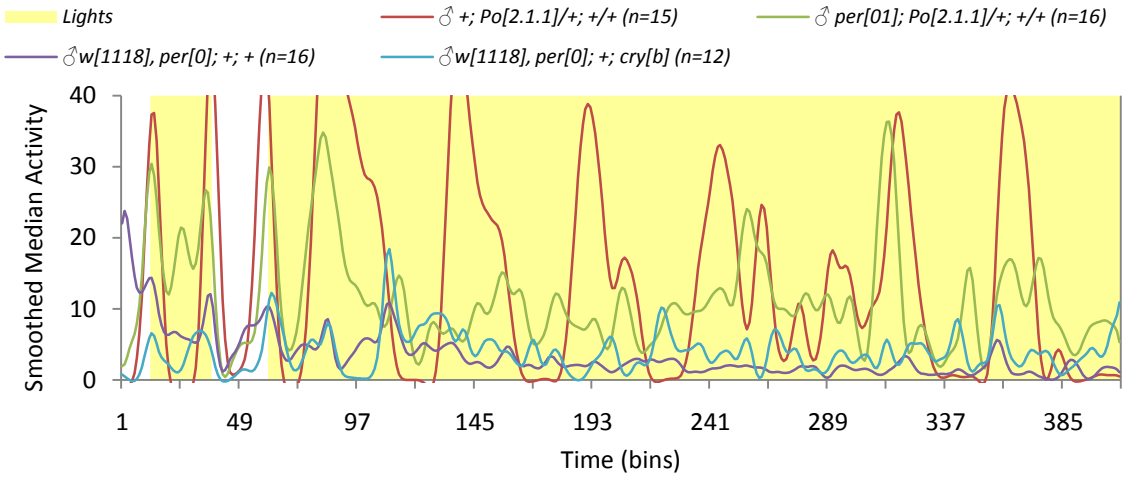


Figure 4-20 The LL rhythms shown by the *Po*$^{2.1.1}$ isolate depend on a functioning circadian clock. Of the lines tested, only the F1 control flies from the reciprocal cross between *Po*$^{2.1.1}$ and *per*^{01} not carrying the *per*^{01} allele showed significant long period rhythmicity (red trace). *per*^{01}*; cry*^{b} flies **were run as a control**

This data suggested that the rhythms seen in the *Po*$^{2.1.1}$ strain in LL were dependent on *per*, however consideration of individual double plots showed that some degree of *per*-independent ultradian rhythms were evident in the data:

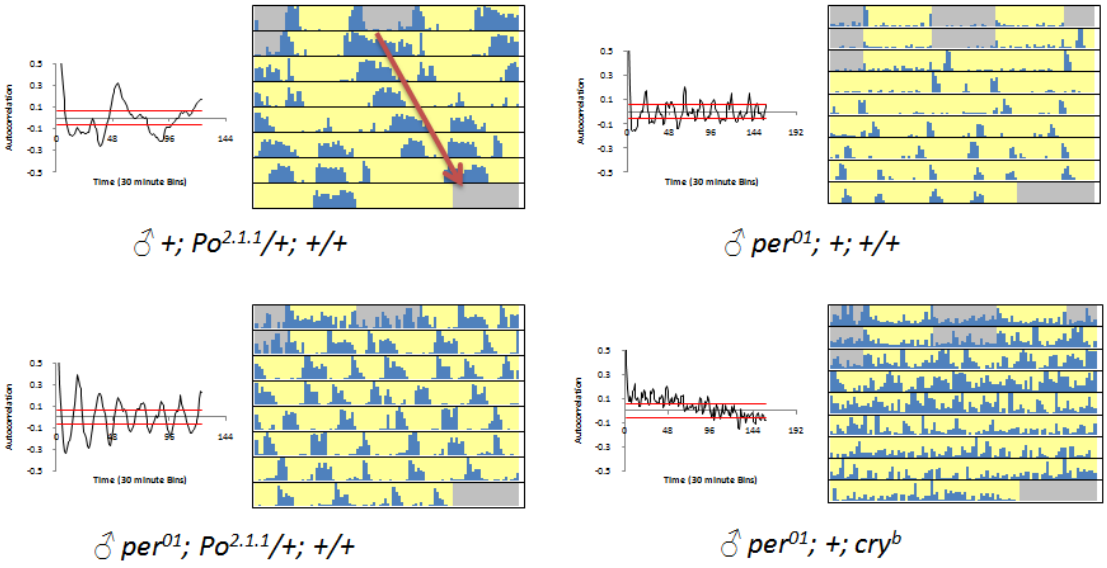


Figure 4-21 Individual behavioural records for flies carrying *per*^{01} and *Po*$^{2.1.1}$ alleles. Both *per*^{01}*; +; +* and *per*^{01}*; Po*$^{2.1.1}$*; +* showed ultradian rhythms, however the combination of *per*^{01} and *cry*^{b} elicited total behavioural arrhythmicity.

These results are not entirely surprising; the existence of a *per* independent, ultradian clock mechanism has been long theorised (Dowse, 2008; Dowse and Ringo, 1987), and can arguably be detected as complex ultradian rhythmicities in DD datasets (Power et al., 1995) or in minimalist 'skeleton' lighting regimes (Dowse, pers. comm.). Such ultradian rhythms might underlie the difficulties inherent in the meiotic mapping of *Po*. Interestingly the ultradian rhythms in the per^{01} strains shown in Figure 4-21 emerge almost immediately in LL conditions, precluding that they might result from emergent oscillator desynchrony.

To confirm that *per* is indeed required for *Po*'s rhythmicity therefore, a further experiment was performed testing whether *per* alleles affecting period might also modify LL rhythm periodicity (Figure 4-22).

The results show that the *Po* LL rhythmicity, like that in cry^b, is sensitive to *per* alleles that affect period length, implying both are mediated by the known molecular clock mechanism. However, combining *Po* (period length +2.6 hours relative to circadian) with per^L (+4.5 hours in the rhythmic per^L*; +;* cry^b background) gives a period of 34.5 hours (+10.5), showing evidence of epistasis rather than simple additivity between these alleles. Given that the LL period of per^L; *+;* cry^b is the same as the DD period of per^L, one can assume that the underlying periodicity of a LL rhythmic stock carrying a per^s allele is 19 hours (as in DD conditions - demonstrated in Yoshii et al., 2004). However, combining per^s (-5) with *Po* (+2.6) gives a period of 25.1 (+1): longer than the average of the two, supporting the hypothesis that there is an epistatic interaction between these alleles that increases the period by 3 hours relative to strict additivity.

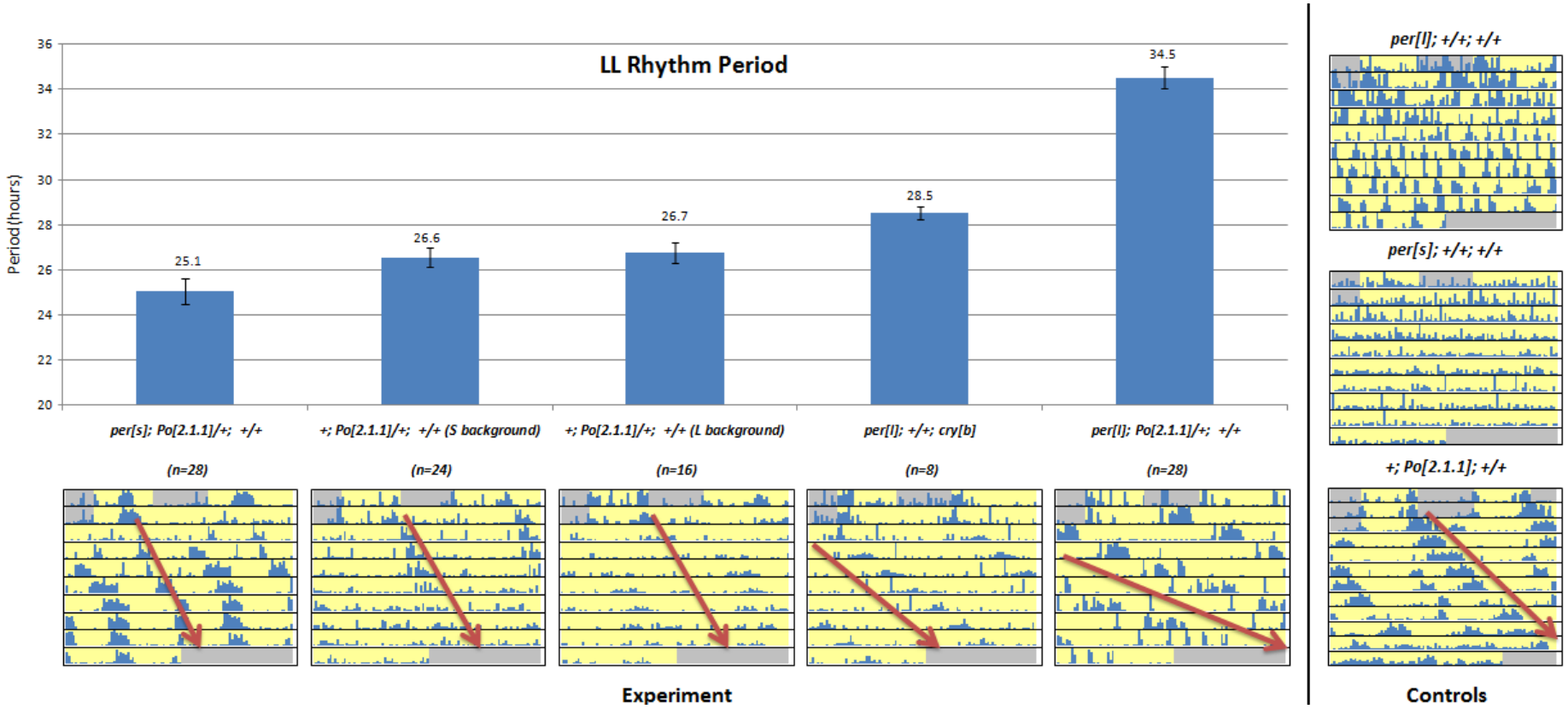


Figure 4-22 Effect of DD period modifying *per* alleles on $Po^{2.1.1}$'s LL rhythmicity. **Red arrows** show predominant rhythm. **'Controls' panel:** representative double plots for control genotypes per^{L}; +; +, per^{s}; +; + and +; $Po^{2.1.1}$; +. Interestingly a small number of per^{L}; +; + flies show an abundance of discrete ultradian activity bouts as in per^{0} (see Figure 4-21), rather than the more usual dispersed, low amplitude noise typical of LL arrhythmicity shown in the per^{s}; +; + example. **'Experiment' upper panel:** all stocks carrying either a single copy of $Po^{2.1.1}$ or two copies of cry^{b} show LL rhythmicity of different periods, shown as genotype mean values ± standard error of the mean. **'Experiment' lower panel:** representative double plots for each genotype are shown beneath the bar chart. Note that in all cases, LL activity peaks are derived from the evening peak.

4.3.6.3 Clk^{JRK}

As the *Po* rhythmicity is sensitive to clock mutations, a further experiment was run to determine whether LL rhythmicity could persist in the presence of the more severe *Clk*^{JRK} circadian mutant known to cause almost fully dominant behavioural arrhythmicity (Allada et al., 1998):

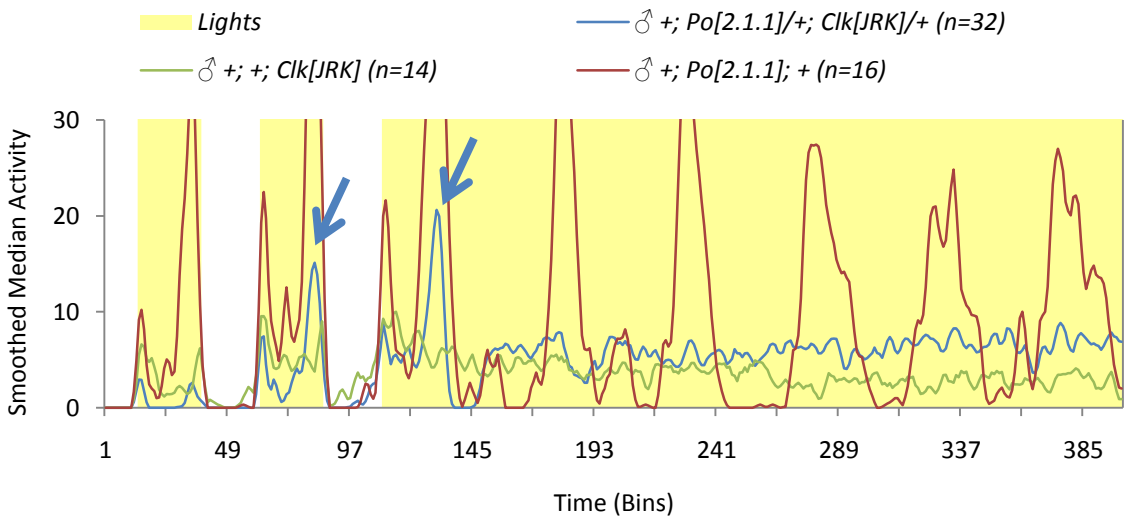


Figure 4-23 Effect of *Clk*^{JRK} on *Po*$^{2.1.1}$ LL rhythmicity. 100% of Clk^{JRK} homozygotes are behaviourally arrhythmic. Whilst flies heterozygous for *Po*$^{2.1.1}$ and *Clk*^{JRK} show some evidence of evening anticipation under LD conditions (**blue arrows**), >80% are arrhythmic under LL conditions.

The results clearly suggest that the *Po*$^{2.1.1}$ LL rhythms depend upon a functioning *Clk* gene, although an alternative explanation might be that LL rhythms require neural pathways that may be morphologically compromised in *Clk*^{JRK} mutants due to the pleiotropic effects of this mutation during development on projections from the l-LN$_v$ and s-LN$_v$ clock cells (Park et al., 2000). However, taken together with the *per* data, one can conclude that LL rhythmicity is dependent on a functioning circadian clock mechanism.

4.3.6.4 *norpA*^{p41} *(phospholipase C)*

The split oscillator phenotype occasionally observed in *cry*^{b} mutants can be prevented by combining the *cry*^{b} and *norpA* mutant alleles, therefore an experiment was performed to

test for any interactions between *Po* and the *norpA*^{p41} mutant. In an effort to simplify the behavioural profiles, the experiment was carried out at 18°C:

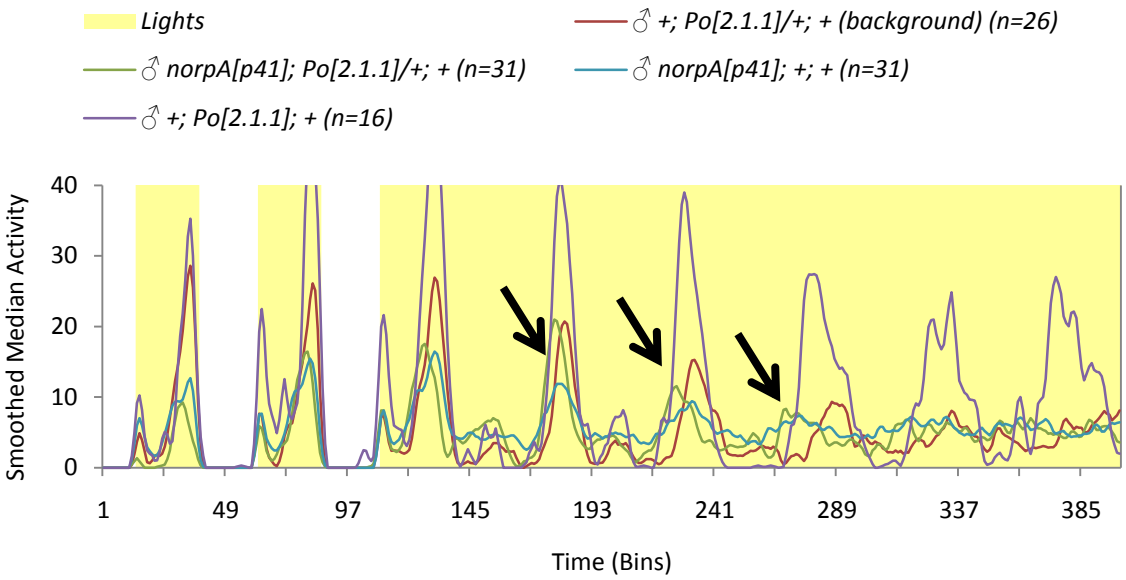


Figure 4-24 The effects of *norpA*^{p41} on *Po*$^{2.1.1}$ LL activity rhythms at 18°C. A reciprocal cross was performed generating flies heterozygous for *Po*$^{2.1.1}$ and *norpA*^{p41}, and a background control ('**background**') heterozygous for *Po*$^{2.1.1}$ but not carrying *norpA*^{p41}. Flies carrying the *norpA*^{p41} mutation are significantly less rhythmic than *Po*$^{2.1.1}$ flies, but both *norpA*^{p41} and *norpA*^{p41}; *Po*$^{2.1.1}$ flies show some evidence of short period LL rhythmicity (**black arrows** indicate activity peaks of this rhythm). For further details see Figure 4-25.

Somewhat surprisingly both the *norpA*^{P41}*; +; + and norpA*^{P41}*; Po*$^{2.1.1}$*/+; +* heterozygotes showed some degree of short period LL rhythmicity. Individuals of these lines were subject to further analysis:

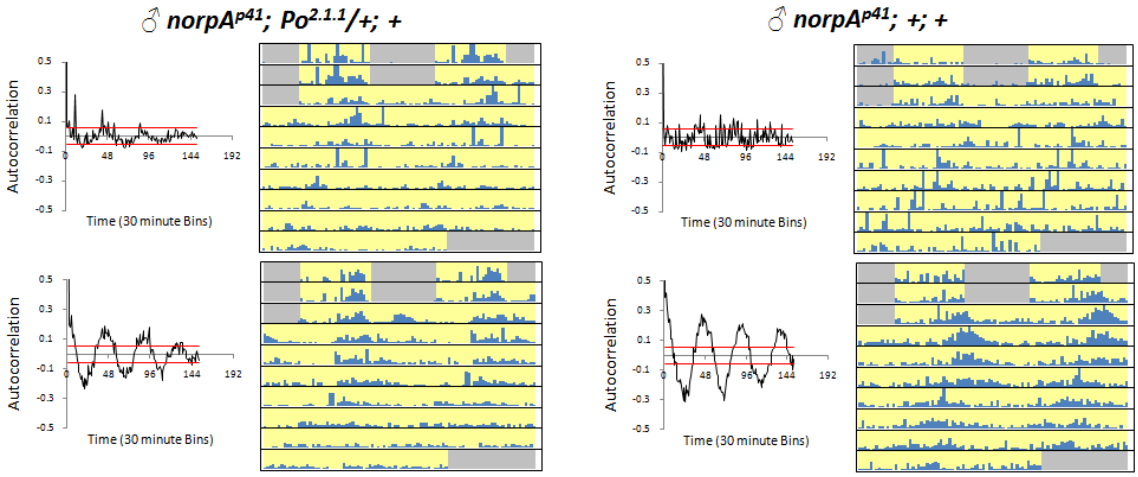


Figure 4-25 Examples of *norpA*^{P41} and *norpA*^{P41}; *Po*$^{2.1.1}$/+ activity in LL at 18°C. Both genotypes show evidence of both strong (**bottom**) and weak (**top**) short period LL rhythms.

Closer analysis of the behavioural profiles for these genotypes revealed that 61% of *norpA*^{P41}; *Po*$^{2.1.1}$/+ flies showed LL rhythms – similar to the 70% rhythmicity shown by *norpA*^{P41}; +; *cry*^{b} (Yoshii et al., 2004). Furthermore, 48% of *norpA*^{P41} flies showed some degree of LL rhythmicity of a similar period to that of the heterozygote:

Table 4-1 Rhythmicity profiles for *norpA*^{p41} and *Po*$^{2.1.1}$ flies in LL at 18°C. B/C = *Po*$^{2.1.1}$ crossed into the *norpA*^{p41}background as a control. See text for details.

Genotype	Arrythmic	Multiple Rhythms	Strong Rhythms	Weak Rhythms	n	τ
♂ +; *Po*$^{2.1.1}$; +	7%	0%	87%	7%	16	24.8 ±0.27
♂ *norpA*^{p41}; +; +	45%	7%	17%	31%	29	22.1 ±0.61
♂ *norpA*^{p41}; *Po*$^{2.1.1}$/+; +	29%	10%	42%	19%	31	22.0 ±0.70
♂ +; *Po*$^{2.1.1}$/+; + (B/C)	8%	12%	69%	12%	26	26.5 ±0.46

4.3.6.5 Phase response

Given the evidence suggesting a genetic interaction between *Po* and *norpA*, it was decided to further probe the light sensitivity of the *Po*$^{2.1.1}$ mutant using a phase shift experiment to directly assay the clock using the cross-correlation based phase shift analysis implemented in BeFly!:

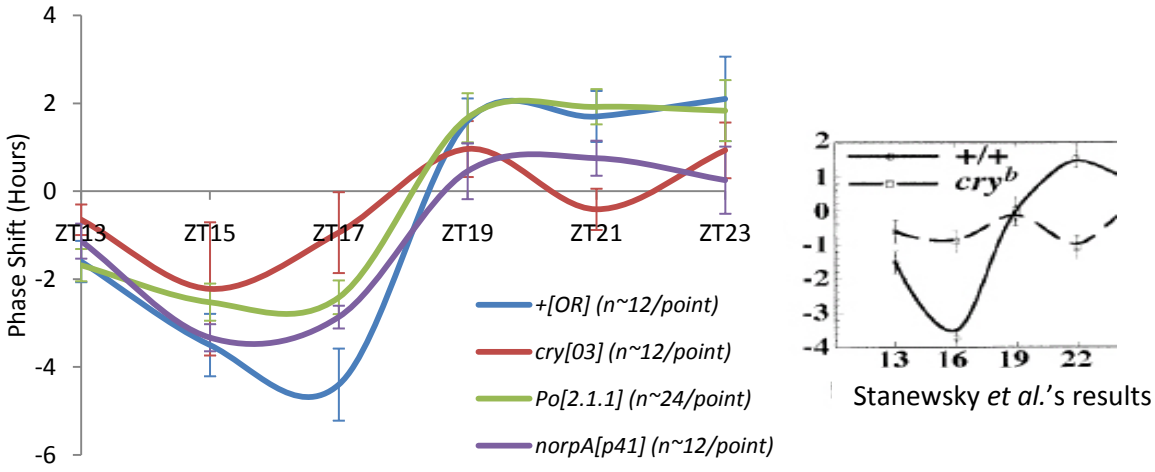


Figure 4-26 Results of *Po*$^{2.1.1}$ anchored phase shifting experiment. The phase of the endogenous clock was shifted using a 15 minute light pulse administered at the times detailed on the (x) axis. **Right panel** shows the results of Stanewsky *et al.*'s phase shift experiment using a 10 minute light pulse (Stanewsky et al., 1998) showing results very similar to those obtained by this author using the Oregon R wild type ($+^{OR}$) and a *cry* deletion strain (cry^{03}).

As can be seen in Figure 4-26, the $norpA^{p41}$ mutant shows reduced light sensitivity relative to the wild type, but a phase shifting profile unlike that of the cry^{03} null mutant. During the early night when light pulses phase advance the clock, $Po^{2.1.1}$ mutants have a phenotype more similar to the $norpA^{p41}$ mutants than the wild type, whilst during the phase delay portion of the night they respond in a manner indistinguishable from wild type controls. The statistical significance of the differing phase shifts elicited at ZT17 was therefore assessed using ANOVA followed by posterior unequal N HSD tests:

Table 4-2 Statistical significance of differing phase shifts elicited by a ZT17 15 minute light pulse relative to wild type Oregon R flies

Unequal N HSD Approximate Probabilities for Post Hoc Tests Error: Between MS = 5.9860, df = 54.000				
	OR	***Po$^{2.1.1}$***	***cry^{03}***	***norpa^{p41}***
OR		0.002563	0.000161	0.028668
Po$^{2.1.1}$	0.002563		0.017430	0.823434
cry^{03}	0.000161	0.017430		0.003029
Norpa^{p41}	0.028668	0.823434	0.003029	

The results clearly show that the *Po$^{2.1.1}$* mutant flies differ significantly from both the Oregon R wild type and the *cry^{03}* null mutant flies in their response to a 15 minute light pulse administered at ZT17. The profiles of the OR flies and *norpa^{p41}* mutants are significantly different, suggesting that the phase shift elicited by a light pulse might also be partially dependent on the canonical phototransduction cascade in addition to the cell autonomous mechanism involving CRY. However, this possibility contradicts evidence in the literature, and one possibility is that this difference emerges as a result of different *tim* s*/ls* alleles in the OR and *norpa^{p41}* stocks.

Due to the similarity in phase shift profiles between *Po$^{2.1.1}$* flies and *norpa^{P41}* flies, aged *Po$^{2.1.1}$* flies were subject to an optomoter behavioural test (Campesan et al., 2001); no evidence was found to suggest that the *Po$^{2.1.1}$* flies did not have a fully intact canonical phototransduction cascade.

4.3.6.6 Response to varying LD conditions

Given that the PRC data suggested that the *Po$^{2.1.1}$* strain might differ in light sensitivity during the early night phase, it was determined to examine the response of *Po$^{2.1.1}$* flies to LD regimes of different lengths. As can be seen in Figure 4-27 there is only limited evidence that *Po$^{2.1.1}$* flies differ from controls in their response to such lighting regimes:

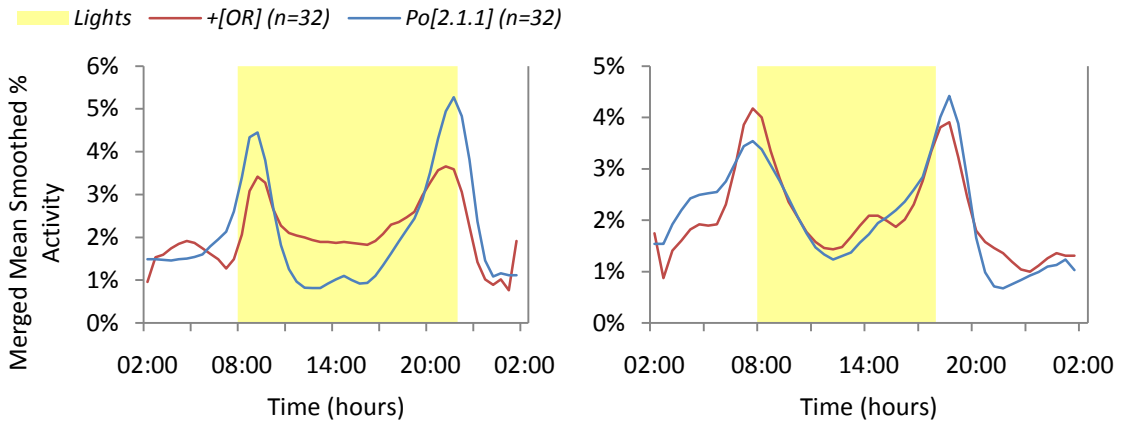


Figure 4-27 Response of *Po*$^{2.1.1}$ and OR flies to a long (left panel) and short (right panel) day entrainment regimes. The genotype mean was calculated over 4 days, which were then merged to form an 'average' day. To normalise for activity differences this 'average' day was then expressed as a % of activity per day.

4.4 Discussion

Circadian biology in *Drosophila* is undergoing a seismic shift in understanding as our knowledge of the molecular and neural basis of the clock increases. With the novel tools being developed the field must move beyond the stark distinction between 'rhythmic' and 'arrhythmic' flies, and come to recognise not only the existence, but the diagnostic potential, of complex behavioural phenomena such as decoupled oscillators (Yoshii et al., 2004) and ultradian rhythmicities (Dowse, 2008). Within this chapter I have identified and partially mapped a novel '*Po*' isolate of *Drosophila* that shows long period rhythmicity in LL conditions, generating data that challenges the orthodox assumption that flies are entirely behaviourally arrhythmic under LL conditions.

4.4.1 Mapping the *Po* strain

The semi dominant inheritance of *Po* phenotype makes it particularly challenging to map genetically. Whilst it is possible that there is a degree of variable penetrance or behavioural plasticity inherent to the *Po* phenotype (such as might be expected if the mutation

were a mild lesion) the observation that outcrossed heterozygous flies generally show an increased proportion of complex rhythmicities suggests that the *Po* phenotype is subject to many background dependent modifying genes. This result is not particularly surprising given the sensitive nature of behaviour (discussed in Chapter 7) and the discovery of many genes playing roles in CRY degradation after this work was completed (Sathyanarayanan et al., 2008; Murad et al., 2007).

Although mapping does not yet suggest a gene candidate for *Po*, data generated in this chapter suggested that further mapping may be facilitated by performing experiments at 18°C, as this lower temperature appears to enhance quality of *Po*'s LL rhythms, probably by suppressing bimodality and therefore eliminating complex interactions between morning and evening peaks (see Figure 4-19).

Crucially the mapping data in Figure 4-18 suggest that *Po* might lie near *curved* on the right arm of chromosome 2 (recombination position 2-75.5), ruling out the possibility that the *Po* strain might represent a novel allele of the known clock genes *tim* or *jet* which lie on the left arm. An attempt was made to sequence coding regions of both *tim* and *jet,* and no sequence change that correlated with the *Po* phenotype could be found (data not shown).

Further mapping might be performed using an appropriate mapping stock carrying markers proximal to *curved*, and by using a robust meiotic mapping scheme more appropriate for mapping a partially dominant behavioural phenotype (see appendix 10.2). As the sequencing work attempted in this chapter was by no means comprehensive, DNA samples were retained from all the flies that survived to the end of long LL experiments in this chapter to facilitate SNP mapping and candidate gene sequencing.

Future mapping studies might be facilitated by the development of novel tools for dissecting complex LL rhythms, a topic considered in detail in the next chapter.

4.4.2 A role for *Po* in the intracellular clock?

Despite the failure to clone *Po*, the data generated allow informed speculation as to the likely function of *Po*. The rhythmic behaviour seen in *Po* flies in LL conditions is clearly dependent upon a functioning clock (as shown by experiments using the severe Clk^{JRK} mutant in Figure 4-23), and is sensitive to *per* alleles that affect period length. However, the results also indicated that some rhythmic behaviour might persist in the per^{0} clock mutant background, and this adds to a growing body of work suggesting that *per* may be at least partially dispensable for rhythmic behaviour in DD conditions (Dowse et al., 1987; Collins et al., 2005). This chapter shows that this is also the case in LL conditions (Figure 4-21).

Interestingly the *Po* long period rhythms in LL are poorly temperature compensated, running much closer to 24 hours at low temperature. A similar effect is seen in a number of known *per* (Konopka et al., 1989) and *tim* (Matsumoto et al., 1999; Rutila et al., 1998a) period changing alleles, suggesting PO's activity might affect the activity, abundance or stability of *per* or, perhaps more likely, *tim* (the key target for CRY mediated light induced degradation).

An intriguing result is the observation that combining period altering *per* alleles with *Po* results in an epistatic three hour period lengthening effect (Figure 4-22). The per^{s} phenotype is thought to occur as a result of accelerated nuclear degradation of PER, and the per^{L} phenotype is likely to occur as a result of delayed nuclear accumulation of PER, possibly due to disruption of PER-PER dimer formation (Huang et al., 1995; Landskron et al., 2009). Given the epistatic interaction between these alleles and *Po*, this suggests that the *Po* mutation's effects on circadian period are mediated by processes independent of and preceding those causing the per^{s} and per^{L} phenotypes – exactly as would be expected if PO interacted with TIM

In trying to elucidate this function one can consider the phase response curve for *Po* (Figure 4-26) which shows the strain has a decreased sensitivity to light during the delay

portion of the night, notably at ZT17 (at which point PER is predominantly cytoplasmic), but normal response during the advance phase beginning at ZT19 when PER and TIM levels peak. Although the phase shifting profile of *Po* during the early night appears similar to that of the *norpA*^{p41}strain, evidence in the literature suggests that the canonical phototransduction pathway does not play a major role in phase shifting (Suri et al., 1998), whilst evidence from optomotor experiments shows that *Po* flies retain an intact visual pathway - unlike *norpa*^{p41} mutants. The difference between *norpA*^{p41} and the OregonR control could therefore be an artefact of small sample sizes, or be due to biologically meaningful differences between the strains affecting their light sensitivity (such as the carrying different isoforms of *tim*).

Both the known LL rhythmic mutants *jet* and *cry* show only limited phase shifts (~one hour) in response to both phase advancing and phase delaying light pulses, reflecting their reduced ability to degrade PER and TIM, so the normal phase advancing behaviour of *Po* is intriguing. This result supports published data suggesting that the early night (phase delay zone) contains several key phosphorylation steps that control the rate of the clock, and that clock lesions (notably *tim*^{SL}) can have phase specific effects (Rutila et al., 1998a).

The author therefore favours a hypothesis in which the likely role for PO is in PER/TIM cytoplasmic degradation processes (which may include phosphorylation steps that potentiate such degradation), as experiments have shown that simply impeding such degradation, even by increasing PER levels in certain clock cells (Murad et al., 2007), is sufficient to confer LL rhythmicity. These results suggest that failure to degrade PER and or TIM is the key molecular step in potentiating such rhythmicity. With the emerging complexity of the degradation pathway (Sathyanarayanan et al., 2008), it is not unreasonable to suggest a gene such as *Po* may only play a role in the early night. Clearly this question can only be properly resolved by immunocytochemical investigation of PER and TIM levels.

4.4.3 A role for Po in the intercellular clock mechanism?

The discussion so far has been largely restricted to consideration of the action of PO in terms of the canonical, cell autonomous clock mechanism, however one should not discount the importance of non-autonomous light input to the clock. Indeed, 'night time' illumination at one quarter the level of moonlight is sufficient to shift the typical morning and evening peaks of *Drosophila* activity into the night in a process independent of CRY, as well as change overall activity levels independently of the circadian clock (Kempinger et al., 2009; Bachleitner et al., 2007). Similarly Figure 4-25 shows that disrupting retinal phototransduction using the *norpA*^{p41} mutant seems - in some cases - sufficient to elicit ultradian LL rhythmicity. It is disappointing that this experiment does not provide more illuminating results, especially given the recent demonstration (both in chapter 6 and by Wijnen and colleagues) of a substantial light dependent transcriptional program that is sensitive to mutations affecting the retinal phototransduction cascade (such as *norpA*) but not to disruptions in *cry* (Wijnen et al., 2006).

Returning to the clock itself, recent studies (emerging after this work) suggest that LL rhythmicity is in many cases the result of manipulations that disrupt specific components of the intracellular clock network. One particularly striking example is that the over-expression of *per* or *morgue* in the evening cells (LNds, DN1s, and DN3s using *tim>GAL4/Pdf>GAL80* to drive *UAS-per* expression) causes LL rhythmicity, an effect not seen when *per* is over expressed in all clock cells (Murad et al., 2007). Similarly over-expression of *shaggy* in the evening cells also induces LL rhythmicity (Stoleru et al., 2007). Surprisingly the LL rhythms observed in such transgenic flies are generated by DN1s themselves, which display robust cycles of PER localisation (though possibly not rhythms of per expression levels - Stoleru et al., 2007) and PDP1ε levels (Murad et al., 2007). These rhythms are therefore independent of the LNvs (which are generally considered to represent the 'core' clock cells), though both studies suggest that the presence of these PDF +ve cells improves LL rhythm quality. The observation that LL rhythms emerge from the CRY-ve DN1s and DN2s (the evening cells) suggests an as yet

undescribed role for CRY as a regulatory gene – as in the mammalian system - in addition to its role in photoreception, a possibility that has been entertained by our laboratory (Collins et al., 2006) and supported by observations that *cry* mutants have marginally shorter periods than wild type flies (Stanewsky et al., 1998).

Stoleru and colleagues have developed the intercellular model further, generating some evidence to suggest that the relationship between the morning and evening oscillator is dependent on the prevailing lighting conditions (Stoleru et al., 2007):

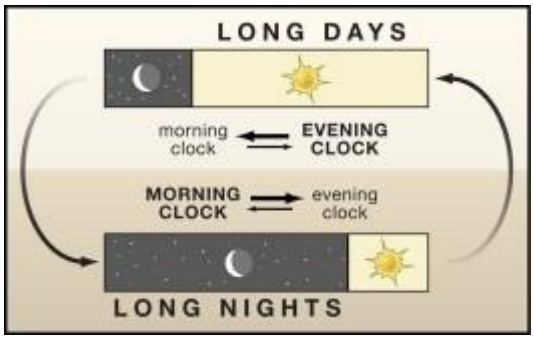


Figure 4-28 A model to show how interactions within the circadian intracellular clock might underlie seasonal adaptation. Under short day conditions the morning clock controls the phasing of the evening clock, whilst under long day conditions the evening clock controls the morning clock. Figure from Edery (Edery, 2007).

Stoleru *et al.* propose that such a model represents a seasonal timer; in this model, the more light sensitive DN cells play a dominant role in controlling the phase of both the morning and evening peak under long-day conditions, whilst the LN_v cells that act as key pacemakers and control phasing under short day conditions, and consequently also in DD. The authors also speculate that the differing anatomical positions of the LN_v and DN cells might underlie their differing photodependencies.

Given the long period rhythmicity of the *Po* strain, and the observation that the LL rhythms seen in the *Po* strain predominantly emerge from the evening peak (in accordance with the general observations that underpin Stoleru *et al.*'s model), one might expect that *Po*

flies would exhibit a delayed morning anticipation peak under long day conditions in which the evening clock is dominant. Unfortunately the evidence in Figure 4-27 is equivocal; whilst *Po* flies clearly anticipate morning at around 05:00, wild type controls show a 'bulge' of activity during the night beginning around 02:00, and again raise their activity at 06:30, therefore no clear conclusion can be drawn.

One possibility not addressed in this chapter is that *Po* mutants have gross neuroanatomical defects which might underlie the LL rhythmicity phenotype. It is known that ablation of large parts of the optic lobes can cause split rhythms under DD conditions (Helfrich, 1986), whilst transgenic disruption of neuronal communication can cause complex rhythmicities (Nitabach et al., 2006). Although this could be resolved using immunocytochemistry, the observation that *Po*'s LL rhythms are poorly temperature compensated suggests that *Po*'s effects on rhythmicity are the result of changes in molecular thermodynamics rather than developmental deficits.

4.4.4 *Po* heterozygotes – evidence for multiple oscillators, or complex rhythmicities?

A key issue arising in this chapter is the discrimination between complex circadian rhythmicities (which may arise due to disrupted neuronal communication - Nitabach et al., 2006)(such as those that can be elicited in DD conditions by disrupting communication between clock neurons - Nitabach et al., 2006)(such as those that can be elicited in DD conditions by disrupting communication between clock neurons - Nitabach et al., 2006), and high frequency 'noise' activity (perhaps controlled by homeostatic processes). Whilst such complex rhythms are generally dismissed as evidence of a damaged clock mechanism, the residual ultradian rhythmicities found in genotypes such as the per^{01} clock mutant have long been known in the literature (Dowse and Ringo, 1987) and are re-emerging as a topic of interest (Kempinger et al., 2009). It appears therefore that PER might play a role in

constraining rhythms to their proper phase, as in its absence rhythms become more variable. Similarly knocking down the HSP90 chaperonin in flies increases behavioural variability both at the population level, and during the lifetime of individual flies (Hung et al., 2009). Taken together, one might hypothesis that chaperonin-like molecules act as 'integrators' within the clock mechanism to ensure its accuracy – indeed several *per* and *tim* mutations are temperature sensitive, suggesting that in these mutants a putative integrating function of *per* is disrupted. The analysis of such complex rhythms in the context of an ultradian clock model is considered in detail in the next chapter.

4.5 Conclusion

A number of further experiments may prove particularly informative in determining the identity and function of *Po.* The main foci should be mapping the *Po* locus and studying the distribution of PER and TIM in *Po* mutants' clock cells during both DD and LL conditions. However, the considerable variation in period and phase between *Po* individuals (even in the inbred $Po^{2.1.1}$ line) makes such an investigation a particularly challenging prospect. Perhaps more accessible would be further experiments probing for interactions between the *Po* gene and *cry*, *norpA* and the key clock gene *Pdf*.

What this chapter has made abundantly clear is that LL rhythmicity is a complex phenomenon much neglected within circadian biology, yet worthy of further investigation at both the intracellular and intercellular level. Further dissecting such rhythmicity, as well as identifying novel genes that might mediate such effects, is the focus of the next chapters of this thesis.

5 Complex rhythms in constant light

5.1 Introduction

In the previous chapter considerable attention was given to LL rhythms controlled by evening cells, which are thought to arise either as a result of compromised light mediated degradation of TIM, or due to disruption of the proper phasic relationship between the morning and evening oscillators. However, such hypotheses do not account for the emergence of complex ultradian rhythmicities in many genotypes under LL conditions. Such LL rhythmicities fall within a wide range of periods, therefore genotypes which display such behaviours appear arrhythmic when considered as group averages (see examples in previous chapter), and consequently are much neglected in the field.

Such observations are easier to explain if one considers an alternative model of the circadian clock based on the structure of mechanical clocks consisting of three parts: a central oscillator of standard frequency, an integrator/differentiator that converts the oscillator's output into useful units of time, and a mechanism by which the integrator can be reset to environmental time. Crucially the higher the frequency of the oscillator, the more accurate the clock at measuring a given interval of time, driving the development of ever more high frequency atomic clocks (Allan et al., 1997).

The measurable accuracy of the circadian clock exceeds that possible using an inherently biphasic transcription/repression timing system; might the circadian clock not therefore use an ensemble of high frequency ultradian oscillators as its time base (Dowse and Ringo, 1987; Pavlidis, 1971)? *Drosophila* exhibit a number of high frequency rhythmic behaviours which might depend upon high frequency oscillators, notably rhythmic pulses comprising courtship 'songs' (Campesan et al., 2001)) and heartbeat rhythms, as shown in Figure 5-1:

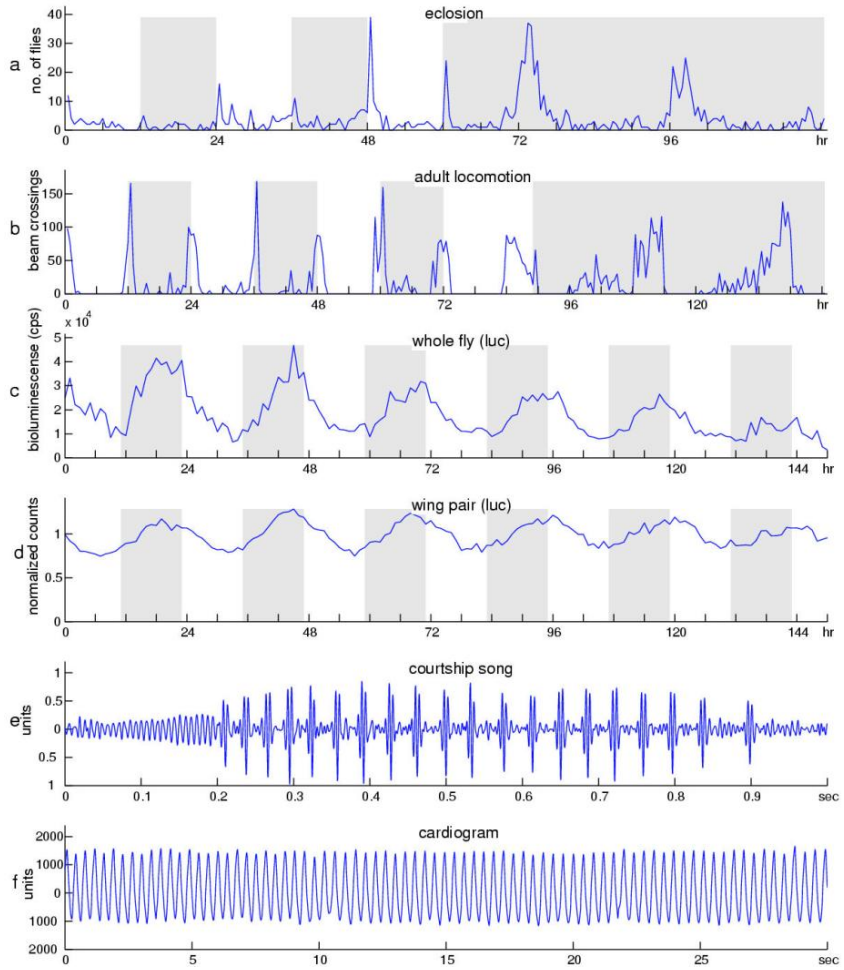


Figure 5-1 Rhythmic processes in Drosophila. Rhythms of circadian (**a-d**) and ultradian (**e,f**) timescales in *Drosophila.* Environmental light dark conditions are plotted as white and shaded blocks on each diagram. **a**: population eclosion rhythm in LD and DD conditions. **b**: individual adult male locomotor activity rhythm. **c,d**: Normalized activity of a firefly luciferase transgene driven by the *timeless* promoter in the whole fly (**c**) and in a dissected wing pair (**d**). **e**: a one-second bout of male courtship song, showing "sine" singing (between 0 and 0.2 seconds) and a train of pulses. This pulse train has a species specific interpulse interval (~35 msec). **f**: a pupal cardiogram. Figure and text adapted from Levine *et al.* (Levine et al., 2002b).

Evidence is beginning to emerge that these rhythmic behaviours are linked; *per* mutants are known to affect circadian rhythms, ultradian courtship song rhythms (Kyriacou and Hall, 1989) and infradian developmental rhythms (Kyriacou et al., 1990), whilst mutants in the Ca^{2+} channel *cacophony* affect both courtship songs and heartbeat rhythms (Ray and Dowse, 2005), and the CCAP neuropeptide is known to play roles in both ecdysis (Park et al., 2003) and in cardiac function (Dulcis et al., 2005).

Supposing therefore that *Drosophila* have a high frequency oscillator system that underlies the lower frequency circadian oscillation (Dowse, 2008) and therefore beyond the resolution of current microarray studies, how might this explain the complex rhythmicities observed in LL? One hypothesis has been advanced that the observed mid-range complex ultradian rhythms in *Drosophila* activity records reflect artifactual outputs produced (or rendered visible) when the integrator system which steps down high frequency rhythms is damaged (Dowse, 2008). As described in the previous chapter, molecular chaperonins such as HSP90 are good candidates to act as integrators (Hung et al., 2009). In the absence of such an integrator, ultradian periods are detected fairly uniformly through the 4 to 18 hour region of the spectrum. As there are no biological or geophysical rhythms within this range, perhaps such ultradian rhythms reflect artefacts of an ultradian clock rather than fluctuating outputs of uncoupled circadian oscillators. Whether these ultradian rhythms are themselves harmonics of even higher frequency oscillators is a question which remains open to debate.

This chapter presents the results of a number of novel analytical techniques not previously applied to complex behavioural rhythms in an attempt to further dissect the nature of such ultradian rhythms in the context of the high frequency clock hypothesis.

5.2 Methods

5.2.1 Wavelet analysis

Wavelet analysis was run within The R Project for Statistical Computing v2.9.0 (http://www.r-project.org/) using WAVECLOCK - Nonparametric wavelet regression for oscillating time series data v1.0-3 (http://sgdp.iop.kcl.ac.uk/tprice/software.html). Further details are available in the BeFly! manual within appendix D1.

5.3 Results

5.3.1 Analysing complex activity patterns

Both the diversity and complexity of *Drosophila* behaviour under LL conditions ensures that analysis of such behaviour is subjective, time consuming and often inconclusive. Although a number of filters can be applied to activity records prior to analysis, these cannot compensate for the limitations of Fourier or autocorrelation based analyses when considering datasets with changes in period or phase – changes which are particularly evident in LL activity records. Therefore an effort was made to analyse LL activity using a novel wavelet analysis developed for cell culture studies which require considerable data filtering to generate meaningful results (Levine et al., 2002b).

The continuous wavelet transformation (Torrence and Compo, 1998) can be used to determine local changes in period and phase by projecting the unidimensional time-series data into two-dimensional time-frequency space. This time-frequency space can be easily visualised as a heat map, in which 'hotter' regions give an indication of the period of the data at any given moment (shown in Figure 5-2 A). As random noise (stochastic activity bursts) is spread over the time-frequency space, wavelet based analyses do not require data to be de-

trended or filtered, and is therefore both robust and informative in the face of transient perturbations to rhythms. Like Fourier analysis, data can be reverse transformed to reconstruct the modal frequencies and give an estimate of phase in the original signal (shown in Figure 5-2 B).

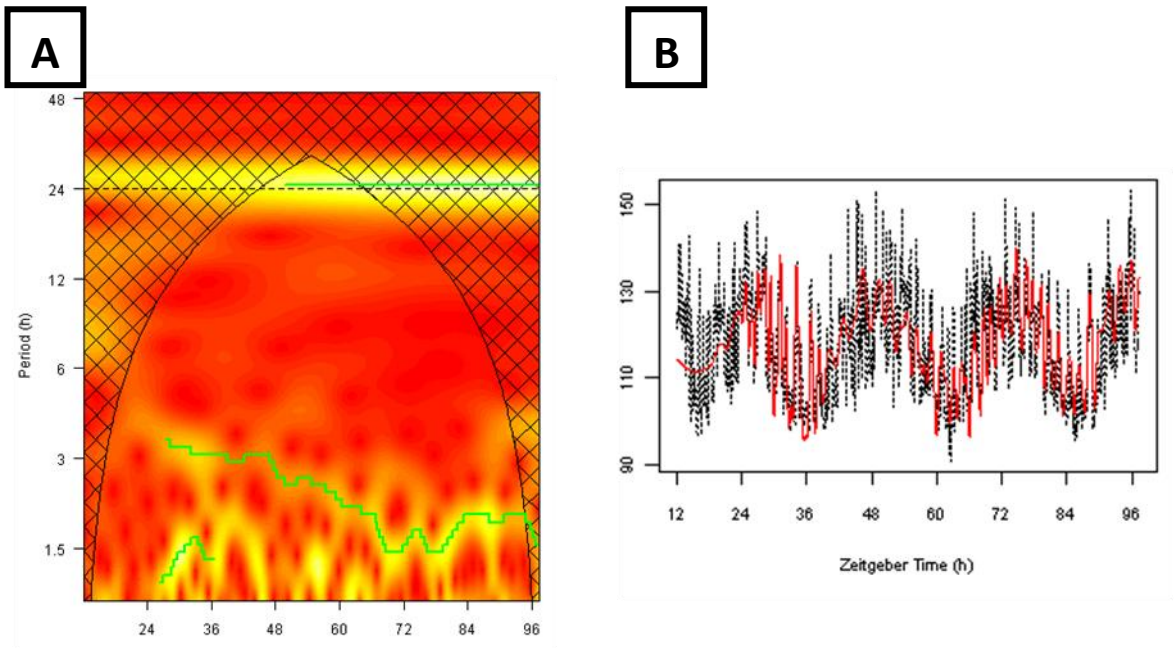


Figure 5-2 Example of WAVECLOCK outputs. (**A**) wavelet scalogram showing 'heat' around the 24 hour period mark, identified as a continuous band by the crazy climbers chain forming algorithm (**green lines**). The original data series (**B**, dotted **black line**) can be reconstructed (**red line**) without the high frequency noise. Figure from Price *et al.* (Price et al., 2008).

An alternative to the continuous wavelet transformation is the discrete wavelet transformation, which can be used to determine the probability that a time series has a period within a given range. Such analysis is useful when seeking to discriminate rhythmic and arrhythmic signals, or when reconstructing a signal using only its significant wavelengths.

There are currently two groups implementing wavelet-based analysis of circadian data; Price *et al.*'s WAVECLOCK package implemented in the free statistical package *R* (Price et al., 2008), and an as yet unpublished package implemented in Matlab (Harang *et al.* in prep,

Dowse, H.B. pers. comm.). Both use the complex-valued Morlet wavelet which preserves phase information and allows greater resolution in the frequency domain (Price et al., 2008).

The only published use of wavelet analysis (Price et al., 2008), utilises cell culture data with a sampling frequency set at 10 minutes. As wavelet analysis has not - to the author's knowledge - been employed for *Drosophila* behavioural analysis, the effect of changing the resolution of behavioural records (normally recorded using 30 minute bins) on WAVECLOCK's output was tested:

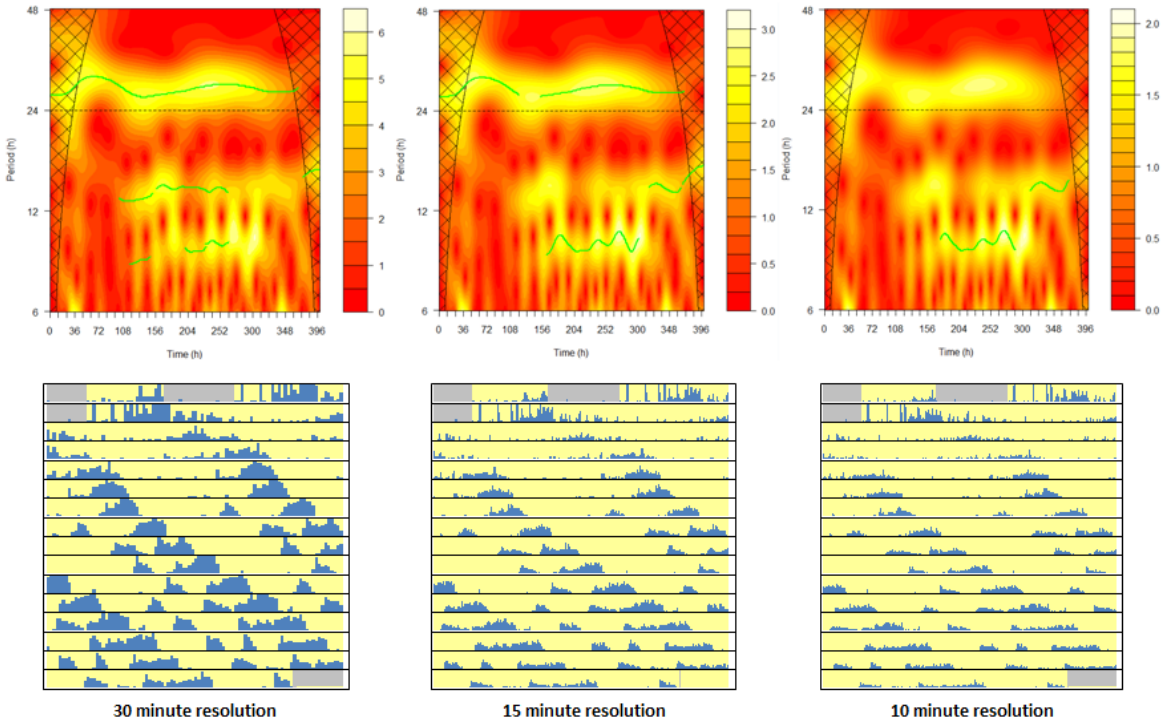


Figure 5-3 Determining the ideal sampling frequency for WAVECLOCK analysis of *Drosophila* activity data. Top panels: scalograms in which **green lines** represent the output of the period chaining algorithm applied to the results to generate regions of similar periodicity, **hatched areas** lie outside the analysis' 'cone of influence', in which there is insufficient data to make a reliable estimate of period. **Bottom panels**: double plots of the data subject to WAVECLOCK analysis. It is clear that the additional resolution of 15 or 10 minute bins makes little difference to time-frequency scalogram (though it does affect the period 'chains' created by the 'crazy climbers' algorithm), probably because the increased temporal resolution at activity bout onsets and offsets is largely inconsequential compared with the periodicity of the dataset.

As the results suggested that activity data recorded at 30 minute resolution was appropriate for WAVECLOCK analysis, several *Po* flies showing unusual LL rhythmicities were analysed using both WAVECLOCK's continuous wavelet transformation, and the established spectral and autocorrelation analysis tools used in the lab (Figure 5-4).

The data in Figure 5-4 panel A shows a fly demonstrating a long period rhythm (1) which speeds up somewhat (2) before splitting into two components (3) which free run with different periodicities (4) as indicated by red arrows. The time frequency plot (more correctly a continuous wavelet transform scalogram) accurately depicts the changing behavioural motifs; in contrast spectral analysis provides a much less clear result (though it does detect both long and short period rhythmicities). The autocorrelation data supports an emerging ultradian component to the data (the closely spaced blue arrows show the relevant peaks), but this analysis is not particularly informative with respect to the dominant periodicity of the dataset.

Panel B shows a fly with a long ~27 hour period slowing to a ~24 hour period (1), a change which is captured by WAVECLOCK, but not by the spectral or autocorrelation analyses, both of which suggest a longer periodicity within the dataset.

Panel C shows a fly exhibiting a single rhythmic component of ~24 hour periodicity (1) splitting to form two oscillators (2), which almost immediately coalesce to form a new peak (3), again showing a periodicity of ~24 hours but with a different phase relationship to the original. In this case the autocorrelation analysis is of limited use as the signal rapidly degrades. Spectral analysis correctly identifies two rhythmicities, but the periods of 23.05 and 27.85 hours are not supported by the double plot data. The time-frequency plot gives a better interpretation of the data, as careful inspection of the heat map around the time that the oscillators split (2) clearly shows both a long and short period component (white arrows), the period chaining algorithm favouring the stronger short period component.

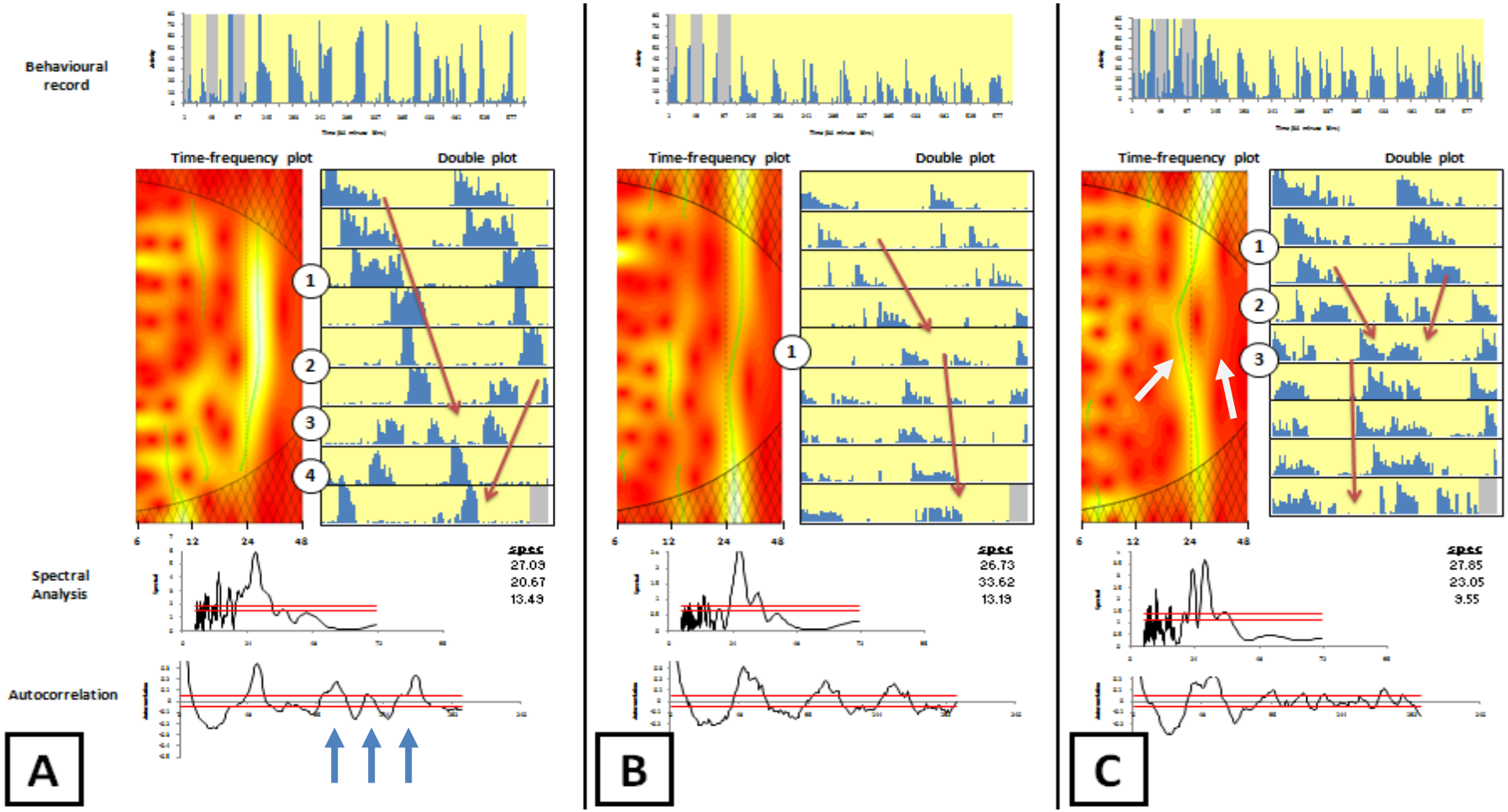


Figure 5-4 Testing different analysis methods against flies showing unusual rhythmic patterns. Each panel shows an individual *Po* fly's activity, including a linear plot (the '**behavioural record**', including a period of DD activity), as well analyses of the LL activity including a **double plot, time-frequency plot** (scalogram), **spectral analysis** and **autocorrelation**. For further details of each fly's activity, see text above.

It is clear from the examples in Figure 5-4 that wavelet analysis can recapitulate and quantify some of the transient changes in rhythmicity that can be qualitatively observed using a double plot.

5.3.1.1 Discriminating ultradian and circadian components in complex rhythms

Drosophila exhibit a crepuscular pattern of activity in LD conditions, which in some genotypes persists in constant conditions, whilst in others the morning peak rapidly damps to form a unimodal pattern of activity (Majercak et al., 1999). Flies that show a sustained bimodality under high temperature constant conditions often show a strong ultradian rhythm with a period of ~12 hours when visualised on a spectrogram, and in general this is accepted as being half the true period of the fly (Levine et al., 2002b).

However, when considering complex behavioural records a question arises; how can one distinguish between long period flies showing bimodal activity patterns, and short period ultradian flies showing a unimodal pattern of activity? For example, both a unimodal pattern with τ=16 hours, and a bimodal pattern with τ=32 hours will exhibit a peak every 16 hours, and spectral analysis will identify both 16 and 32 hour components in each case. This is particularly pertinent when considering tidal rhythms, in which the 12 hour ultradian component of a circadian rhythm and the 12.4 hour tidal rhythms must be distinguished (Lin Zhang pers. comm.).

Similarly in LL, *Po* flies show a mixture of unimodal and bimodal activity patterns (though the majority of flies show a unimodal pattern derived solely from the evening peak – Figure 4-10). Outcrossed *Po* flies clearly show discrete peaks of activity under LL conditions (Figure 4-12), a number of which appear as complex rhythms (which can be inhibited by low temperature by suppressing bimodal behaviour). In an effort to further understand the complex rhythmicities of outcrossed flies, data from a number of the flies in Figure 4-12 were reanalysed using WAVECLOCK:

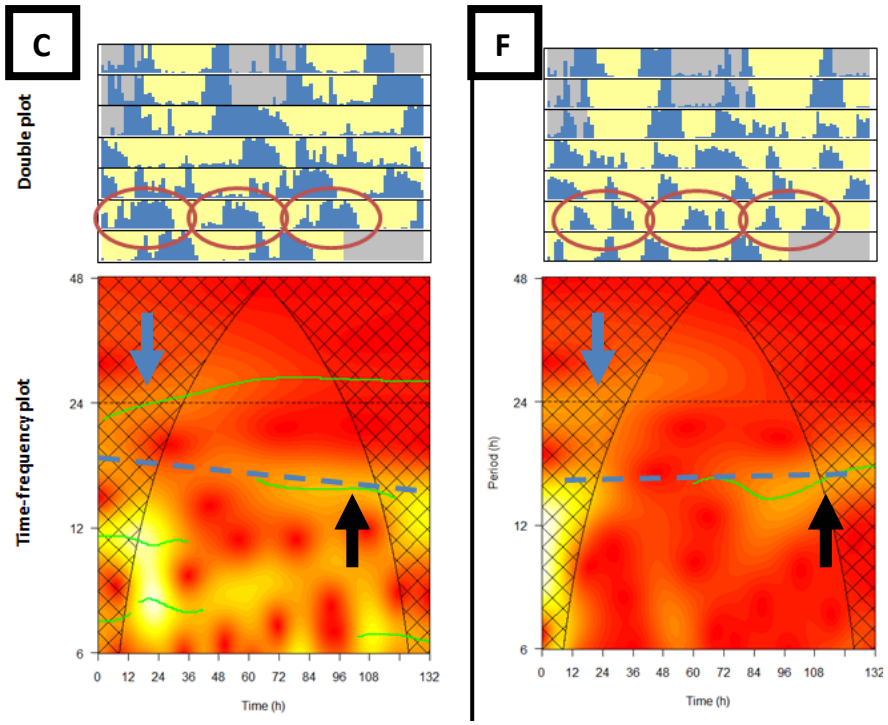


Figure 5-5 Using WAVECLOCK to discriminate between ultradian and circadian rhythmicities in outcrossed *Po* flies in which neither spectral or autocorrelation analysis is informative. Data replotted from Figure 4-12 panels **C** and **F**. **Green lines** indicate regions of similar periodicity, as detected using the 'crazy climbers' chaining algorithm. **Left panels**: automated analysis detects a circadian component running through the data, as well as an emerging ultradian component (**black arrow**). However, a 'by eye' fit (**dashed blue line**) also reveals an ultradian periodicity running through the data. **Right panels**: automated analysis detects an ultradian period in the later part of the behavioural record (**black arrow**). Setting aside the results of the chaining algorithm, the circadian rhythmicities identified early in the both datasets (**blue arrows**) appear to give way to ultradian rhythmicities later (**black arrows**).

The results of wavelet analysis are somewhat equivocal, indeed results seem very sensitive to the parameters used by the chaining algorithm used to visualise the dominant period (shown by green lines). If one considers the data 'by eye' then there is some qualitative evidence that in the latter part of the experiment, both flies showed ultradian rather than complex circadian periodicities. Such a conclusion might be quantitatively expressed by using a discrete wavelet transformation to assess the probability of circadian versus ultradian rhythmicities, however this is currently not possible using WAVECLOCK.

An additional method to distinguish ultradian and circadian rhythms is the use of autocorrelation. In bimodal activity traces, the 'morning' and 'evening' peaks of activity arise

through the action of different neuronal populations (Grima et al., 2004; Stoleru et al., 2004), and as such have subtly different profiles. As a result of this, a morning peak correlates much better with another morning peak than with an evening peak; in some cases this can be used to identify the dominant rhythmicity in a dataset:

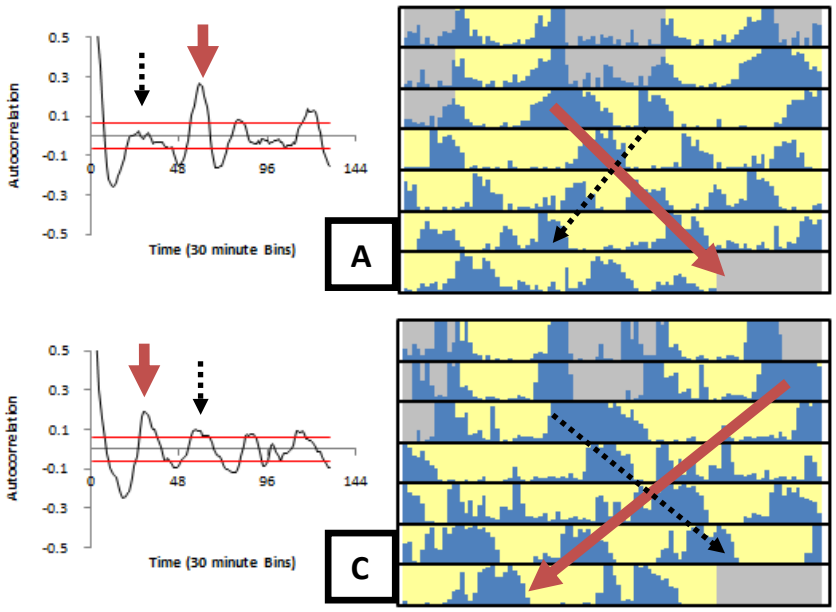


Figure 5-6 Discriminating ultradian and circadian rhythmicities using autocorrelation. Data replotted from Figure 4-12, panels **A** and **C**. **A**: autocorrelation shows that the circadian period (**red arrow**) is a better fit to the activity data than an ultradian period (**dashed black arrow**). The converse is true in **C**.

The use of correlational techniques to examine data is considered in further detail in the BeFly! manual within the appendix (D1). One limitation to such approaches is that although morning and evening peaks are clearly different when considered as group averages, individual flies' activity is subject to many stochastic factors which may obfuscate the true morning/evening relationship. One solution advanced by Dowse *et al.* is to examine the data using double plots. As described in the methods, double plots are generally plotted such that each row represents two days of data (i.e. 96 30 minute bins). In such plots, repeating behavioural motifs with a period of 24 hours will appear as vertical columns, whilst repeating motifs with long or short periods will appear as slanted columns. Given the difference between morning and evening peak profiles, it was feasible that by iteratively changing the

double plot interval until similar peaks line up to form a column, the periodicity of the oscillator controlling that peak can be determined (Dowse, 2007):

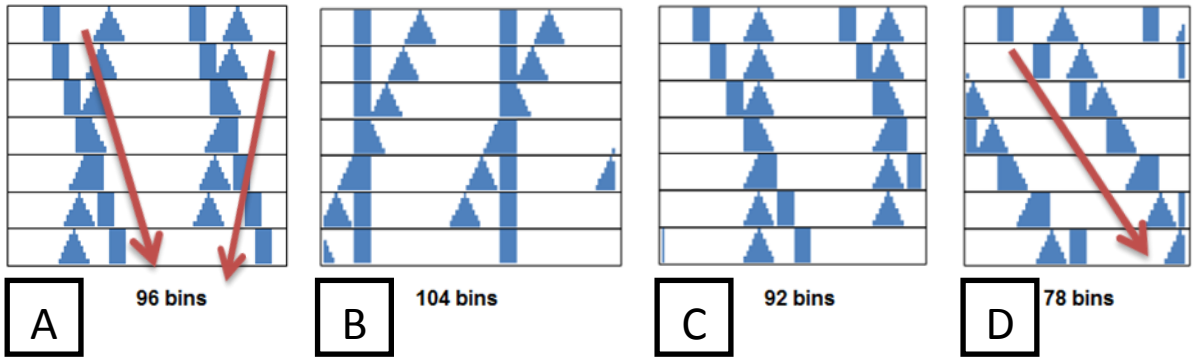


Figure 5-7 The use of visual inspection and an iterative change in double plot interval to identifying rhythms in complex data series. **A**: In this example, double plotting around the normal interval of 2 days (96 bins) shows the presence of both long (τ=26, rectangular blocks) and short (τ=23, triangular blocks) rhythmic components in the dataset. Changing the double plot interval until one component lines up to form a column allows the period of each component to be estimated with some accuracy (**B**, **C**). However, use of an inappropriate double plot interval (**D**) can obscure the presence of multiple oscillators, as they appear to be components of the same output (**red line**) with some degree of stochastic disruption.

Although such a technique was implemented in BeFly! and an attempt was made to elucidate the true periodicity of outcrossed *Po* files using this method, the results did not significantly add to the interpretations presented above.

5.3.2 Evidence for widespread LL rhythmicity

5.3.2.1 per⁰; timGAL4; and LL

During the course of this PhD, accumulating evidence suggested that the orthodox assumption that LL conditions induce behavioural arrhythmicity was a considerable over-simplification. Contemporaneously with this work, a colleague in the laboratory working on tidal rhythms serendipitously observed prominent ultradian rhythms in LL at 25°C in a number of fly lines in which the endogenous *per* gene had been replaced transgenically with the *ePer* homolog isolated from *Eurydice pulchra*, a cirolanid isopod that exhibits robust circatidal

swimming rhythms in constant conditions (Wilcockson and Zhang, 2008). Although this data initially excited speculation that the tidal clock might be linked to the *per* gene, re-examination of the data suggested that the most parsimonious explanation was that the combination of the alleles w^{1118}, per^{01} *and timGAL4* uncovered a latent LL rhythmicity in flies, as the *ePer* was unable to rescue the per^{01} arrhythmicity in DD conditions:

Table 5-1 LL rhythm profiles for flies carrying a transgenic *per* gene derived from *Eurydice pulchra*. **Chr1, 2, 3** refers to the genetic composition of that chromosome. The '**strong rhythm**' category includes flies which show strong ultradian rhythmicity. Strains which show predominantly rhythmic behaviour in LL are highlighted in blue, all other strains showed similar rhythmicity profiles to the w^{1118}, per^{01}; *+/+; +/+* controls (see also Figure 4-21 for comparison). Source data courtesy of L. Zhang.

Chr1	Chr2	Chr3	Strong rhythm	Complex rhythms	Arrhythmic
w^{1118}, per^{01}	*+/CyO*	+/+	0%	0%	100%
w^{1118}, per^{01}	+/+	+/+	1%	37%	62%
w^{1118}, per^{01}	*UAS-ePer-10/CyO*	+/+	20%	20%	60%
w^{1118}, per^{01}	*UAS-ePer-6/CyO*	+/+	25%	21%	54%
w^{1118}, per^{01}	*tim>GAL4/UAS-ePer-10*	+/+	56%	26%	18%
w^{1118}, per^{01}	*tim>GAL4/CyO*	+/+	62%	15%	23%
w^{1118}, per^{01}	*tim>GAL4/UAS-ePer-6*	+/+	65%	25%	11%
w^{1118}, per^{01}	*tim>GAL4/CyO*	+/+	67%	15%	19%
w^{1118}, per^{01}	*tim>GAL4/+*	+/+	73%	17%	9%

The LL rhythmicity observed was qualitatively similar to the ultradian LL rhythmicity of per^{01} observed by the author in the previous chapter:

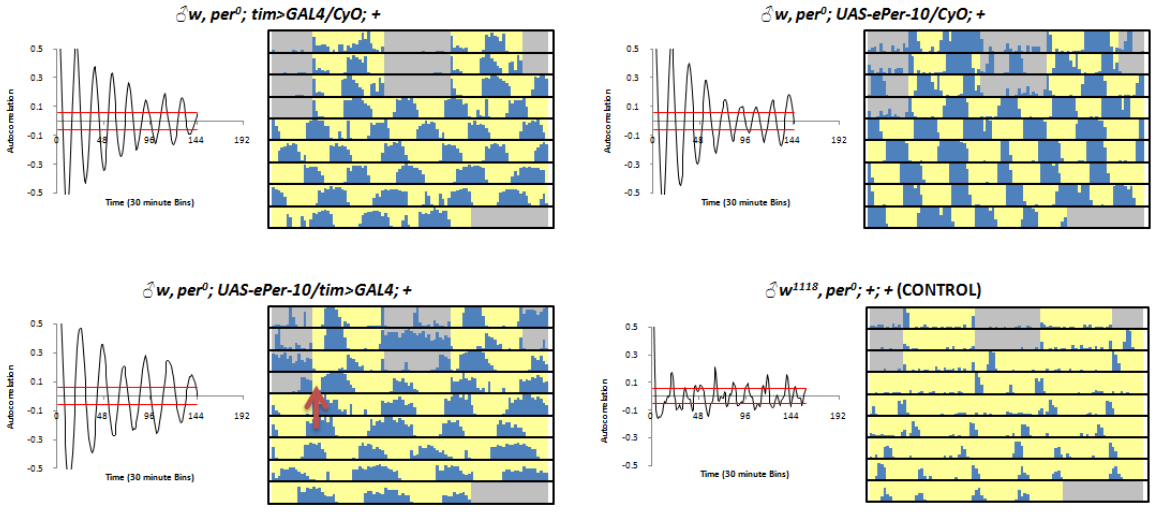


Figure 5-8 Examples of LL rhythms in a per^{01} background. The control genotype (**bottom right**) comes from an experiment run by the author, all other examples run by L. Zhang. All flies run by L. Zhang show

very high quality ultradian rhythms which do not segregate with any one transgene (*tim>GAL4* and *UAS-ePer-10*), but are all qualitatively better than the rhythms seen in the w^{1118}, per^{01} flies run by the author (see autocorrelograms), suggesting that the w^{+} genetic background used by L. Zhang is more permissive for per^{01} mediated LL rhythms. For further details see text.

The rhythmicity apparent in Figure 5-8 is surprising given that both the per^{01} mutation and LL conditions normally induce arrhythmic behaviour. Careful examination of the double plots also shows that activity patterns seem better co-ordinated in LL rather than LD conditions, particularly in the case of the fly illustrated in the lower left panel, which has clear difficulty entraining to LD conditions and may even be blind (note the lack of a startle response to 'lights on' shown by red arrow).

Although the presence of *tim>GAL4* seems to potentiate LL rhythmicity (increasing the proportion of flies within a genotype showing LL rhythms as shown in Table 5-1), a number of flies not carrying this transgene also showed very high quality LL rhythms (Figure 5-10). In an effort to clarify this situation, a further experiment was performed in which the LL behaviour of the w^{1118}, per^{01}; +; + driver line was assessed at 18°C, which had previously been shown to reduce the complexity of LL rhythms. Surprisingly the previous results could not be replicated:

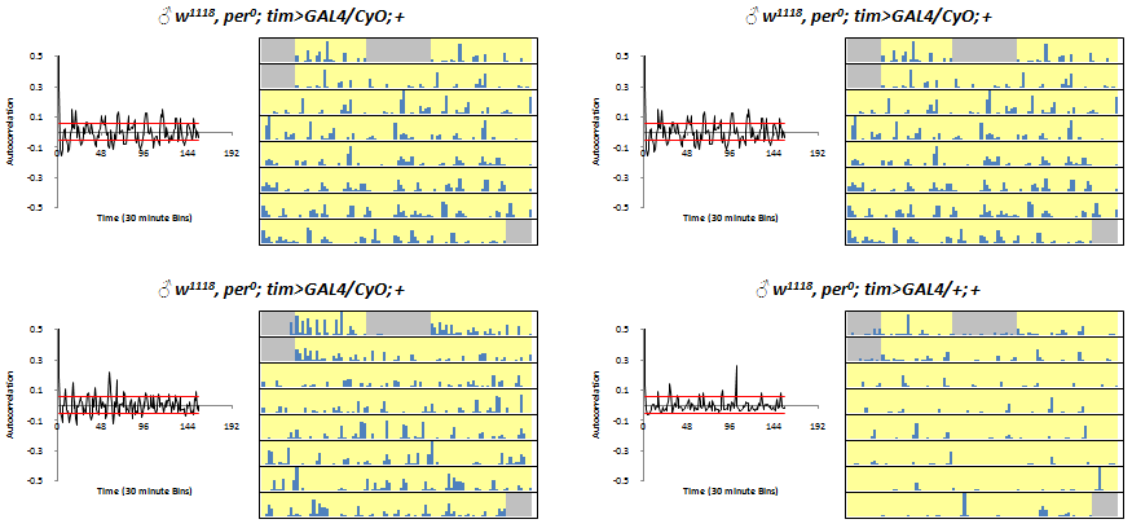


Figure 5-9 LL behaviour of transgenic flies in a per^{01} background at 18°C. Although flies show evidence of discrete, dispersed bouts of activity under LL conditions, the autocorrelation results show that these are the result of a much lower quality rhythmicity than that illustrated in Figure 5-8.

As the transgenic LL rhythmicity phenotype had previously been observed in three replicate experiments performed at 25°C, the relatively poor LL rhythmicity observed at 18°C must either be the result of the lower temperatures, or the genetic background. One possibility is that the *tim>GAL4* transgene insertion used in the 18°C experiment differed from that used by Lin Zhang, as our laboratory uses a number of different *tim>GAL4* insertion lines which are phenotypically hard to distinguish but which are known to elicit subtly different phenotypic effects (possibly due to differing transgene expressivity or insertional effects, K. Garner pers. comm.).

Given that GAL4 has been shown to have a direct neurotoxic effect (Haywood et al., 2002; Kramer and Staveley, 2003), it was decided to screen the activity of a number of GAL4 drivers to determine whether GAL4 expression or insertion effects were potentiating LL rhythmicity in the per^{01} background.

5.3.2.2 GAL4 stocks and LL

A number of lines expressing GAL4 in neurons implicated in the circadian clock were crossed into the same w^{1118}, per^{+}; +; + background and the behaviour of male heterozygous F1 progeny was assayed in DD conditions:

Table 5-2 DD periods of drivers expressing GAL4 in cells playing a role in the circadian clock.

Genotype	Period (hours)	SEM	n
♂ w^{1118}; +; GMR>GAL4/+	23.21	0.10	12
♂ w^{1118}; per>GAL4(yw BS line)/+; +	23.33	0.11	12
♂ w^{1118}; +; cry>GAL4/+	23.51	0.18	5
♂ w^{1118}; +; + (hs/exelixis)	23.66	0.16	10
♂ w^{1118}; tim>GAL4(A3 line)/+; +	23.73	0.08	12
♂ w^{1118}; pdf>GAL4/+; +	23.90	0.13	4
♂ w^{1118}; tim>GAL4(27 lethal)/+; +	24.05	0.09	9
♂ w^{1118}; tim>GAL4(yw line)/+; +	24.07	0.12	8
♂ w^{1118}; tim>GAL4(AR line)/+; +	24.10	0.13	15

The data in Table 5-2 show that although there is statistically significant variation in period between lines (ANOVA $p<0.0001$), the limited resolution of CLEAN when using small

sample sizes, combined with the natural variation in period length in different strains of *Drosophila* makes it difficult to ascertain which lines (if any) have truly 'aberrant' DD periods.

As all GAL4 lines tested retained functioning circadian clocks in DD conditions, their activity was further tested in LL conditions:

Table 5-3 LL activity of circadian GAL4 drivers. Although the *w*^{1118}*; +; +* control and both the *per>GAL4* and *cry>GAL4* driver lines show the expected degree of arrhythmicity typical of LL conditions (**blue highlight**), all other lines show unexpected rhythmicity under these conditions.

Genotype	Arrhythmic	Multiple Rhythms	Strong Rhythm	n
♂ *w*^{1118}*; tim>GAL4(27 lethal)/+; +*	0%	33%	67%	15
♂ *w*^{1118}*; tim>GAL4(yw line)/+; +*	7%	20%	73%	15
♂ *w*^{1118}*; tim>GAL4(AR line)/+; +*	13%	40%	47%	15
♂ *w*^{1118}*; tim>GAL4(A3 line)/+; +*	20%	60%	20%	15
♂ *w*^{1118}*; +; GMR>GAL4/+*	25%	31%	44%	16
♂ *w*^{1118}*; pdf>GAL4/+; +*	33%	33%	33%	6
♂ *w*^{1118}*; +; cry>GAL4/+*	50%	13%	38%	8
♂ *w*^{1118}*; per>GAL4(yw BS line)/+; +*	63%	25%	13%	16
♂ *w*^{1118}*; +; + (hs/exelixis)*	69%	6%	25%	16

The results were very surprising; a number of different insertions of *tim>GAL4* and *GMR>GAL4* elicited LL rhythmicity (closer inspection of the *pdf>GAL4* results suggested that due to the small number of surviving flies it was unclear whether this genotype was rhythmic in LL or not), contrary to expectation. Furthermore, unlike the previous experiment, the genetic background used contained a functioning copy of *per*, showing that the mutant *per*^{01} allele is dispensable for LL rhythms.

Both *GMR>GAL4* and *tim>GAL4* express in the compound eyes of the fly, suggesting that disrupting non-cell autonomous photic entrainment might have a similar effect in potentiating LL rhythmicity as disrupting the cell autonomous light input mediated by CRY in the *cry*^{b} mutant (although the resultant rhythms were ultradian rather circadian as in the case of *cry*^{b}). Examples of individual activity traces for these rhythmic genotypes are presented in the appendix (10.3).

5.4 Discussion

The data presented in this chapter flies in the face of the conventional assumption within the circadian literature that constant light elicits behavioural arrhythmicity in 'wild type' flies, a conclusion which is becoming more widely recognised by a number of groups (Poster session at SRBR 2008). In general LL rhythms are complex, leading to two competing hypotheses as to their origin; either that they result from oscillator desynchrony, or that they reflect an underlying ultradian rhythmicity upon which the circadian clock depends.

5.4.1 The use of novel analyses to describe complex rhythmicities

In an effort to discriminate between such hypotheses, a number of novel analyses were developed and applied to complex LL datasets. The results suggest that wavelet analysis shows promise as a novel technique for analysing complex behavioural records, as the time frequency scalogram accurately recapitulated the often complex rhythmicity of flies under LL conditions. This was particularly the case when the periodicity of a fly changed over time, as existing Fourier and autocorrelation techniques struggle to reflect such changes (as shown in Figure 5-4).

However, in particularly complex cases, such as when attempting to discriminate between unimodal ultradian and bimodal circadian rhythmicities, the current WAVECLOCK implementation of the continuous wavelet transformation proved little better than current autocorrelation analyses (Figure 5-5), though the discrete wavelet transformation may show greater promise in dissecting such overlapping features.

5.4.2 Neuronal damage induces LL rhythmicity

This study found that transgenic flies carrying *GMR>GAL4* or *tim>GAL4* (but most likely not *Pdf>GAL4*) showed LL rhythms (Table 5-3); indeed the *tim>GAL4* driver has been known for some time to cause 'unusual' behavioural phenotypes in LL (B. Collins, K.Garner pers. comm.). As both *GMR>GAL4* and *tim>GAL4* express the neurotoxic GAL4 transgene in the compound

eyes, this suggests that compromising the cell autonomous input might in some way potentiate LL rhythmicity. Indeed, links are beginning to emerge between eye development and the circadian clock; the development of both may be stimulated by the *eyeless* transcription factor (N. Glossop pers. comm.), whilst the circadian RNA-binding protein LARK has also been shown to play a role in eye development (Sofola et al., 2008).

Intriguingly the cell autonomous photoreceptor CRY has recently been demonstrated to be expressed in the compound eyes as well within a subset of clock neurons; together with the GAL4 data this raises the possibility that the LL rhythmicity seen in cry^b mutants is not solely the result of attenuation of the cell autonomous light sensing pathway, but also as a result of compromising light input mediated by the canonical phototransduction cascade.

Might the eyes therefore play a role in the clock as an integrator of rhythmicity? There is certainly evidence that mutants lacking large parts of the optic lobes show high proportions of complex rhythmicities even in LD conditions (Helfrich, 1986). Clearly this is a topic for further research; the logical next experiment would be to subject cry^b and GAL4 driver lines to optomotor experiments testing the integrity of the visual pathway, and indeed the role of cry^b in the eyes and dorsal cells is already a subject of research interest in our laboratory (E. Rosato, pers. comm.)

5.4.3 LL rhythmicity is not the result of experimental error

The widespread LL rhythmicity described in this chapter has not otherwise been widely reported in the field. This may well be because the tendency in the field is to describe behaviour in terms of group averages, rather than focussing on individual behaviour. Nevertheless, one must question the integrity of the results in this chapter. As has been shown, flies are sensitive to Zeitgebers other than light, and indeed the addition of temperature cycles to LL conditions bypasses the normal arrhythmicity seen in such conditions

(Yoshii et al., 2005). Might the equipment used in these experiments therefore generate entraining stimuli of some form?

Although several designs of light boxes are in use within our laboratory, those used for experiments in this chapter were fitted with a 'cool' LED light source; accurate temperature measurements using a TriKinetics environmental monitor suggested that the temperature differences between 'lights on' and 'lights off' in such boxes were smaller than 1°C, below the minimum level of 3°C required for temperature mediated entrainment (Yoshii et al., 2005).

The light intensity of the boxes was also assessed, and though there was found to be some degree of inter-box variation, the mean light intensity was of the order of 300 lux, considerably brighter than the 10 lux reported to induce arrythmicity in Canon-S flies (Yoshii et al., 2004). Furthermore, both in this chapter and in the previous chapter, 'wild type' control flies were found to be predominantly arrhythmic (e.g. the '♂ w^{1118}; +; + *(hs/exelixis)*' line in Table 5-3), suggesting that the equipment employed is sufficient to induce arrhythmic behaviour, and is not a contributing factor in LL rhythmicity.

5.5 Conclusion/further work

Recent microarray studies have shown that many genes' transcription is directly responsive to light (Wijnen et al., 2006), that light and temperature information is integrated by the clock (Boothroyd et al., 2007), and that light can have clock-independent effects on behaviour (Kempinger et al., 2009). Given that organisms have adapted to living at extreme latitudes in which there are annually exposed to periods of constant light, it is of great interest to more fully understand the response of the circadian clock to LL conditions.

Both this and the previous chapter suggest that there is still a great deal to be learned about the circadian clock, particularly its response to LL conditions. Attempting to identify further clock components by various methods is the central theme of the rest of this thesis.

6 A circadian meta-analysis

6.1 Introduction

The circadian clock model - though well understood - is plainly incomplete, as described in the thesis introduction and evidenced by data in the previous two chapters. Given that many of the core clock genes are transcribed in a rhythmic fashion, a number of groups attempted to identify novel components of the circadian clock by performing microarray experiments. To date there have been five microarray studies in *Drosophila* which have attempted to identify circadian cycles in the transcriptome (Ueda et al., 2002; McDonald and Rosbash, 2001; Ceriani et al., 2002; Lin et al., 2002; Claridge-Chang et al., 2001), each study using Affymetrix GeneChip *Drosophila* Genome 1.0 arrays, but interpreting data using differing statistical techniques and with different emphases. Consequently each study identified an independent list of genes judged to be regulated in a circadian manner.

Table 6-1 Summary of 5 circadian microarray studies. ADDER= amplification of double-stranded cDNA end restriction fragments, qRT-PCR = quantitative real time PCR, Northern = Northern blot. Adapted from Duffield (Duffield, 2003), additional detail regarding data processing strategies can be found in Keegan *et. al.* supplementary figure 1 (Keegan et al., 2007).

Study	Tissue studied	Timepoints	Timepoint interval	Experimental paradigm	Analysis type	Validation type	Genes found
Ueda *et al.*	Head	12	4	1 cycle LD, 1 cycle DD	Cross-correlation with cosine waves, Cluster analysis	qRT-PCR	115
McDonald and Rosbash	Head	6	4	1 cycle DD	Cross-correlation with cosine waves, Cluster analysis	ADDER	158
Claridge-Chang *et al.*	Head	12	4	1 cycle LD, 1 cycle DD	24-h Fourier component (LD and DD combined), Cluster analysis	*in situ* hybridization	134
Lin *et al.*	Head	24	4	3 cycles LD, 2 cycles DD	Autocorrelation analysis, Cluster analysis	qRT-PCR	22
Ceriani *et al.*	Head	12	4	1 cycle LD, 1 cycle DD	Cosine wave fitting (COSOPT)	Northern	116
	Body					Northern	173

At the time of writing, reviews of circadian microarray data had focused on the contradictory nature of these studies by drawing attention to the limited overlap between datasets (Etter and Ramaswami, 2002; Lin et al., 2002), all too often concluding that microarray technology produces too much experimental noise to provide replicable circadian data. However, a recent technical study has shown that Affymetrix microarrays used by different laboratories performing the same experiments can produce results with a correlation as high as 0.91 (Bammler et al., 2005), suggesting that a meta-analysis of the existing work would not only be technically valid, but might also generate valuable insights into circadian transcription not evident when considering single studies alone.

A meta-analysis of this sort was first attempted by Lin *et al.*(Lin et al., 2002), however the analysis in this chapter seeks to go further by incorporating data generated in Ceriani *et al.*'s most recent study (Ceriani et al., 2002), by using the latest annotations, and by presenting more information for each gene, allowing investigators to decide whether their gene of interest shows a circadian pattern of expression using all the data available. Additionally, this study sought to present mean peak expression times for genes found to cycle in more than one study, a resource of considerable use when determining the possible function of a gene within the context of the clock mechanism.

6.2 Methods

6.2.1 Construction of the spreadsheet

Results from the existing circadian microarray studies (detailed in Table 6-1) were imported into Microsoft Excel 2003 worksheets, observing where possible the inherent problems Excel has in handling large genomic datasets as described by Zeeberg *et al.* (Zeeberg et al., 2004).

In order that gene expression data from each study could be compiled, the unique gene identifier used in each study (e.g. CG number or Affymetrix probe ID) was converted to its matching FlyBase annotation number (using FlyBase's version 4 annotations) using batch processing tools available at www.FlyBase.org. Results generated by the batch tool were inspected and, where necessary, annotated manually using both FlyBase and the most recent Affymetrix GeneChip Drosophila Genome 1.0 Array annotations (available at http://www.affymetrix.com/support/technical/byproduct.affx?product=fly).

Once associated with their correct FlyBase identifier, the genes identified in each study were collated to form a comprehensive list, from which duplicate entries were removed using the ASAP toolset for Excel available from http://www.asap-utilities.com. The peak phase of gene expression was rounded to the nearest hour in studies that reported phase more accurately. Excel 'LOOKUP' functions were then used to present all data available for each gene on one summary worksheet. If a gene was identified as cycling in more than one study, the peak phases of expression were averaged using the circular statistics tools available in Oriana v2.0 (http://www.kovcomp.co.uk/oriana/).

In order to increase the utility of the dataset, FlyBase tools were used to annotate each gene with its known symbols and synonyms and cytological location.

6.2.2 Characterising genes identified in the spreadsheet

To facilitate further characterisation of genes of interest, protein domain homology predictions were retrieved from FlyBase and appended to the database, and a number of these predictions were verified using Interpro (http://www.ebi.ac.uk/InterProScan/). Protein interaction information was retrieved from the FlyGRID database using links from FlyBase. Gene ontology enrichment analysis was performed using tools available at the GOEAST website (http://omicslab.genetics.ac.cn/GOEAST/) using the default parameters.

6.2.3 Exelixis FLP/FRT mediated deletion

An effort was made determine the function of a number of candidate clock genes by generating lines hemizygous for those genes. Due to the high efficiency of FLP mediated recombination between Flip Recombinase Target sites (FRTs) not naturally found in the *Drosophila* genome sequence (Golic and Golic, 1996), a number of public and private initiatives have sought to saturate the *Drosophila* genome with *P* elements carrying FRT sites, allowing researchers to generate molecularly defined deletions between FRT sites inserted in the genome:

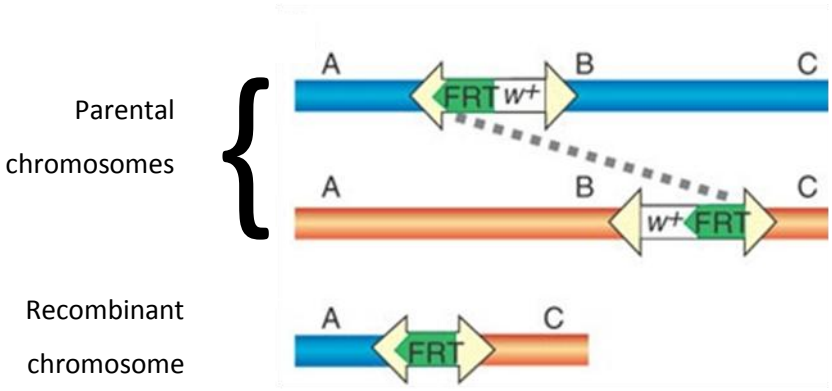


Figure 6-1 FLP mediated recombination between FRT sites to create deletions. Lines carrying appropriately orientated FRT sites flanking the gene of interest (B) are brought together in a background in which the flippase enzyme (FLP) can be inducibly expressed using a heat shock promoter. Induction of flippase catalyses recombination between FRT sites, creating a recombinant chromosome in which the gene of interest has been deleted. This recombinant chromosome can then be recovered using balancers. Figure adapted from Parks et al. (Parks et al., 2004).

Deletions made during this project used the stocks, crossing scheme and methods developed by Exelixis, detailed in Parks *et al.* (Parks et al., 2004). All deletions resulted in the loss of the *white*$^{+}$ transgene (as shown in

Figure 6-1), and could therefore be followed by changes in eye colour. Deletions were verified by using the hybrid PCR strategy shown in Figure 6-2, which relies on recombination between FRT sites to bring sequences in the flanking *P* elements together on the same chromosome, allowing the deletion to be verified using primers specific for the identity and orientation of

the *P* elements used to generate the deletion. Genetic material to perform PCR was recovered from fly wings as detailed in the general methods. For PCR conditions and additional details see Parks *et al.* (Parks et al., 2004); the primers used to verify each deletion used in this study are reprinted in Table 6-2.

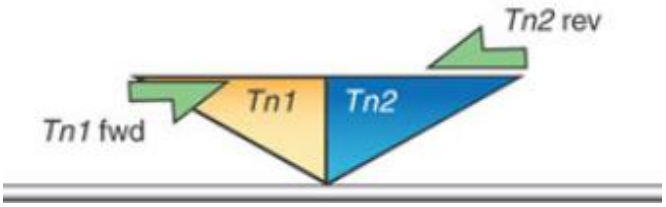


Figure 6-2 Validating FRT mediated deletions. Following FRT mediated deletion (see Figure 6-1) sequences in transposon 1 (**Tn1**) and 2 (**Tn2**) are brought together in close proximity on the same chromosome, such that performing a hybrid PCR using the unique forward (**Tn1 fwd**) and reverse (**Tn2 rev**) primers for each *P* element generates a fragment of known length that can be visualised on a gel. .

Table 6-2 Primers used to verify both DrosDel (DD) and Exelixis (Ex) deletions using a hybrid PCR strategy. For details of PCR conditions, see general methods.

Deficiency	Primer1	Primer2	Fragment size (bp)
Df(DD)CG10553	*CTTCTCACCCGCAGCAAC*	*CAATCATATCGCTGTCTCACTCA*	200
Df(DD)CG11891	*CTTCTCACCCGCAGCAAC*	*CAATCATATCGCTGTCTCACTCA*	200
Df(DD)CG4784	*CACCCGCTGGTACTTCGT*	*CAATCATATCGCTGTCTCACTCA*	212
Df(DD)CG5156	*AGTGGGCAAGCAAAGCAC*	*CAATCATATCGCTGTCTCACTCA*	228
Df(DD)Ugt35b	*TGCTGCTCATCCTGTCCA*	*CAATCATATCGCTGTCTCACTCA*	187
Df(DD)CG17386	*CAAAACCAAGAGAACTTCGGA*	*CAATCATATCGCTGTCTCACTCA*	293
Df(Ex)CG10553	*AATGATTCGCAGTGGAAGGCT*	*TGCATTTGCCTTTCGCCTTAT*	1700
Df(Ex)CG11891	*GACGCATGATTATCTTTTACGTGAC*	*AATGATTCGCAGTGGAAGGCT*	1800
Df(Ex)CG4784	*TGCATTTGCCTTTCGCCTTAT*	*AATGATTCGCAGTGGAAGGCT*	1700
Df(Ex)CG5156	*AATGATTCGCAGTGGAAGGCT*	*GACGCATGATTATCTTTTACGTGAC*	1800
Df(Ex)Ugt35b	*GACGCATGATTATCTTTTACGTGAC*	*AATGATTCGCAGTGGAAGGCT*	1800

6.2.3.1 Copyright/Technology Transfer notes

The stocks used to generate FRT mediated deletions carry technology transfer notices which must be published in any work using the stocks. The appropriate notices are found in the appendix (10.4).

6.3 Results

6.3.1 A comprehensive spreadsheet database

The results from the five *Drosophila* circadian microarray studies were re-annotated using the latest FlyBase gene numbers, then collated to form a spreadsheet database, an excerpt of which is shown below. The complete version is found in appendix D2.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AB	AE
1	FlyBase ID	KEY	DD Phase	Cycling level in *Clk/CS*	Relative levels *(Clk/JRK/CS)*	LD Phase	DD Phase	LD Phase clock mutant *(per[01])*	LD phase (head)	DD Phase (head)	Levels in Clk[JRK] (head)	Ceriani body LD phase	LD Phase	Levels in Clock mutant (Clk[JRK])	Ueda LDDD LD	Ueda LDDD DD	mRNA levels in *Clk[JRK]* mutant	DD Phase	mRNA levels in *Clk[JRK]* mutant	DD Phase	Studies identifying gene as cycling in LD	Mean peak phase of expression in LD	Studies identifying gene as cycling in DD	Mean peak phase of expression in DD	Gene Symbol	Synonyms	Cytological Location	
2	Identifier		Rosbash			Lin			Ceriani				Ueda							Claridge-Chang	Summary							
199	FBgn0031495	11599																		11:00	1	11:00	1	11:00	*CG17257*		23C5;	
200	FBgn0027571	11305																7:00	Not Changed		0		1	7:00	*CG3523*	*Fatty acid synth*	23C5--D1;	
201	FBgn0011648	10444							15:00												1	15:00	0		*Mad*	*CG12399, mac*	23D3;	
202	FBgn0014906	10445							15:00				15:00	0.63							2	15:00	0		*CG3488*	*c24, anon-23D*	23D4;	
203	FBgn0031515	10865											12:00	0.5							1	12:00	0		*CG9664*		23E1;	
204	FBgn0021967	10866											22:00	0.92							1	22:00	0		*Pdsw*	*CG8844, l(2)k1*	23F3;	
205	FBgn0014396	10012	14:00	low		12:00	8:00		16:00	17:00		18:00			15:00	15:00	-0.19			15:00	5	15:13	5	14:14	*tim*	*CG3234, rit, Rii*	23F6;	
206	FBgn0022073	10555										23:00									1	23:00	0		*Thor*	*CG8846, CG8*	23F6;	
207	FBgn0031559	10227		high																	0		0		*CG3513*		24B1;	
208	FBgn0031558	10464							15:00		Up										1	15:00	0		*CG16704*	*BcDNA:RH311*	24B1;	
209	FBgn0031560	10465							10:00		Up										1	10:00	0		*CG16713*		24B2--3;	
210	FBgn0031562	10424				12:00		8:00													1	12:00	0		*CG3604*	*BcDNA:LP040*	24B3;	
211	FBgn0031561	10577										9:00									1	9:00	0		*CG16712*	*BcDNA:RH054*	24B3;	
212	FBgn0000256	10466							19:00												1	19:00	0		*capu*	*CG3399, CG15*	24C8--9;	

Figure 6-3 Excerpt from the spreadsheet database of cycling gene expression, sorted by cytological location (far right column). Red triangles at the top of each column contain explanatory comments for that column. Columns containing data pertaining to DD rhythmicity are coloured **grey**, to LD rhythmicity **yellow**, and to clock mutants **orange**. The clock gene *tim* is highlighted; note the concordance between studies in assessing its peak phase of *tim* expression. The spreadsheet has been visualised using Excel's split screen function (**blue bar** below row 2) to facilitate interpretation of results, see Figure 6-4 for an alternate visualisation.

By presenting the data in Excel, a user can take advantage of the powerful data sorting and filtering options available in this software:

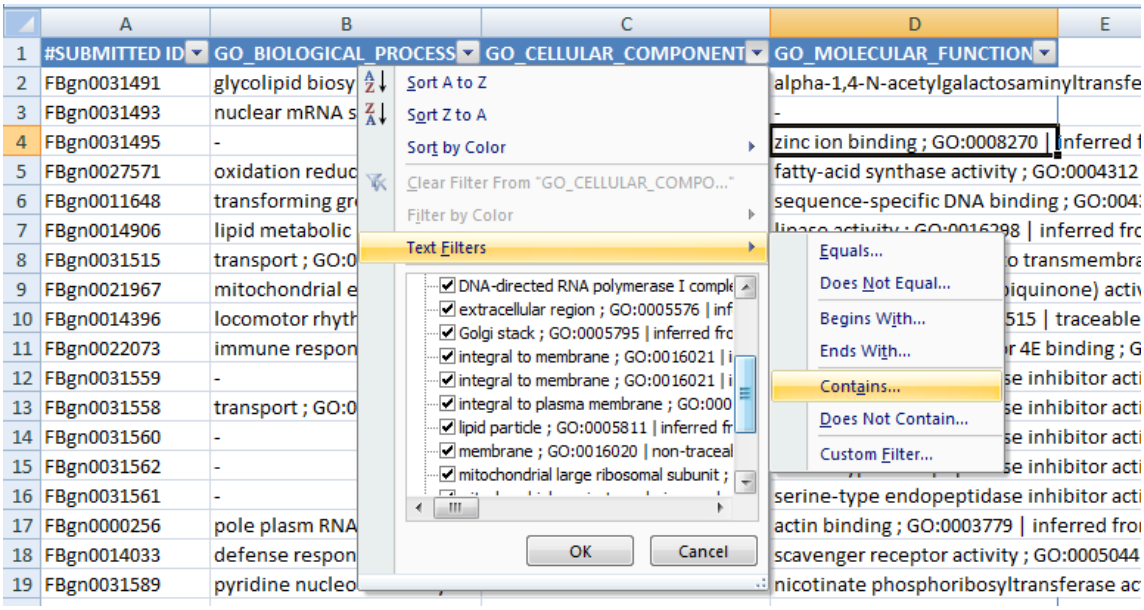


Figure 6-4 Alternate visualisation of results using Excel 2007's table tools to filter the spreadsheet database. In this example genes are filtered using gene ontology criteria. A user can easily import new data into the spreadsheet by submitting the FlyBase gene numbers (Figure 6-3 column A) using FlyBase's 'Batch Download' tool, ensuring the latest annotation is always available when examining genes.

By using FlyBase gene numbers as the primary key identifying genes, it is easy to retrieve up-to-date annotation for genes of interest using the FlyBase 'Batch Download' tool, which can then be fed into further analysis. One powerful tool is gene ontology enrichment analysis:

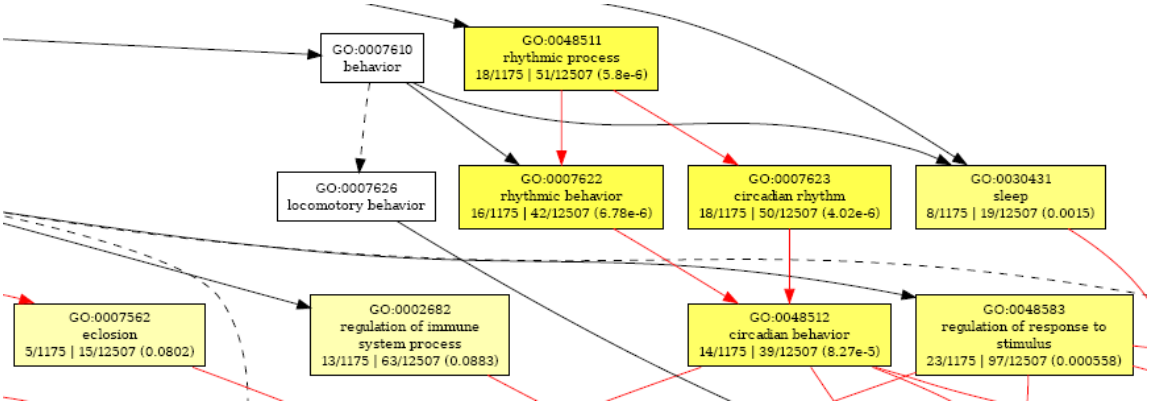


Figure 6-5 GOEAST Gene ontology enrichment analysis showing the biological process of cyclically expressed or clock regulated genes. Significantly enriched gene ontology (GO) terms are marked

yellow, the degree of colour saturation of each node reflecting the significance of enrichment of the corresponding GO term. Non-significant GO terms within the hierarchical tree are shown as white boxes. Edges represent the hierarchical connections between GO terms; red edges show relationships between two significantly enriched GO terms, black solid edges stand for relationships between enriched and un-enriched terms, and black dashed edges stand for relationships between two un-enriched GO terms.

By using FlyBase gene numbers, the spreadsheet database facilitates the export of data for further analysis using tools such as GOEAST, a powerful ontology analysis tool that can detect and visualise categories of genes enriched in lists of genes, as shown in Figure 6-5.

Encouragingly GOEAST analysis of all the rhythmically expressed genes identified by the five microarray studies (as well as those whose levels are dependent upon the circadian clock) shows not only highly – as one might expect - significant enrichment for known circadian genes, but also enrichment for genes involved in eclosion and sleep - processes known to be partially regulated by the circadian clock.

6.3.2 Bioinformatics

By collating genes identified in published studies it was possible to determine how consistently a given gene was identified as being rhythmically expressed:

Table 6-3 Results of collating the five existing microarray studies. Table includes both those genes identified as having a circadian transcription profile, and those which show significantly different levels of expression in clock mutants relative to wild type flies, and are presumably therefore under clock control (though pleiotropic effects of clock mutations cannot be excluded). '**DD**' refers to genes only identified under free running experimental conditions, '**LD**' refers to genes identified under entrainment conditions (LD12:12). The LD total is greater, containing both genes regulated in a light dependent, as well as circadian, fashion.

Number of studies identifying a gene	5	4	3	2	1	Grand Total
DD	7	6	19	51	612	695
LD	3	10	25	91	914	1043

It is clear from Table 6-3 that inter-study concordance is very low; only 13 genes are identified by four or more of the original microarray studies under DD conditions, including the

canonical clock genes *tim*, *per*, *Clk*, *vrille* and *Slob* (see Table 6-5 for further details). Due to the enrichment of core clock genes in this category, and on the basis of similar results in the mouse (in which comparison of tissue specific microarray studies revealed that the canonical clock genes were enriched in all cases: Delaunay and Laudet, 2002), the eight novel cycling genes identified in this study are likely to play a role in the clock mechanism itself, rather than mediating the output of the central clock. These genes were therefore designated 'candidate canonical clock genes' (CCCGs).

Table 6-3 also reveals the surprising result that almost 400 genes are detected as being rhythmically expressed in LD but not DD conditions. These genes are likely to be transcribed or repressed as a direct response to light, showing that the light dependent transcription pathway plays almost as large a role in the transcriptome as clock regulated transcription. Consideration of the overlap between these two modes of transcription allows genes to be divided into a number of categories:

Table 6-4 Overlap between circadian and light dependent transcription.

Meta analysis		
Studies identifying gene as cycling in DD	Studies identifying gene as cycling in LD	Total
1	0	355
	1	208
	2	43
	3	6
2	0	6
	1	20
	2	19
	3	5
	4	1
3	1	2
	2	6
	3	6
	4	5
4	3	6
5	4	4
	5	3
Grand Total		695

Mostly genes regulated by light

Unreliable circadian clock genes

Reliable circadian clock genes

Candidate canonical clock genes

As *bona fide* circadian genes should show transcriptional cycles in both LD and DD, Table 6-4 allows more than half the genes identified by only a single circadian microarray study

to be excluded as type 1 errors (false positives). Many genes may be identified as showing rhythmic expressed due to the inherent inaccuracy of the microarray experimental platform.

Although there is still relatively limited knowledge as to the genes and mechanisms governing clock output, it was decided to focus attention on the CCCGs as disrupting such genes should affect all circadian outputs, making them the best candidates for further analysis.

An effort was made to further characterise the CCCGs by assessing their position within the genome and likely function. Table 6-5 shows that the CCCGs *CG11891* and *CG10553* lie in close proximity on the chromosome, have a similar peak phase of expression, and contain the same protein domain of unknown function (DUF227). Closer examination shows that the genome region proximal to these genes is enriched for the DUF227 domain, suggesting these genes may have arisen by duplication, may be functionally redundant, and as such are unlikely to play canonical roles in the clock mechanisms. These genes also lie very close to the known clock gene *takeout* (So et al., 2000), as shown in Figure 6-6. Many of the genes in the region illustrated in Figure 6-6 show circadian patterns of expression peaking during the early night, suggesting this region of the chromosome is co-ordinately regulated, an observation also made by McDonald and Rosbash (McDonald and Rosbash, 2001).

Table 6-5 Features of candidate canonical clock controlled genes. Table details the name of each gene and the number of studies identifying it as showing a circadian expression pattern. Additionally the mean peak phase of gene expression, cytological location and protein domains associated with that gene (which indicate possible function) are listed. Known clock genes are marked with an *. Mean peak expression phase is a circular mean of the results generated in each study, thus is expressed to the nearest minute, despite each study individually having a temporal resolution of 4 hours.

Symbol	Studies identifying gene	Mean peak expression phase	Cytogenetic location	Protein domains
tim*	5	14:14	23F6;	TIMELESS protein
vri*	5	12:19	25D4--5;	Basic-leucine zipper (bZIP) transcription factor
Slob*	5	12:27	28C1;	Protein kinase
Ugt35b	5	02:10	86D5;	UDP-glucuronosyl/UDP-glucosyltransferase
CG5798	5	12:21	93C1;	Peptidase C19, ubiquitin carboxyl-terminal hydrolase 2
CG11891	5	21:56	96C8;	Domain of unknown function 227
CG10553	5	19:28	96D1;	Domain of unknown function 227
CG5156	4	03:14	21F3;	Acyltransferase 3
per*	4	12:09	3B1--2;	PAS
CG17386	4	19:20	51A2;	RNA-binding protein Lupus La
CG15093	4	03:58	55F2;	6-phosphogluconate dehydrogenase
Clk*	4	00:09	66A12;	Nuclear translocator, Basic HLH dimerization, HLH DNA-binding; PAC motif
CG4784	4	03:44	72E2;	Insect cuticle protein

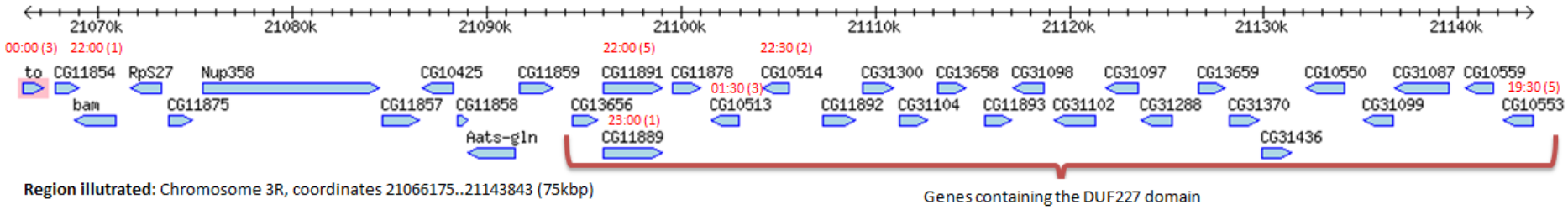


Figure 6-6 75kb of the *takeout* genomic region. The mean phase of peak gene expression is displayed above genes identified by this meta-analysis as a red superscript including the number of studies identifying the gene as 'circadian' in parentheses. All genes to the right of *CG11859* contain the DUF227 domain.

Of the remaining genes, *CG4784* has significant homology with insect cuticle proteins, which are likely to play structural rather than functional roles, suggesting *CG4784* may act as a scaffolding molecule for other circadian interactions, in a similar manner to the role hypothesised for *dAxin* (K. Garner, pers. comm.). A further possibility considered was that the deposition of insect cuticle is regulated in a circadian manner, a hypothesis which has recently been confirmed (Ito et al., 2008). Perhaps most surprising is their observation that whilst the peripheral rhythms in cuticle deposition depend on functioning copies of the core clock genes *per*, *tim*, *cyc* and *Clk*, they do not require *cry* (contrary to previous work suggesting that CRY might play a central role in peripheral clocks: Collins et al., 2006), despite *CG4784* being a major binding partner for CRY in pulldown assays (Rosato, E. unpublished observations).

The *CG17386* RNA binding protein is of interest given that other RNA binding proteins such as LARK have been shown to play a role in the clock (Sofola et al., 2008), whilst a number of core clock genes might be regulated by miRNAs (Yang et al., 2008; Pegoraro and Tauber, 2008). *CG5156's* role as an acyltransferase may also be of functional significance given the recent discovery of CLK's histone acetlyation activity (Doi et al., 2006), and the demonstration of epigenetic changes accompanying rhythmic binding of clock proteins to promoters (Ripperger and Schibler, 2006).

Given the intriguing functions of the majority of the novel CCCGs, an effort was made to further link them to the clock mechanism by attempting to identify putative binding partners using the BIOGRID repository:

Table 6-6 Possible interaction partners of the CCCGs. Possible interactions as collated by BioGrid (Stark et al., 2006). Phenotypic enhancement or suppression assays are considered to be more reliable than yeast two hybrid interactions. *Ugt35b, CG11891, CG10553, CG17386, CG4784* and *Clk* did not have any listed interaction partners when this work was performed despite known associations for at least some of these genes in the literature, revealing the limitations of this approach.

Gene	Phenotypic enhancement/suppression	Yeast two hybrid
tim*	*per*, tim**	*CG4778*
vri*	*dpp, ea, Pdp1ε**	
Slob*		*CG12426, CG9025*
CG5798		*CG3713, CG10510, CG10882, ATP synthase-β, CG1962, Mcm5, Set, Rack1, Rho1*
CG5156		*Bic*
per*	*tim*, dco*, GAP1, per**	
CG15093		*CG18128*

Table 6-6 shows the putative interaction partners for the CCCGs known at the time this work was performed. *CG5798* has many listed interaction partners, suggesting either that the gene represents a network hub, or that this profligate binding does not reflect its cellular function, but instead shows the limitations of yeast two hybrid technology. Of the interactions listed, the association with *Mcm5* (*minichromosome maintenance 5*), a gene involved in chromosome condensation and the creation and resolution of structures important for DNA replication, is of greatest interest given the putative links between the cell cycle and circadian rhythms (Tauber et al., 2004; Chan et al., 2003). Other associations gave no obvious clue as to the function of their associated CCCG.

Given the intriguing diversity of functional domains and binding partners within the CCCGs, it was decided to elucidate the function of the genes using reverse genetics techniques to generate small deletions uncovering the CCCGs.

6.3.3 Verification of DrosDel lines hemizygous for CCCGs

As a preliminary step, large deletions which uncovered the CCCGs (created by the DrosDel project for complementation mapping purposes) were ordered from the (now defunct) Szeged stock center. These lines were analysed whilst crosses were performed to

generate smaller, more targeted deletions using the Exelixis collection of FRT insertions. Behavioural studies are particularly sensitive to genetic background; the DrosDel and Exelixis collections are therefore ideal for studying behaviour as these collections originate from isogenic lines, and are not subject to modifier effects which may otherwise preclude the detection of subtle phenotypic differences.

Table 6-7 Hemizygous deletion strains used in this study. Table shows deletion names, size of the deletion and the number of genes deleted in each case. Partial gene deletions are indicated in parentheses. All hemizygous strains are white eyed due to the presence of the w^{1118} allele.

	DrosDel deletions			Exelixis deletions		
Target CCCG	Deletion	Size (bp)	Genes deleted	Deletion	Size (bp)	Genes deleted
CG10553	*Df(3R)ED6230*	518419	80+1	*FDD-0047199*	50864	17+2
CG11891	*Df(3R)ED10948*	70912	24+1	*FDD-0317950*	26129	10+2
CG4784	*Df(3L)ED220*	324193	88+2	*FDD-0153023*	26068	7+2
CG5156	*Df(2L)ED104*	301338	40+1	*FDD-0009197*	159438	19+1
Ugt35b / CG6649	*Df(3R)ED5506*	287750	17+0	*FDD-0272917*	10324	3+2

Due to the complex naming of deletions, deficiency strains used in this study are referred to in the format 'Df' denoting 'deficiency', 'Ex' or 'DD' to denote whether they are created in the Exelixis or DrosDel backgrounds, and carry the name of the CCCG uncovered by the deletion. Therefore *Df(3R)ED6230* is referred to as *Df(DD)CG10553*, and its matching Exelixis deletion *FDD-0047199* as *Df(Ex)CG10553*. In all cases deletion lines were not homozygous viable, therefore the names refer to hemizygous strains.

The DrosDel deletions received from Szeged were validated using a hybrid PCR strategy as detailed in Figure 6-2 using the primers listed in Table 6-2:

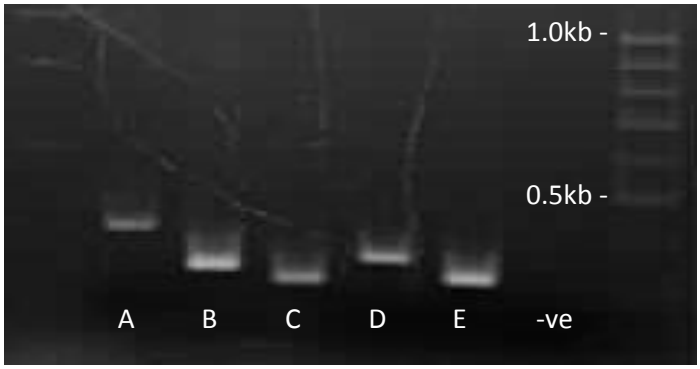


Figure 6-7 Hybrid PCR verifying creation of deficiencies using DrosDel stocks. Each lane contains a band at around 200bp, confirming the presence of the deletion (see **Table 6-2** for details). **A**: *Df(DD)CG17386*, **B**: *Df(DD)CG5156*, **C**: *Df(DD)Ugt35b*, **D**: an additional deletion of *Df(DD)CG5156*, **E**: *Df(DD)CG4784*, **-ve**: no DNA PCR control.

6.3.4 Activity profiles of DrosDel lines hemizygous for CCCGs

Having confirmed the validity of the hemizygous DrosDel deletions, their behaviour was assessed by backcrossing into the DrosDel w^{1118}*; +; +* background to limit any confounding effects arising from alleles carried on the balancer chromosomes used to maintain the homozygous lethal deficiency stocks (notably the e1 allele of ebony on the TM6B balancer, which affects circadian locomotor behaviour as a homozygote: Newby and Jackson, 1991). Flies were entrained to LD conditions and their activity recorded as detailed in the general methods:

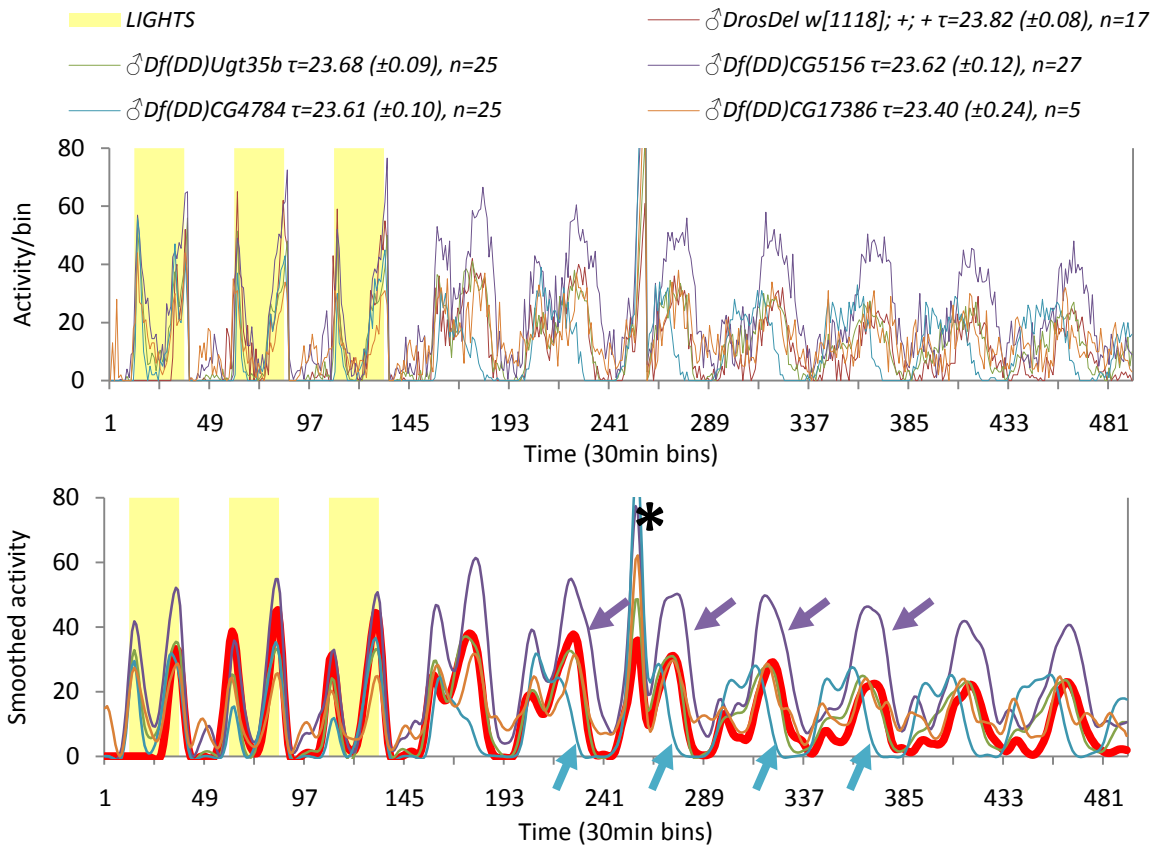


Figure 6-8 DD activity profiles for DrosDel deficiency stocks. **Top**: median activity profiles for the 4 deletions and isogenic control line. Periods for each genotype are listed in the legend. **Bottom**: Data as above but following application of Butterworth smoothing filter. The control genotype is highlighted in red. Note the activity profile change in *Df(DD)CG4784* (**pale blue arrows**) in which the activity offset point falls progressively earlier than in the w^{1118} control. Note also the higher activity levels in *Df(DD)CG5156* (**purple arrows**). Qualitative consideration of the activity of this strain suggests that it might have a long period, but such a difference is not apparent in the algorithmic determination of period length (**legend**). Note that the DAM system data recorder suffered a fault during the experiment at the point indicated (*), causing abnormally high activity counts. As a precaution the periodicity of each genotype was assessed using only beam crossing events recorded after this point.

Although Figure 6-8 suggests that there might be phase and activity profile changes between *Df(DD)CG4784*, *Df(DD)CG5156* and the isogenic control strain, further analysis using a normalised activity metric (Figure 6-9) suggested the differences in *Df(DD)CG5156's* activity arose as a result of its higher activity levels, causing it to appear to have a prolonged period of evening activity (purple arrows):

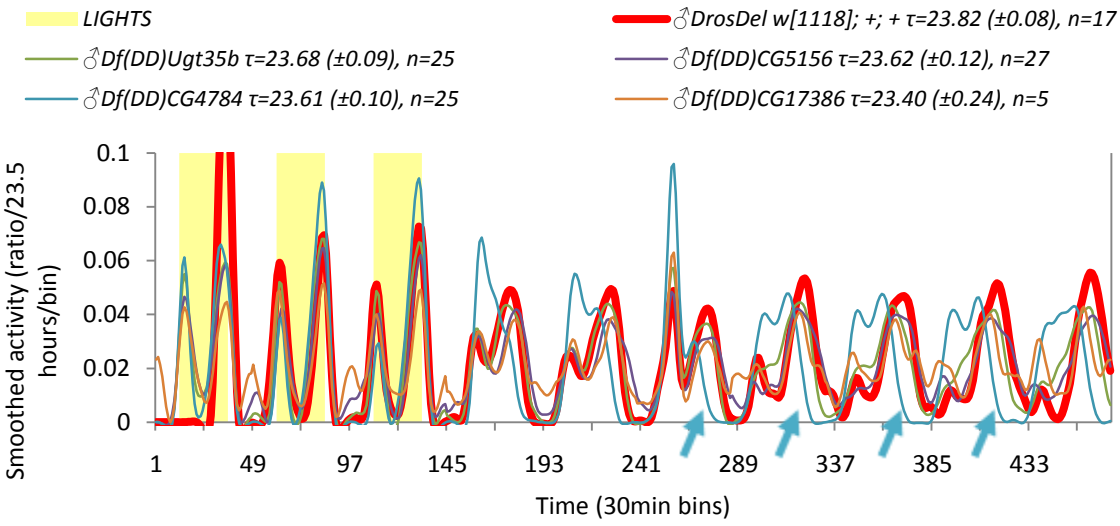


Figure 6-9 Normalised activity profiles for DrosDel hemizygotes in DD. Note that after normalising activity to a % per cycle value, the activity profile differences between *Df(DD)CG5156* and the isogenic w^{1118}*; +; +* control are no longer evident, however the early activity onset phenotype of the *Df(DD)CG4784* strain remains (**pale blue arrows**).

Despite the *Df(DD)CG4784* and isogenic control lines having almost indistinguishable periods and activity onsets following normalisation, the activity offset for the deficiency line still appeared to be moving forwards, suggesting that the morning and evening oscillators in this strain might be running at different rates. Due to the subtle nature of this mutation the experiment was repeated with a larger sample size:

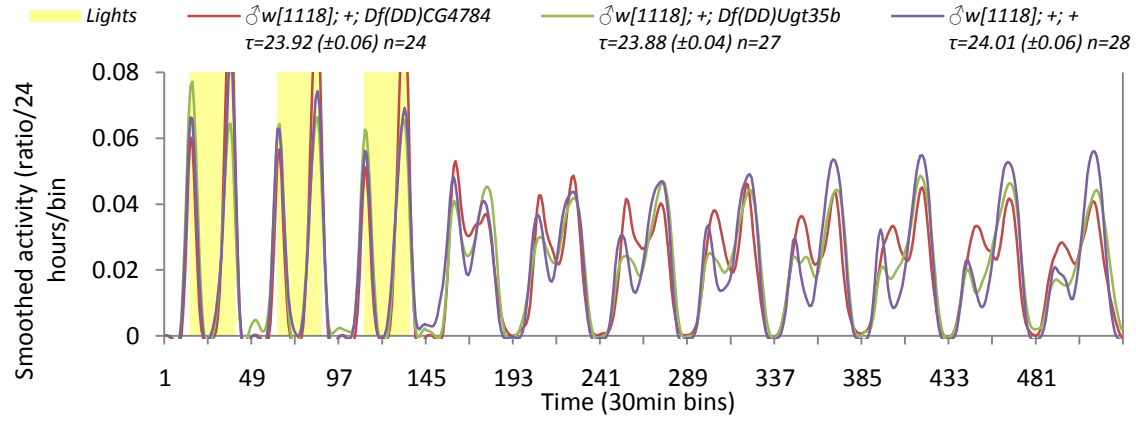


Figure 6-10 Repeat of the *Df(DD)CG4784* DD activity experiment. Sample sizes and estimates of periodicity are shown in the legend. This experiment finds no evidence for any difference between the activity profiles of the deficiency strains and the isogenic control, in contrast to the data shown in Figure 6-9.

The failure to repeat the previous result suggested that it might have arisen as an artefact of the relatively small sample size tested in the first experiment. As the CCCGs displayed no differences from controls in DD conditions, the lines were further tested in LL conditions (in part due to the known association between *CG4784* and the cell autonomous photoreceptor CRY):

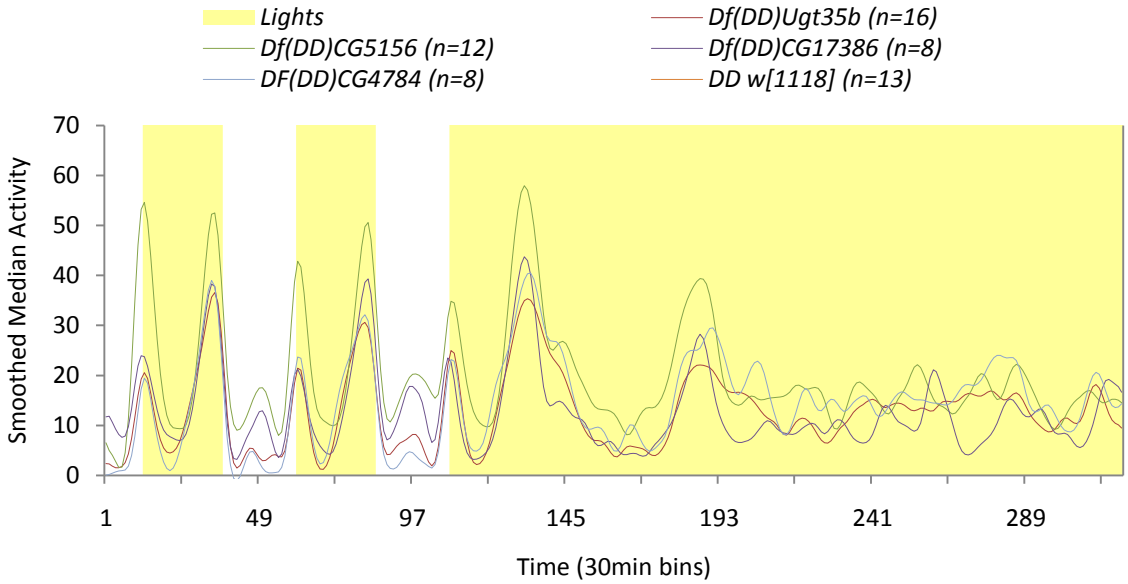


Figure 6-11 LL activity of DrosDel hemizygous deficiency lines. All lines rapidly exhibit arrhythmicity in LL conditions.

Again there was no evidence that lines hemizygous for CCCGs were significantly different from controls. Although these preliminary experiments did not provide any evidence that the novel CCCGs might play a role in the canonical clock mechanism, as many *Drosophila* genes are haplosufficient, it was hoped that generating smaller deletions using the Exelixis stocks might create homozygous viable deficiencies more likely to uncover the function of CCCGs.

6.3.5 Activity profiles for CCCG deletions generated using Exelixis FRT sites

After following the appropriate 5 generation crossing scheme, potentially recombinant white eyed flies were tested to verify whether they carried a deficiency using the hybrid PCR strategy described previously:

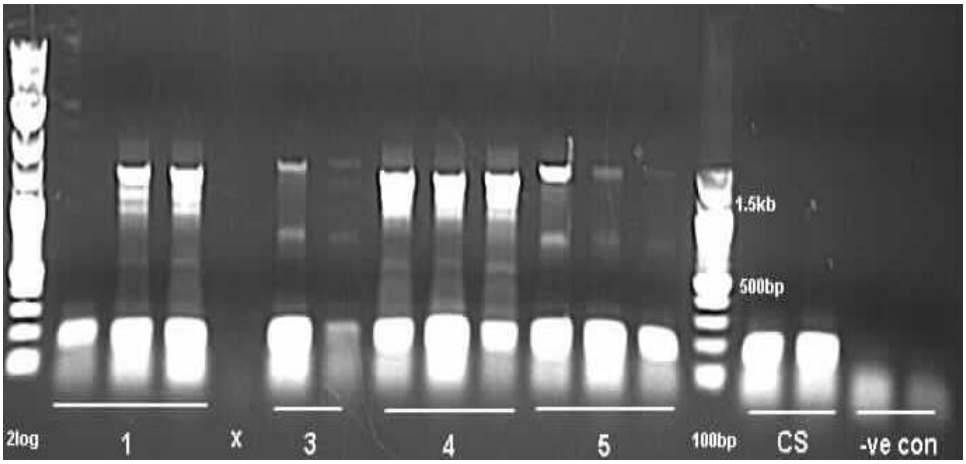


Figure 6-12 Hybrid PCR verification of newly manufactured Exelixis deletion lines. White eyed flies were tested by PCR to verify the presence of the appropriate deletion using primers listed in Table 6-2. Several flies were tested for each genotype (indicated by white bars). Stocks were established using flies in which the deficiency had been confirmed. Genotypes are coded such that **1**: *Df(Ex)CG10553*, **3**: *Df(Ex)CG11891*, **4**: *Df(Ex)CG4784*, **5**: *Df(Ex)CG5156*, **CS**: w^{1118} control strain, **-ve con**: no DNA (negative control), **x** = blank lane.

This experiment failed to generate two deletions; in one case no white eyed recombinant flies were observed, whilst in the other white eyed flies were recovered, but PCR failed to confirm the presence of a deletion. 'False positive' deletions have been observed by other groups, and remain a subject of investigation by the DrosDel consortium, therefore the recovery of a false positive deletion is unlikely to be the result of stock contamination.

Unfortunately none of these smaller deficiencies generated proved to be homozygous viable. However, as these deletions had been newly synthesised, it was possible that they might still exhibit phenotypes different from the larger DrosDel deletions as a result of the DrosDel stocks collecting genetic modifiers over time (though this would be limited to regions not balanced in the isogenic stocks). However, preliminary experiments suggested that the hemizygous Exelixis mutants did not show significant differences from control genotypes in either DD or LL (data not shown).

It was therefore decided to repeat experiments by crossing hemizygous Exelixis deletion lines into a background in which the circadian clock mechanism had been sensitised

using the alleles w^{1118} (affecting non cell autonomous sensing of light), per^{0} (affecting the core clock transcriptional oscillator) and cry^{b} (affecting the cell autonomous detection of light). The activity of virgin F1 flies was tested in both DD and LL conditions:

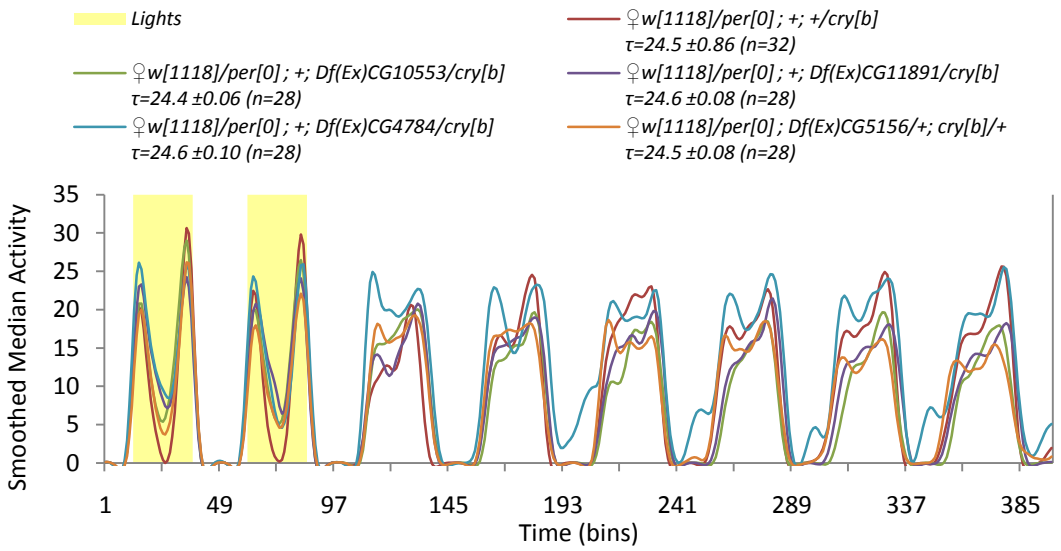


Figure 6-13 Activity profiles for hemizygous Exelixis deletion mutants in a sensitised w^{1118},per^{0}/+; +; cry^{b}/+ background under DD condtions. Despite flies carrying several alleles which might sensitise the clock to further disruption by hemizygous deletion strains, there is no evidence for any effect, either in terms of activity profile (in both entraining and free running conditions), or free running period (τ).

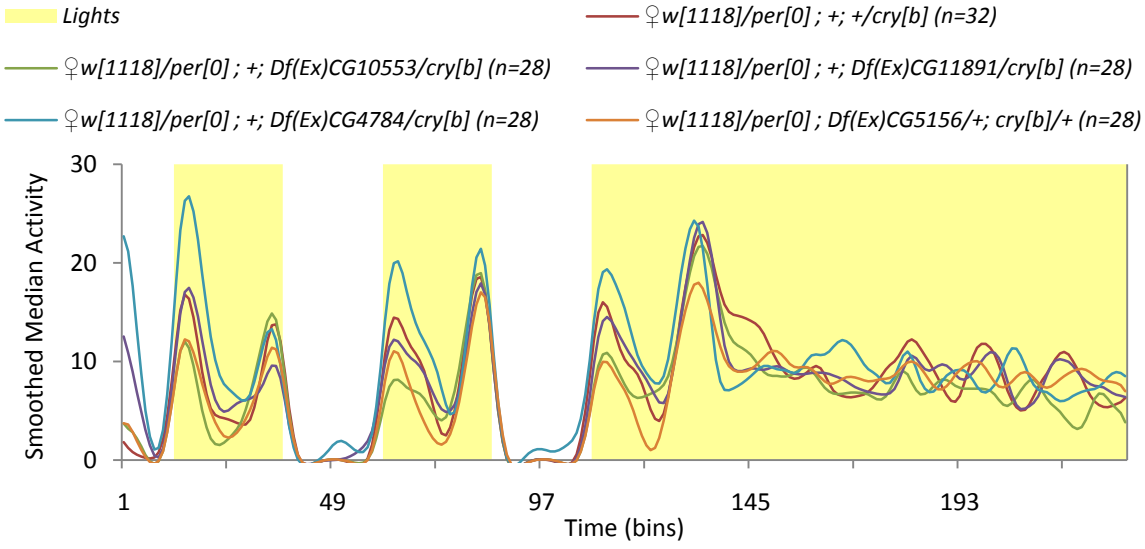


Figure 6-14 Activity profiles for hemizygous Exelixis deletion mutants in a sensitised per^{0}/+; +; cry^{b}/+ background under LL condtions. Qualitative examination of both the group average actograms presented above and individual activity traces (data not shown) provides no evidence for any effect.

In contrast to the previous work using hemizygous DrosDel deletions, activity experiments using Exelixis deficiencies were recorded at high resolution, allowing sleep metrics to be calculated for each line whilst under entraining conditions (see chapter 8 for an in depth discussion of sleep and the development of sleep analysis software). As sleep metrics show some variability in young flies, the sleep patterns of both young and old flies were recorded.

Table 6-8 and Table 6-9 show the probabilities that sensitised deficiency lines differ from the control strain. As expected, immature flies assessed at the start of the experiment had not settled into their adult sleep patterns, and consequently showed greater variability, reflected in the discovery of statistically significant differences between lines. However, mature flies showed almost no significant differences in 11 different sleep metrics; the significant differences in total activity and the correlated mean waking activity metric for the w^{1118}, per^{0}/+ ; +; Df(Ex)CG4784/cry^{b} strain were shown to be the result of two outlier flies showing abnormally high activity levels (possibly as a result of grooming behaviours).

Table 6-8 Probability that sleep metrics (averaged over 2 days LD) for 2 day old sensitised hemizygous lines were significantly different from *w^{1118}, per^{0} ; +; cry^{b}/+* sensitised background controls. n~28 for each genotype. Probabilities calculated using a post hoc conservative Dunnett test against the control genotype.

Genotype (young flies)	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Bout Length (Mins/Dark)	Mean Bout Length (Mins/Light)	Mean Waking Activity/Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/12hrs)	Total Sleep in Light (Mins/12hrs)
w^{1118}, per^{0}/+ ; Df(Ex)CG5156/+; cry^{b}/+	0.985	0.996	0.990	0.964	0.939	0.031	0.989	0.611	0.543	0.819	0.060
w^{1118}, per^{0}/+ ; +; Df(Ex)CG11891/cry^{b}	0.061	0.965	0.445	0.638	0.084	0.000	0.953	0.871	0.119	0.316	0.001
w^{1118}, per^{0}/+ ; +; Df(Ex)CG4784/cry^{b}	0.000	0.990	0.812	0.114	0.002	0.000	1.000	0.002	0.000	0.000	0.000
w^{1118}, per^{0}/+ ; +; Df(Ex)CG10553/cry^{b}	0.357	0.988	0.567	0.254	0.307	0.000	1.000	0.653	0.069	0.329	0.012
w^{1118}; +; +	0.194	0.000	0.998	0.179	0.081	0.044	0.011	0.400	0.733	0.996	0.682

Table 6-9 Probability that sleep metrics (averaged over 2 days LD) for 14 day old sensitised hemizygous lines were significantly different from *w^{1118}, per^{0} ; +; cry^{b}/+* sensitised background controls. Probabilities calculated using a post hoc conservative Dunnett test against the control genotype. Closer examination of results showed that the higher mean and total activity levels for *w^{1118}, per^{0} ; +; Df(Ex)CG4784/cry^{b}* flies was the result of 2 flies showing a single bin of abnormally high activity (~400 counts in 30 mins), possibly as a result of repetitive grooming behaviour whilst underneath the infrared recording beams. The sensitising alleles alone have a limited effect on flies; *w^{1118}, per^{0} ; +; cry^{b}/+ and w^{1118}, +; +* controls differ only in the number of sleep bouts occurring during the light part of the day.

Genotype (older flies)	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Bout Length (Mins/Dark)	Mean Bout Length (Mins/Light)	Mean Waking Activity/Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/12hrs)	Total Sleep in Light (Mins/12hrs)
w^{1118}, per^{0} ; Df(Ex)CG5156/+; cry^{b}/+	1.000	0.369	0.829	0.998	0.947	1.000	0.992	0.172	0.774	0.846	0.963
w^{1118}, per^{0} ; +; Df(Ex)CG11891/cry^{b}	0.505	1.000	0.224	1.000	0.996	1.000	1.000	0.106	0.996	0.997	0.998
w^{1118}, per^{0} ; +; Df(Ex)CG4784/cry^{b}	0.075	0.952	0.782	0.922	0.520	0.246	0.002	0.143	0.000	0.416	0.963
w^{1118}, per^{0} ; +; Df(Ex)CG10553/cry^{b}	1.000	0.996	0.597	1.000	1.000	0.990	0.996	0.102	0.999	0.598	0.988
w^{1118}; +; +	0.921	0.000	0.273	0.917	0.730	1.000	0.705	0.052	0.981	0.940	0.073

Although hemizygous deletion strains exhibited no activity differences relative to controls in LD, DD and LL, the possibility remained that the role of the CCCGs might be elucidated by considering circadian outputs other than locomotor activity, as lesions in the clock mechanism can have distinct effects on outputs; classically the mutant *ebony* was identified due to its effects on activity rhythms rather than its more subtle effects on eclosion rhythms (Newby and Jackson, 1991).

The eclosion of hemizygous mutants was therefore assessed manually as described in the general methods:

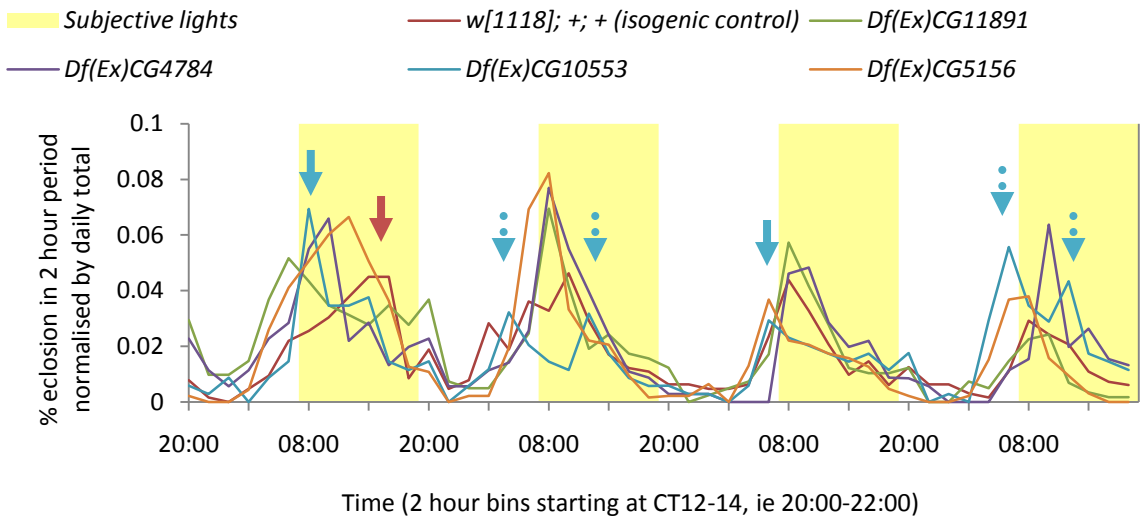


Figure 6-15 Eclosion rhythms for hemizygous deletion lines in DD conditions (note lighting data represents subjective lighting). Approximates 1000 flies were collected for each genotype shown. The *Df(Ex)CG10553* strain shows some evidence of short eclosion period (**blue arrows**), but the double peaks on the 2nd and 4th days (**hashed blue arrows**) makes this difficult to confirm. Note also the delayed eclosion peak on the first day for the w^{1118}; +; + control (**red arrow**), possibly due to the smaller numbers of flies emerging on the 1st day (data not shown).

Visual inspection of Figure 6-15 shows that all hemizygous strains retain rhythmic eclosion behaviour. The short datasets of eclosion studies are always problematic for spectral analysis of periodicity, therefore the eclosion rhythm for each genotype was assessed using two different spectral tools; CLEAN and MESA (which was integrated into the BeFly! package specifically for this purpose):

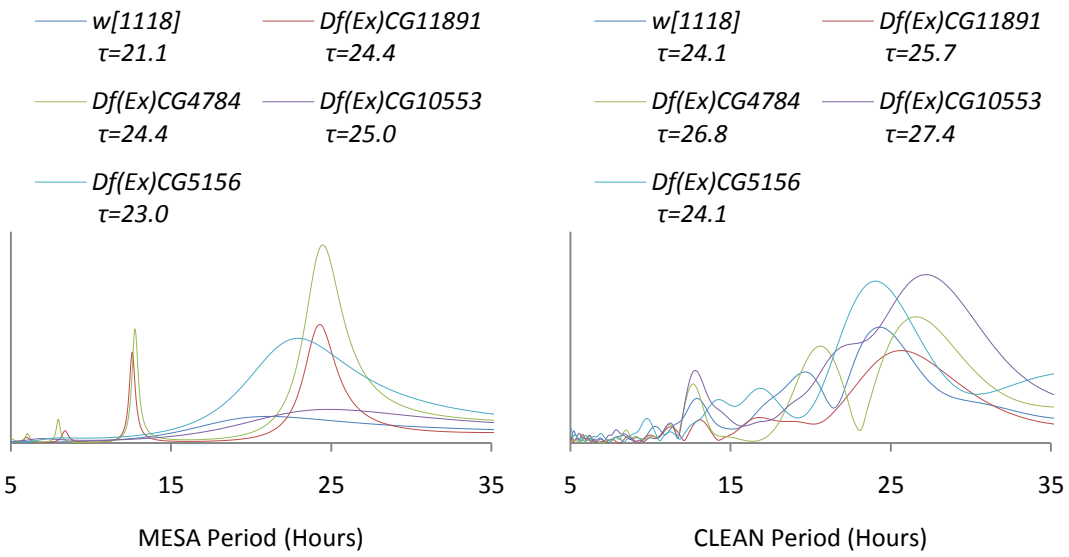


Figure 6-16 Spectral estimates of eclosion rhythm periods. CLEAN and MESA, despite both using Fourier analysis, suggest different periods for the genotypes due to the short, low resolution dataset.

Figure 6-16 shows considerable differences between the MESA and CLEAN analyses, though both assign a long period to *Df(Ex)CG10533*, suggesting that both algorithms are sensitive to the double peaks of eclosion this strain displays (hashed lines in Figure 6-15). In conclusion therefore there is no compelling evidence that eclosion rhythms are dramatically perturbed in these hemizygous strains.

6.3.6 RNAi knockdown of CCCGs

Circadian clock genes have been shown to be expressed in a variety of tissues including specific neurons within the brain (Park and Hall, 1998), glial cells (such as is the case for ebony - Suh and Jackson, 2007), and in peripheral tissues (Glaser and Stanewsky, 2005). As such an advantage of the CCCG deletion approach is that no assumption must be made as to the likely expression of each gene, as levels are depleted systemically. However, the failure to generate homozygous viable deficiency lines, and the apparent haplosufficiency of the CCGs even in clock sensitised backgrounds, suggested an alternative approach to determining gene function was required.

Using the yeast derived GAL4/UAS system (Duffy, 2002) and RNAi inducing hairpin constructs, the expression of genes can be knocked down in specific tissues. A preliminary experiment focusing on the gene *Epac* (data not shown) showed that generating *in situ* localisations and creating RNAi knockdown constructs for all the CCCGs identified in this study would be impractical, however as this work was being performed, RNAi stocks become available at the Vienna Drosophila RNAi Centre (VDRC) which allowed the further screening of the CCCGs using targeted knockdown (Dietzl et al., 2007). Hairpin constructs were expressed using the strong *tim>GAL4* driver to reduce CCCG levels in the known clock cells of the brain (as well as within the eyes):

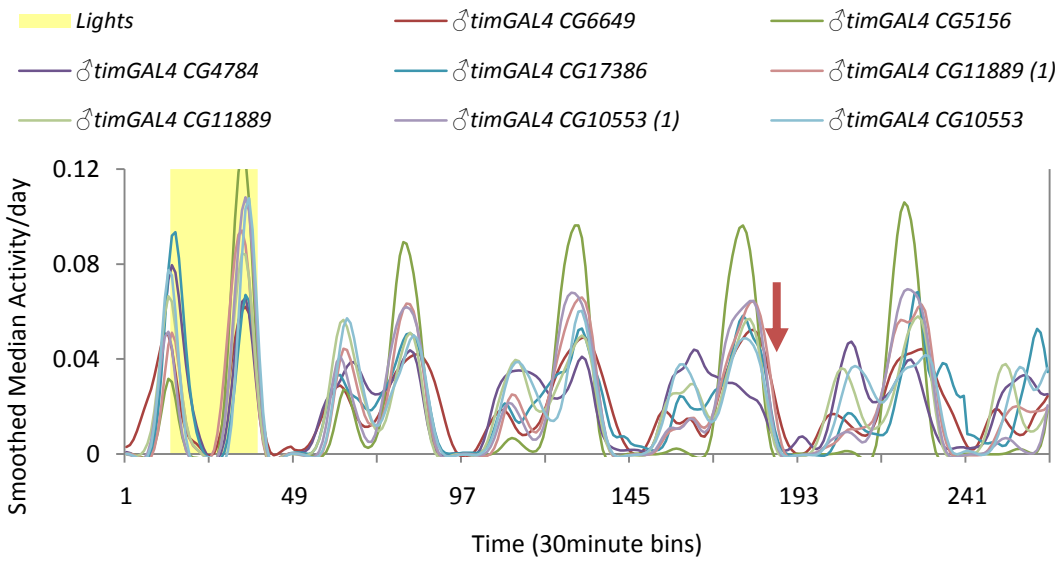


Figure 6-17 Normalised, smoothed median activity profiles for isogenic VDRC RNAi lines knocking down CCCGs in DD conditions. For several genes multiple RNAi constructs were tested; in such cases the number of additional lines is indicated in parentheses in the legend. After a number of days in free running conditions, all knockdown lines retain some degree of bimodal activity, and activity offsets appear to remain in phase (**red arrow**).

The results in Figure 6-17 were subject to CLEAN analysis to determine the periodicity of the lines:

Table 6-10 Period of isogenic VDRC RNAi lines knocking down CCCGs in DD conditions. Different insertions of the same construct (indicated by parentheses) give very similar periods in both of the cases tested. Significance tested using a conservative *post hoc* honest HSD test relative to the *timGAL4 w^{1118}* control.

Genotype	Period	SEM	Rhythmic	Arrhythmic	Dead/Hyperactive	p
♂***timGAL4 CG10553***	23.8	0.4	13	2	1	1
♂***timGAL4 CG10553 (1)***	23.9	0.31	14	0	2	1
♂***timGAL4 CG17386***	24.9	0.54	4	4	8	0
♂***timGAL4 CG4784***	24.1	0.28	7	2	7	1
♂***timGAL4 CG5156***	23.7	0.39	14	0	2	1
♂***timGAL4 CG6649***	23.9	0.43	15	0	1	1
♂***timGAL4 CG11889***	23.9	0.25	14	2	0	1
♂***timGAL4 CG11889 (1)***	24	0.41	13	2	1	1
♂***timGAL4 w^{1118} control***	23.8	0.27	15	0	1	-

CLEAN analysis showed that knockdown of *CG17386* in *tim* expressing neurons lead to one half of surviving flies showing a significant increase in circadian period, the other half being arrhythmic. Only 8 of the 16 *CG17386* knockdown flies survived the experiment, suggesting that *CG17386* knockdown might compromise viability, and that the circadian effects might arise secondarily as a result of general reduced health.

6.4 Discussion

6.4.1 A high quality dataset for identifying circadian genes

The collation of five microarray experiments to form a new, comprehensive spreadsheet meta-analysis detailing the circadian expression of genes has opened a number of new avenues for circadian research, some of which have since been pursued by a number of independent groups (Wijnen et al., 2006; Keegan et al., 2007). This chapter therefore fits into a now established body of meta-analyses using various techniques to reliably identify candidate cycling genes, making it easier than ever for researchers to determine whether their

genes of interest cycle in expression, and whether it has a putative role in the central clock mechanism or output.

Of this body of work, perhaps the most comprehensive re-analysis of data was performed by Keegan *et al.*, reanalysing the raw data generated by each microarray study using multiple rhythm detecting algorithms (which are themselves still subject to considerable research: Zhao et al., 2009). In addition, since this work was performed, considerable progress has been made in elucidating the role of light dependent transcription (Wijnen et al., 2006), temperature induced cyclic expression (Boothroyd et al., 2007) and the interactions between sleep and the circadian clock (Cirelli et al., 2005b) – all factors which may confuse the identification of *bona fide* circadian genes.

Despite these recent advances, the spreadsheet presented in this PhD contains a number of salient features that differentiate it from published databases and reviews. Though it may not represent a complete list of cycling genes, the aim of this work was to identify new canonical clock genes, and collating existing data (rather than re-analysing the source data) proved sufficient to address this question. A number of conclusions could be drawn from the dataset generated:

6.4.2 Light dependent transcription

The program of light dependent transcription is revealed to be almost as large as the program of true circadian expression in DD conditions (see Table 6-3). This suggests that although the circadian clock provides an adaptive advantage by allowing the anticipatory transcription of genes, light dependent transcription also plays an important role in allowing the fly to adjust to its environment. Such transcription may play a role in synchronising the circadian clock to environmental conditions independent of, but parallel to, the clock dependent entrainment mechanism mediated by CRY protein – a subject of considerable research interest in our laboratory.

The overlap between these methods of regulation shown in Table 6-4 has been further elucidated by Wijnen *et al.,* who were able to show that two of the CCCGs identified by this study (*CG5798* and *CG17386*) show a pattern of expression that is not only regulated by the circadian clock, but also directly responsive to light in a process requiring the phospholipase C component of the visual transduction pathway, *norpA* (Wijnen et al., 2006). *CG5798's* transcription is promoted by light, whilst *CG17386's* is repressed; presumably this dual regulation causes them to exhibit particularly robust rhythmicity in LD conditions. Wijnen *et al.* suggest this mode of regulation allows direct responses to environmental light intensity – information that is not encoded by the circadian clock.

It was therefore particularly disappointing that homozygous Exelixis deletions could not be recovered for either of the CCCGs showing dual regulation, given that hemizygous lines proved uninformative. RNAi knockdown of the light-repressed *CG17386* transcript in a sensitised genetic background appeared to have a period lengthening effect in DD (when the levels of this gene are already low – see Table 6-5). Knockdown of this gene also increased the proportion of arrhythmic flies recovered (), a result consistent with it playing a role in the central clock mechanism. Unfortunately this RNAi line was lost by both our laboratory and the VDRC stock center before further work could be carried out to confirm this observation (possibly due to the insertion causing a significant degree of position effect lethality, reflected in part by the large number of *CG17386* knockdown flies dying early in the RNAi experiment, see

Table 6-10).

Further work to confirm this result might use novel *CG17386* knockdown stocks generated by the National Institute of Genetics (NIG) consortium (http://www.shigen.nig.ac.jp/fly/nigfly) to reassess behaviour in both LL and DD conditions. In addition, phase shift experiments might also elucidate the relationship dually regulated CCCGs and light (although small scale preliminary experiments produced limited evidence for any effect - data not shown). Further experiments in both cry^b and *norpA* mutant backgrounds might also help elucidate the contribution of these genes to clock dependent and independent light responses.

6.4.3 The failure to uncover circadian phenotypes in CCCG hemizygote deficiencies and RNAi knockdown lines

A key premise of this chapter was that genes robustly detected as having cycling expression levels were likely to represent canonical clock genes (CCCGs), in part because many known canonical genes fell into this category. As such, disrupting the levels of such genes would be expected to affect all the circadian outputs (in contrast with genes downstream of the clock such as ebony which are more likely to affect only a single circadian output - Newby and Jackson, 1991). Although this study goes further than any other published meta analysis in assaying rhythmically expressed genes for clock-related phenotypic effects, considering locomotor behaviour in LD, DD and LL conditions, examining eclosion and sleep rhythms and testing lines in sensitised backgrounds, *CG17386* knockdown aside there was only limited evidence for any replicable phenotypic effect.

The main reason for this apparent failure might be that the degree of knockdown achieved by these experiments is insufficient to uncover phenotypes; it has been shown that knockdown of the core clock gene *per* has a period lengthening effect proportional to the

degree of *per* knockdown, a result that can be used to test the relative 'strength' of different GAL4 drivers and the efficiency of UAS-hairpin constructs (Martinek and Young, 2000).

Indeed, it has been shown that circadian RNAi screens are sensitive to hairpin construct design; only 15 of the 25 RNAi constructs designed against the core clock gene *tim* by Matsumoto *et al.* elicited behavioural arrhythmicity, whilst extending the screen to knockdown 133 candidate cycling transcripts revealed only 5 showing rhythmic activity defects – *cwo, E23, Iswi, prod* and *CG5273* (Matsumoto et al., 2007).

It is entirely possible therefore that the CCCGs identified in this study have well buffered transcriptional mechanisms, and are therefore haplosufficient in a circadian context. Such genes would not be amenable to study using hemizygous deletions, and would only show phenotypes in RNAi screens if the hairpin constructs very efficiently reduced gene expression (as might be assessed using RT-PCR).

As this study only used single RNAi constructs (in some cases testing different insertions to guard against position effects, see Figure 6-17), further experiments might seek to use constructs designed by the NIG consortium or the improved 'KK' 2nd generation lines (which do not suffer from insertion effects) available from the VDRC. Another solution might be to repeat experiments in the presence of a *UAS-dicer2* enhancer to increase the degree of gene knockdown. An alternative might be to use *P* element insertions that disrupt the CCCGs, however such lines are not kept in isogenic backgrounds and therefore would not be appropriate for behavioural studies without time consuming backcrossing (Rosato and Kyriacou, 2006). Since this work was performed there have been rapid advances in developing other reverse genetics techniques available to disrupt gene function, notably the refining of

recombination systems that allow accurate single gene deletions to be created in *Drosophila* (Venken and Bellen, 2007), which might also be employed.

Clearly therefore there is potential for further work on CCCG function by reducing gene expression levels still further, one caveat being that functionally redundant genes (such *CG11891* and *CG10553* in Figure 6-6) would not be amenable for study even in homozygous null backgrounds. However, given that no other known clock gene in *Drosophila* has a redundant copy (though functional degeneracy does occur), the most parsimonious explanation is that *CG11891* and *CG10553* arose through gene duplication, and are likely to be non functional.

6.4.3.1 Alternative roles for CCCGs

The alternative hypothesis to that advanced above is that the CCCGs identified by this study do not play canonical roles in the clock mechanism, but instead act in clock output processes. Indeed, around 10% of proteins identified in the murine hepatic proteome have been demonstrated to cycle at the transcriptional level (Reddy et al., 2006), whilst a growing body of literature shows that key components of, for example, the visual (Gorska-Andrzejak et al., 2009) and olfactory systems (Saifullah and Page, 2009) in *Drosophila* show cyclical expression. Such cycling expression should not be regarded as an artefact, as the peak phases of gene expression occur throughout the day (Table 6-5), suggesting that the circadian network is particularly deeply rooted as a result of its early evolution.

Given the identity of the CCCGs revealed in this study, the most likely candidate for a clock output gene is *CG4784* (an insect cuticle protein) owing to the recent discovery of rhythms in cuticle deposition (Ito et al., 2008). However, one might question this, given that *CG4784* is a major binding partner for CRY (E. Rosato, pers. comm.), a gene not required for the cuticle deposition rhythm.

6.4.4 Gene clustering – coordinate replication or enhancer duplication?

It is clear from the spreadsheet database that in a number of cases cytologically close genes show similar patterns of transcription (as shown in Figure 6-6), a subject of continuing research interest in many systems (Cohen et al., 2000).

Although this effect has also been noted in previous microarray studies (Ueda et al., 2002; McDonald and Rosbash, 2001), and requires formal statistical verification, data from this and the other published meta-analyses should prove fertile ground for studies investigating the functional significance inherent in spatial clustering of circadian genes within specific regions of the chromosome such as that performed by Ueda and colleagues (Ueda et al., 2005). Although several regions can be identified in the database as showing coordinate regulation, the region close to the known clock gene *takeout* illustrated in Figure 6-6 remains one of the best examples (also shown in McDonald and Rosbash, 2001). This leads one to question whether transcriptional regulation might have any inherent efficiency advantages over post-translational regulation processes, a question which this study cannot illuminate.

Whether coordinate regulation occurs as a result of 'leaky' transcription, duplication of enhancer motifs or as a result of higher order regulation of regions of the genome remains unclear. However the demonstration that in the mammalian system CLOCK protein is a histone acetyltransferase (Doi et al., 2006) whose activity is antagonised by the SIRT1 histone deacetylase (Belden and Dunlap, 2008) lends credence to the latter model.

The possibility of clock gene regulation at the histone level was examined by screening a panel of histone modification mutants in collaboration with Christina Tuffarelli, though the work was abandoned due to the difficulty of excluding background effects in diverse mutant stocks without performing long term inbreeding experiments. Notably in a study performed after this work, Matsumoto *et al.* used a systems biology approach to elucidate the functions of cycling transcripts by generating RNAi constructs against 133 cycling genes, two of which

were found to play roles in chromatin modification; *Iswi* (acting in the maintenance of chromatin architecture), and *prod* (acting in chromatin binding). Taken together with the results generated in this study suggesting that the CCCGs *CG5156* might act in histone acetylation, whilst *CG5798* might be associated with chromatin maintenance proteins, one can conclude that co-ordinate gene regulation at the histone level is likely to be a feature of the *Drosophila* as well as the mammalian clock mechanism.

6.5 Conclusion/further work

Despite the rapid advances in technology and annotation during the course of this PhD, the spreadsheet generated in this chapter is by no means superseded by later work. Indeed, many avenues remain to be explored in analysing meta datasets, including investigations into the enhancer motifs underlying circadian rhythmicity (such as that performed in the mouse - Ueda et al., 2005) and those seeking to reveal novel forms of circadian gene regulation. With the rapid adoption of RNAi technology for further screening, and the elucidation of links between the fields of sleep and circadian research, one could argue that an even more comprehensive meta analysis is now overdue.

The observation that peak phases of gene expression occur throughout the day suggests that the circadian clock is significantly more complex than the biphasic oscillator model might suggest (for review see Dowse, 2008). Whilst evidence is emerging to suggest that much of this complexity might be generated at the post translational level (Reddy et al., 2006), results garnered both in this study and in other meta-analyses (and the original microarray papers) suggest that genes of diverse ontology are associated with the circadian clock mechanism (Figure 6-4). Such a result suggests that the clock mechanism might yet be

regulated in a number of novel ways, despite the failure of experiments in this study to confirm such a hypothesis.

The next chapter details a novel method for determining gene function that reveals novel genes playing in the clock mechanism that do not show transcriptional rhythms.

7 Transheterozygote study (untargeted screening for gene function)

7.1 Introduction – the paradox of forward genetic screening

In the past four decades, the forward genetic screening (mutagenesis) approach developed by Benzer to reveal the genetic basis of the circadian clock (Konopka and Benzer, 1971) has become the paradigm for the directed identification of gene function. To make such screens more efficient, the Berkeley Drosophila Genome Project (BDGP) was set up to disrupt every gene in *Drosophila*, initially using *P* element-mediated mutagenesis (Bellen et al., 2004), but including more recent FLP/FRT recombination and site specific integration technologies (or Venken and Bellen, 2007; for review see Venken and Bellen, 2005). By screening the BDGP stock collection *(http://flypush.imgen.bcm.tmc.edu/pscreen/)* using high throughput, quantitative phenotype assays, it was thought that the function of every gene in the genome could be determined within in a short period of time (Friedman and Perrimon, 2004).

However, *in vivo* analysis of gene disruption strains has uncovered a paradox; most genes appear to have no known function (despite considerable screening efforts), whilst a small number of genes such as *cryptochrome* have been observed to show a highly promiscuous degree of functional pleiotropy. *cry* mutations have been implicated in the circadian clock as a key blue-light photoreceptor important for light input and phase shifting the central clock (Stanewsky et al., 1998), as a core component of the peripheral tissue clock (Stanewsky et al., 1998; Dolezelova et al., 2007; Myers et al., 2003), as a molecule involved in magnetosensitivity (Gegear et al., 2008), and as a molecular correlate of geotaxis (Toma et al., 2002). This paradox remains despite the development of sensitive genetic tools providing information in addition to that gleaned from gene disruption alone, including P element mediated gene mis-expression, GFP localisation and protein tagging (Greenspan, 1997).

7.1.1 Networks, not pathways

Explaining this apparent paradox requires a re-examination of the classic model of gene function, which is essentially an elaboration of the linear biochemical pathways first conceptualised by Morgan and Fisher's pioneering genetics experiments (Fisher, 1918; Morgan, 1911) in which genes act sequentially in discrete steps:

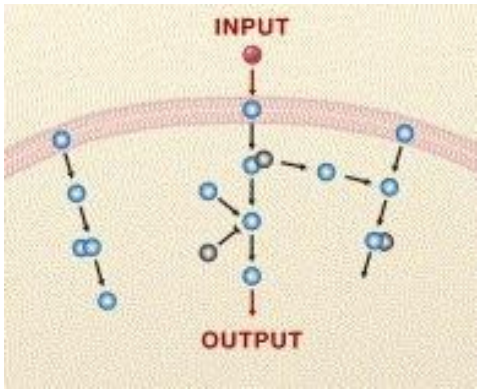


Figure 7-1 Classical genetics model of a signalling pathway.

Can such a model be applied more generally to complex phenotypic outputs such as behaviour? One possibility is that coherent expression of a behavioural phenotype represents the integration of the outputs of many individual linear pathways such as that shown in Figure 7-1. Independent linear pathways might for example ensure the correct development and connectivity of the neural substrate, others control the integration of different stimuli, and still more play roles in mediating motor outputs.

Due to the number of such linear pathways (or 'bio-behavioural modules': Tully, 1996) required to generate a complex behaviour, one might expect that mutations affecting behaviour should be easy to identify, and indeed many genes of diverse functional classes have been found to have pleiotropic effects on behaviour (Hall, 1994), supporting the argument that behaviour represents the summation of many disparate linear pathways.

It was therefore a considerable surprise when Fedorowicz *et al.* were able to show significant epistatic interactions within a panel of functionally diverse mutations affecting

olfactory behaviour (Fedorowicz et al., 1998), suggesting a significant degree of overlap or interaction between pathways previously thought to be functionally independent. Even more surprisingly, van Swinderen *et al.* were able to show that the epistatic interactions between genes controlling even a simple behaviour were not fixed, but changed depending on genetic background (van Swinderen and Greenspan, 2005).

The emerging data show that mutations affecting behaviour have complex, non-exclusive and context dependent interrelationships. The inference of such studies is that complex phenotypes (of which behaviour is the pre-eminent example) are not determined by a plethora of linear genetic pathways, but by overlapping, wide ranging networks of genes. Whilst the properties of such networks are beyond the scope of this introduction (for review see Stelling et al., 2004; and Barabasi and Oltvai, 2004), the general consensus from both theoretical and experimental studies is that genes form scale free networks, in which certain 'hub' genes are connected to many other, less well connected genes as shown below:

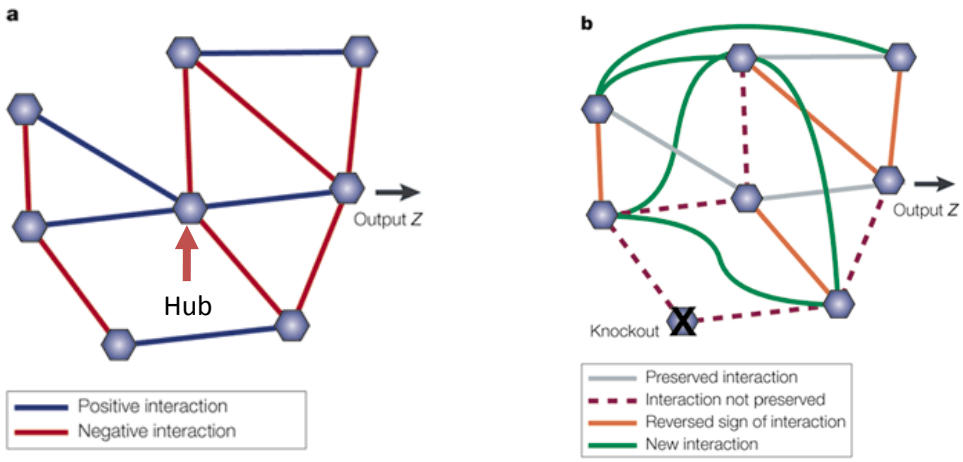


Figure 7-2 Effects of network disruption on a scale free network. Left panel shows the interactions occurring within a gene network, genes (or 'nodes') illustrated as linked **blue hexagons** and the central hub indicated by a **red arrow**. **Right panel** shows the response of the network to the disruption of a peripheral node; although flux through the system is altered by new interactions, the output (**Z**) may be identical as the hub is not affected. Figure adapted from Greenspan (Greenspan, 2001).

One of the properties of scale free networks is that they are only sensitive to disruption of hub genes, as shown in Figure 7-2; disruption of poorly connected nodes is

mitigated by architectural features of the network such as modularity, redundancy and feedback control. Scale free networks therefore can compensate for disruption to non-hub genes, achieving the same output in a degenerate fashion (Greenspan, 2001).

This robustness plays an important role in the development of complex organisms; whilst simple organisms such as *C. elegans* rely on a lineage dependant developmental process, such an approach cannot accommodate disruptions during development (such as might occur due to copy number variation between parents, or as a result of stochastic environmental perturbation). The development of complex organisms must therefore be buffered, network degeneracy being one of the main methods by which the outcome of development may be canalised (Dworkin, 2005).

However, in the context of determining gene function by forward genetic screening, degeneracy limits such screens to only uncovering hub genes within a network:

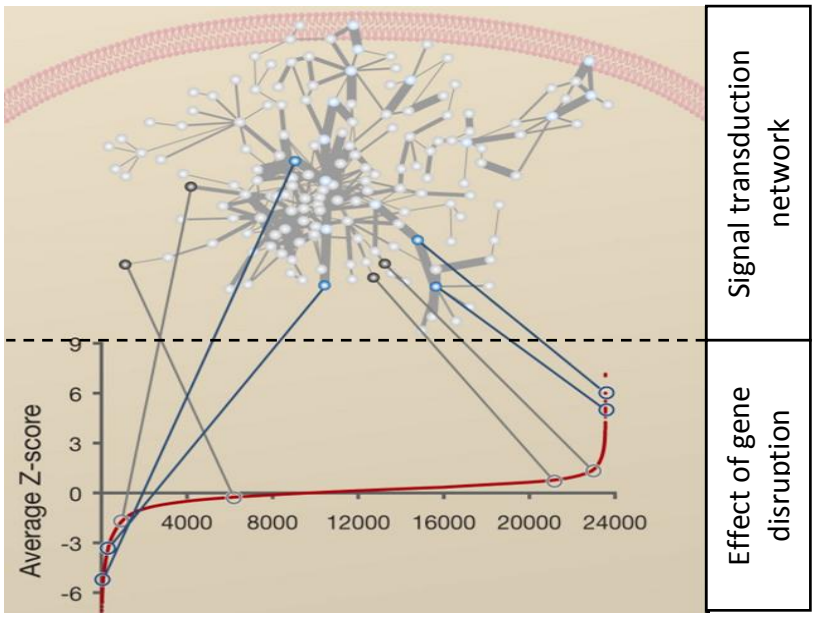


Figure 7-3 The effect of disrupting different elements of gene networks. Top panel shows a network model for signal transduction, in which a stimulus propagates information throughout the network, leading to multiple outputs. The greater the contribution of each protein to the output, the thicker the line weight linking it to other components. Known canonical components are indicated by **blue dots**. **Bottom panel** shows that chance that disrupting any given distribution affecting the phenotypic output follows a sigmoidal distribution. Only those genes lying at the extremes of the distribution can be

detected by forward genetic screening, these genes are generally hubs. Figure adapted from Friedman and Perrimon (Friedman and Perrimon, 2007).

Efforts to identify non hub genes have focussed on using 'sensitised' genetic backgrounds in which flies carrying a mutation in a given pathway are further mutagenised (second site mutagenesis: Byrne et al., 2007). However, the mechanics of network connectivity dictate that a disrupted network may be equally robust as the wild type network, and may even possess emergent properties which confuse identification of *bone fide* components of wild type network (Friedman and Perrimon, 2007).

A further complication in determining gene function is the observation that hubs are more likely than other nodes to represent essential genes (Barabasi and Oltvai, 2004), therefore although a hub may play an important role in a given phenotype, its pleiotropic role in development may preclude its identification in screening:

Figure 7-4 The pleiotropic roles of genes. Genes (**circles**) involved in behaviour may either have specific effects (**yellow** circles), or may have pleiotropic roles in numerous processes (**orange**, **green** and **black** circles). The essential role of the orange and black genes in early development therefore precludes their identification as *bone fide* behavioural genes. Figure adapted from de Belle (de Belle, 2002).

The properties of gene networks and functional pleiotropy therefore limit the effectiveness of forward screening approaches.

7.1.2 Systems genetics approaches to identifying gene function

Efforts to overcome the inherent limitations of both quantitative genetics and Benzerian screening in identifying gene function have largely focussed on the development of cell culture based RNAi gene knockdown (Fire et al., 1998). This approach has depended upon the development of dynamic, quantitative reporters, the measurement of which can easily be automated (e.g. CRE-luciferase: Iijima-Ando and Yin, 2005). In a pioneering study, Boutros *et al.* used this approach to identify 438 genes affecting cell viability, 80% of which had not been identified in previous forward genetics screens (Boutros et al., 2004). Although such approaches are valid for relatively simple, cell autonomous phenotypes, they do not scale to more complex phenotypes such as behaviour.

Toma *et al.*, reasoning that genes showing significant differences in expression between bidirectionally selected lines might have a causative relationship with the phenotype (Toma et al., 2002), assayed gene expression levels using microarrays in two lines showing extreme geotaxis (Ricker and Hirsch, 1985). Further analysis in a controlled genetic background allowed the identification of a number of genes not previously known to be involved in geotaxis, complementing traditional forward genetic screens (Armstrong et al., 2006) in the dissection of geotaxis.

7.2 The object of this study

In many ways Toma *et al.*'s study of geotaxis has become the paradigm for recent efforts to identify gene function. As a result, the Mackay group is currently sequencing a panel of 192 inbred *Drosophila* lines (http://www.hgsc.bcm.tmc.edu/projects/drosophila/dgrp.html) in the hope that typing these lines for dozens of phenotypes might shed new light on how genes interact to give rise to phenotypes. However, as the networks underlying behavioural phenotypes are deeply rooted, it is likely that the variation inherent in inbred natural lines will disrupt gene networks at many points, making it very difficult to determine the contribution of

individual genes to phenotypes. Indeed, the most appropriate approach to study gene networks must be to assess the transcriptomic changes resulting from a single perturbation of that network in an otherwise equilibrated genetic background. Such experiments allow accurate estimation of network depth (how great a change in phenotype a single mutation can induce relative to a wild type control) and breadth (how many phenotypes are changed by a single mutation).

This study goes further still; rather than studying the effects of a single mutation on a network, a novel approach was developed to study the effects of changing a single genetic interaction by combining phenotypically mild mutations of diverse function in a controlled genetic background to form transheterozygous flies (i.e. flies carrying one copy of each mutant allele). As the mutations are subtle and recessive, *a priori* one would assume that transheterozygous flies would not show any phenotypic differences from appropriate controls. However, should the mutations affect the same genetic network, the cumulative effect of the two mutations might be sufficient to elicit an effect.

As the results of such changes are likely to be subtle, the state of the network must be assessed using a sensitive output. Behaviour, by virtue of its complexity, is particularly sensitive to genetic perturbation; furthermore, due to the evolutionary advantage behavioural flexibility confers upon an individual, it is likely to be resistant to canalising effects.

This novel approach therefore promises to not only describe the depth and breadth of genetic networks, but also to determine the relationships between different behaviours, and possibly present a new paradigm by which gene function can be determined. Given that many human diseases and disorders are behavioural in nature (e.g. depression, schizophrenia), there is significant value in determining the networks, genetic factors and interactions modulating different behaviours.

7.3 Methods

7.3.1 Mutant alleles and diallel cross design

Eight ontologically diverse, phenotypically subtle mutations were backcrossed into a Canton-S background for at least 6 generations so as to minimise the effect of genetic background upon sensitive behavioural analyses.

Table 7-1 Alleles used in this study. All alleles are the result of *P{w⁺}* insertions backcrossed into a CS background. *Itp, Pen, trf,* and *cnk* alleles are recessive lethal mutations kept as balanced stocks.

Allele	Gene Name	Gene Function
Adf^{nal}	*Adh transcription factor 1 (nalyot)*	Transcription factor; enhances *Adh* transcription. Mutants have defects in learning, memory and larval locomotor activity.
Csp^{P1}	*Cysteine string protein*	Exocytosis; stabilizing and controlling release of neurotransmitters at neuromuscular junction. Mutants have defective courtship behaviours.
mth^{1}	*methuselah*	G-protein coupled receptor activity; plays role in longevity through response to environmental stresses; mutants have neurophysiology defects.
trf^{1}	*TBP-related factor*	Transcription regulation. Mutant males sterile and show behavioural defects and hyperactivity.
Itp-r83a^{j5B4}	*Inositol tri-P receptor*	Signal transduction and ion transport; mutants have neuroanatomy and flight behaviour defects.
Pen^{k14401a}	*Pendulin*	Protein transporter; mutants may be sterile and show defects in geotaxis behaviours.
nmo^{P1}	*Nemo*	Serine/threonine kinase. Plays role many biological processes including eye development, geotaxis, wing development and regulation on WNT signalling.
cnk^{k16314}	*connector enhancer of ksr*	Protein binding; RAS protein signal transduction. Involved in eye and wing development

For clarity, allele numbers are omitted in the text, i.e. *Csp* refers to *Csp^{P1}*, whilst *Adf^{nal}* is referred to by its common synonym *nal*. Backcrossed mutant strains were then subject to a diallel crossing scheme generating non-reciprocal transheterozygote F1 flies as described by Griffing (Griffing, 1956) and illustrated in Table 7-2. To prevent any maternal effect bias, the crosses were balanced such that any one mutant line would be the male parent in generating one half its seven transheterozygote F1 progeny, and the female parent in the other half:

Table 7-2 Diagram of the diallel crossing scheme showing the direction of crosses used to generate F1 flies. Each F1 transheterozygote progeny is assigned a unique number, and is properly named in the fashion 'maternal allele/paternal allele' as in Greenspan (Greenspan, 1997). For example, F1 progeny 4 is properly labelled '*Itp*/*nal*'. F1 progeny carrying a *mth* allele were generated reciprocally (*).

Male/Female	*nal*	*Csp*	*mth*	*trf*	*Itp*	*Pen*	*nmo*	*cnk*
nal		1			4		6	
Csp			*	9	10			13
mth	2	8		*	15	*	17	
trf	3		14			20	21	
Itp			*	19		23		25
Pen	5	11	16				26	
nmo		12			24			28
cnk	7		18	22		27		

To bolster the power of the scheme, as well as to minimize and detect potential errors, several additional lines were tested;

- The Canton-S background strain used for backcrossing (i.e. *+/+).*
- Backcrossed heterozygote F1 flies, which can be used as an independent measure of the effect an allele has on a phenotype (e.g. *nal/+*).
- Reciprocally crossed transheterozygotes containing the *mth* mutant allele to check for maternal inheritance effects (e.g. *trf/mth* and *mth/trf*)
- *nal*/*mth* transheterozygote F1 flies were tested twice to test the reproducibility of behavioural characteristics (i.e. *nal/mth* and *nal/mth (2)*).
- Some experiments were repeated at a later date to assess the magnitude of genetic drift on sensitive behavioural phenotypes.
- Some tests were performed on each sex separately to assess sex-specific effects, as up to two thirds of genes may show significant sex-specific transcription (Wayne et al., 2007)

Due to time constraints, these additional lines were not assessed for all phenotypes.

7.3.2 Fly Keeping

Crosses were performed by J. Wagner (JW) according to the scheme illustrated in Table 7-2. Flies were kept in 12:12 LD conditions, and handled according to published guidelines (Greenspan, 1997).

7.3.3 Behavioural paradigms

Drosophila have been used to establish an experimental paradigm in almost every field of biology, including the investigation of complex learning and memory tasks long thought to be the preserve of research in higher vertebrates (Sokolowski, 2001). During this study the analysis of various behaviours was split between a number of specialist investigators:

Table 7-3 Division of work between investigators involved in the transheterozygote study. JW = Jenée Wagner, EG = Edward Green, RA = Rozi Andretic, HR = Helen Roe, DE = David Evans, BS = Bruno van Swinderen, SB = Senait Bekele, HD = Herman Dierick.

Behaviour	Investigator	Equipment	Published paradigm
Geotaxis	SB	Vertical Maze	(Toma et al., 2002)
Optomotor Maze	JW	Horizontal Maze	(van Swinderen and Flores, 2006)()()
Phototaxis	SB	Illuminated Maze	(Hirsch and Erlenmeyer-Kimling, 1961)
Fainting at 39C	BS	Variable Temperature Tube	(van Swinderen and Greenspan, 2005)
DD Period	HR/RA/EG	Triketics Activity Tube	(Rosato and Kyriacou, 2006)
Sleep	RA/BS/EG	Triketics Activity Tube	(Andretic and Shaw, 2005)
Waking Activity	RA/JW/EG	Triketics Activity Tube	-
Courtship	HR/HD	Activity Arena	(Campesan et al., 2001)

7.3.4 Unifying group annotations

Due to the balanced nature of the diallel crossing scheme, it was important to perform crosses in the same direction for all phenotypic analyses to minimise the contribution of maternal effects. On collating the data generated by different investigators, it was found that differing annotation systems had been used to describe each genotype; an example of which is shown in Table 7-4. It was therefore necessary to create a congruency table that allowed results from different groups to be amalgamated:

Table 7-4 Extract from the congruency table used to translate annotations used by different investigators to describe the same F1 lines. First column contains the correct genotype, other columns contain the annotation used by different investigators. Column headings refer to experiments, e.g. 'Fainting' refers to the high temperature induce fainting behaviour assayed in this study. Each incidence of incorrect annotation was checked with the investigator to distinguish annotation errors from biological errors.

Genotype	Phototaxis	Arrays	Song	Mating	Maze	Fainting
cnk/+	-	R10206	*canton s f x cnk m*	*cnk/+*	*cnk/+*	-
cnk/Csp	-	R10207	*cnk f x csp m*	*Csp cnk*	*cnk Csp*	*csp/cnk*
cnk/Itp	-	R10208	*cnk f x Itp m*	*Itp cnk*	*cnk Itp*	*itp/cnk*
cnk/nmo	-	R10212	*cnk f x nmo m*	*nmo cnk*	*cnk nmo*	*nmo/cnk*
CS	*cs*	R10254	*canton s f xcanton s m*	*CS*	-	-
Csp/+	*csp/+*	R10211	*canton s f x csp m*	*Csp/+*	*Csp/+*	-

7.3.5 Microarrays

Flies were collected by J. Wagner at ZT3. Gene expression data was generated from these samples in-house by the FlyChip group at the Cambridge Systems Biology Centre using FlyChip FC004 arrays (http://www.flychip.org.uk/services/core/FC004/). Protocols for cDNA library creation and spotting, sample processing, sample hybridisation and image acquisition can be obtained from the FlyChip website. FlyChip probes can be mapped to the more common Affymetrix probes using the annotation files provided at http://www.flychip.org.uk/services/core/FC004/#AFF. Further details regarding array annotation and interpretation of results files is available online at the FlyChip website; http://www.flychip.org.uk/services/analysis/downloads/.

7.3.6 Genetic architecture

Several techniques were used to quantify the depth and breadth of the genetic networks underlying behaviour in *Drosophila*:

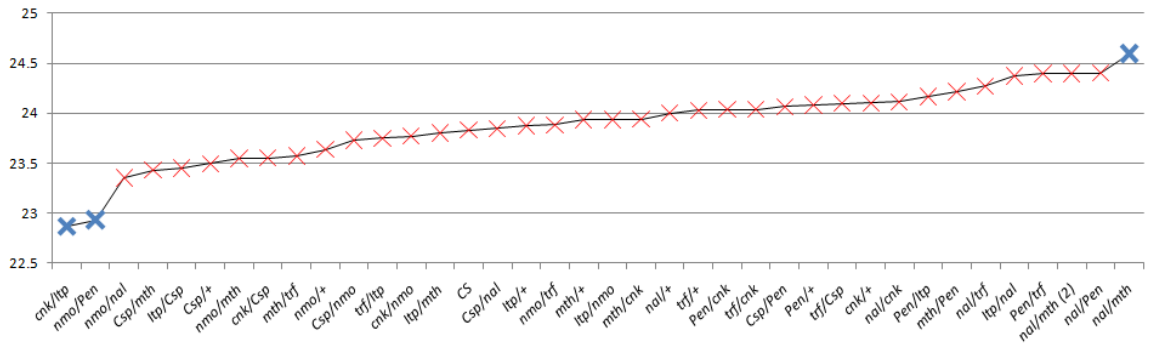
7.3.6.1 Calculation of phenotype/gene expression correlation coefficients

Phenotypic scores for each genotype were mapped onto microarray expression datasets using the annotation congruency table and a series of Excel LOOKUP functions. Two

dynamic data mining tools (chosen from a drop down menu) were created to calculate the correlation between gene expression and a given phenotypic score:

Linear correlation - the correlation between gene expression and phenotypic score is calculated for all available transheterozygotes for each gene. Optionally the user can set a filter specifying the minimum number of transheterozygote genotypes in which a gene must be detected to prevent the detection of spurious correlations in cases where the gene is found to be expressed in only a small number of transheterozygotes.

Maximum difference - there is no *a priori* reason for assuming gene expression must correlate linearly with phenotype; instead genes showing extremely high or low expression in phenotypic outliers might represent genes which significantly contribute to that phenotype. Accordingly the gene expression workbook (see appendix D3) contains a spreadsheet to facilitate such analysis. Phenotypic outliers can be selected from a dynamic plot of phenotypic score:



	Largest					Smallest
Phenotype	24.60	24.40	24.40	23.35	22.93	22.87
Include	Yes	No	No	No	Yes	Yes

FBGN	Clone	FB	L1	L(-1)	L(-2)	S(-2)	S(-1)	S	Soup	Hi-Lo	Lo-Soup	Hi-Soup
0	L22585	0	0.0000				0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	L22585	0	0.0000				0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	L22585	0	0.0000				0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	L22585	0	0.0000				0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	L22585	0	0.0000				0.6571	0.4475	0.0000	-0.5523	0.5523	0.0000
0	L22585	0	0.0000				0.6060	0.4041	0.0000	-0.5050	0.5050	0.0000

Figure 7-5 Calculating non-linear regression between phenotypic outliers and gene expression data. A plot of phenotypic data (top) shows 2 clear outliers at the lower end and 1 at the top end (**blue markers**).

As shown in Figure 7-5, it is possible in the spreadsheet to select one or more of the transheterozygote genotypes having the highest (**Hi**) and lowest (**Lo**) phenotypic scores, as well as the level of gene expression averaged across all genotypes (the '**soup'** expression). The expression of each gene in the selected transheterozygote lines can then be compared using a number of metrics; the difference in a gene's expression between phenotypically 'high' and 'low' lines (**Hi-Lo**), and the difference relative to the 'soup' score. The results can be sorted by magnitude, allowing the genes which show the largest expression difference in the most phenotypically distinct lines to be identified.

The statistical significance of correlations can be calculated using a Monte Carlo randomisation approach, in which the experimental data for each gene is randomised such that each phenotypic score is matched with a random gene expression value, and the expression/phenotype correlation recalculated. By performing 1000 such randomisations, a distribution of possible correlations can be calculated. If the experimentally determined correlation exceeds the 95th percentile of the calculated distribution, the result is taken to be significant at the 5% level.

7.3.6.2 Epistatic Networks

In breeding studies it is valuable to know quantitatively how one line or mutation affects a phenotype, and whether specific combinations of lines act additively or show evidence of epistasis when brought together.

Such questions can be addressed using a diallel crossing scheme, in which each line is reciprocally crossed to each other line, so as to generate all possible F1 progeny, including selfed F1 lines. More commonly a half diallel crossing scheme is used in which reciprocal crosses are omitted (as was used in this study, shown in Table 7-2 and Table 7-5).

The degree of phenotypic variance attributable to any one line - or allele if using isogenised backgrounds - is referred to as a line's general combining ability (GCA). GCA scores

for a line are calculated by comparing the mean phenotypic score of all its F1 progeny to the mean phenotype of all F1 progeny in the diallel scheme. Expressed algebraically, GCA may be calculated as:

$$GCAx = \frac{Tx}{n-2} - \frac{\sum T}{n(n-2)}$$

Where $GCAx$ = Is the GCA for line x,

Tx = The mean value for line x

ΣT =twice the sum of the mean values for each line in the diallel scheme

n =number of lines in the diallel scheme

GCA scores can only be applied in the context they are calculated, as the values are relative to the genetic background chosen for the diallel crossing scheme. Consider the following example half diallel crossing scheme in which the geotaxis behaviour was measured in a number of F1 transheterozygote lines to give a geotaxis score between 1 and 9:

Table 7-5 Example calculation of GCA using a half diallel. Each score represents the mean of three experimental replicates, performed with at least 70 flies per run. SUM and GCA scores are calculated for the allele detailed in the first column of the row.

	Csp	*trf*	*ltp*	*Pen*	*mth*	SUM	GCA
nal	5.17	5.11	5.35	4.49	4.82	24.94	-0.54
Csp		5.47	6.34	5.84	6.08	28.90	0.45
trf			5.89	5.14	5.61	27.22	0.03
ltp				5.30	5.51	28.38	0.32
Pen					5.22	25.99	-0.28
mth						27.23	0.03
						162.66	**0**

To calculate the GCA for the *nal* mutation in the geotaxis phenotype, we substitute values into the formula detailed above;

$$GCAnal = \frac{Tnal}{n-2} - \frac{\sum T}{n(n-2)}$$

$$GCAnal = \frac{24.94}{4} - \frac{162.66}{24}$$

$$GCAnal = 6.24 - 6.78$$

$$GCAnal = -0.54$$

As $Gnal$ has a negative value, this indicates F1 progeny carrying the *nal* mutation will have a tendency to have a geotaxis score 0.54 units lower than the F1 average. One can therefore predict the geotaxis score of F1 individuals by adding the GCA values for each parental line ($GCAx$ and $GCAy$) to the general mean ($\bar{X}$):

$$e^{xy} = \bar{X} + GCAx + GCAy + SCAxy$$

Clearly such estimates are not always accurate, both due to within group variance (i.e. error), and as a result of genetic epistasis or dominance effects (Fisher, 1918), collectively referred to as the specific combining ability (SCA) of a given cross. SCA is therefore calculated the difference between expected and observed phenotypic scores; in the case the *nal*/*Csp* flies' geotaxis behaviour:

$$e^{xy} = \bar{X} + GCAx + GCAy + SCAxy$$

$$e^{nal/Csp} = \bar{X} + GCAnal + GCACsp + SCAnal/Csp$$

$$e^{nal/Csp} = 5.42 - 0.54 + 0.45 + SCAnal/Csp$$

As the *nal/Csp* geotaxis score can be experimentally determined value to be 5.17, the SCA value for this cross is -0.15. Repeating such calculations for every F1 genotype one can create a table of SCA values:

Table 7-6 Example table of SCA values calculated for geotaxis behaviour. Notably large deviations from additivity are highlighted in blue.

SCA	*Csp*	*Trf*	*Itp*	*Pen*	*mth*
nal	-0.15	0.21	0.15	-0.11	-0.09
Csp		-0.43	0.15	0.25	0.18
trf			0.12	-0.03	0.13
Itp				-0.16	-0.26
Pen					0.05

The statistical significance of the SCA values can be calculated using ANOVA as detailed in Griffing's original paper (Griffing, 1956). Such calculations are particularly complex, subject to ongoing research concerning their validity, and not easily implemented without dedicated software packages such as SAS. For the purposes of this study, the Excel spreadsheet collating results was used to calculate GCA and SCA values, and the largest SCA values were then studied without recourse to statistical verification of the significance of such values. For further details, see Falconer and Mackay (Falconer and Mackay, 1995).

7.3.7 GAL4 driver mediated gene knockdown

In an effort to confirm that genes correlating with phenotypic scores played a causative role in that phenotype, candidate gene expression levels were knocked down using RNAi hairpin constructs created by the VDRC (Dietzl et al., 2007) using the transgenic GAL4/UAS expression system (Brand and Perrimon, 1993):

$$\male\ w^{1118}\ ;\ \frac{tim{>}GAL4}{tim{>}GAL4}\ ;\ +\quad \times \quad \female\ \frac{w^{1118}}{w^{1118}}\ ;\ \frac{P\{UAS{>}hairpin\}}{P\{UAS{>}hairpin\}}\ ;\ \frac{+}{+}$$

Figure 7-6 Typical crossing scheme for gene knockdown in which a GAL4 driver is brought together with a P element containing UAS responder driving the expression of an RNAi inducing hairpin construct.

7.3.8 *UAS-Dicer2* enhanced gene knockdown.

Certain cell types, particularly neural tissue, have been found to be refractory to RNAi mediated gene knockdown (Dietzl et al., 2007). Overexpression of the *Dicer-2* (*Dcr-2*) component of the RNAi machinery in *Drosophila* can enhance the RNAi effect in approximately 50% of cases – though at the cost of increasing the false positive detection rate by 6% (Dietzl et al., 2007). As behaviour is particularly dependent on the nervous system, a number of RNAi knockdown experiments were carried out using a line in which the *UAS>Dicer2* transgene had been recombined onto a chromosome carrying the *tim>GAL4* transgene (kind gift of S.Dissel).

7.4 Results

7.4.1 Inbred transheterozygote lines show considerable phenotypic variation

The transheterozygote progeny generated in this study showed considerable variation in all behaviours studied (see appendix D3), an excerpt of which is shown below:

Table 7-7 Excerpt of table in appendix D3 showing phenotypic variation in transheterozygous and heterozygous lines. n/a = genotype not scored for this behaviour.

Phenotype	*cnk/Csp*	*cnk/ltp*	*cnk/nmo*	*CS*	*Csp/+*
Geotaxis score	n/a	n/a	n/a	5.2	5.3
Phototaxis score	n/a	n/a	n/a	3.1	4.5
$t_{1/2}$ fainting at 39°C	104	135	161	n/a	n/a
Circadian period	24	23	24	24	23
Mean waking activity	5.5	5.6	7.1	5.8	5.7
Total activity per day	871	863	1070	685	811

As the alleles used to generate transheterozygotes had been backcrossed for a minimum of six generations into the same Canton-S background, the variation evident in Table 7-7 must arise as a direct result of interactions between transheterozygous alleles. The inference of this result is that the genetic networks underlying behaviours are sensitive even to very minor disruption (as considerable variation was observed despite the use of single copies of phenotypically mild alleles), and run particularly deep (as all behaviours studied showed considerable variation, as shown graphically in Figure 7-7).

7.4.2 Creation of spreadsheet gene expression/phenotype spreadsheet

Given the extensive phenotypic variation evident in transheterozygous lines, it should be possible to identify genes whose expression correlates with the phenotypic variation following the approach used by Toma *et al.*

cDNA samples from all transheterozygous were microarrayed by the FlyChip consortium, and normalised results were incorporated into a spreadsheet in which phenotypic data (such as that describing circadian period in Figure 7-7) could be correlated with gene expression data for each transheterozygote, as shown in Figure 7-8:

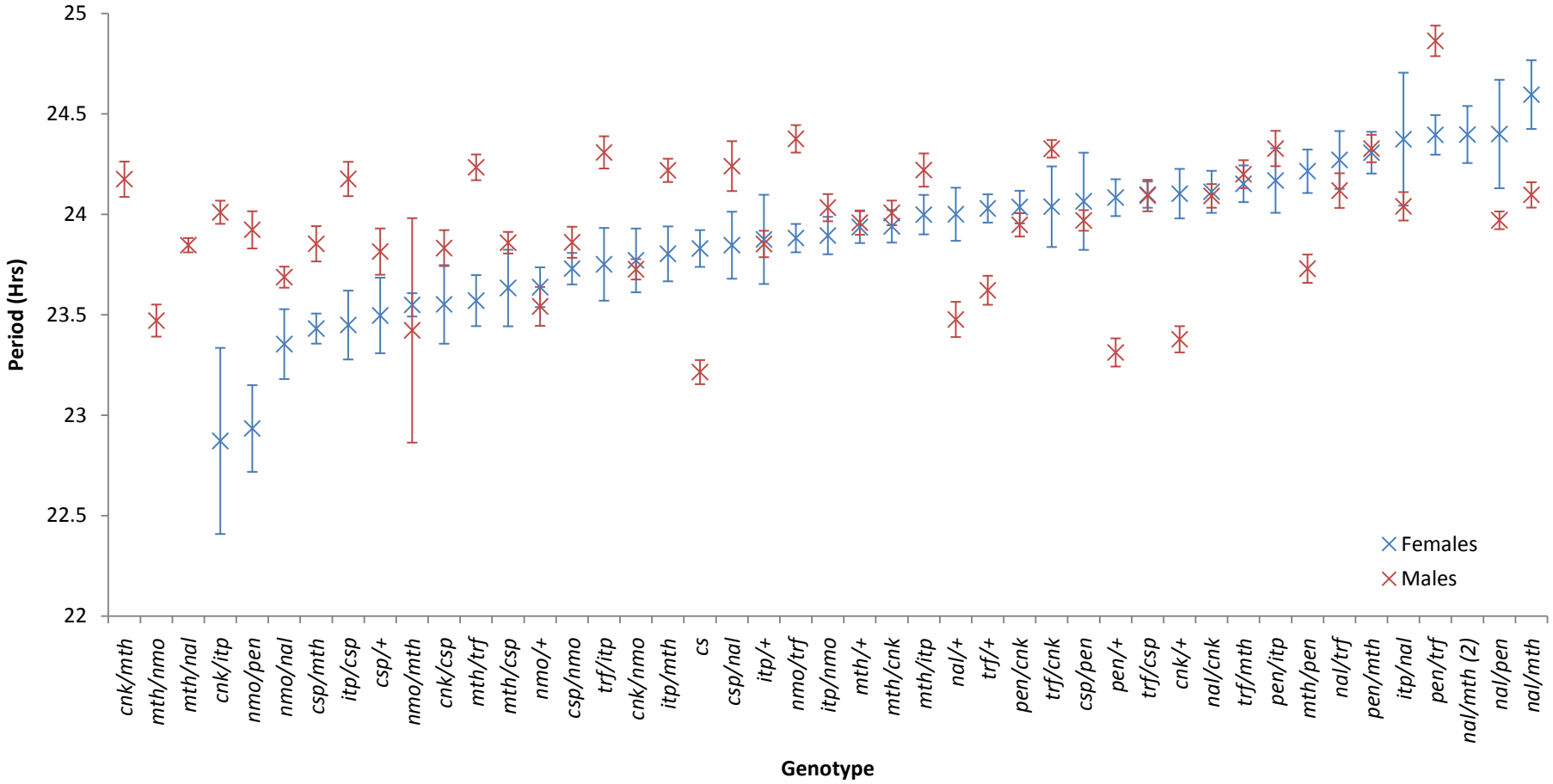


Figure 7-7 Free running periods of heterozygous and transheterozygous flies, together with the Canton-S (CS) background control. For each point n=32. Female data recorded on Trikinetics system, Male data on Drosophix system. Correlation coefficient between male and female data = 0.23.

	A	B	C	D	AZ
1				Phenotype	Total Sleep Bins (/Day)
2					
3				Minimum Spots	30
4					
5					
6				Minimum Correlation	0.6
7					
8					
9				Number of strains analaysed	38
10					
11					
12				Update Table	
24					
25	Flybase_ID	ID	CloneID	FlyBase_symbol	Correlation
2891	FBgn0034659	1000018075	CG4021-RA	CG4021	0.688
2910	FBgn0002645	1000016534	CG1483-RA	Map205	0.617
4270	FBgn0000024	1000015978	CG17907-RA	Ace	0.610
4287	FBgn0039203	1000013342	CG13618-RA	CG13618	-0.639
4295	FBgn0039205	1000013350	CG13623-RA	CG13623	-0.666
4576	FBgn0045862	1000013071	CG12878-RB	btz	-0.622
5523	FBgn0039005	1000020107	CG6969-RA	CG6969	-0.629

Figure 7-8 Excerpt of the Excel spreadsheet used to linearly correlate quantitative phenotype and gene expression data. Full spreadsheet can be found in appendix D3.

Figure 7-8 shows a number of candidate genes showing significant correlations between expression and circadian period in female transheterozygote flies using the linear correlation model (see methods section for discussion).

7.4.3 Candidate gene knockdown

In order to assess whether such correlative relationships were truly causative, candidate gene expression levels were knocked down using RNAi stocks available from the VDRC as detailed in the methods. As there was no *a priori* way to ascertain where each candidate gene was expressed, gene expression was knocked down using the pan-neuronal *elav>GAL4* driver in an effort to maximise the chance that gene knockdown occurred in relevant tissues. However, performing a control experiment to test the relative 'strength' of *elav>GAL4* by knocking down *per* expression (a core clock gene whose levels inversely correlate with circadian period: Martinek and Young, 2000) showed that *elav>GAL4* could only elicit a period of 26.3 hours, compared with 29.8 hours using the more clock-cell-specific *tim>GAL*. Given that all known clock genes are found in *tim* expressing cells, it was decided to preferentially use the more powerful *tim>GAL4* driver in experiments.

Due to the long running nature of circadian experiments, not all genes showing significant expression correlations could be tested, therefore a subset were selected based upon gene ontology criteria, in addition to a number of genes showing weaker correlations acting as negative controls. The results of gene knockdown are shown in Table 7-8.

Although almost half of genes in *Drosophila* are thought to have an essential role (Young and Judd, 1978), very few knockdown genotypes were inviable, suggesting either that gene expression was being reduced in tissues in which candidate genes played non essential roles, or that the extent of expression knockdown was incomplete – a common source of false negative results in RNAi screening studies (e.g. failure to induce arrhythmicity, see Matsumoto et al., 2007). Although the degree of gene knockdown was not directly assessed, the extent of knockdown using the VDRC constructs has been assessed by Dietzl *et al.*, showing that in more than half of cases gene expression is reduced to less than 25% of wild type levels using an *actin>GAL4* driver (Dietzl et al., 2007).

7.4.4 Confirmation that knockdown caused significant effects

Despite construct dependent variation in the efficiency of gene knockdown, reducing genes' expression using both the *tim>GAL4* and *elav>GAL4* drivers elicited significant or nearly significant effects on circadian period as shown in Table 7-8 and Table 7-9 (using nearly significant results is appropriate due to the limited resolution of algorithmic period analysis - see D1 appendix manual for detailed discussion of the limitations of algorithmic period analysis).

Crucially no genotypes were fully arrhythmic, suggesting that the *tim* expressing clock cells remained intact, and that results represented genuine disruption of the clock mechanism rather than ablation of clock cells.

Table 7-8 The role of genes implicated in the circadian clock using a novel systems biology approach was assessed by knocking down gene expression in clock cells (tim>GAL4) and pan-neuronally. Genes were selected based on the magnitude of their expression/phenotype Pearson product-moment correlation coefficient (PMC) and their biological process ontology categorisation (GO). The probability (**p**) that gene knockdown elicited a significant effect on the period (**τ**) given the number of flies tested (**n**) was calculated by ANOVA using posterior Dunnett significance tests (significant or nearly significant results **highlighted in blue**).

♂		*elav>GAL4*			*tim>GAL4*			
Gene	**PMC**	**τ**	**n**	**p**	**τ**	**n**	**p**	**GO Biological Process**
+	-	23.4 ±0.10	15	-	23.8 ±0.10	15	-	-
+(2)	-	23.5 ±0.06	16	-	-	-	-	-
CG10118	0.38	23.4 ±0.09	16	1.00	24.0 ±0.07	13	1.00	dopamine metabolic process ; courtship behaviour ; locomotor behaviour
CG10553	-0.10	23.3 ±0.10	12	1.00	-	-	-	-
CG1064	-0.32	23.8 ±0.17	3	0.99	23.9 ±0.11	13	1.00	regulation of transcription ; development
CG11440	0.41	23.3 ±0.11	12	1.00	24.5 ±0.23	15	0.00	dephosphorylation ; phototransduction ; response to light intensity
CG11889	-	23.3 ±0.09	26	1.00	-	-	-	-
CG12817	-0.37	-	-	-	23.9 ±0.09	24	1.00	-
CG13305	-0.12	23.4 ±0.11	14	1.00	-	-	-	-
CG1519	-0.48	23.7 ±0.14	12	0.38	23.9 ±0.06	9	1.00	ubiquitin-dependent protein catabolic process ;
CG15431	0.52	-	-	-	23.9 ±0.09	15	1.00	-
CG1632	0.41	23.5 ±0.05	16	1.00	23.8 ±0.05	8	1.00	proteolysis
CG17348	0.35	23.5 ±0.07	14	1.00	23.8 ±0.07	14	1.00	signal transduction ; learning or memory
CG2201	-0.37	23.8 ±0.14	11	0.24	24.1 ±0.11	13	0.85	phosphorylation
CG2812	0.31	-	-	-	23.5 ±0.06	13	0.74	-
CG30456	0.45	23.5 ±0.07	14	1.00	24.1 ±0.08	14	0.85	regulation of Rho protein signal transduction
CG3077	0.34	-	-	-	24.0 ±0.10	14	0.99	cell communication
CG31200	0.40	23.4 ±0.06	15	1.00	24.2 ±0.17	4	0.95	proteolysis
CG31793	0.41	-	-	-	23.8 ±0.07	15	1.00	transport
CG32147	0.38	-	-	-	24.1 ±0.11	14	0.87	proteolysis
CG4250	0.42	-	-	-	23.8 ±0.15	7	1.00	-
CG4328	0.39	23.6 ±0.09	15	0.85	-	-	-	regulation of transcription, DNA-dependent

♂ Gene	PMC	elav>GAL4 τ	n	p	tim>GAL4 τ	n	p	GO Biological Process
CG4656	-0.49	23.8 ±0.14	15	0.31	23.3 ±0.15	12	0.19	negative regulation of signal transduction
CG4806	0.32	23.4 ±0.14	16	1.00	24.1 ±0.03	14	0.60	-
CG4917	0.29	23.6 ±0.09	16	1.00	23.8 ±0.07	12	1.00	-
CG4933	-0.44	-	-	-	24.0 ±0.14	11	0.95	proteolysis
CG5156	0.18	23.4 ±0.06	16	1.00	-	-	-	-
CG5549	0.59	23.6 ±0.07	15	0.94	24.0 ±0.07	15	0.98	-
CG6181	0.47	23.4 ±0.09	15	1.00	-	-	-	-
CG6515	0.42	23.7 ±0.15	15	0.93	23.7 ±0.10	16	1.00	G-protein coupled receptor protein signalling pathway
CG6641	-0.27	-	-	-	24.2 ±0.15	7	0.74	sensory perception of chemical stimulus
CG6649	0.44	23.3 ±0.15	9	1.00	-	-	-	-
CG7018	0.49	23.7 ±0.24	9	0.99	23.7 ±0.11	11	1.00	regulation of transcription, DNA-dependent
CG7156	-0.25	23.6 ±0.06	15	1.00	24.2 ±0.07	15	0.11	cell communication
CG7449	0.46	23.2 ±0.11	10	1.00	23.7 ±0.06	6	1.00	regulation of striated muscle development
CG7717	0.56	23.7 ±0.20	10	1.00	23.9 ±0.07	8	1.00	protein amino acid phosphorylation ; response to oxidative stress
CG7749	0.18	-	-	-	24.1 ±0.29	9	0.81	calcium-dependent cell-cell adhesion
CG7887	0.51	23.7 ±0.13	19	0.83	24.1 ±0.14	5	0.94	G-protein coupled receptor protein signalling pathway
CG8426	0.33	-	-	-	27.3 ±0.36	11	0.00	transcription
CG8548	-0.47	24.0 ±0.14	13	0.05	23.9 ±0.11	11	1.00	protein import into nucleus
CG8772	-0.37	23.9 ±0.12	14	0.11	24.1 ±0.11	9	0.84	locomotor behaviour
CG9153	0.56	23.4 ±0.17	11	1.00	23.6 ±0.14	8	1.00	protein modification process
CG9610	0.27	-	-	-	23.8 ±0.18	12	1.00	regulation of transcription
CG9753	0.38	23.3 ±0.11	15	0.94	24.0 ±0.12	14	0.93	G-protein coupled receptor protein signalling pathway
CG9784	0.44	23.9 ±0.12	14	0.06	24.0 ±0.08	13	1.00	dephosphorylation

In cases in which the hairpin construct had inserted onto the X chromosome, RNAi experiments were performed using virgin female flies:

Table 7-9 Columns as in Table 7-8 but period assessed using virgin female flies as UAS-RNAi transgenes lie on the X chromosome.

♀			*elav>GAL4*			*tim>GAL4*			
Gene	**PMC**	**Transformant**	**τ**	**n**	**p**	**τ**	**n**	**p**	**GO Biological Process**
+	-	6000	24.1 ±0.05	14		24.3 ±0.12	6	-	-
CG10697	0.52	3329	23.9 ±0.11	9	0.38	24.2 ±0.11	28	0.98	learning or memory; dopamine biosynthesis; courtship; eclosion
CG1171	0.37	11352	-	-	-	25.1 ±0.15	10	0.11	neuropeptide signalling pathway
CG17818	-0.51	19089	-	-	-	24.1 ±0.20	14	0.13	signal transduction
CG4259	0.40	2541	24.0 ±0.15	4	0.95	23.8 ±0.09	11	0.30	proteolysis
CG6919	0.41	47895	-	-	-	23.6 ±0.11	8	0.83	G-protein signalling

A number of lines were shown to have significant or nearly significant effects on period as assessed by the conservative Dunnett *post hoc* test (i.e. after normalising for repeated testing), validating our approach for determining gene function. The likely expression of these genes can be assessed using data from the FlyAtlas project.

Table 7-10 Likely expression patterns for genes showing nearly significant effects on circadian period when knocked down. '**Brain**' and '**Eye**' value represent mRNA enrichment in those tissues compared to whole flies. Values called by the Affymetrix analysis software as being up regulated shaded in **green**, down regulated in **red**. The expression pattern of off-target genes was also assessed.

Gene	***elav>GAL4* p**	***tim>GAL4* p**	**Gene symbol**	**Affymetrix probe ID**	**Brain**	**Eye**	**OFF target**	**OFF Brain**	**OFF Eye**
CG1171	-	0.11	*Akh*	1631816_at	23.8	0.03	-	-	-
CG17818	1.00	0.13	*RdgB-beta*	1629552_at	1.6	0.41	*CG5807*	1	0.25
CG8426	-	0.00	*lethal (2) NC136*	1634881_at	1.2	0.4	*CG11695*	2.3	3.12
CG11440	1.00	0.00	*laza*	1635757_at	0.1	15.95	*CG16791*	2.2	2
CG8548	0.05	1.00	*karyopherin-alpha1*	1629920_at	1.5	1.81	*CG13599*	0.1	0.07
CG8772	0.11	0.84	*nemy*	1634658_a_at	0.6	0.14	-	-	-
CG4656	0.31	0.19	*rassf*	1631122_at	1.2	0.48	*CG13109*	2.2	1.12
CG7156	1.00	0.11	-	1628552_at	0.9	0.6	-	-	-
CG9784	0.06	1.00	*IPP*	1636632_at	1.3	1.32	-	-	-

Of the genes causing significant effects in Table 7-8, a number were selected for further analysis to test the accuracy of our candidate. Although a knockdown of a number of genes showed particularly large phenotypic effects (e.g. *CG8426*), further analysis was limited to stocks immediately available, therefore candidates were selected based on their expression patterns within the fly (assessed using data generated by the FlyAtlas gene expression mapping project: Chintapalli et al., 2007).

Three genes were therefore selected; *CG1171* (correlation 0.37, highly enriched in the brain) was deemed the most likely to play a genuine role in the clock, *CG9784* (correlation 0.44, expressed almost constitutively) was selected as a likely output genes as only *elav>GAL4* mediated knockdown had any effect on period, and *CG11440* (correlation 0.41, expressed predominantly in the eyes) as a negative control due to its limited expression pattern. These confirmation experiments were performed using a *UAS>Dicer2* enhancer to further reduce gene expression, thereby potentially magnifying any period differences relative to controls.

Table 7-11 Repeated gene knockdown using the UAS>Dicer2 RNAi enhancer. Although few *w; tim>GAL4, UAS>Dicer2/+; +* control flies were recovered, the *CG1171* knockdown effect is of such a magnitude that one can safely assume it.

Genotype	τ	n	p
♀ ***UAS> RNAi-CG1171/w; tim>GAL4, UAS>Dicer2/+; +***	30.1 ±0.61	12	0.000001
♀ ***w; tim>GAL4(A3)/+; MKRS/+***	24.4 ±0.18	10	0.688172
♀ ***w; tim>GAL4, UAS>Dicer2/+; +***	25.0 ±0.52	5	-
♂ ***w; tim>GAL4(A3)/+; MKRS/+***	23.8 ±0.11	13	0.008655
♂ ***w; tim>GAL4, UAS>Dicer2/+; +***	24.3 ±0.11	23	-
♂ ***w; tim>GAL4, UAS>Dicer2/UAS>RNAi-CG11440; +***	24.3 ±0.14	16	0.970334
♂ ***w; tim>GAL4, UAS>Dicer2/UAS> RNAi-CG9784; +***	24.8 ±0.15	14	0.03432

The results in Table 7-11 are encouraging, showing that knockdown of *CG1171* – the gene judged most likely to play a role in the clock by virtue of its significant enrichment within the brain - caused a highly significant period lengthening effect, individual examples of which are shown in Figure 7-9:

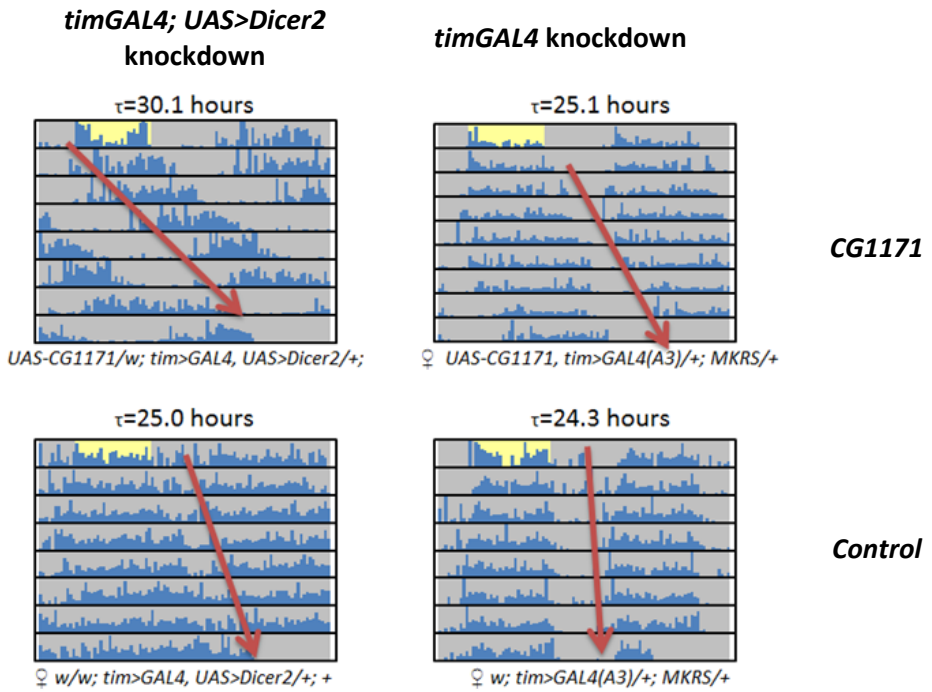


Figure 7-9 Example double plots showing the effects of CG1171 knockdown on individual fly activity in DD conditions. **Red arrows** show period trend, **Top panels** represent experimental lines with (**left**) and without (**right**) the UAS>Dicer2 enhancer, **bottom panels** show the appropriate controls. **MKRS** represents a balancer used in crosses.

The period lengthening effect of knocking down *CG1171*, also known as *Adipokinetic hormone-like (Akh)*, is likely to be specific for two reasons; the hairpin construct has no homology with other known genes (shown by the lack of an OFF target effect as calculated by e-PCR in Table 7-10), and the *Akh* neuropeptide is not part of a cleaved precursor sequence, as shown in Figure 7-10:

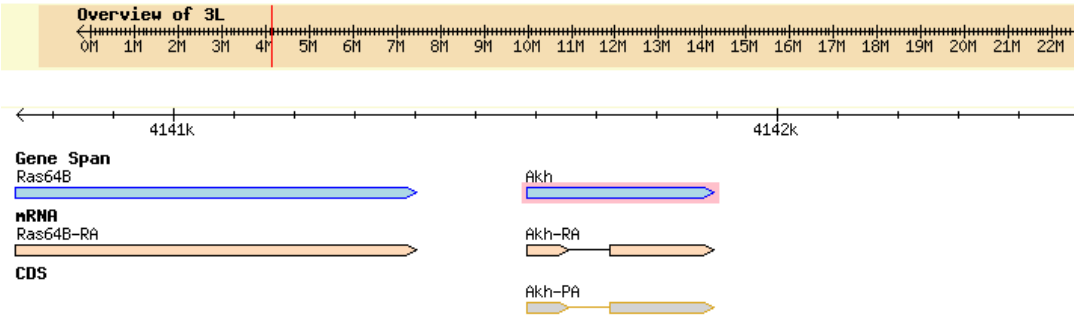


Figure 7-10 The *Akh* peptide is not derived from a precursor mRNA. Figure from FlyBase.

7.4.5 The genetic architecture of *Drosophila* behaviours

This study was conceived not only to identify gene function, but also to address deeper questions concerning how gene networks interact to give rise to behaviour. Behaviours range in complexity depending both on the number of sensory and regulatory stimuli that must be integrated, and the intricacy of the behavioural output. Simple behaviours include geotaxis, phototaxis and high temperature fainting; basic locomotor responses to environmental stimuli. In contrast, complex circadian and mating behaviours require the integration of many sensory and regulatory stimuli and the control of many output modules; in the case of mating, a female fly must hear a male's song, feel his taps and licks, smell his odours and visually evaluate his stature (Sokolowski, 2001).

The diallel crossing scheme employed allows the proportion of phenotypic variance attributable to general (GCA) and epistatic (SCA) effects to be estimated, allowing informed speculation as to the genetic architecture underlying a given behaviour. A pattern emerges on comparing GCA and SCA values for a sample of the behaviours evaluated in this study:

Table 7-12 % variance for a sample of behavioural metrics in transheterozygote flies attributable to GCA and SCA effects.

Phenotype	Geotaxis	Phototaxis	Fainting t1/2	Total activity	Mean waking activity	Mating duration	Circadian period	Mating latency
% variance attributable to GCA	83	69	55	39	26	22	19	7
% variance attributable to SCA	17	31	45	61	74	78	81	93

Table 7-12 shows that the more 'complex' a phenotype, the greater the proportion of the phenotypic variance accounted for by SCA rather than GCA effects. The inference is that complex behaviours such as mating latency are more sensitive to genetic perturbation than simple behaviours as more genetic networks contribute to the expression of such behaviour. The 'broader' the range of networks underlying a behaviour, the greater the chance that any

two mutant alleles will have an epistatic interaction at some point within those networks, therefore the higher the proportion of variance attributable to SCA rather than GCA effects.

Epistatic interactions between alleles are generally specific to a given behaviour, even in the case of geotaxis and phototaxis, two behaviours which are both 'simple' and related to locomotor output. Whilst these two behaviours might therefore be assumed to utilise similar gene networks to generate behavioural output, Table 7-13 shows that the pattern of epistatic interactions differs in each behaviour (note that the magnitude of the epistatic interaction can be directly compared for these two behaviours as both are scored from 1 to 9):

Table 7-13 Comparison of SCA values calculated in half diallel scheme for both phototaxis and geotaxis. Correlation between matrices values = 0.13. Values gauged to be 'significant' (as described in methods) highlighted in blue.

Phototaxis SCA					
	Csp	***trf***	***Itp***	***Pen***	***mth***
nal	0.02	0.03	-0.09	0.08	-0.04
Csp		-0.25	0.02	0.21	0.01
trf			0.22	0.35	-0.35
Itp				-0.58	0.43
Pen					-0.06

Geotaxis SCA					
	Csp	***trf***	***Itp***	***Pen***	***mth***
nal	-0.15	0.21	0.15	-0.11	-0.09
Csp		-0.43	0.15	0.25	0.18
trf			0.12	-0.03	0.13
Itp				-0.16	-0.26
Pen					0.05

One can find only two 'significant' interactions that are preserved in both behaviours: the *Pen/Csp* interaction (which is of similar magnitude in both behaviours) and the *trf/Csp* interaction (which increases from -0.25 in the case of phototaxis to -0.43 in the case of geotaxis). The inference of these results is that the network disruption caused by *Csp*, *trf* and *Pen* alleles must affect locomotor output - the presumed common factor between geotaxis and phototaxis behaviours. However, whilst in geotaxis the *Pen/trf* combination of alleles shows evidence for epistasis, this is not evident when considering phototaxis behaviour, underlining the different connectivity even in these similar behaviours. This is further evidenced by the *Itp/mth* transheterozygote, which shows a positive interaction in phototactic behaviour, but a negative interaction for geotactic behaviour.

7.4.6 The application of architectural information to forward screening

Such information regarding network connectivity can be used to refine the gene expression/phenotype correlation model employed to indentify new genes affecting behaviours. Rather than employing either linear or maximum difference correlation models, knowledge of epistatic interactions suggests that comparisons between *Csp/trf*, *Csp/Pen, trf/Pen* and the grand mean 'soup' expression of genes might be most informative for unravelling the control of directed locomotor behaviour, whilst the strains showing epistatic interactions for phototactic but not geotactic behaviour (*trf/Itp*, *trf/Pen*, *Itp/Pen*) might be most informative for identifying genes playing specific roles in phototaxis, presumably those acting in phototransduction.

The different methods of identifying candidate geotaxis genes were compared to determine whether there might be any overlap in the genes identified by different methods, allowing one to confidently predict a role in geotaxis (in a similar manner to that employed comparing circadian microarray studies in chapter 6). Unfortunately none of the methods employed were able to identify any of the genes known to act in geotaxis:

Table 7-14 Genes identified as correlating with geotaxis using different methodologies. '**Published**' genes are those genes already implicated in geotaxis, as assessed experimentally by Toma *et al.* (Toma et al., 2002) and Armstrong *et al.* (Armstrong et al., 2006), '**Informed**' genes are those showing maximally different expression in the *Pen/Csp* and *trf/Csp* lines shown to interact in Table 7-13, '**Linear**' genes are those showing the best correlation when considering all transheterozygotes, and '**Max difference**' are those genes showing the best correlations when considering only the transheterozygotes falling into the top or bottom 3 by phenotypic score. Only the top 26 candidates are presented here. No one gene is identified by more than one method.

Published	Informed	Linear	Max difference
FBgn0000210	FBgn0051938	FBgn0035205	FBgn0039882
FBgn0004839	FBgn0038902	FBgn0035206	FBgn0034510
FBgn0003715	FBgn0052815	FBgn0001108	FBgn0052233
FBgn0003996	FBgn0004513	FBgn0021796	FBgn0032127
FBgn0039464	FBgn0033810	FBgn0040227	FBgn0038053
FBgn0031079	FBgn0034966	FBgn0034858	FBgn0011746
FBgn0034570	FBgn0052685	FBgn0030377	FBgn0033811

Published	Informed	Linear	Max difference
FBgn0025680	FBgn0026170	FBgn0034258	FBgn0038156
FBgn0011817	FBgn0033633	FBgn0035207	FBgn0024923
FBgn0023178	FBgn0050408	FBgn0035979	FBgn0002531
FBgn0260753	FBgn0043534	FBgn0000241	FBgn0031853
FBgn0011823	FBgn0032612	FBgn0034643	FBgn0038016
FBgn0052423	FBgn0023537	FBgn0036907	FBgn0038202
FBgn0045842	FBgn0040522	FBgn0034420	FBgn0031249
FBgn0035142	FBgn0050145	FBgn0037249	FBgn0001079
FBgn0023520	FBgn0004243	FBgn0032513	FBgn0020429
FBgn0001624	FBgn0031901	FBgn0038574	FBgn0039313
FBgn0014859	FBgn0029769	FBgn0037303	FBgn0011769
FBgn0052598	FBgn0028855	FBgn0015277	FBgn0032045
FBgn0001297	FBgn0011836	FBgn0000256	FBgn0037615
FBgn0035989	FBgn0015567	FBgn0034504	FBgn0030846
FBgn0001981	FBgn0040256	FBgn0022702	FBgn0038036
FBgn0044323	FBgn0029922	FBgn0015576	FBgn0040971
FBgn0052100	FBgn0013988	FBgn0000416	FBgn0016762
FBgn0039553	FBgn0058081	FBgn0014877	FBgn0039733
FBgn0086686	FBgn0032290	FBgn0035633	FBgn0034885

Selecting the top 26 genes implicated by three different methods of correlating geotaxis behaviour with gene expression failed to re-identify any of the known geotaxis genes. Whilst there is no reason to suppose that the list of 26 known geotaxis genes is in any way comprehensive, nor that there is any statistical significance in using the top 26 candidates identified by each method, the lack of overlap between each method raises the question as to which is the most accurate and appropriate, a question that can only be answered by further experimentation.

7.5 Discussion

From the outset it was unclear whether the creation of transheterozygote lines in controlled genetic backgrounds would sufficiently disrupt genetic networks to allow the structure and components of such networks to be elucidated. The results in this chapter, representing as they do analysis of only a small subset of the data generated in this study, are a comprehensive vindication of the methodology employed, revealing not only the underlying patterns of connectivity between various *Drosophila* behaviours and identifying novel components of such behavioural networks, but developing the methodology by which the function of yet more genes might be determined.

7.5.1 The more complex a phenotype, the broader the network

A key tenet of the original hypothesis was that analysis of behaviour, being exquisitely sensitive to genetic variation due to the complex integration of multiple pathways required to generate a coherent output, would be most informative when using only the mildest genetic perturbation in otherwise equilibrated genetic backgrounds (Greenspan, 1997). Not only did the results confirm this, but also revealed that not all behaviours are equal; Table 7-11 showing that logical *a priori* assumptions as to the likely 'complexity' of a behaviour are borne out by experimental evidence. This suggests that the more complex a behaviour, the more bio-behavioural modules must be co-ordinated to generate the response (Tully, 1996), therefore the larger that behaviour's genetic hinterland. The larger the hinterland, the greater the chance that network perturbation brought about by two random genetic mutations might overlap in an epistatic fashion; the sensitivity of behaviour (and its resistance to canalising factors) then ensures that this interaction is captured experimentally.

This study also confirms the general observation that epistatic interactions (as calculated by GCA/SCA statistics) are specific with respect to the behaviour being studied and the genetic background in which the work is performed. However, this study also reveals that - in the case of behaviours assumed to regulate the same bio-behavioural modules (the

example being geotaxis and phototaxis) - some epistatic interactions are preserved; information which can inform the identification of candidate genes. Although several variants of Toma *et al.'s* candidate identification strategy were used to analyse transheterozygote geotaxis data generated in this study, none were able to identify known geotaxis genes amongst the top 26 candidates.

7.5.2 Evaluating different methods of identifying gene function

To distinguish between the various methods by which gene expression and behaviour might best be correlated, the roles of the candidate genes identified by each method (such as those listed in Table 7-14) must be assessed experimentally using knockdown or knockout of gene function to determine the percentage of false positive results inherent to each method. Although this chapter has placed emphasis on the simple behaviours of geotaxis and phototaxis, measurement of both behaviours is both laborious and slow, therefore neither behaviour is an appropriate system suitable to discriminate between the theoretical models by which gene function can be identified.

The ideal test bed system must be a behaviour that is rapidly measured in an automated fashion, as well as being particularly sensitive to genetic perturbation. Ideally several behaviours might be identified that overlap to a greater or lesser extent, allowing interactions between the behaviours to be studied as a measure of network breadth.

One such behaviour is sleep. The analysis of sleep is both conceptually and mathematically simple, requiring only that a researcher (or software script) examines a data series in chronological order to identify episodes of consolidated rest. 'Sleep' comprises many behaviours such as sleep bout number and total sleep which, whilst correlated, are genetically distinct (Harbison and Sehgal, 2008), representing a number of overlapping networks. In addition to the technical advantages of analysing sleep, the field of sleep research in *Drosophila* is still very young, holding significant potential for the discovery of novel genes.

The development and application of a novel sleep analysis package to answer such questions is discussed in detail in the next chapter.

7.5.3 Gene expression/phenotype correlations can identify novel genes

Despite there being reservations as to the most appropriate method to be used for identifying gene function, by using a linear correlation method this study was able to identify a number of candidate genes involved in the circadian clock. Candidate gene knockdown suggested that a number of the correlations observed had a causative relationship with changes in the period of the circadian clock. It is difficult to assess the false positive rate as the degree of gene knockdown was not measured; further experiments must evaluate this using real time PCR. One further source of error was that period/gene expression correlations were calculated using data from female flies, while gene knockdown experiments were performed in males (due to the difficulty of collecting sufficient virgin flies to perform a large scale screen); as is evident in Figure 7-7, circadian period is subject to strong sex-specific effects.

In order to further test the reliability of our screening approach, three genes were re-tested by enhancing the degree of gene knockdown using a *UAS>Dicer2* transgene. By incorporating information as to the likely expression pattern of each gene, it was confirmed that both those genes expressed in the head showed statistically significant period differences from controls (particularly in the case of *Akh*), further validating our approach to assaying gene function.

7.5.4 A circadian role for *Akh*?

Akh is highly enriched in the fly brain compared to the body as a whole, and like other neuropeptides is thought to be multifunctional, acting in the regulation of both intermediary metabolism in the fat body (in a fashion functionally analogous to vertebrate glucagon: Van der Horst et al., 2001), but also in mediating the hyperactive behavioural response to starvation (Lee and Park, 2004). AKH has been shown to have direct excitatory effects on

motor neurons in moths, and projections emanating from the AKH neurons indicate that AKH has multiple target tissues, including the crop and brain (presumably near motor centres) in adult flies (Lee and Park, 2004). It is therefore reasonable to suppose that *Akh*, although it shows no evidence for rhythmic expression (Wijnen et al., 2006), mediates some aspects of circadian locomotor behaviour. However, there are a number of contraindications, notably data from a previous study by Lee and Park examining the distribution and function of AKH (Lee and Park, 2004). In accordance with earlier work (Noyes et al., 1995), Lee and Park, using a mixture of *in situ* hybridisation, imunnocytochemistry and transgenic expression of *LacZ* using an *Akh*>GAL4 promoter construct, showed that *Akh* expression in both the larva and the adult is restricted to the corpora cardiac, an important endocrine structure in the cyclorraphous Diptera.

Such results are difficult to reconcile with the *tim>GAL4* mediated knockdown data generated in this study. The identity of the RNAi construct used in this study must be confirmed by direct sequencing, and the levels of *Akh* mRNA should be directly assessed by RT PCR in knockdown strains. RNAi experiments must also be repeated using different hairpin constructs to experimentally rule out off target knockdown effects, whilst *Akh*'s role in the circadian clock might be further probed by knocking down expression using the more specific *Pdf>GAL4* which is restricted to expression in the core clock cells (PDF being the only neuropeptide currently known to be involved in the clock, see introduction). These confirmation experiments have already been initiated.

Although Lee and Park showed in their paper that ablation of *Akh* expressing cells is neither lethal, nor sufficient to induce behavioural arrhythmicity, they were able to show that *Akh* expressing cells synapse in the brain, though the resolution of images precluded more accurate identification. Such synapses might be the key to reconciling the data from Lee and Park with the results in this chapter.

7.6 Conclusion/further work

The elucidation of gene function remains one of the central problems of the genomic era, particularly in light of evidence showing that genes form robust expression networks refractory to study using conventional forward screening techniques (as is perhaps evidenced by the difficulty in assigning a function to CCCG's in the previous chapter).

This chapter has developed the techniques first used by Toma *et al.* to determine gene function, and by using only a small sample of the full dataset has determined that the neuropeptide *Akh* is likely to play a role in the circadian clock. However, neither circadian nor geotaxis behaviours are appropriate to test the accuracy of the new methods to determine gene function detailed in this chapter. To do this, one requires a sensitive behavioural output that can be recorded quickly and reliably, the most obvious candidate being sleep associated behaviours.

The development of sleep analysis software and its application to both the transheterozygote dataset and the problems of determining gene function are described in the following chapter.

8 Sleep Follow-up

8.1 Introduction

In the previous chapter, a novel approach to determining gene function was developed in which gene networks were disrupted by subtle mutations, causing changes in patterns of gene expression which resulted in disrupted behavioural output. By correlating these changes, it was possible to generate lists of genes likely to play a role in mediating a given behaviour. A number of alternate correlation strategies were developed, the relative utility of which could be tested by experimental validation of candidate gene function.

The most appropriate behaviour for such a study must be easy to measure, sensitive, and amenable to high throughput studies. Sleep not only fulfils these criteria, but due to the small number of groups working in the field also offers great potential for the discovery of novel genes. Due to the dearth of reviews of sleep in *Drosophila*, this topic will be introduced in some depth.

8.1.1 What is sleep?

Despite decades of research, there is not yet any functional definition for the state of 'sleep', in part due to ongoing debate as to the true purpose of sleep. Instead sleep is defined using both behavioural and electrophysiological criteria, drawing on Pieron's early physiological studies of sleep (Pieron, 1913). The generally accepted criteria for identifying sleep (Campbell and Tobler, 1984) include:

1. Sleep is a period of behavioural quiescence (which in some - but not all - species is consolidated into long bouts).
2. During a period of sleep an organism has an increased arousal threshold (*i.e.* a decrease in sensory responsiveness).
3. Sleep, as opposed to hibernation or trauma-induced coma, is rapidly reversible.
4. Sleep is homeostatically regulated such that a period of sleep loss is later compensated for by a period of increased, or more intense, sleep.

5. Sleep may involve the adoption of a stereotypical posture and/or the use of a specific sleep site.

Sleep is regulated both by the circadian clock and by a homeostatic sleep drive. When a subject is deprived of sleep for a sufficient length of time, the homeostatic 'sleep pressure' builds up until brain activity characteristic of sleep will start to 'leak' into periods when the subject appears to be awake. Such periods of sleep are termed 'microsleeps' and, together with the symptoms of "sleepiness", are thought to be a large contributory factor to accidents in the workplace and on the road (Rosekind, 2005).

Indeed, whilst we are accustomed to thinking of sleep as period of around eight hours, insomnia affects approximately one third of the adult American population (Roth, 2005). Sleep disorders form a heterogeneous group, and insomnia may arise though compromised sleep onset, sleep maintenance or terminal insomnia, implying that the genetic networks underlying sleep are complex. Although the genetic basis for sleep phenotypes is confirmed by the existence of inherited sleep disorders such as familial advanced sleep phase syndrome (FASP), there is a high degree of inter-individual variation in sleep metrics (Andretic and Shaw, 2005), emphasising the role of environmental and stochastic factors in controlling sleep.

8.1.2 Is rest in *Drosophila* an appropriate model for sleep?

Although *Drosophila* has long been a model for both simple and complex behaviours (Sokolowski, 2001), the differences in brain architecture and neurochemistry between higher vertebrates and invertebrates (Hartse, 1994) lead many to conclude *Drosophila* were an inappropriate model for sleep research, lacking as it does neurotransmitters such as hypocretin which are essential for mammalian sleep (Lin et al., 1999).

However, in a landmark paper, Hendricks *et al.* showed that rest in *Drosophila* is a sexually dimorphic period of decreased sensory responsiveness driven by a homeostatic pressure (Hendricks et al., 2000). Flies, like humans, are active during the day, and the circadian system is responsible for consolidating most of their sleep during the night.

Furthermore, using video evidence they showed that sleeping *Drosophila* adopt stereotypical postures during sleep, prefer to sleep close to a food source, and show evidence of small, sporadic movements of no apparent purpose during sleep in a similar way to mammals (Hendricks et al., 2000). They concluded that rest in *Drosophila* is indeed a sleep-like state, and by showing that caffeine decreased sleep, suggested that it shared aspects of mammalian sleep neurochemistry.

Contemporaneously Shaw and colleagues revealed further neurochemical similarities between *Drosophila* and mammalian sleep by showing that the drug Hydroyzine could increase sleep without changing the baseline sensitivity of flies to arousing stimuli (Shaw et al., 2000). This work also showed that, as in mammals, old flies sleep less than young flies, and later work has shown that older flies also show increased sleep fragmentation and less robust sleep/wake cycles (Koh et al., 2006).

More recent studies have elaborated the similarities between mammalian and *Drosophila* sleep, showing that sleep deprivation leads to an increase in sleep intensity (measured using a sleep fragmentation index - Huber et al., 2004) and induces cognitive impairment as in mammals (assessed using a T-maze learning paradigm: Seugnet et al., 2008). Although flies do not show the slow wave and rapid eye movement sleep traces typical of mammalian sleep electroencephalogram recordings (Aserinsky and Kleitman, 1953), they do show clear electrophysiological correlates of rest and activity within the mushroom bodies (Nitz et al., 2002). Taken together therefore, one can conclude that rest in *Drosophila* shares many of the phenotypic and neurochemical characteristics of mammalian sleep, and is therefore an appropriate model in which to study sleep.

8.1.3 Is sleep essential?

Given the extent of inter-individual variation in sleep, is sleep truly essential? In their comprehensive review of sleep literature, Cirelli and Tononi make a convincing case for the

essential role of sleep by refuting three commonly used counter arguments (Cirelli and Tononi, 2008):

1. There is no good evidence for animals that do not require sleep; those which appear continually conscious show evidence of microsleeps or are capable of resting alternate parts of the brain (unihemipsheric sleep). Indeed, even simple organisms such as *C. elegans* display a sleep-like 'lethargus' state coupled to important developmental stages (Raizen et al., 2008), suggesting that sleeping is a behaviour deeply rooted in evolution.
2. It is impossible to completely deprive an animal of sleep for more than 24 hours (Cirelli and Tononi, 2005), and any period of sleep deprivation is always compensated for by a period of homeostatically controlled sleep 'rebound', the implication being that there is a physiological requirement for sleep.
3. Sleep deprivation is lethal in a number of models, including in flies (Shaw et al., 2002) and rats (Rechtschaffen et al., 1983), but not pigeons (Berger and Phillips, 1994), though it has not been conclusively proven that this is not the result of the arousing stimulus causing physiological stress.

Accumulating evidence links sleep with longevity and health; *Drosophila* lifespan has been shown to correlate with sleep time (Pitman et al., 2006), and independent mutations causing short sleep phenotypes display significantly reduced longevity (Cirelli et al., 2005a; Yuan et al., 2005), suggesting a causative rather than correlative or pleiotropic relationship. Furthermore, pharmacological induction of sleep slows cognitive decline in a transgenic mouse model of Huntington's disease (Pallier et al., 2007). There are therefore many lines of evidence to support the conclusion that sleep is indeed essential for proper bodily function.

8.1.4 Does sleep have a function dissociable from rest?

Given the seemingly essential nature of sleep, a number of questions emerge as to the function of sleep. Did such a function emerge early in evolution (as suggested by its apparent

universality even in simple organisms), or has 'sleep' evolved independently several times (supported by the dramatically different sleep phenotypes observed in different species). Additionally, what features of sleep's function are dissociable from waking rest?

Early taxonomic surveys revealed a negative correlation between body weight and sleep (Campbell and Tobler, 1984), suggesting that sleep might have evolved as a mechanism to balance metabolic budgets. However, sleep saves little energy compared to minimal waking activity (Zepelin and Rechtschaffen, 1974), therefore alternate models have been developed considering cost/benefit analysis and predator/prey dynamics in the context of sleeping behaviour. None of these models are entirely convincing; the environmental disconnection essential to sleep remains a stumbling block not explained by such hypotheses.

More recent studies have used microarrays in an effort to directly determine the function of sleep. In mice, rats and hamsters sleep upregulates the expression of genes involved in synaptic depotentiation and lipid metabolism (Cirelli and Bushey, 2008 and references therein), processes required for the synthesis and maintenance of cell membranes. A recent publication supports these observations by showing that synaptic strength correlates with the amount an animal sleeps in the 6 hours prior to sampling in both mice (Vyazovskiy et al., 2008) and flies (Gilestro et al., 2009).

Similarly, two studies measuring gene expression in waking, sleeping, sleep deprived and mechanically stressed controls in *Drosophila* (Cirelli et al., 2005b; Zimmerman et al., 2006), revealed between 12 and 23 transcripts whose levels were upregulated during sleep, including genes involved glial and lipid metabolism, membrane recycling, nervous system development and actin reorganisation.

Behavioural evidence also links sleep related changes in synaptic plasticity with learning and memory; sleep deprivation is associated with cognitive impairment in Humans (Belenky et al., 2003), flies (Seugnet et al., 2008), birds (Rattenborg et al., 2004) and rodents

(Tartar et al., 2006), whilst performance in certain types of memory tasks improves following even short periods of sleep (Korman et al., 2007).

The 'synaptic homeostasis' hypothesis (an important complement to the rules of Hebbian learning in neurobiology - Turrigiano and Nelson, 2000; Tononi and Cirelli, 2006) suggests that neuronal plasticity is a finite resource, therefore during sleep there is a general depotentiation of synapses within the brain in order to prevent saturation of the brain's capacity to learn (Griffith and Rosbash, 2008). Environmental disconnection during sleep (as opposed to rest) is therefore the essential component that prevents runaway formation of new synapses. Under such a model, only the strongest synapses are preserved during sleep, whilst dreaming might be interpreted as exercising longer term memory (Kavanau, 1997). Supporting such a model, wild type flies that have been socially enriched or subjected to a memory formation experiment require more daytime sleep than control flies (Seugnet et al., 2008), yet learning and memory mutants do not show the same increases in sleep. Crucially mammals and flies show many similarities at the synaptic level (Gilestro et al., 2009 and references therein).

Failure to depotentiate synapses after a period of waking activity (i.e. sleep deprivation) would therefore be expected to represent a cellular stress due to the energy cost of synthesising and maintaining new synapses, as has been partially demonstated in the mouse model (Benington and Heller, 1995). Extended waking activity in flies elevates transcription of immune response genes (Cirelli et al., 2005b; Zimmerman et al., 2006) and the stress marker *BiP/GRP78* (Shaw et al., 2000), levels of which accumulate in response to sleep loss (Naidoo et al., 2007). Similarly in the mouse model sleep deprivation induces the unfolded protein stress response pathway within the cerebral cortex (Naidoo et al., 2005). The hypothesis that sleep deprivation represents a cellular stress has been developed in the 'free radical flux' model of sleep (Reimund, 1994; Inoue et al., 1995), and is supported in the

literature by some evidence showing that oxidative stress may be a key factor affecting age dependent changes in sleep profiles (Koh et al., 2006).

Although it is attractive to conclude that sleep's primary function is to regulate synaptic plasticity, microarray data shows that sleep has much wider transcriptional ramifications, and may therefore have a pleiotropy of functions, as evidenced by the differences in thermoregulation, brain metabolism and activity observed between REM and NREM sleep in mammals.

8.1.5 Sleep and the circadian clock

As might be expected, the sleep and circadian systems have considerable neuroanatomical and neurohormonal overlap, as both must be integrated to control locomotor output. Consequently mutations in the cAMP and EGFR signalling pathways generally affect both sleep and circadian phenotypes (Yuan et al., 2005; Foltenyi et al., 2007b).

The functions of sleep are clearly gated by the circadian clock, and conversely arrhythmic animals retain a homeostatic need for sleep (Larkin et al., 2004). Intriguingly arrhythmic *Drosophila* females carrying mutations in the core clock gene *cyc* are particularly sensitive to the lethal effects of sleep deprivation, and permanently increase baseline levels of sleep in response to even short periods of sleep deprivation. However, this has shown to be a clock independent effect whereby *cyc* mutants are unable to induce the expression of the heat shock protein *Hsp83* (which normally mediates the response to sleep deprivation, further supporting the hypothesis that extended wakefulness causes some degree of cellular damage), though the sex specific nature of this effect remains unexplained.

It appears that the key region of neuroanatomical overlap may be the l-LN_v cells, which express the canonical components of the circadian clock but have limited known function within the clock mechanism itself. These cells have recently been shown to express the transcription factor *Atf-2* (which positively regulates sleep time as well as affecting sleep

rebound and arousal thresholds, possibly by a CRE mediated mechanism - Shimizu et al., 2008), as well as the GABA receptor subunit *Rdl,* also shown to play a role in sleep regulation (Sheeba et al., 2008b). l-LN_v cells also express the circadian neuropeptide *Pdf*, mutants of which are no longer able to restrict sleep predominantly to the subjective night phase, reinforcing the conclusion that the distribution of sleep is mediated by the circadian clock rather than directly responsive to light. The exact function of these neurons in regulating sleep is not yet clear, and represents a field of intense research.

A recent study has also implicated the LN_vs as an important centre mediating the effects of sleep deprivation on learning and memory (Donlea et al., 2009), though misexpression studies have long implicated the mushroom bodies as the nexus for regulation of learning and memory (McBride et al., 1999), sleep (Pitman et al., 2006; Joiner et al., 2006), and dopamine mediated changes in arousal (Andretic and Shaw, 2005; Andretic et al., 2005; Kume et al., 2005). These data suggest that the regulation of sleep depends on a plethora of neurons throughout the brain.

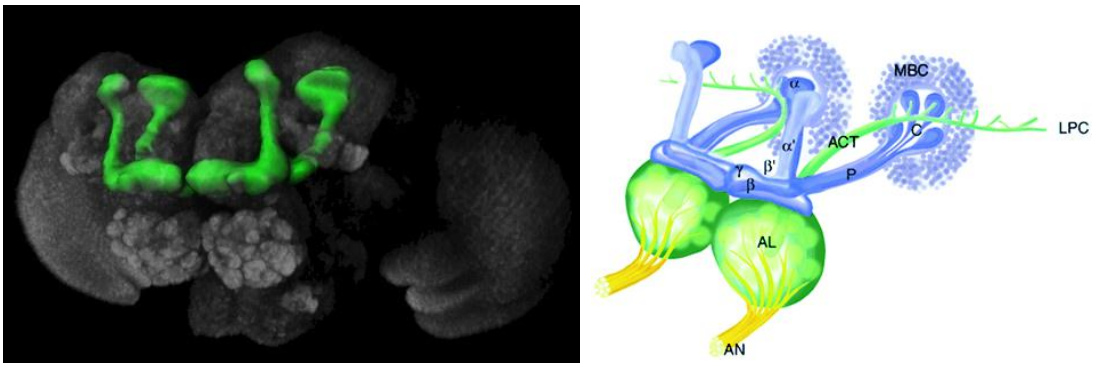


Figure 8-1 The *Drosophila* mushroom bodies. Left panel: confocal image of the *Drosophila* brain showing the mushroom body complexes (MBCs) in **green**, from Baier *et al.* (Baier et al., 2002). **Right panel**: a schematic of the region highlighted in left panel showing the MBC (**blue**) are heterogeneous structures, forming a number of lobes designated α, β, γ and δ. Figure from McGuire *et al.;* for further details see source paper (McGuire et al., 2001).

8.1.6 Neurochemical control of sleep in the *Drosophila* Brain.

It is clear that *Drosophila* share some of the neurochemical features of sleep with mammals, including sensitivity to drugs such as caffeine and histamines (Shaw et al., 2000).

The main neurochemical implicated in the regulation of *Drosophila* sleep is dopamine; flies carrying the *fumin* mutation of the *dopamine transporter* gene show enhanced sensitivity to mechanical stimuli and prolonged activity once aroused, and unusually do not show a significant sleep rebound in response to sleep deprivation (Kume et al., 2005). Further experiments have shown that sleep-loss induced learning impairments can be mitigated by activation of the dopamine receptor dDA1 within the mushroom bodies (Seugnet et al., 2008), suggesting sleep, arousal and learning processes are all regulated within this region of the brain by dopamine.

In addition to dopamine, a number of other neurotransmitters are known to play a role in sleep, including amylase (levels of which correlate with sleep need - Seugnet et al., 2006), the cAMP-PKA-CREB pathway (Hendricks et al., 2001), the EGFR pathway (Foltenyi et al., 2007b), serotonin (receptors for which are required in the mushroom bodies for normal sleep - Yuan et al., 2006) and GABA (Parisky et al., 2008; which may control sleep latency - Agosto et al., 2008).

Electrochemical signalling between cells also plays an important role in regulating sleep, as evidenced by mutations in the fast-inactivating potassium channel *Shaker* (Cirelli et al., 2005a), and its ß-modulatory subunit *Hyperkinetic* (Bushey et al., 2007) showing short sleeping phenotypes. A recent report identified *Sleepless (sss)* as a signalling molecule that may positively regulate *Shaker* expression (Koh et al., 2008).

8.1.7 The genetic architecture of sleep

Long running forward mutagenesis screens have identified many mutants that affect sleep length, but very few that affect sleep rebound (Harbison and Sehgal, 2008; Cirelli and Bushey, 2008). Furthermore, mutations that affect total sleep do not necessarily affect sleep rebound or latency, and *vice versa* (for discussion, see Bushey et al., 2007). This suggests both that sleep involves the coordination of many disparate biological processes, and that while the

functions performed during sleep can vary in speed or efficiency (therefore affecting sleep time), the homeostatic drive for sleep is essential to the organism.

8.2 Methods

8.2.1 The Analysis of Sleep

8.2.1.1 Development of a new sleep analysis software tool

Different groups have approached the analysis of sleep in different ways, including the development of Matlab scripts (Yuan et al., 2006), Statsoft add-ins (Huber et al., 2004), C++ programs (Harbison and Sehgal, 2008), Microsoft Excel visual basic scripts (Andretic and Shaw, 2005), and most recently the pySolo Python package (Gilestro and Cirelli, 2009). With the exception of the pySolo package, the existing sleep analysis packages have poor graphical output capabilities and/or restrictive requirements on the format of input data that make them difficult to use. The field has not yet established a common software tool.

For these reasons it was decided to develop a set of sleep analysis tools within the BeFly! Package that would be both easy to use, and provide a clear graphical output within the Microsoft Excel environment most users are familiar with.

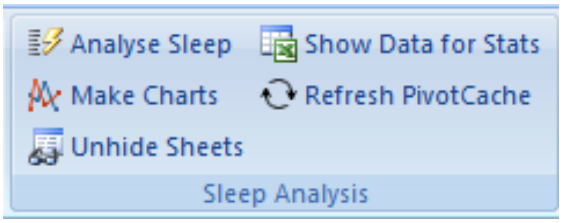


Figure 8-2 The BeFly! Sleep Analysis tools group on the BeFly! RibbonUI tab.

Due to the diversity of sleep analysis tools, there is also a diversity of metrics used within the field for analysing sleep. Many of these metrics describe very similar aspects of sleep, and are therefore very closely correlated, though not interchangeable (Harbison and Sehgal, 2008). The BeFly! Package implements most of these metrics, including;

1. ***Bouts of sleep (/unit time)*** – This is a measure of the number of sleep episodes or 'bouts' within the specified time interval. If a bout extends over two or more time intervals, it will be counted as a separate bout of sleep in each interval rather than one

complete bout - the standard way to treat sleep bouts in the field (Paul Shaw, pers. comm.). For this reason it is useful to analyse sleep not only on a 'per hour' basis, but also at coarser resolution (e.g. per day) to get a more accurate measure of sleep bout number when sleep bouts are long.

2. ***Longest Sleep Duration (Mins)*** – This metric is a measure of sleep regulation as it reflects the ability of a fly to maintain sleep.
3. ***Longest Sleep Start (Bin)*** – This metric is used by some investigators in the field as a read-out of sleep regulation and its relationship to the circadian clock and environmental light/dark factors.
4. ***Mean Sleep Bout Length (/unit time) –*** This metric measures sleep fragmentation and is more accurate when considered at coarser resolutions due to the problems associated with analysing sleep episodes that are longer than one hour.
5. ***Mean Inter-bout interval (Min/unit time)*** – An alternate measure of sleep fragmentation.
6. ***Mean Waking Activity/Bin (/unit time)*** – This does not correlate with metrics measuring sleep, and is therefore used as a indication of fly health, being used to identify hyper- and hypo-active flies. It is a particularly useful metric when studying neurodegeneration phenotypes.
7. ***Night Offset (Min)*** – This measures the time between the environmental lights being turned off, and the first bout of sleep beginning. This metric is thought to assess aspects of the regulation of sleep (Agosto et al., 2008), and is also be referred to as 'sleep latency'.
8. ***Total Activity (/unit time)*** – This is the standard metric for locomotor activity analysis and is used to draw actograms.
9. ***Total Inactivity (/unit time)*** – This metric can be used to analyse periods of inactivity as well as sleep (which is a sustained period of inactivity).

10. ***Total Sleep (/unit time)*** – This is the standard metric for sleep analysis; in this implementation if a bout of sleep starts whilst the lights are on and continues into the dark period without interruption, then the entire bout will be counted as occurring during the lights on period.

8.2.1.2 Methodological considerations when examining sleep in Drosophila

When using Trikinetics equipment to study sleep, care must be taken not to use sum data over too large or too small a timeframe, as this can lead to both either under- or over-representation of the true sleep phenotype compared to high resolution video analysis of sleep (or Zimmerman et al., 2008; i.e. type 1 and 2 errors - for discussion see Andretic and Shaw, 2005). After a five minute period of inactivity *Drosophila* exhibit the increased arousal threshold indicative of sleep (Shaw et al., 2000), therefore in this study data were recorded at a resolution of 5 minutes.

Existing work has shown that a number of factors must be controlled when performing sleep experiments in *Drosophila*:

8.2.1.2.1 Biological factors

The age (Shaw et al., 2000), sex (Helfrich-Forster, 2000) and genetic background (Andretic and Shaw, 2005) of flies are all important factors that determine *Drosophila* activity and sleep. As sleep experiments are carried out under LD conditions, mutations which change flies' responses to light (e.g. naturally occurring rhodopsin Rh6 null mutants:Cook et al., 2003), or alter their baseline activity (e.g. couch potato: Bellen et al., 1992) may easily be confused with *bona fide* sleep mutants. The degree of anaesthesia used to collect flies should therefore also be minimised; ideally a mouth pipette should be used to transfer flies to activity tubes without anaesthesia (Andretic and Shaw, 2005).

8.2.1.2.2 Environmental Factors

Social enrichment (Ganguly-Fitzgerald et al., 2006), lighting conditions, ambient temperature and ambient humidity all affect *Drosophila* entrainment and activity, and must therefore be maintained as constants during an experiment.

8.2.2 Measuring sleep

A complete, step-by-step manual explaining the use and pitfalls of sleep analysis software is included in appendix D4.

In this experiment, transheterozyous flies were aged for at least 3 days prior to activity experiments. Flies were loaded into activity tubes and were subjected to 2 full days of LD 12:12 conditions by Rozi Andretic in the Greenspan laboratory between October and November 2002. Activity data was recorded using Trikinetics Drosophila Activity Monitors (DAM5, http://www.trikinetics.com/). Beam crossing data was collected in five minute bins. Follow-up experiments were performed in the same manner by the author.

Sleep was analysed using the BeFly! Behavioural analysis program v7.15 by the author. Sleep metrics were calculated for individual flies to identify aberrant and dead flies, and such flies were excluded from further analysis. The default sleep analysis parameters were modified to ignore flies showing less than 250 beam crossing events per day, and the analysis repeated to calculate the mean of baseline sleep for each transheterozygote genotype.

Genotypic mean data was correlated with microarray gene expression data as detailed in the previous chapter. Statistics were performed using STATISTICA.

8.2.3 GAL4 driver mediated gene knockdown

The function of genes correlating with sleep metrics was examined using GAL4/UAS mediated RNAi as described in the previous chapter. The expression of RNAi inducing hairpin constructs was mediated by a number of drivers expressing GAL4 in neurons implicated in the regulation of sleep, as shown in Table 8-1. Although the *Ddc>GAL4* driver is known to reduce

lifespan in flies (Haywood et al., 2002), this effect occurred at timescales longer than those used in these experiments, and is not thought to affect results.

Table 8-1 GAL4 drivers known to drive expression in sleep-associated neurons used in this study.

Driver	Expression Pattern	Stock Source	Creator	Sleep Publication
Ddc	Dopaminergic neurons	Bloomington 7009	(Li et al., 2000)	(Andretic and Shaw, 2005)
c687	*Pars intercerebralis*	D.Armstrong	www.fly-trap.org	(Foltenyi et al., 2007a)
Th	Serotonergic neurons	J.Wagner	(Friggi-Grelin, F. *et al.*, 2003)	(Yuan et al., 2006)
c767	*Pars intercerebralis*	D.Armstrong	www.fly-trap.org	(Foltenyi et al., 2007a)
c309	Mushroom bodies	Bloomington 6906	www.fly-trap.org	(Nitz et al., 2002)
tim	Clock cells, eyes	J.Blau	(Blau and Young, 1999)	(Parisky et al., 2008)

As behaviour is particularly sensitive to genetic background effects, all hairpin constructs were obtained from the isogenic VDRC collection (Dietzl et al., 2007), and all GAL4 driver lines (except *tim>GAL4*) were backcrossed into the same inbred Canton-S line for a minimum of six generations. All GAL4 experiments were analysed relative to their respective GAL4/+ control.

8.2.4 *UAS>Dicer2* enhanced gene knockdown.

The *UAS>Dicer2* transgene was crossed into the GAL4 driver backgrounds listed in Table 8-1 using standard crosses and balancers appropriate to the chromosomes being manipulated (example shown in Figure 8-3).

♂ w^{1118} ; *c309>GAL4* ; + x ♀ $\frac{w^{1118}, P\{UAS\text{-}Dcr\text{-}2.D\}1}{w^{1118}, P\{UAS\text{-}Dcr\text{-}2.D\}1}$; $\frac{hs\text{-}hid}{CyO}$; $\frac{+}{+}$

♂ $\frac{w^{1118}, P\{UAS\text{-}Dcr\text{-}2.D\}1}{}$; $\frac{c309>GAL4}{CyO}$; $\frac{+}{+}$ x ♀ $\frac{w^{1118}, P\{UAS\text{-}Dcr\text{-}2.D\}1}{w^{1118}, P\{UAS\text{-}Dcr\text{-}2.D\}1}$; $\frac{hs\text{-}hid}{CyO}$; $\frac{+}{+}$

Figure 8-3 UAS-Dicer2 crossing scheme for 2nd chromosome GAL4 driver. The progeny of each cross were heat-shocked whilst at the larval stage to activate the *hs-hid* killer transgene, facilitating the recovery of the desired genotypes.

The resulting lines were verified by PCR (data not shown) using generic primers targeting sequences within the pUAST vector (J. Gesto, pers. comm.) used to create the *UAS-Dicer2* transgene:

Primer name	Sequence (5'-3')	Melting temp (°C)
5uasA	TGTCCTCCGAGCGGAGACTCTAG	64.8
3uasA	TTCTTGGCAGATTTCAGTAGTTGCAG	64.8

Knockdown experiments were then performed using the same crossing scheme as that illustrated in the previous chapter.

8.3 Results

8.3.1 Results obtained using BeFly! are comparable those using published tools

The BeFly! sleep analysis module is able to calculate many metrics, not all of which are normally distributed, as shown in Figure 8-4. Such metrics may be more appropriately analysed following log transformation as demonstrated by Wu *et al.* (Wu et al., 2008a), though such a procedure is not currently automated in BeFly!'s sleep analysis.

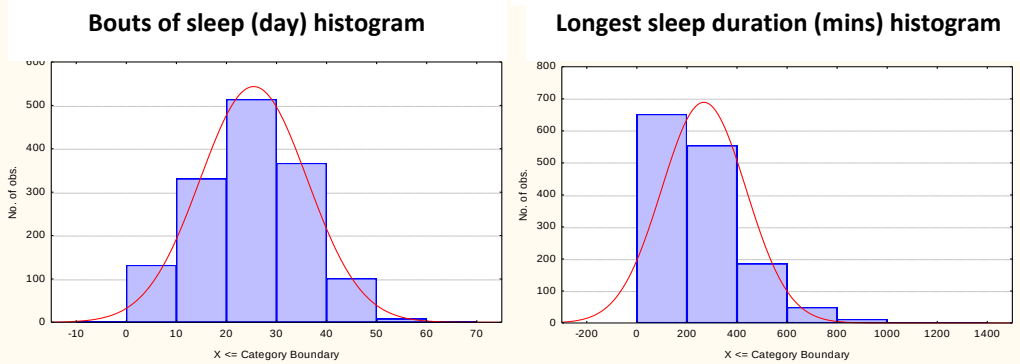


Figure 8-4 The distribution of individual transheterozygous flies as measured using two different sleep metrics. Bout number metrics such as those illustrated in the left panel are normally distributed, those measured in terms of time (right panel) show evidence of skew.

The various different metrics with which sleep can be analysed are correlated, but not interchangeable, reflecting the shared genetic architecture underlying sleep:

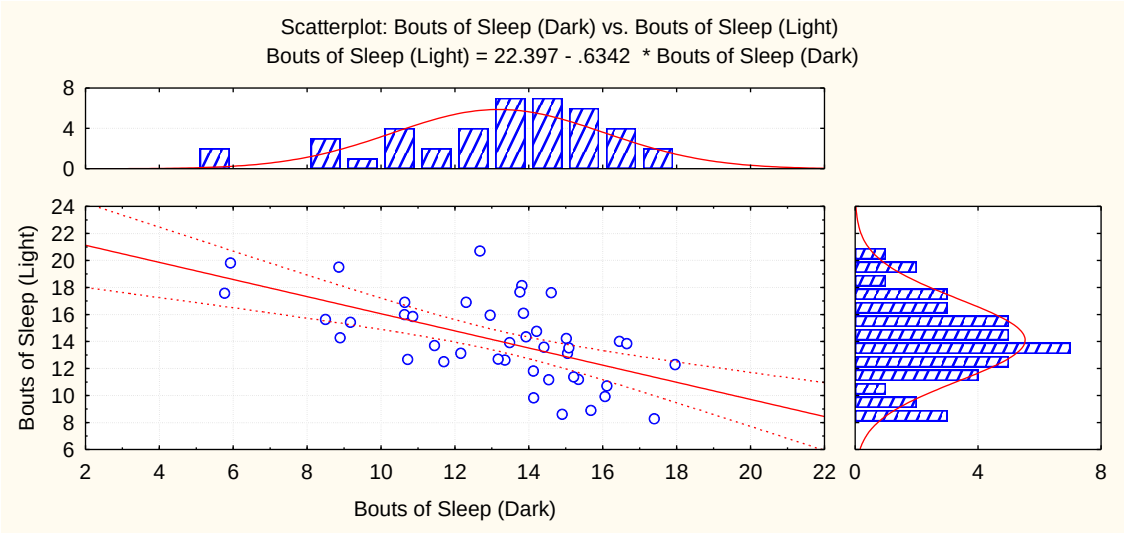


Figure 8-5 Correlation between daytime and nighttime sleep bouts. r = -0.60, significant at p<0.05. **Red line** shows normal distribution**, dashed lines** show 95% confidence intervals.

The correlations between sleep metrics identified during the analysis of transheterozygous flies in this study were compared with those identified by Harbison and Sehgal in their study of the genetic architecture of sleep using a range of stocks carrying *P* element insertions (Harbison and Sehgal, 2008):

Table 8-2 Comparison of the correlations found between sleep metrics in this study and in the study of Harbison and Sehgal. * indicates non-significant results at p<0.05.

Harbison and Sehgal, 2008	24-hr sleep (min)	Nighttime sleep (min)	Daytime sleep (min)
24-hr bout number	0.45	0.32	0.52
Nighttime bout number	-0.16*	-0.12*	-0.18
Daytime bout number	0.77	0.54	0.87
Average bout length (min)	0.81	0.83	0.63
This study	**Total Sleep (Bins)**	**Total Sleep (min/Dark)**	**Total Sleep (min/Light)**
Bouts of Sleep (Day)	0.00*	-0.12*	0.15*
Bouts of Sleep (Dark)	-0.87	-0.85	-0.75
Bouts of Sleep (Light)	0.82	0.70	0.84
Mean Bout Length (Mins/Day)	0.85	0.88	0.67

There is considerable agreement between the correlations found by both studies, suggesting that the genetic architecture underlying the various facets of sleep is similar in different genetic backgrounds, and therefore amenable to further study.

Due to the diallel nature of the transheterozygote crossing scheme, the genetic architecture of sleep could also be assessed by measuring epistatic interactions both in the light (Table 8-3) and in the dark (Table 8-4):

Table 8-3 Epistatic interactions affecting total daytime sleep in transheterozygous lines. A value of 15 indicates that the given genotype sleeps 15 minutes more during the 12 hour light phase than would be expected given the general combining abilities of its parental alleles (GCA score).

Total Sleep in Light (Mins/12hrs) SCA values							
Allele	***Csp***	***trf***	***Itp***	***Pen***	***mth***	***nmo***	***cnk***
nal	15	-24	-5	-14	8	17	2
Csp		-21	1	-8	-1	-12	27
trf			52	-33	32	-16	9
Itp				-8	-23	-11	-6
Pen					14	71	-22
mth						-35	5
nmo							-15

Table 8-4 Epistatic interactions affecting total sleep in dark for transheterozygous lines

Total Sleep in Dark (Mins/12hrs) SCA values							
Allele	*Csp*	*trf*	*Itp*	*Pen*	*mth*	*nmo*	*cnk*
nal	52	-69	27	-32	-11	65	-32
Csp		-22	36	70	-79	-72	15
trf			52	-116	89	27	38
Itp				28	-38	-57	-47
Pen					33	42	-25
mth						-26	31
nmo							20

Comparison of the epistatic interactions evident in Table 8-3 and Table 8-4 reinforces the conclusion that the genetic networks underlying daytime and nighttime sleep, though overlapping to some extent (as revealed by the phenotypic correlation data, Figure 8-5), are clearly independent. One example is the *mth/Csp* strain, which sleeps 79 minutes more than expected during the night, but shows no epistatic interaction affecting daytime sleep. Such conclusions are reinforced by the recovery of mutations affecting daytime and nighttime sleep independently of each other (R. Greenspan pers. comm.).

8.3.2 Sleep metrics are more robust than total activity counts

Analysing the sleep patterns of transheterozygous flies at high resolution showed that the pattern of sleep on the first day of analysis was similar to that on the second (Figure 8-6), implying that the metrics calculated to describe sleep were sufficiently stable to support further analysis. Indeed, not only were sleep metrics found to be stable over time, but also were more accurate and robust than the measurements of total activity commonly used in the circadian field, as illustrated in Figure 8-7.

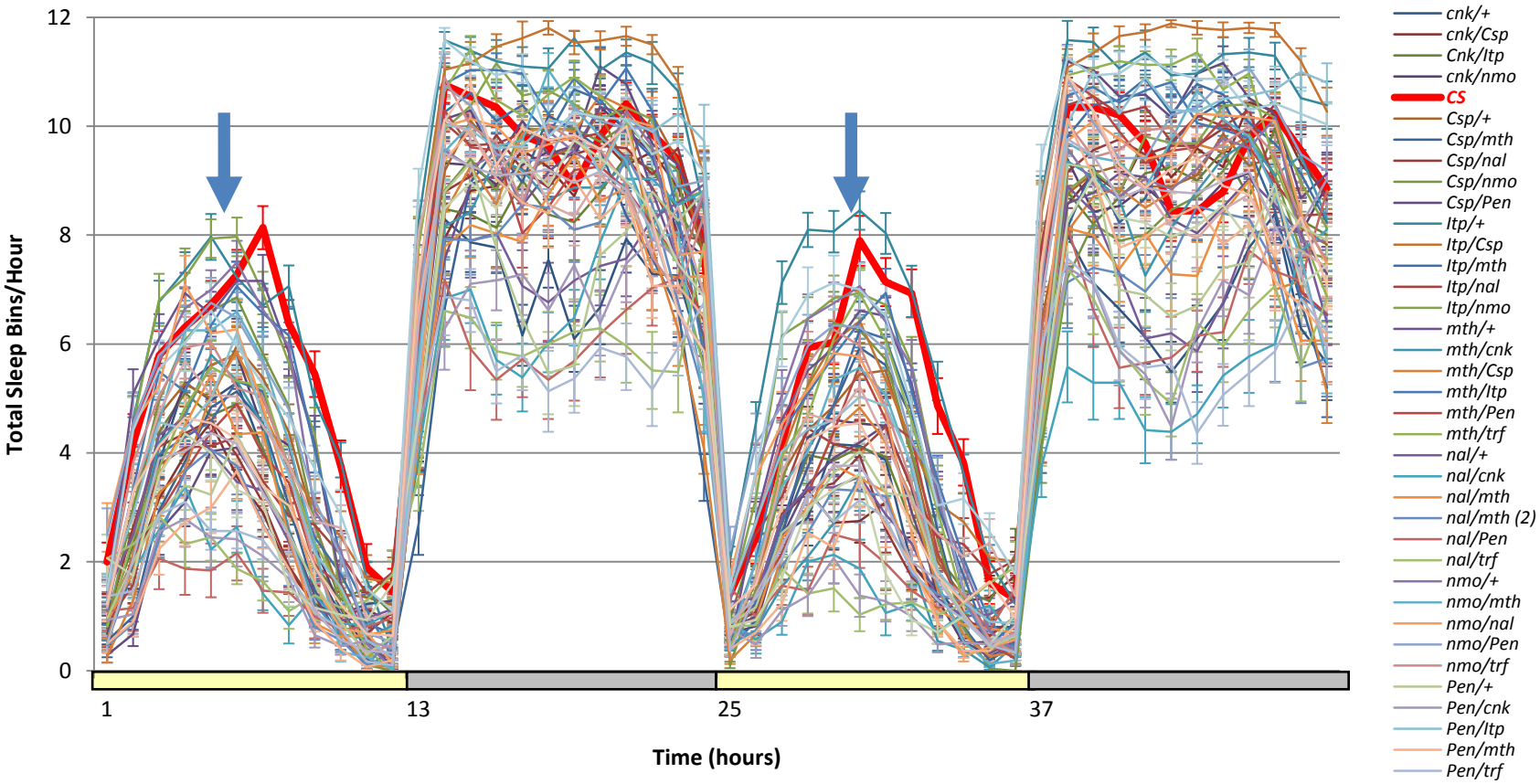


Figure 8-6 'Per Hour' Sleep Metric Chart for mated transheterozygous female flies. Lines represent the mean (±SEM) for each genotype. Environmental lighting shown by coloured bar on (x) axis. **Blue arrows** represent the mid day 'siesta' sleep. The (y) axis is measured in 5 minute bins, i.e. 12 bins = 1 hour. Note that the Canton-S (**CS**) background control (**highlighted in red**) is amongst the highest sleeping lines both during the day and night.

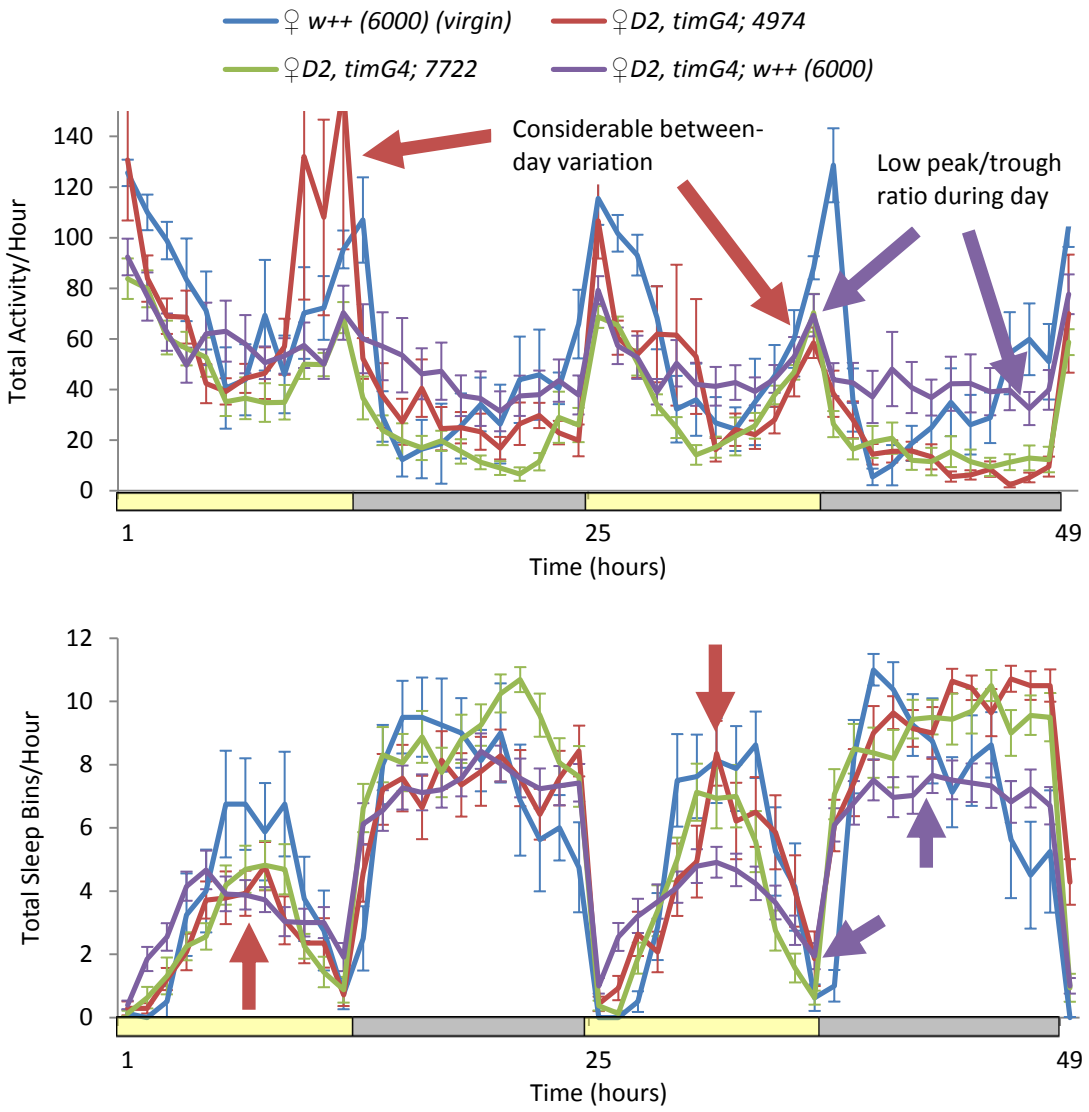


Figure 8-7 Locomotor behaviour described using total activity counts (top panel) and total sleep (bottom panel). Comparison of the top and bottom panels shows that aberrantly high locomotor activity (**red arrows** show the evening peak on days 1 and 2) has less effect on measures of sleep than total activity, as per hour sleep metrics must fall between 0 and 12. In cases where there is a low ratio between peak and trough activity (**purple arrows**), sleep metrics also provide better resolution, more clearly delineating daytime and nighttime behavioural patterns.

As sleep metrics were both accurate and robust, the qualitative analysis of transheterozygote sleep patterns (Figure 8-6) was quantified using 'by day' analysis. Mean sleep metrics for each transheterozygote genotype were calculated by treating each fly and each day's activity as independent. Conservative *post hoc* Dunnett's tests were performed to assess the statistical significance of the variation between transheterozygous lines.

8.3.3 Transheterozygote sleep metrics measured 'by day'

Table 8-5 'By Day' Sleep metrics for transheterozygotes. Each cells represents the mean value for mated females of each genotype (n=32) over two days of recording. Genotypes were analysed relative to the Canton-S control using a conservative *post hoc* Dunnett's test on values for individual flies (n=1344). Significance levels are shown using asterisks: * denotes P<0.05, ** denotes P<0.01, *** denotes P<0.001. It is clear that sleep metrics differ most significantly during the light portion of the LD cycle.

Genotype	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Bout Length (Mins/ Dark)	Mean Bout Length (Mins/ Light)	Mean Waking Activity/ Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/ 12hrs)	Total Sleep in Light (Mins/ 12hrs)
cnk/+	16.5 ±0.6	14 ±0.7 ***	101 ±8 *	199 ±6	29.2 ±2.5	10.3 ±0.5	6.5 ±0.2	32 ±3	1157 ±48 ***	406 ±14 ***	151 ±11 ***
cnk/Csp	14.4 ±0.9	13.6 ±0.8 ***	190 ±22	207 ±6 *	55.4 ±6	14.1 ±5.3	5.5 ±0.2	29 ±3	871 ±38	518 ±17	139 ±17 ***
Cnk/Itp	14.2 ±0.7	14.8 ±0.8 ***	190 ±19	200 ±6	48 ±5	11.1 ±0.7	5.6 ±0.2	29 ±2	863 ±37	503 ±15	167 ±12 ***
cnk/nmo	14.5 ±0.8	11.2 ±0.6 ***	167 ±16	198 ±7	64 ±14.3	14.3 ±1.1	7.1 ±0.2 **	30 ±2	1070 ±34 ***	518 ±13	160 ±12 ***
CS	12.7 ±0.8	20.7 ±0.6	189 ±13	179 ±6	68.6 ±12.4	14.1 ±0.6	5.8 ±0.2	33 ±6	685 ±28	556 ±12	290 ±14
Csp/+	13.8 ±0.9	18.1 ±0.6	193 ±17	198 ±5	67.5 ±12.9	10.5 ±0.5	5.7 ±0.1	26 ±2	811 ±30	536 ±15	193 ±10 **
Csp/mth	15 ±0.6	14.2 ±0.6 ***	179 ±14	195 ±5	40.5 ±5.2	10.5 ±0.5	6.7 ±0.1	45 ±4	1085 ±33 ***	479 ±12	153 ±10 ***
Csp/nal	12.2 ±0.9	13.1 ±0.6 ***	223 ±22	200 ±6	78.9 ±15.7	9.9 ±0.6	5.9 ±0.2	29 ±3	908 ±42	541 ±18	133 ±10 ***
Csp/nmo	13.5 ±0.8	13.9 ±0.4 ***	198 ±18	188 ±7	69.2 ±16.4	14.3 ±1	7.5 ±0.2 ***	39 ±4	1109 ±40 ***	505 ±14	197 ±12 **
Csp/Pen	11.7 ±0.8	12.5 ±0.7 ***	237 ±17	193 ±5	73.4 ±7	9.6 ±0.4	5.8 ±0.1	26 ±2	846 ±27	587 ±11	122 ±8 ***
Itp/+	5.9 ±0.7 ***	19.8 ±0.5	310 ±21 ***	170 ±4	254.9 ±30 ***	16.5 ±0.7	5.8 ±0.2	21 ±2	563 ±25	650 ±9	313 ±7
Itp/Csp	5.8 ±0.6 ***	17.6 ±0.8	380 ±26 ***	171 ±4	233.9 ±29.5 ***	11.9 ±0.6	6.2 ±0.2	19 ±2	691 ±27	664 ±8 *	209 ±13
Itp/mth	12.3 ±0.8	16.9 ±0.6	261 ±20	178 ±4	72 ±12.2	11.4 ±1	6.6 ±0.2	35 ±11	898 ±37 *	566 ±13	192 ±14 ***
Itp/nal	13 ±0.8	16 ±0.9 **	225 ±16	184 ±4	62.1 ±6	10.6 ±0.5	6.2 ±0.2	21 ±2	867 ±38	562 ±12	174 ±13 **
Itp/nmo	9.2 ±0.6	15.4 ±0.5 **	252 ±18	188 ±5	127.8 ±22.8	17.3 ±0.7	6.4 ±0.1	32 ±2	783 ±31	566 ±13	260 ±9
mth/+	13.8 ±0.7	17.7 ±0.4	183 ±16	196 ±5	51.5 ±6.8	13.8 ±0.6	5.6 ±0.1	35 ±3	804 ±32	495 ±15	240 ±10
mth/cnk	15.3 ±0.7	11.2 ±0.7 ***	152 ±11	203 ±5	41.7 ±3.7	10 ±0.6	6.2 ±0.1	34 ±6	993 ±33 **	509 ±12	125 ±12 ***
mth/Csp	15.1 ±0.8	13.1 ±0.7 ***	181 ±16	191 ±5	48.3 ±6.9	10.6 ±0.5	7.1 ±0.2 ***	36 ±4	1156 ±40 ***	488 ±13	141 ±10 ***
mth/Itp	10.6 ±0.7	16.9 ±0.6	281 ±19 *	178 ±4	83 ±8.6	13.7 ±0.6	6.7 ±0.2	25 ±2	854 ±34	574 ±8	230 ±11
mth/Pen	13.9 ±0.8	16.1 ±0.9 *	220 ±16	180 ±5	50.8 ±4.7	9.1 ±0.5	5.5 ±0.2	25 ±1	842 ±35	526 ±12	153 ±12 ***
mth/trf	10.6 ±0.6	16 ±0.7 *	291 ±20 ***	173 ±3	79.3 ±8.7	11.9 ±0.6	6.3 ±0.2	22 ±2	843 ±35	588 ±10	188 ±10 ***
nal/+	13.4 ±0.7	12.6 ±0.8 ***	141 ±14	191 ±7	53.8 ±11.6	10.8 ±0.7	6.2 ±0.2	36 ±3	1071 ±49 ***	438 ±19 ***	140 ±11 ***
nal/cnk	15.7 ±0.7	8.9 ±0.9 ***	103 ±10	219 ±6 ***	28.6 ±2.8	8.8 ±0.5	6.6 ±0.2	57 ±13	1321 ±67 ***	378 ±17 ***	85 ±11 ***
nal/mth	14.1 ±0.8	11.8 ±0.7 ***	160 ±15	203 ±6 *	60.6 ±12.5	11.2 ±1.2	8.2 ±0.2 ***	37 ±4	1394 ±60 ***	453 ±17 **	134 ±13 ***

Genotype	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Bout Length (Mins/ Dark)	Mean Bout Length (Mins/ Light)	Mean Waking Activity/ Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/ 12hrs)	Total Sleep in Light (Mins/ 12hrs)
nal/mth (2)	15.2 ±0.8	11.4 ±0.7 ***	156 ±14	196 ±5	41.1 ±5	8.8 ±0.6	7.2 ±0.2 ***	38 ±6	1278 ±47 ***	443 ±18 **	107 ±11 ***
nal/Pen	14.9 ±0.8	8.6 ±0.8 ***	99 ±10 *	201 ±6	32 ±3.5	8.7 ±0.7 *	6.1 ±0.1	61 ±15 *	1162 ±43 ***	392 ±21 ***	88 ±14 ***
nal/trf	16.1 ±0.7	9.9 ±0.9 ***	98 ±11 *	202 ±7	29.9 ±5.5	9.2 ±0.6	5.8 ±0.2	38 ±3	1170 ±54 ***	361 ±15 ***	96 ±11 ***
nmo/+	10.9 ±0.6	15.9 ±0.4 **	212 ±16	195 ±6	77.3 ±12.2	16.1 ±0.6	6.8 ±0.1 *	31 ±3	912 ±34 *	526 ±15	249 ±8
nmo/mth	13.2 ±0.6	12.7 ±0.6 ***	220 ±18	186 ±4	57.4 ±7	14.4 ±1	7.1 ±0.2 ***	33 ±3	1050 ±44 ***	526 ±11	183 ±13 ***
nmo/nal	10.7 ±0.8	12.7 ±0.6 ***	254 ±24	184 ±6	98 ±16.7	15.2 ±0.8	6.8 ±0.1 *	31 ±2	937 ±40 *	549 ±15	198 ±13 **
nmo/Pen	8.9 ±0.7	14.3 ±0.4 ***	244 ±21	185 ±5	123.2 ±22.6	18.5 ±1.5	6.8 ±0.2 *	45 ±12	834 ±31	554 ±18	265 ±24
nmo/trf	11.5 ±0.8	13.7 ±0.6 ***	219 ±17	180 ±5	95.7 ±18.9	14.3 ±0.8	7.2 ±0.2 ***	23 ±2	1006 ±39 ***	545 ±14	196 ±12 ***
Pen/+	15.1 ±0.7	13.6 ±0.7 ***	125 ±9	189 ±6	35.9 ±3.1	9.3 ±0.4	6.6 ±0.2	34 ±3	1139 ±45 ***	450 ±17 **	129 ±9 ***
Pen/cnk	17.4 ±0.8 **	8.3 ±0.7 ***	88 ±7 *	210 ±6 **	27.4 ±2.5	8.2 ±0.5 *	6.1 ±0.2	45 ±11	1174 ±43 ***	414 ±16 ***	74 ±8 ***
Pen/ltp	8.5 ±0.8 *	15.7 ±0.7 **	216 ±19	182 ±6	148 ±22.6	11.4 ±0.9	5.4 ±0.1	29 ±2	717 ±26	592 ±14	183 ±14 ***
Pen/mth	14.1 ±0.8	9.8 ±0.8 ***	184 ±18	182 ±5	58.8 ±12.4	8.6 ±0.4 *	6.3 ±0.2	23 ±2	1088 ±37 ***	485 ±15	92 ±9 ***
Pen/trf	18 ±0.7 ***	12.3 ±0.8 ***	75 ±6 **	199 ±6	19 ±1.2	7.5 ±0.3 **	6.8 ±0.2 *	34 ±4	1371 ±55 ***	342 ±19 ***	100 ±8 ***
trf/+	14.6 ±0.7	17.6 ±0.6	154 ±11	184 ±5	47.4 ±5.4	11.3 ±0.6	6.0 ±0.1	27 ±5	879 ±35	514 ±12	200 ±12 **
trf/cnk	16.1 ±0.7	10.7 ±0.7 ***	127 ±12	202 ±5	36.9 ±3.1	11.3 ±0.9	7.2 ±0.2 ***	35 ±4	1189 ±35 ***	483 ±14	123 ±11 ***
trf/Csp	16.7 ±0.9 *	13.9 ±0.8 ***	140 ±15	185 ±5	45.3 ±6	8.8 ±0.4	6.6 ±0.1	25 ±2	1077 ±40 ***	501 ±15	127 ±10 ***
trf/ltp	8.9 ±0.7	19.5 ±0.8	273 ±19 *	164 ±4	134.2 ±20.8	13.7 ±0.9	6.0 ±0.1	19 ±3	664 ±27	622 ±9	261 ±14
trf/mth	13.9 ±0.7	14.4 ±0.9 ***	189 ±11	174 ±3	50.8 ±5.2	10.5 ±0.6	5.8 ±0.2	22 ±2	848 ±29	533 ±9	160 ±14 ***

As is clear in Figure 8-6, the CS control background used in the transheterozygote study is a particularly high sleeping line, therefore most transheterozygote lines show statistically significant differences in terms of daytime sleep bout number and length. As the CS isolate has been kept under laboratory conditions for many years, it is possible that extending daily sleep might represent an adaptation to the crowded environment within vials, therefore the sleep pattern of the domesticated CS line was compared to that of the recently isolated Houton 'HU' line using activity data obtained by Supriya Bhutani:

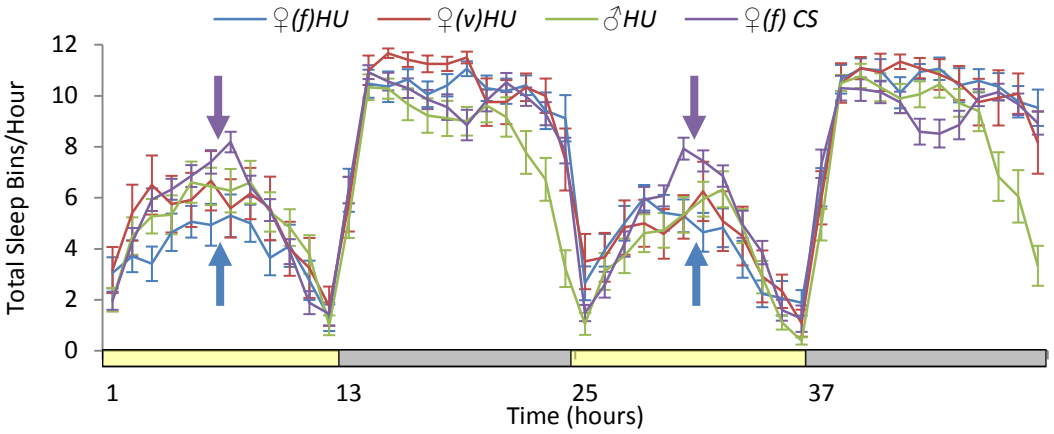


Figure 8-8 Comparison of the sleep pattern of recently isolated 'HU' flies with the domesticated 'CS' strain. **f** = fertilised, **v** = virgin. Fertilised CS flies (**purple**) clearly sleep almost 15 minutes per hour more than fertilised HU flies (**blue**) during the daytime siesta (**arrows**).

The results in Figure 8-8 show that domesticated flies sleep considerably more than newly isolated flies. This suggests that the high sleeping phenotype may well be adaptive in the laboratory (but not in the wild), and is disrupted by the presence of transheterozygous alleles (Figure 8-6).

8.3.4 Identification of genes correlating with sleep

As in the previous chapter, candidate genes linearly correlating with total sleep were identified in a preliminary study by Bruno van Swinderen using sleep analysis software from the Shaw laboratory (see table in appendix 10.5). The author then selected a sample of those candidates (biased towards genes acting as kinases or playing roles in cell signalling) for confirmation of function by targeted gene knockdown in neural centres known to be implicated in sleep regulation (for further details of drivers and expression patterns, see Table 8-1 in the methods section).

8.3.5 Candidate gene knockdown in sleep centers

The results of knocking down candidate sleep genes using the *c687>GAL4* driver (which expresses GAL4 in the *Pars intercerebralis)* are shown below; tables detailing the results of *tim>GAL4, c767>GAL4, th>GAL4, Ddc>GAL4* and *c309>GAL4* mediated knockdown of the same genes are shown in appendix D5.

Table 8-6 Results of c687>GAL4 driver mediated knockdown of candidate genes involved in sleep. Values shown are genotype means (n=16 per genotype per sex), ± the standard error of the mean. Statistics were calculated using a conservative *post hoc* two tailed Dunnett's test relative to the congenic no-knockdown control using individual scores (i.e. not using genotype means). Significance levels are shown using shading and indicated by asterisks: * denotes p<0.05, ** denotes p<0.01, *** denotes p<0.001.

c687>	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Bout Length (Mins/Dark)	Mean Bout Length (Mins/Light)	Mean Waking Activity/Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/12hrs)	Total Sleep in Light (Mins/12hrs)
+											
Female	19.3 ±1.1	12.6 ±1.6	91.3 ±0.9	173 ±10	19.1 ±17.6	14.4 ±1.5	5.52 ±1.64	45.8 ±2.2	1,008 ±2	351 ±0	178 ±5
Male	13.3 ±1.0	8.5 ±1.2	178.9 ±0.5	114 ±17	43.4 ±16.6	49.3 ±6.2	8.83 ±5.03	35.3 ±5.3	1,079 ±6	478 ±0	353 ±2
CG10460											
Female	19.9 ±1.2	11.8 ±1.6	94.3 ±1.0	219 ±10*	45.5 ±8.1	8.51 ±24.99	5.57 ±2.30	30.5 ±0.7	1,024 ±2	426 ±0	102 ±4
Male	11.4 ±0.7	11.7 ±1.2	168.1 ±1.1	188 ±18***	51.5 ±10.8	24.96 ±4.95***	7.97 ±3.27	39.0 ±3.1	1,069 ±3	508 ±0	263 ±5
CG10799											
Female	18.1 ±1.3	14.5 ±1.7	133.6 ±0.6	197 ±12	32.1 ±6.8	10.96 ±3.88	5.86 ±1.94	25.3 ±0.9*	930 ±3	477 ±0**	161 ±3
Male	10.2 ±1.1	10.1 ±1.6	289.9 ±1.0**	148 ±39	107.9 ±7.1	39.49 ±33.66	8.58 ±6.05	25.9 ±4.7	957 ±5	548 ±0	352 ±2
CG1093											
Female	18.9 ±1.0	18.7 ±1.6**	109.7 ±1.0	179 ±14	30.0 ±15.1	15.08 ±2.33	5.44 ±1.22	15.8 ±1.2***	697 ±2	500 ±0***	278 ±3*
Male	14.4 ±0.9	10.7 ±1.1	168.2 ±0.6	164 ±17	42.4 ±11.0	31.05 ±4.00**	8.10 ±2.77	21.1 ±2.3	986 ±4	527 ±0	306 ±1
CG10955											
Female	20.1 ±0.9	10.0 ±1.8	103.6 ±1.8	231 ±16**	20.4 ±8.0	7.66 ±1.53	5.92 ±0.84	37.5 ±0.5	1,141 ±2	388 ±0	84 ±7*
Male	13.6 ±1.1	14.4 ±1.6***	154.1 ±0.8	183 ±12**	42.6 ±14.0	25.84 ±4.99***	7.93 ±3.21	29.8 ±2.7	975 ±3	491 ±0	332 ±3
CG11121											
Female	16.4 ±1.1	14.0 ±1.6	129.8 ±1.2	200 ±12	36.4 ±7.7	11.14 ±3.85	5.34 ±1.60	24.7 ±1.1*	819 ±2	517 ±0***	156 ±3
Male	12.5 ±0.7	12.2 ±1.2	236.1 ±1.0	167 ±21*	72.2 ±10.4	28.69 ±22.77***	8.67 ±3.26	24.3 ±3.7	1,051 ±3	550 ±0	298 ±1
CG11173											
Female	20.3 ±1.5	13.1 ±1.7	115.8 ±1.2	199 ±19	31.1 ±7.0	7.74 ±4.69	5.31 ±1.71	22.0 ±0.8**	893 ±2	488 ±0**	105 ±3
Male	13.3 ±0.6	12.6 ±1.1*	172.8 ±1.0	164 ±13	46.3 ±9.5	25.34 ±4.05***	8.87 ±2.56	17.5 ±4.6*	1,132 ±3	523 ±1	263 ±1
CG11228											
Female	20.9 ±0.9	14.9 ±0.9	98.4 ±0.5	189 ±9	26.0 ±8.0	11.17 ±3.04	4.84 ±1.40	31.4 ±0.7	781 ±1	474 ±0**	165 ±3
Male	15.3 ±1.5	11.6 ±1.7	160.9 ±0.5	162 ±25	60.6 ±13.7	25.41 ±22.95***	8.80 ±4.93	26.9 ±2.7	1,181 ±3	472 ±0	277 ±1
CG11301											

Female	11.0 ±0.7***	10.2 ±1.4	278.7 ±0.8***	137 ±22	91.1 ±12.0***	51.23 ±16.77***	5.89 ±4.94	25.8 ±6.0**	543 ±7***	591 ±0***	380 ±1***
CG11807											
Female	17.1 ±1.2	13.9 ±2.2	134.2 ±1.8	198 ±16	40.3 ±10.5	10.68 ±11.50	7.48 ±1.83	27.2 ±1.0**	1,431 ±3	471 ±2**	152 ±8
Male	12.1 ±1.0	14.5 ±1.1***	204.8 ±0.7	168 ±23*	101.6 ±8.3	25.21 ±30.60***	8.18 ±4.75	22.3 ±2.3	875 ±2	569 ±0*	337 ±1
CG11841											
Female	19.4 ±1.0	14.9 ±1.7	105.5 ±1.3	231 ±8**	25.5 ±5.4	9.20 ±2.43	5.82 ±1.19	36.6 ±0.5	1,007 ±2	445 ±0	135 ±4
Male	11.8 ±0.8	12.3 ±1.0	210.9 ±0.8	140 ±18	71.2 ±15.6	33.07 ±20.42**	9.15 ±2.49	36.4 ±3.0	979 ±3	535 ±0	366 ±3
CG12163											
Female	15.7 ±0.8	14.9 ±1.1	159.6 ±0.7	191 ±11	47.5 ±4.2	10.97 ±11.00	6.59 ±1.40	17.9 ±0.8***	1,002 ±2	504 ±1***	166 ±1
Male	12.0 ±0.7	12.4 ±1.0**	189.5 ±0.7	170 ±18**	77.0 ±5.9	21.17 ±16.02***	8.00 ±3.23	22.8 ±2.0	1,053 ±2	546 ±0	243 ±2**
CG12163 (x)											
Female	13.0 ±0.6***	11.8 ±0.9	180.5 ±0.6**	144 ±12	42.4 ±8.5	25.77 ±2.80*	6.46 ±2.07	31.3 ±2.5	926 ±2	503 ±0***	274 ±3**
CG12163 (y)											
Female	14.4 ±0.6	9.9 ±1.1	144.7 ±0.6	191 ±17	34.7 ±10.5	24.31 ±3.35	9.18 ±2.54***	35.3 ±2.9	1,423 ±3**	452 ±0	214 ±2
Male	18.0 ±1.2*	10.6 ±1.2	113.0 ±1.0	204 ±16***	27.1 ±8.8	9.39 ±2.80***	6.25 ±1.52***	32.0 ±0.8	1,139 ±3	429 ±0	103 ±3***
CG12163 (z)											
Female	13.1 ±0.9**	12.4 ±1.2	175.3 ±0.6*	176 ±24	45.6 ±13.6	26.50 ±6.18*	8.51 ±3.08***	31.1 ±3.0	1,116 ±3	490 ±1**	299 ±2***
Male	18.1 ±1.3*	14.5 ±1.7***	159.8 ±0.9	171 ±27*	40.1 ±8.1	10.15 ±7.96***	5.37 ±2.85***	26.1 ±0.9	851 ±5	507 ±0	149 ±2***
CG17034											
Female	19.1 ±1.5	14.1 ±1.9	144.8 ±1.3	195 ±19	31.0 ±7.1	7.88 ±4.45	4.76 ±1.85	25.2 ±0.5*	815 ±2	473 ±0**	115 ±3
Male	11.8 ±1.0	12.7 ±1.3*	208.2 ±1.0	172 ±28*	93.0 ±10.1	23.70 ±31.31***	7.49 ±2.51	32.0 ±2.2	935 ±2	545 ±0	274 ±3
CG18582											
Female	18.3 ±1.1	11.0 ±1.6	116.1 ±1.3	198 ±13	28.1 ±11.3	10.16 ±4.14	5.83 ±1.47	30.3 ±0.9	1,046 ±2	439 ±0	117 ±10
Male	14.6 ±1.2	13.5 ±1.6**	163.1 ±0.8	181 ±21**	39.8 ±9.9	19.36 ±5.01***	8.76 ±2.62	27.7 ±1.9	1,226 ±3	489 ±0	251 ±3*
CG18859											
Female	18.7 ±1.2	20.5 ±2.6***	32.3 ±1.6	126 ±7*	9.8 ±12.7	8.52 ±1.14	3.97 ±0.77	42.0 ±0.6	848 ±4	190 ±0***	184 ±7
Male	16.7 ±1.5	17.1 ±2.4***	60.3 ±1.1**	86 ±6	12.6 ±12.0	20.42 ±1.29***	6.53 ±1.35**	43.5 ±2.0	1,168 ±4	213 ±0***	332 ±19
CG1924											
Female	18.7 ±1.0	12.4 ±1.1	131.3 ±0.9	212 ±14	30.4 ±7.7	9.71 ±2.85	5.61 ±1.19	25.2 ±1.1*	932 ±2	501 ±0***	122 ±4
Male	11.3 ±0.9	9.3 ±0.7	202.5 ±0.6	158 ±17	64.2 ±14.6	33.87 ±10.26*	8.83 ±2.32	23.9 ±3.3	1,057 ±2	556 ±0	294 ±2
CG2041											
Female	19.4 ±1.4	14.5 ±1.8	150.9 ±0.9	190 ±29	36.9 ±9.6	13.14 ±9.45	7.48 ±4.80*	19.2 ±3.7***	1,208 ±5	437 ±1	169 ±1
Male	12.2 ±1.0	15.0 ±1.2***	161.6 ±1.0	162 ±14	109.0 ±8.6	22.82 ±34.99***	8.95 ±1.90	24.1 ±2.0	1,048 ±2	543 ±0	317 ±4
CG2219											
Female	21.3 ±0.8	13.0 ±1.6	91.2 ±1.1	180 ±14	20.7 ±14.0	9.70 ±2.38	5.34 ±1.46	34.3 ±1.7	963 ±2	398 ±0	133 ±4
Male	13.4 ±1.2	12.2 ±1.8	210.5 ±0.9	188 ±38***	64.8 ±10.6	30.55 ±15.61**	9.13 ±8.70	24.1 ±7.1	1,138 ±7	531 ±0	272 ±2
CG30404											
Female	13.3 ±0.7**	11.9 ±1.3	138.4 ±1.0	161 ±12	37.8 ±7.0	25.97 ±2.48*	6.85 ±2.49	33.7 ±3.3	983 ±3	456 ±0*	259 ±3*
Male	18.4 ±0.8**	14.6 ±1.3***	124.4 ±1.0	187 ±10***	30.0 ±5.2	11.02 ±2.24***	5.31 ±1.14***	28.9 ±0.8	851 ±2	477 ±0	166 ±3***
CG32091											
Female	15.4 ±2.0	18.8 ±2.8*	100.9 ±1.6	205 ±13	78.0 ±7.9	13.04 ±43.07	5.09 ±3.24	26.6 ±1.4	692 ±3	520 ±0**	246 ±7
CG32146											
Female	17.3 ±1.4	16.6 ±1.5	155.8 ±0.7	197 ±24	39.8 ±9.9	10.50 ±6.74	4.48 ±2.06	12.5 ±0.7***	677 ±2	521 ±0***	173 ±1
Male	12.4 ±0.7	14.7 ±1.5***	226.7 ±1.1	196 ±23***	75.6 ±12.1	19.87 ±21.56***	7.44 ±3.46	17.2 ±3.5*	900 ±4	590 ±0***	244 ±1*
CG32147											

Female	21.6 ±0.9	14.1 ±1.1	75.7 ±0.5	200 ±7	22.6 ±7.3	10.38 ±2.85	5.03 ±1.28	24.1 ±0.4**	901 ±2	415 ±0	145 ±3
Male	14.9 ±0.4	12.7 ±0.7**	152.9 ±0.5	175 ±11***	35.0 ±11.1	22.43 ±1.36***	9.30 ±1.26	25.1 ±1.8	1,286 ±1	496 ±0	250 ±1**
CG32149											
Female	17.0 ±0.8	12.7 ±1.4	141.9 ±1.1	197 ±9	31.4 ±7.5	9.58 ±2.25	5.77 ±0.96	20.5 ±0.5***	951 ±2	492 ±0**	124 ±1
Male	9.3 ±1.2	11.7 ±1.0	257.1 ±0.8	167 ±32*	155.1 ±8.5***	22.89 ±48.56***	8.16 ±4.15	18.0 ±1.9*	956 ±2	601 ±0***	264 ±1
CG32296											
Female	12.4 ±1.3**	17.0 ±1.5	224.9 ±0.8***	181 ±28	81.2 ±7.4**	12.11 ±22.29	4.36 ±2.23	15.9 ±1.3***	574 ±2**	585 ±0***	197 ±1
Male	11.6 ±1.3	12.8 ±1.0*	247.0 ±0.7	162 ±32	111.0 ±8.8	20.95 ±36.07***	7.70 ±2.60	16.6 ±1.6**	936 ±2	580 ±0**	263 ±1
CG32491											
Female	20.9 ±1.2	11.3 ±2.0	98.3 ±1.5	206 ±12	23.4 ±9.9	6.89 ±2.84	5.00 ±1.68	29.7 ±0.4	919 ±3	443 ±0	81 ±4
CG3533											
Female	20.1 ±1.0	13.0 ±1.4	104.8 ±1.3	217 ±8	26.9 ±8.4	9.89 ±2.07	4.60 ±0.98	21.5 ±0.7**	765 ±2	493 ±0**	128 ±3
Male	16.1 ±0.9	12.6 ±1.0*	140.0 ±0.6	177 ±18**	34.6 ±11.0	18.29 ±2.73***	8.41 ±1.36	25.9 ±1.6	1,218 ±2	507 ±0	220 ±2***
CG4147											
Female	18.1 ±1.3	14.8 ±2.1	133.2 ±1.3	186 ±21	27.5 ±13.6	10.17 ±4.10	4.95 ±1.89	42.3 ±0.9	839 ±5	442 ±0	149 ±15
Male	14.7 ±0.9	14.6 ±1.2***	188.3 ±0.8	164 ±27	39.1 ±13.1	23.13 ±4.98***	6.65 ±2.31**	22.8 ±1.7	855 ±2	486 ±0	315 ±2
CG4422											
Female	18.5 ±1.1	12.0 ±1.4	121.3 ±0.6	160 ±14	27.2 ±9.6	21.55 ±2.40	6.74 ±1.99	40.7 ±2.6	1,002 ±2	444 ±0	232 ±3
CG4721											
Female	18.9 ±1.4	11.4 ±1.7	128.0 ±1.5	206 ±15	35.5 ±7.6	13.11 ±6.26	5.92 ±2.65	29.2 ±2.0	949 ±2	507 ±0***	148 ±4
CG5131											
Female	17.3 ±1.1	14.1 ±1.2	164.5 ±1.0	187 ±26	40.5 ±7.5	11.52 ±11.54	5.68 ±2.27	22.2 ±1.6**	920 ±1	478 ±0**	163 ±3
CG5341											
Female	17.7 ±1.2	17.5 ±2.2*	132.8 ±1.2	134 ±17	28.5 ±8.3	23.77 ±3.72	5.46 ±3.16	30.1 ±3.2	735 ±3	387 ±0	316 ±3***
Male	19.3 ±0.9***	18.1 ±1.3***	109.9 ±0.8	130 ±10	24.9 ±12.8	20.37 ±1.97***	5.04 ±1.22***	24.8 ±2.0	701 ±1**	420 ±0	328 ±3
CG6051											
Female	17.2 ±1.5	14.0 ±2.0	116.9 ±1.0	188 ±21	37.2 ±8.2	7.89 ±8.64	5.61 ±2.11	27.3 ±0.3	984 ±2	451 ±0	109 ±3
Male	11.8 ±1.2	11.6 ±1.2	256.0 ±0.6	153 ±27	59.6 ±11.5	33.11 ±6.27*	7.40 ±3.91	26.0 ±3.7	864 ±4	537 ±0	327 ±1
CG6562											
Female	10.7 ±0.8***	11.0 ±1.0	241.7 ±0.5***	146 ±16	110.7 ±7.1***	31.32 ±19.87***	6.55 ±3.38	26.5 ±2.8**	681 ±4*	597 ±0***	304 ±1***
CG6574											
Female	16.2 ±0.9	15.0 ±1.4	146.2 ±1.1	177 ±11	38.3 ±7.4	12.74 ±3.94	4.55 ±1.30	22.7 ±1.1**	665 ±2*	528 ±0***	186 ±2
Male	12.7 ±1.1	11.5 ±1.8	199.5 ±0.9	144 ±19	54.7 ±15.1	32.24 ±7.88**	7.15 ±4.18	28.4 ±3.5	812 ±5	551 ±0	319 ±2
CG7220											
Female	16.5 ±1.0	15.0 ±1.1	176.7 ±0.8**	179 ±20	52.2 ±6.3	13.00 ±7.98	4.77 ±1.79	20.9 ±1.4***	695 ±2*	527 ±0***	183 ±1
Male	11.3 ±0.7	11.8 ±1.2	223.0 ±0.8	158 ±14*	81.4 ±7.2	26.50 ±14.28***	8.35 ±3.57	25.1 ±3.7	973 ±4	585 ±0***	263 ±1*
CG7891											
Female	17.2 ±1.1	15.6 ±1.7	138.0 ±1.2	192 ±12	37.3 ±8.6	10.11 ±6.77	5.18 ±2.31	20.3 ±1.2***	800 ±2	501 ±0***	158 ±1
Male	12.8 ±0.9	11.3 ±1.0	187.3 ±0.8	165 ±26*	83.1 ±9.1	25.38 ±28.05***	9.08 ±2.60	28.0 ±2.0	1,236 ±2	510 ±0	260 ±2
CG7945											
Female	21.6 ±0.5	11.8 ±1.2	79.7 ±1.2	200 ±8	19.5 ±10.2	8.28 ±1.09	4.88 ±0.53	30.8 ±0.8	906 ±1	409 ±0	106 ±6
Male	14.0 ±0.7	13.4 ±1.5**	156.9 ±1.0	173 ±13**	45.6 ±12.3	23.28 ±7.93***	7.49 ±3.07	24.7 ±2.8	1,004 ±3	488 ±0	291 ±1
CG8604											
Female	21.1 ±1.1	13.4 ±1.5	106.8 ±1.1	189 ±13	21.6 ±10.1	11.42 ±2.35	5.57 ±1.34	30.5 ±1.5	990 ±1	417 ±0	147 ±4
Male	14.4 ±0.7	13.0 ±1.1*	156.9 ±0.7	156 ±17	39.3 ±9.8	24.50 ±3.25***	8.09 ±2.27	28.3 ±2.2	998 ±3	505 ±1	303 ±2

CG8768											
Female	18.5 ±1.0	10.5 ±1.3	120.8 ±1.0	211 ±12	28.6 ±7.0	8.34 ±2.46	5.31 ±1.49	32.2 ±0.7	922 ±2	481 ±0**	93 ±5
Male	12.5 ±1.0	9.3 ±1.4	206.4 ±0.7	149 ±26	57.2 ±18.8	41.55 ±12.07	8.94 ±7.80	26.1 ±6.4	1,085 ±11	529 ±0	321 ±3
CG8983											
Female	16.0 ±1.2	10.8 ±1.7	150.3 ±1.1	195 ±17	41.7 ±6.9	10.16 ±6.26	6.70 ±4.32	26.6 ±1.4*	1,100 ±6	512 ±0***	116 ±2
CG9293											
Female	17.9 ±1.6	15.4 ±2.0	183.3 ±0.9*	176 ±30	37.2 ±14.9	14.59 ±5.01	4.50 ±2.24	21.5 ±0.9**	614 ±3*	544 ±0***	213 ±2
Male	13.9 ±0.9	8.3 ±1.2	154.8 ±0.7	171 ±13*	48.5 ±10.6	25.16 ±4.66***	8.18 ±3.14	30.8 ±2.6	1,121 ±4	562 ±0*	196 ±2***
CG9890											
Female	18.2 ±1.5	12.5 ±1.7	122.8 ±1.0	204 ±17	35.6 ±8.0	16.36 ±5.99	5.25 ±3.67	29.5 ±3.9	778 ±3	490 ±0**	195 ±5
Male	13.1 ±0.7	9.5 ±0.9	198.0 ±0.5	173 ±24*	45.4 ±8.5	29.61 ±4.70***	9.97 ±1.91	23.0 ±2.0	1,309 ±2	521 ±0	265 ±1
CG9906											
Female	13.7 ±1.4*	11.7 ±1.6	182.3 ±1.3*	191 ±37	109.1 ±8.1***	9.23 ±36.00	6.36 ±3.11	34.4 ±1.1	1,093 ±3	462 ±0*	119 ±5

The tables clearly show that most of the effects of gene knockdown were sex specific (as has recently been shown in the literature: Harbison and Sehgal, 2008), except in the case of the *c767>GAL4* driver (see appendix D5). This is surprising; given that both *c767>GAL4* and *c687>GAL4* drivers express GAL4 in the *Pars intercerebralis* region of the brain (having been chosen to serve as biological replicates), the results for these drivers should be similar. The evidence of sex specific effects in the *c687>GAL4* but not in *c767>GAL* suggests that these driver lines might have subtly different expression patterns, and cannot therefore be treated as redundant controls.

Knocking down a single gene in an otherwise isogenic background resulted in many more significant changes in sleep metrics than might have been expected *a priori*; most knockdown lines showed significant effects on at least one sleep metric, despite there being no evidence that the candidate genes were even expressed in the brain (and therefore amenable to knockdown). Data in Figure 8-9 suggests that this may be an artefact; the control line showed particularly high levels of sleep, possibly as a result of the GAL4 driver lines being backcrossed into the same 'high sleeping' CS line used in the transheterozygote study. This had two effects on the study; it precluded the identification of knockdown lines that might cause an increased sleep

phenotype due to ceiling effects, and it questioned the value of assessing the significance of the results using a *post hoc* Dunnett test relative to the CS line (Table 8-6).

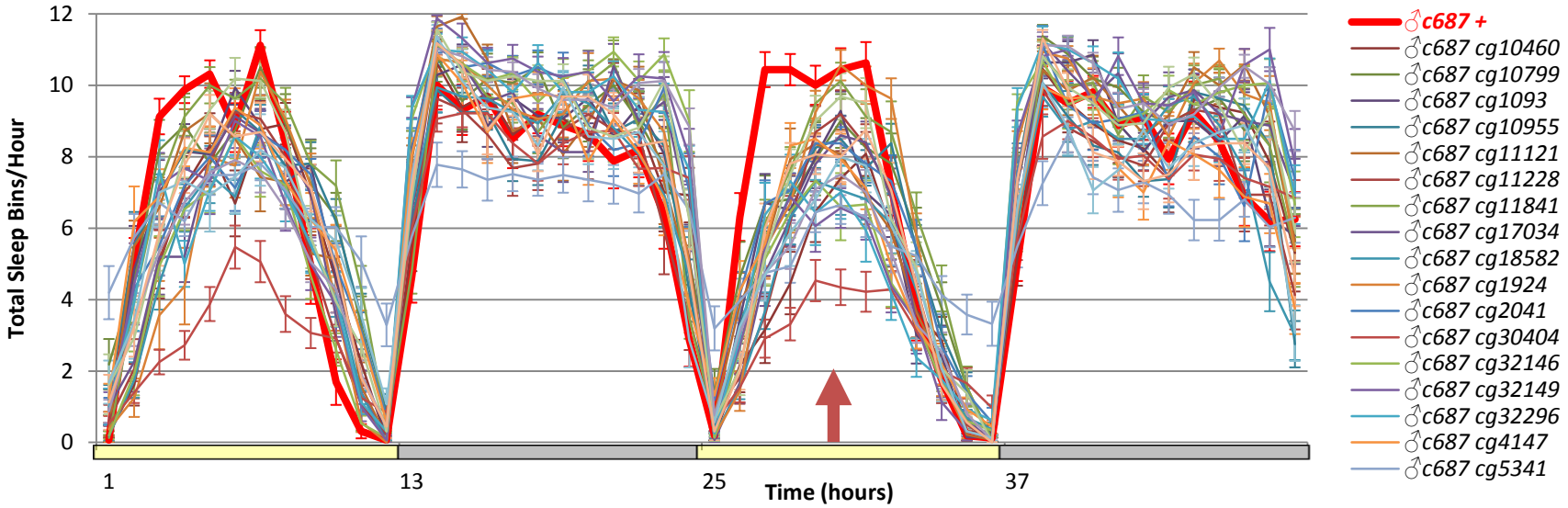


Figure 8-9 Total sleep per hour for a small sample of male flies in which candidate genes were knocked down using the *c687>GAL4* driver. The control genotype (**red highlight**) slept considerably more than other lines, bringing its utility as a control into question. Note the reduced daytime sleep of the *c687 CG30404* line (**vermillion**).

Given these limitations, an attempt was made to identify genes which had significant effects on sleep by collating data generated using different GAL4 lines, thereby identifying genes which - by virtue of their wide expression patterns - might play the most significant roles in regulating sleep. Two problems were envisioned with such an approach; firstly that the importance of a gene to sleep regulation need not be correlated in any way with the breadth of its expression (indeed, the converse may be true: key circadian genes such as Pdf show very restricted expression patterns - Renn et al., 1999), and also that this approach might preferentially identify genes that disrupt the proper functioning of cells rather than those that play roles specific to sleep regulation (although the panel of candidate genes screened was biased towards cells playing roles in communication rather than metabolism, mitigating this risk).

Table 8-7 Excerpt of table collating the probabilities that gene knockdown using a GAL4 driver elicits a significant effect on sleep metrics.

Gene	Bouts of Sleep (dark)						Bouts of Sleep (light)					
Driver	*c309*	*c687*	*c767*	*Ddc*	*th*	*Mean*	*c309*	*c687*	*c767*	*Ddc*	*th*	Mean
CG10236	0.00		0.00	0.98	0.72	0.43	0.97		0.06	0.00	1.00	0.51
CG10460	1.00	1.00		0.02		0.67	0.24	1.00		0.00		0.41
CG10799		1.00		1.00	0.00	0.67		0.99		1.00	1.00	1.00
CG1093	0.99	1.00	0.00	1.00	0.00	0.60	0.02	0.00	0.09	1.00	0.70	0.36
CG10955	1.00	1.00	0.46	1.00	1.00	0.89	0.07	0.83	1.00	1.00	1.00	0.78
CG11121	0.06	0.81	1.00	0.15		0.50	0.03	1.00	0.01	0.03		0.27
CG11173	0.97	1.00	0.99	0.97	0.00	0.79	1.00	1.00	0.15	1.00	0.14	0.66
CG11228	0.23	1.00	0.99	1.00	0.18	0.68	0.97	0.92	0.71	0.99	0.99	0.92
CG11301	0.80	0.00	0.73	1.00	0.00	0.51	0.01	0.73	1.00	1.00	0.25	0.60
CG11807		1.00	1.00	1.00		1.00		0.95	1.00	0.34		0.76

As can be seen in Table 8-7, knockdown of *CG11121* expression in various regions of the brain is likely to cause a change in daytime (light) sleep bout number, suggesting it might play an important generalised role in sleep regulation. However *CG11121*, or *sine oculis*, is a transcription factor important for the development of the compound eyes, therefore knockdown of this gene within the brain may prevent the proper development of the visual system, leading to sleep misregulation during the light phase as a pleiotropic effect of its role in development.

Given such limitations, an alternate approach was developed whereby the assumption was made that gene knockdown would either not elicit a change in a given sleep metric, or have an equal probability of increasing or decreasing that metric. In such a model, the grand mean of the all the results represents the 'normal' phenotypic score, and results of gene knockdown can therefore be compared to that mean, those falling more than two standard deviations from the mean being judged to be 'significant'. Although this approach has been successfully applied in sleep screens by other groups (Koh et al., 2008), it is limited in that a set number of genes will always be identified as 'significant', irrespective of the extent of phenotypic variation in the results.

Table 8-8 Excerpt of results table showing sleep metrics per genotype expressed as deviation from the grand mean. Values falling more than 2 standard deviations from the grand mean (i.e. greater than the value in the top row) are highlighted in **blue**.

Gene	Bouts of Sleep (dark)					Total	Bouts of Sleep (light)					Total
Driver	*th*	*c309*	*c687*	*c767*	*Ddc*		*th*	*c309*	*c687*	*c767*	*Ddc*	
2stdev	4.81	5.86	4.67	7.64	4.78	5.55	5.19	3.53	4.38	7.11	4.62	4.96
+	0.30	1.77	-0.45	4.96	-4.59	0.40	-3.64	-3.28	-4.13	-2.22	-4.75	-3.60
CG10236	-1.38	2.78		6.77	-1.43	1.68	1.36	3.59		0.53	1.60	1.77
CG10460	1.68	-2.44	-2.34		-1.46	-1.14	0.98	-2.46	-0.87		-2.25	-1.15
CG10799	-3.29		-3.61		1.52	-1.79	-2.43		-2.47		-1.49	-2.13
CG1093	0.11	-1.41	0.66	2.52	-0.78	0.22	-0.06	0.90	-1.92	0.03	1.56	0.10
CG10955	-2.12	-3.56	-0.20	-0.92		-1.70	-3.47	-0.32	1.84	-6.28		-2.06
CG11121	-0.38	-1.25	-1.30	-0.73		-0.92	-3.02	0.24	-0.42	-0.53		-0.93
CG11173	0.43	-0.39	-0.51	-3.35	1.16	-0.53	-0.39	1.36	-0.04	3.03	-0.37	0.72
CG11228	-1.23	-1.19	1.52	0.28	2.38	0.35	1.14	-1.69	-1.00	-1.89	-2.18	-1.13
CG11301	-0.04	0.71		-6.66	-1.04	-1.76	-0.92	-1.61		-1.20	0.11	-0.90
CG11807	-2.92		-1.70	9.48		1.62	0.51		1.90	-2.73		-0.11

Although such an approach showed some promise, it was abandoned in favour of visual inspection of 'by hour' results (as shown in Figure 8-9) in an effort to identify the most promising candidates for further study.

8.3.6 Behavioural arrhythmicity in *Ddc>GAL4* knockdown strains

In addition to identifying phenotypic outliers (such the *CG30404* knockdown strain in Figure 8-9), visual inspection of results identified genes which, when knocked down, caused behavioural arrhythmicity in LD conditions. Arrhythmicity in LD conditions is an unusually

severe mutant phenotype that might represent disruption of the circadian clock, sleep controlling neurons or coordinated locomotor output in general, therefore further investigation focussed on these results. Behavioural arrhythmicity was evident predominantly when knocking down expression using the *Ddc>GAL4* driver, but also using *th>GAL4* and - in one case - *c687>GAL4* drivers:

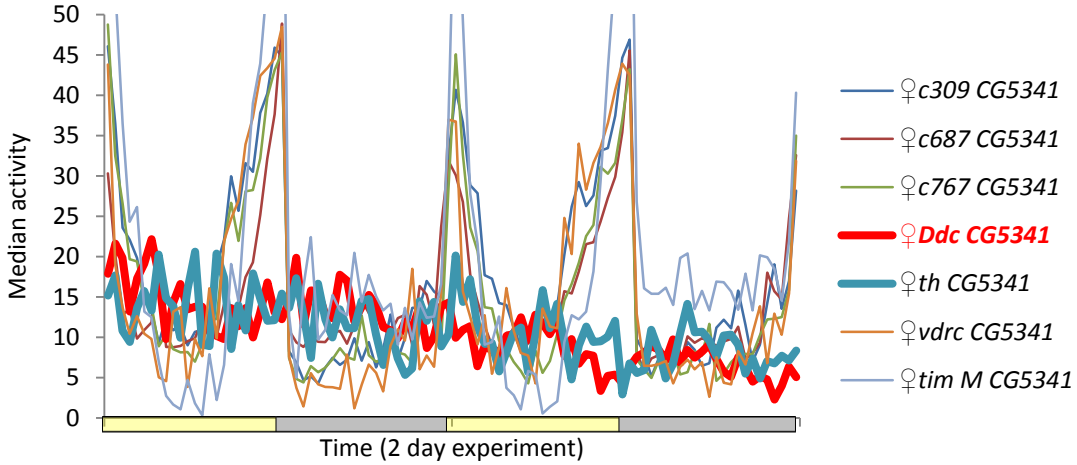


Figure 8-10 Knockdown of *CG5341* using either the *Ddc>GAL4* or *th>GAL4* drivers elicits behavioural arrhythmicity under LD conditions. vdrc = insertion control (i.e. no GAL4 driver). n=16 for each genotype**.**

Examination of individual activity traces for flies showed that the behavioural arrhtyhmicity observed was not the result of phase desynchronisation between flies, but was actually evident in the majority of individual fly activity traces:

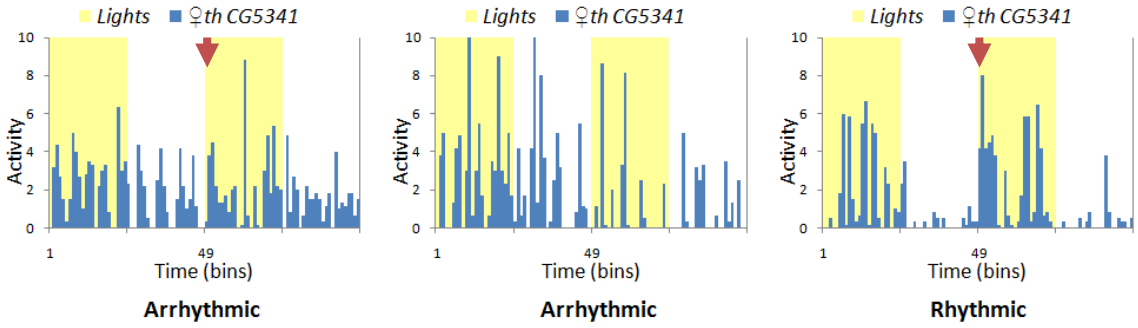


Figure 8-11 Examples of individual activity traces for *th>GAL4 CG5341* knockdown flies showing a mixture of rhythmic and arrhythmic behaviour. Two of the three examples shown above show a startle response to lights on (**red arrows**), suggesting that these lines have normal phototransduction.

Table 8-9 shows the likely function of genes showing such an arrhythmic phenotype, including many genes likely to play a role in the development of the eye or within the phototransduction signalling cascade. However, in most cases flies showed a robust startle response to changes in lighting conditions (Figure 8-11), suggesting that the phototransduction pathway remained intact - contrary what one might suppose based on ontology alone.

Table 8-9 Gene ontology information for genes causing behavioural arrhythmicity when knocked down using *Ddc>GAL4*. Note several genes are involved in eye development and the response to light.

Gene	molecular function	Associations
CG10236	receptor binding/cns development	startle response
CG15556	G-protein coupled receptor activity	-
CG2219	GTPase	-
CG8604	neurotransmitter secretion/endocytosis	Rhabdomere development
CG5341	vesicle targeting/neurotransmitter secretion	Rhabdomere development
CG10460	cysteine-type peptidase	*Drosophila* DMD model, circadian expression

Interestingly, gene knockdown using *tim>GAL4* driver (which expresses in clock cells as well as the eyes) did not cause behavioural arrhythmicity. If gene knockdown alone were neurotoxic, *tim>GAL4* mediated knockdown would cause damage to both the phototransduction input pathway and the circadian oscillator – the most likely causes of behavioural arrhythmicity. Given that this was not the case, this suggested that dopamineric (*Ddc>GAL4*) and serotinergic (*th>GAL4*) neurons must be involved either directly in sleep regulation (as has been established for dopamine), or that these neurons must play a role in controlling locomotor output

To resolve whether knockdown of these genes affected light input, the central oscillator or locomotor output, the experiments were repeated using not only the dopaminergic *Ddc>GAL4* driver, but also highly clock cell specific *Pdf>GAL4* driver (Renn et al., 1999), and the eye specific *GMR>GAL4* driver (Freeman, 1996). Surprisingly the original arrhythmic results could not be repeated, even after addition of a *UAS>dicer2* enhancer to the *Ddc>GAL4* background in an effort to reduce gene expression levels still further:

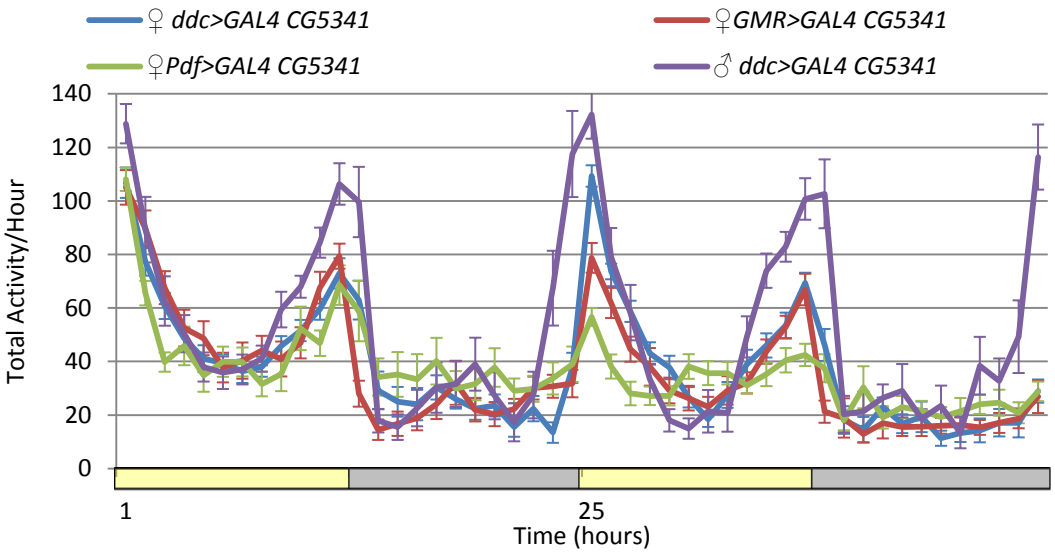


Figure 8-12 Example result illustrating the failure of repeat experiments to induce behavioural arrhythmicity in LD conditions. Results were similarly rhythmic for all candidate genes tested.

This failure was particularly disappointing given that the initial observations were specific to *Ddc>GAL4* and *th>GAL4* mediated knockdown, which suggested that these neurons played a functional role rather than that the results represented some form of experimental error or transient effect.

8.3.7 *UAS>Dicer2* enhanced knockdown

Given the difficulties inherent in identifying significant effects when the control genotype represents an outlier (Figure 8-9), and the failure to repeat results identified by manual inspection of data (Figure 8-12), it was clear that further screening would require the generation of more robust phenotypes in a more 'normal' sleep background.

To accomplish this, previous results were set aside and the *tim>GAL4* (clock cell) and *c309>GAL4* (mushroom body) drivers were crossed into a background containing the *UAS>Dicer2* RNAi enhancer. This served to disrupt the 'high sleeping' genetic background of the GAL4 driver lines, and also to enhance the RNAi effect. As a guard against ceiling effects limiting the identification of knockdown lines sleeping more than controls, further experiments

were performed using mated female flies, which sleep less than virgin female or male flies (Figure 8-8).

Rather than re-analysing existing lines, a more in-depth gene expression/sleep phenotype analysis was performed by the author using the newly developed BeFly! sleep analysis module to identify genes whose expression correlated not only with total sleep (as in the van Swinderen analysis), but also those correlating with sleep bout number and sleep bout length averaged across the whole day. Although daytime and nighttime sleep have been shown to represent independent networks (see previous section), by correlating sleep metrics calculated across the entire day with gene expression it was hoped that only those genes showing the most significant effects on sleep would be identified, facilitating their detection in follow-up studies.

The list of candidate genes was once again restricted to genes of interesting ontology and those showing the most significant correlations, generating the final list of genes (shown in appendix 10.5) that were subject to further analysis by RNAi. This list of genes contained a number of candidates identified by other studies.

Qualitative data also showed that – unlike in previous screens - *UAS>Dicer2* enhanced gene knockdown resulted in lethality in between 20 and 25% of crosses, confirming that the *UAS>Dicer2* transgene elicits a much greater degree of gene knockdown, and that the genes identified by our correlational analysis were indeed expressed in the brain at least at some stage during development.

The results in Figure 8-13 shows that - unlike in previous experiments - the control genotypes (**thick red lines**) lay squarely in the middle of the genotype distribution:

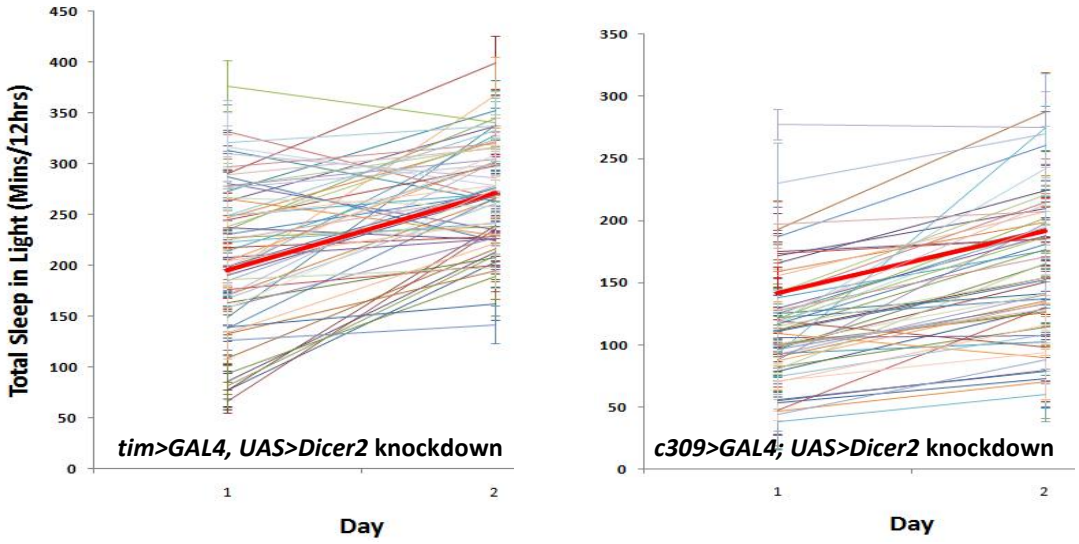


Figure 8-13 RNAi knockdown of gene expression using both *tim>GAL4* and *c309>GAL4* in a *UAS>Dicer2* background elicits significant changes in sleep metrics such as total sleep during the light phase.

Although the data in Figure 8-13 show a general trend towards flies sleeping for an additional hour during the light phase on the second day of analysis, in most lines the trend was of a similar magnitude and direction, therefore the average of both days' sleep was taken to be the mean value for each genotype and the results tabulated (see Table 8-10 and Table 8-11). As the control genotype lay in the middle of the result distribution, it was appropriate to analyse the data using the *post hoc* conservative Dunnet test.

Table 8-10 Results of *c309>GAL4; UAS-dicer2* mediated knockdown of candidate genes involved in sleep. Values shown are genotype means ± standard error of the mean. Genotypes were analysed relative to the no knockdown, congenic control line (*c309>GAL4; UAS-dicer2 x w; +; +*) using a *post hoc* Dunnett's test on values for individual flies (n=960). Significance levels are shown using asterisks: * denotes P<0.05, ** denotes P<0.01, *** denotes P<0.001. *CG11474* expression was knocked down using independent insertion lines (**x** and **y**). Several experiments were repeated, in such cases the repeat number is shown in brackets.

c309>GAL4; UAS>Dicer2	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Bout Length (Mins/Dark)	Mean Bout Length (Mins/Light)	Mean Waking Activity/Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/12hrs)	Total Sleep in Light (Mins/12hrs)
CG13704	25.23 ±1.01	10.4 ±0.9	75 ±7	202 ±10	17.8 ±1.6	9.7 ±0.7	5.15 ±0.17	18.7 ±3.3	961 ±46	413 ±18	104 ±12
CG8317	18.88 ±1.42	14.8 ±1.1	139 ±22	187 ±9	42.2 ±10.9	9.2 ±0.5	4.77 ±0.21	19.4 ±4.2	801 ±53	485 ±21	133 ±12
CG10103	15.96 ±1.63	8.9 ±0.7	211 ±34	167 ±9	95.5 ±34.9 *	15.5 ±1.8	5.49 ±0.16	17.6 ±1.0	896 ±53	504 ±22	131 ±14
CG10842	21.04 ±1.59	14.0 ±1.0	105 ±10	184 ±11	38.4 ±13.8	14.7 ±1.6	4.64 ±0.16	17.9 ±2.5	710 ±37	479 ±20	195 ±18
CG11208	14.93 ±1.09	9.3 ±1.1	201 ±28	189 ±7	62.2 ±15.2	15.6 ±1.4	4.97 ±0.24	24.3 ±2.4	761 ±49	552 ±14 *	133 ±16
CG11386	16.13 ±1.58	13.9 ±1.1	190 ±25	168 ±11	67.4 ±23.4	17.1 ±1.8	5.43 ±0.32	45.0 ±21.9	774 ±76	521 ±25	240 ±32
CG1143	21.41 ±0.89	6.4 ±0.7	112 ±15	200 ±8	23.1 ±1.7	9.5 ±0.9	5.58 ±0.24	28.0 ±4.4	1,035 ±48	452 ±16	63 ±8 *
CG1143	15.37 ±1.07	8.7 ±1.0	184 ±20	197 ±9	50.2 ±11.4	14.3 ±1.5	5.59 ±0.19	19.8 ±1.7	903 ±49	514 ±19	125 ±16
CG11474 (x)	19.53 ±1.03	7.7 ±0.9	118 ±13	188 ±7	26.3 ±2.4	11.8 ±1.5	6.02 ±0.23	32.3 ±5.1	1,086 ±69	453 ±23	98 ±15
CG11474 (y)	16.97 ±1.31	11.6 ±0.7	174 ±22	195 ±11	49.5 ±11.4	16.8 ±2.2	5.77 ±0.48	18.2 ±2.0	807 ±73	540 ±18	191 ±22
CG11501	19.37 ±0.90	6.9 ±0.5	149 ±19	198 ±11	25.5 ±2.3	9.6 ±0.9	5.30 ±0.18	30.7 ±2.9	996 ±49	444 ±18	68 ±8
CG12082	20.07 ±1.03	9.7 ±1.2	110 ±13	200 ±9	25.9 ±2.5	10.3 ±1.2	4.70 ±0.13	27.5 ±4.9	822 ±45	461 ±20	113 ±19
CG1214	20.60 ±1.42	13.0 ±1.0	129 ±18	204 ±9	51.6 ±23.1	10.3 ±0.7	4.84 ±0.20	16.6 ±3.1	810 ±53	487 ±22	132 ±13
CG12878	16.07 ±1.35	9.9 ±0.9	176 ±21	172 ±9	47.3 ±6.2	18.0 ±1.8	5.19 ±0.20	23.5 ±1.8	764 ±56	555 ±17 *	168 ±18
CG13348	20.53 ±1.43	14.5 ±1.0	133 ±19	188 ±11	36.3 ±8.3	9.9 ±0.8	4.54 ±0.22	18.8 ±2.7	765 ±62	479 ±26	145 ±15
CG13623	16.67 ±1.21	11.4 ±1.1	132 ±14	177 ±10	45.8 ±7.7	19.5 ±3.1	4.81 ±0.18	16.0 ±1.6	679 ±32 *	552 ±13 *	180 ±17
CG13742	20.59 ±1.18	10.0 ±0.7	132 ±17	168 ±11	47.7 ±21.5	15.9 ±1.9	5.78 ±0.18	21.8 ±2.1	910 ±45	497 ±16	159 ±18
CG13833	17.84 ±0.97	10.7 ±0.7	125 ±13	218 ±7	34.6 ±5.1	11.4 ±1.1	5.00 ±0.17	24.7 ±3.2	831 ±39	500 ±14	113 ±10
CG1483	19.22 ±1.12	10.6 ±0.8	128 ±19	186 ±11	31.4 ±4.1	14.8 ±1.5	5.15 ±0.29	20.3 ±3.2	837 ±65	491 ±20	149 ±14
CG15016	20.50 ±1.16	12.4 ±1.0	119 ±16	182 ±8	22.0 ±2.4	8.3 ±0.5	5.95 ±0.19	22.2 ±2.3	1,132 ±44	376 ±19	109 ±12
CG15110	17.60 ±1.35	11.6 ±1.0	208 ±24	190 ±7	35.5 ±3.9	9.1 ±0.8	4.84 ±0.19	17.3 ±1.7	809 ±42	494 ±18	114 ±15
CG16761	19.81 ±1.12	13.0 ±1.1	145 ±23	194 ±10	30.3 ±7.0	11.7 ±1.2	5.33 ±0.24	29.2 ±5.0	913 ±53	426 ±21	158 ±19
CG17907	21.59 ±0.80	11.5 ±0.9	79 ±8	215 ±9	20.4 ±1.2	8.4 ±0.4	4.32 ±0.17 *	23.0 ±2.3	801 ±36	415 ±13	98 ±10
CG17921	20.40 ±1.44	12.5 ±0.8	122 ±23	163 ±14	50.5 ±23.9	15.8 ±1.4	5.99 ±0.24	34.3 ±7.2	1,047 ±67	401 ±25	179 ±11
CG18812	15.63 ±0.98	10.6 ±1.0	166 ±15	213 ±6	40.8 ±4.9	11.2 ±1.0	6.89 ±0.38	23.1 ±1.7	1,120 ±77	510 ±10	130 ±18
CG2890	20.53 ±1.10	8.9 ±1.0	90 ±11	209 ±9	19.0 ±1.4	8.9 ±1.0	5.38 ±0.22	19.2 ±2.5	1,078 ±72	371 ±22	89 ±16
CG3172	21.38 ±1.23	13.1 ±1.2	124 ±20	178 ±10	37.0 ±12.8	13.2 ±1.7	4.79 ±0.18	14.4 ±1.6	752 ±41	499 ±20	162 ±17
CG32412	21.38 ±0.94	6.9 ±0.9	104 ±12	230 ±8	21.4 ±2.3	8.6 ±0.9	5.44 ±0.24	42.7 ±6.1	1,061 ±63	405 ±19	67 ±15

c309>GAL4; UAS>Dicer2	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Bout Length (Mins/Dark)	Mean Bout Length (Mins/Light)	Mean Waking Activity/Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/12hrs)	Total Sleep in Light (Mins/12hrs)
CG32475	16.13 ±0.97	11.0 ±0.9	173 ±22	181 ±10	75.5 ±29.4	16.9 ±2.1	5.21 ±0.21	20.8 ±1.3	769 ±39	549 ±12 *	157 ±11
CG3397	18.16 ±1.03	6.6 ±1.0	144 ±20	220 ±8	30.5 ±4.4	8.2 ±0.5 *	5.91 ±0.19	28.9 ±4.2	1,104 ±44	446 ±20	59 ±10 *
CG3500	21.45 ±1.42	17.5 ±1.2 *	119 ±14	168 ±11	28.0 ±4.5	12.7 ±1.1	4.36 ±0.32	13.6 ±1.8	625 ±54 *	496 ±21	224 ±24
CG3924	17.47 ±1.10	11.6 ±0.9	163 ±15	179 ±7	37.2 ±3.7	16.1 ±2.1	4.35 ±0.12 *	21.7 ±2.0	625 ±32 **	543 ±15	179 ±20
CG4250	21.03 ±1.18	12.1 ±0.8	140 ±17	191 ±9	32.4 ±5.8	13.3 ±1.4	5.36 ±0.19	18.0 ±1.6	853 ±45	493 ±19	153 ±13
CG4250 (2)	19.34 ±0.96	9.8 ±0.9	139 ±15	218 ±7	28.9 ±3.2	10.7 ±0.7	5.16 ±0.17	25.9 ±3.6	892 ±42	473 ±17	107 ±12
CG4408	22.57 ±0.95	14.6 ±1.3	93 ±10	175 ±11	20.7 ±1.9	11.6 ±1.1	4.76 ±0.30	19.5 ±5.4	810 ±75	428 ±22	184 ±27
CG4408 (2)	21.34 ±0.91	9.8 ±0.9	124 ±12	194 ±8	23.3 ±1.5	9.7 ±0.7	4.98 ±0.19	15.9 ±1.3	891 ±46	457 ±12	99 ±11
CG4583	19.63 ±1.36	10.3 ±1.2	120 ±16	197 ±11	24.6 ±2.4	10.7 ±0.7	5.33 ±0.24	37.1 ±10.9	982 ±73	429 ±25	115 ±15
CG4686	13.59 ±1.28 *	11.1 ±0.7	224 ±28	188 ±8	78.1 ±22.2	16.9 ±1.7	4.59 ±0.17	21.9 ±1.9	643 ±27 **	554 ±15 *	178 ±15
CG4974	7.77 ±1.25 ***	8.0 ±0.7	280 ±33 ***	155 ±12	241.9 ±50.1 ***	25.8 ±3.7 ***	4.81 ±0.18	20.5 ±1.5	576 ±25 **	649 ±12 ***	182 ±19
CG5067	22.00 ±1.18	12.5 ±1.5	98 ±12	174 ±12	22.6 ±4.3	10.8 ±1.1	5.03 ±0.31	17.3 ±1.4	921 ±88	409 ±26	143 ±22
CG5532	21.83 ±1.04	7.0 ±0.8	123 ±13	221 ±7	24.1 ±2.4	7.0 ±0.4 *	4.10 ±0.18 **	24.2 ±3.2	758 ±34	463 ±17	49 ±6 *
CG5555	20.57 ±1.00	11.2 ±0.9	135 ±14	200 ±8	25.1 ±2.2	9.6 ±0.7	5.61 ±0.22	23.3 ±3.1	983 ±51	458 ±14	111 ±12
CG5590	20.93 ±1.17	8.1 ±1.1	87 ±10	219 ±8	32.4 ±11.5	7.9 ±0.7 *	4.82 ±0.18	33.3 ±5.5	940 ±54	413 ±25	66 ±10
CG7590 (2)	22.19 ±0.93	13.2 ±1.4	104 ±18	192 ±9	21.8 ±1.9	11.2 ±0.8	5.55 ±0.22	15.8 ±2.2	961 ±58	438 ±16	149 ±18
CG7722	19.04 ±1.39	11.8 ±1.2	145 ±17	186 ±10	35.4 ±4.9	10.3 ±2.0	4.49 ±0.16	19.8 ±2.2	726 ±42	513 ±19	123 ±21
CG7851	19.67 ±1.46	11.3 ±1.2	107 ±24	206 ±13	24.2 ±3.4	9.5 ±1.2	5.31 ±0.46	29.6 ±6.9	964 ±110	429 ±28	117 ±26
CG8058	17.13 ±1.27	7.7 ±1.0	182 ±26	204 ±8	38.5 ±5.0	12.3 ±1.9	5.52 ±0.20	20.0 ±1.8	942 ±60	505 ±21	92 ±17
CG8453	19.57 ±1.46	14.7 ±0.8	152 ±29	180 ±10	61.7 ±27.1	12.6 ±0.8	4.90 ±0.22	19.8 ±3.6	798 ±60	467 ±25	181 ±13
CG8680	17.29 ±1.37	15.9 ±0.8	155 ±16	157 ±9	40.7 ±5.5	17.0 ±1.6	4.82 ±0.36	18.6 ±2.2	651 ±64 *	524 ±21	250 ±17 *
CG8768	17.33 ±1.38	14.8 ±1.2	182 ±24	163 ±11	41.2 ±5.9	14.4 ±1.3	4.19 ±0.17 *	13.1 ±0.7	577 ±37 **	556 ±20 *	202 ±16
CG9224	14.23 ±1.28	9.9 ±0.8	198 ±22	179 ±9	74.5 ±22.9	13.4 ±1.2	4.80 ±0.19	20.9 ±1.9	715 ±49	577 ±14 **	134 ±15
CG9320	19.45 ±1.58	13.4 ±1.4	169 ±23	140 ±16	33.8 ±4.6	24.8 ±2.8 **	4.61 ±0.17	20.5 ±1.7	601 ±53 **	527 ±24	276 ±19 **
CG9353	21.13 ±0.91	16.4 ±1.1	105 ±8	185 ±10	25.4 ±2.0	10.5 ±0.7	4.69 ±0.18	17.3 ±2.1	739 ±50	489 ±20	180 ±19
CG9852	19.56 ±1.11	8.8 ±1.3	130 ±22	195 ±11	26.5 ±3.3	9.8 ±0.8	5.90 ±0.29	31.9 ±6.3	1,101 ±71	427 ±25	93 ±15
w + +	17.67 ±1.69	11.8 ±1.4	129 ±17	192 ±9	118.4 ±43.1	15.4 ±1.6	5.81 ±0.24	24.8 ±6.0	966 ±78	468 ±30	167 ±17
w + + (old)	20.72 ±1.29	13.7 ±1.0	108 ±19	177 ±12	24.4 ±4.4	12.1 ±1.0	5.93 ±0.29	25.8 ±3.2	1,079 ±74	374 ±21	168 ±16
CG11386	19.89 ±1.09	11.5 ±1.3	135 ±19	191 ±9	28.8 ±2.7	9.4 ±0.7	5.25 ±0.20	18.2 ±2.4	866 ±48	501 ±19	121 ±18
CG17304	21.77 ±1.34	11.2 ±1.0	88 ±20	170 ±11	23.0 ±4.8	10.2 ±0.7	5.56 ±0.19	36.0 ±8.4	1,064 ±50	375 ±24	114 ±12
CG30115	15.64 ±1.45	11.7 ±1.5	184 ±27	190 ±12	48.7 ±7.6	12.5 ±1.8	5.75 ±0.38	39.5 ±10.9	931 ±88	504 ±26	156 ±29
CG6969	17.22 ±1.46	8.7 ±0.9	141 ±19	179 ±9	52.5 ±22.2	9.6 ±0.9	5.57 ±0.25	26.5 ±3.2	1,041 ±49	419 ±25	83 ±10

Table 8-11 Results of *tim>GAL4; UAS-dicer2* mediated knockdown of candidate genes involved in sleep. Values shown are genotype means ± standard error of the mean. Genotypes were analysed relative to the no knockdown, congenic control line (tim>GAL4; UAS-dicer2 x w; +; +) using a *post hoc* Dunnett's test on values for individual flies (n=864). Significance levels are shown using asterisks: * denotes p<0.05, ** denotes p<0.01, *** denotes p<0.001. Several knockdown experiments were repeated, the repeat number is shown in brackets. Results using the *w; +; +* control line were repeated with old and young flies as indicated.

tim>GAL4, UAS>Dicer2	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Bout Length (Mins/Dark)	Mean Bout Length (Mins/Light)	Mean Waking Activity/Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/12hrs)	Total Sleep in Light (Mins/12hrs)
CG10103	20.69 ±0.77	17.03 ±0.90	86.41 ±10.15	149.81 ±14.80	20.75 ±1.92	14.07 ±1.04	6.56 ±0.36	22.34 ±4.36	1,125.63 ±111.84	402.97 ±25.54	229.69 ±15.66
CG10842	19.89 ±1.03	16.39 ±0.80	116.25 ±13.54	168.75 ±13.34	28.19 ±3.20	19.55 ±1.78	5.06 ±0.22	19.11 ±3.74	656.21 ±56.65	485.89 ±23.78	313.04 ±25.18
CG11208	19.18 ±1.14	13.29 ±1.01	85.64 ±14.75	169.30 ±17.01	17.90 ±2.92	11.56 ±0.88	6.39 ±0.26	63.04 ±11.12 ***	1,269.04 ±78.00 *	295.18 ±25.96 ***	163.57 ±20.41
CG1143	22.16 ±0.68	13.50 ±0.63	69.22 ±7.29 **	212.91 ±9.44	14.57 ±0.99	11.20 ±0.61	5.82 ±0.20	26.88 ±5.37	1,139.31 ±54.78	315.63 ±19.32 **	150.31 ±10.34
CG11474	19.00 ±1.45	17.67 ±1.28	120.28 ±19.37	158.11 ±15.38	38.21 ±11.98	21.57 ±2.65	5.80 ±0.30	20.00 ±11.33	703.83 ±72.70	489.44 ±29.30	345.00 ±26.93
CG11501	17.63 ±2.02	10.75 ±1.28	150.31 ±39.26	199.88 ±13.54	37.55 ±10.27	11.83 ±1.96	5.63 ±0.20	21.88 ±6.83	981.88 ±106.50	443.75 ±42.12	145.63 ±34.84
CG1214	22.90 ±0.96 *	17.47 ±0.76	86.67 ±8.75	140.43 ±14.23	20.85 ±2.10	17.87 ±1.50	5.23 ±0.23	29.83 ±4.85	747.50 ±57.07	430.17 ±22.28	300.33 ±22.30
CG12878	18.64 ±1.12	16.29 ±0.81	149.46 ±17.29	157.50 ±10.41	32.55 ±3.53	18.48 ±1.54	5.37 ±0.43	14.29 ±3.14	697.07 ±78.78	510.54 ±19.43	288.04 ±19.14
CG13618	11.28 ±0.73	13.34 ±0.74	185.56 ±12.46	173.87 ±6.08	59.12 ±4.49	17.55 ±1.44	5.68 ±0.23	13.59 ±1.33	755.47 ±52.04	575.47 ±10.61	218.28 ±14.87
CG13623	19.04 ±1.11	16.00 ±0.88	103.08 ±13.31	190.96 ±12.46	26.82 ±2.98	17.83 ±2.05	5.95 ±0.26	19.23 ±2.61	892.38 ±56.29	439.62 ±20.38	255.38 ±16.30
CG13704	20.47 ±1.22	15.66 ±0.84	112.97 ±11.45	155.72 ±12.81	33.15 ±5.70	18.02 ±1.38	4.67 ±0.13	25.16 ±4.36	623.34 ±33.87	504.53 ±16.59	271.56 ±19.24
CG13742	20.37 ±1.05	15.70 ±0.79	112.17 ±13.93	171.40 ±15.05	27.35 ±3.28	19.37 ±1.57	5.90 ±0.23	24.83 ±4.20	829.13 ±62.91	463.33 ±22.33	289.83 ±20.39
CG13833	20.32 ±1.00	15.46 ±1.06	103.57 ±12.11	196.89 ±11.74	24.72 ±2.19	15.20 ±1.06	5.21 ±0.27	36.07 ±4.80	796.64 ±59.25	450.00 ±19.90	231.07 ±18.96
CG1483	17.07 ±1.11	13.13 ±0.80	167.17 ±21.47	161.97 ±12.07	38.89 ±6.77	21.70 ±2.60	5.16 ±0.16	31.00 ±7.52	739.50 ±65.42	488.83 ±26.03	260.67 ±20.50
CG15110	14.43 ±1.14	16.21 ±0.98	198.75 ±23.48	171.46 ±8.93	53.80 ±7.78	18.54 ±1.57	5.40 ±0.28	14.11 ±2.09	630.14 ±43.92	570.18 ±19.19	284.29 ±19.75
CG16761	20.90 ±1.10	14.70 ±0.85	103.17 ±11.30	179.17 ±12.07	24.73 ±3.58	13.69 ±1.29	5.34 ±0.18	24.67 ±3.71	890.70 ±60.66	420.33 ±22.01	202.33 ±20.79
CG17907	21.31 ±1.02	15.00 ±0.85	88.27 ±10.48	155.42 ±14.83	20.73 ±3.20	15.31 ±0.95	6.19 ±0.29	19.23 ±3.13	1,053.85 ±80.44	387.12 ±22.01	223.08 ±14.76
CG17921	22.44 ±1.13 *	17.22 ±0.80	73.13 ±6.86 *	179.09 ±15.50	18.05 ±1.59	13.22 ±0.94	5.64 ±0.19	38.28 ±7.53	959.66 ±58.28	365.00 ±19.67	232.50 ±18.77
CG18812	23.13 ±1.07 *	15.93 ±0.91	86.83 ±8.58	168.33 ±13.06	19.98 ±1.94	14.46 ±1.70	4.72 ±0.17	19.17 ±3.10	773.17 ±45.16	412.83 ±17.03	217.17 ±19.84
CG3172	19.23 ±1.36	13.85 ±0.95	122.92 ±18.82	157.04 ±14.16	31.33 ±6.05	18.84 ±2.33	5.27 ±0.14	40.96 ±11.33	776.38 ±62.49	455.77 ±27.05	257.88 ±31.35
CG32412	19.88 ±0.93	14.53 ±0.83	101.56 ±7.58	168.59 ±13.84	26.06 ±2.64	17.64 ±1.63	5.26 ±0.13	25.63 ±4.03	793.97 ±52.75	453.13 ±18.85	247.34 ±20.30
CG32412 (2)	17.13 ±1.56	14.75 ±0.89	127.92 ±12.35	186.63 ±11.01	43.01 ±6.45	17.62 ±1.29	5.10 ±0.18	16.67 ±2.60	672.88 ±57.72	542.50 ±23.84	251.04 ±18.48
CG32412 (3)	17.00 ±1.37	15.55 ±0.95	159.00 ±19.49	176.20 ±8.75	36.71 ±5.66	13.05 ±1.83	5.04 ±0.18	20.25 ±2.16	758.35 ±40.39	493.75 ±24.56	188.50 ±16.05
CG32475	15.32 ±1.01	17.64 ±0.85	187.86 ±24.61	160.21 ±9.16	48.15 ±7.86	21.50 ±1.63	4.82 ±0.12	23.21 ±6.50	491.29 ±19.50 *	567.86 ±11.60	358.39 ±17.55 *
CG3397	18.35 ±1.44	16.46 ±1.16	174.77 ±26.87	186.04 ±11.09	39.27 ±6.31	18.51 ±3.24	5.30 ±0.21	15.77 ±1.59	683.85 ±50.43	529.23 ±20.15	257.88 ±25.78
CG3924	16.63 ±1.12	16.30 ±0.55	169.33 ±21.61	191.13 ±9.26	43.92 ±7.52	16.17 ±0.75	5.59 ±0.12	14.33 ±2.25	710.27 ±27.12	543.17 ±16.64	260.00 ±12.51
CG4250	21.87 ±1.00	16.30 ±1.03	97.00 ±10.59	134.50 ±12.60	19.00 ±1.67	16.35 ±1.62	5.68 ±0.34	28.17 ±6.64	938.17 ±97.94	395.50 ±25.24	253.50 ±22.69

tim>GAL4, UAS>Dicer2	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Bout Length (Mins/Dark)	Mean Bout Length (Mins/Light)	Mean Waking Activity/Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/12hrs)	Total Sleep in Light (Mins/12hrs)
CG4250 (2)	20.10 ±0.94	11.00 ±0.84	100.83 ±9.97	196.07 ±11.96	24.93 ±4.28	12.33 ±0.82	5.61 ±0.25	27.83 ±4.27	1,018.80 ±64.10	411.50 ±21.13	134.00 ±11.75
CG4408	20.97 ±0.88	15.06 ±0.79	105.94 ±10.17	127.00 ±13.12	24.46 ±2.60	19.86 ±1.62	5.11 ±0.16	16.72 ±1.89	713.69 ±57.42	454.53 ±22.99	298.44 ±22.40
CG4408 (2)	19.00 ±1.42	17.59 ±0.94	112.73 ±19.49	169.18 ±18.20	31.85 ±5.29	16.96 ±1.39	5.19 ±0.21	23.41 ±4.92	729.55 ±59.18	458.41 ±34.56	281.14 ±17.23
CG4583	21.38 ±0.81	15.29 ±1.22	84.38 ±8.35	210.08 ±11.98	23.04 ±1.76	13.12 ±0.94	4.74 ±0.14	18.96 ±2.23	749.58 ±39.59	463.33 ±16.10	192.92 ±17.19
CG4686	17.70 ±1.44	15.90 ±1.06	128.00 ±18.03	177.75 ±14.45	43.65 ±12.43	15.01 ±0.90	5.19 ±0.26	16.50 ±2.15	763.50 ±64.81	486.25 ±26.32	232.25 ±15.07
CG4974	21.39 ±1.00	13.82 ±1.02	124.64 ±16.28	141.46 ±15.40	23.70 ±1.73	21.54 ±2.57	5.36 ±0.19	25.00 ±5.76	768.96 ±65.47	467.50 ±20.71 ***	279.82 ±27.30
CG5067	17.69 ±0.97	14.34 ±0.71	112.03 ±11.18	157.94 ±12.77	27.93 ±3.04	16.63 ±1.04	5.76 ±0.17	39.06 ±7.84	912.63 ±59.19	433.13 ±24.80	232.03 ±14.39
CG5532	13.75 ±1.78	12.65 ±0.92	213.75 ±31.03	143.05 ±13.87	72.39 ±13.99 **	29.20 ±4.79 **	5.03 ±0.18	20.00 ±4.63	552.35 ±53.18	585.75 ±23.24	308.50 ±28.07
CG6969	21.37 ±1.02	17.37 ±0.84	99.33 ±13.29	152.57 ±14.45	27.65 ±5.30	17.65 ±1.32	4.89 ±0.17	20.83 ±4.75	659.27 ±39.76	466.67 ±22.10	296.83 ±19.89
CG7143	15.56 ±1.84	17.61 ±0.89	180.56 ±33.41	156.67 ±14.02	57.32 ±13.86	17.27 ±1.70	5.74 ±0.41	23.33 ±5.24	740.72 ±107.23	534.17 ±32.77	291.94 ±23.12
CG7590	21.67 ±1.36	17.33 ±0.94	105.00 ±14.99	134.33 ±14.12	25.95 ±4.31	17.04 ±1.29	5.36 ±0.23	19.38 ±3.93	777.42 ±70.52	450.00 ±26.64	290.83 ±21.11
CG7722	23.83 ±0.99 **	13.71 ±1.01	76.04 ±6.38 *	173.21 ±14.81	19.39 ±2.15	12.64 ±1.32	5.29 ±0.27	28.13 ±4.09	915.63 ±77.72	422.29 ±22.62	177.71 ±20.30
CG7851	17.14 ±1.46	13.82 ±1.22	148.50 ±27.14	179.05 ±13.57	40.54 ±7.83	15.62 ±2.27	5.31 ±0.18	21.36 ±2.98	779.50 ±72.26	501.59 ±29.07	214.55 ±30.60
CG8058	21.72 ±0.95	17.97 ±0.89	94.38 ±10.17	147.91 ±13.79	24.38 ±3.28	14.12 ±0.83	5.41 ±0.21	22.66 ±4.16	819.53 ±52.89	441.72 ±22.84	247.19 ±14.47
CG8317	20.72 ±0.91	14.44 ±0.71	82.97 ±13.40	169.31 ±14.32	19.93 ±2.54	15.50 ±3.37	5.95 ±0.21	40.63 ±7.29	1,064.50 ±67.38	365.63 ±23.93	191.88 ±17.95
CG8453	21.93 ±1.11	16.29 ±0.81	107.50 ±11.45	156.11 ±14.48	22.52 ±2.47	14.55 ±0.98	5.80 ±0.20	21.79 ±4.26	905.21 ±53.88	434.11 ±19.41	236.79 ±20.08
CG8680	19.89 ±1.54	13.67 ±1.12	138.06 ±13.75	157.28 ±13.42	31.39 ±3.04	25.25 ±2.37	4.35 ±0.18	11.39 ±1.61	482.50 ±28.86	548.61 ±15.73	328.89 ±26.34
CG8768	12.37 ±0.97	15.90 ±0.97	240.17 ±22.69 *	182.50 ±6.49	61.41 ±6.83	14.40 ±1.14	4.40 ±0.17 *	17.00 ±1.65	552.97 ±33.32	593.33 ±12.14 *	223.83 ±16.93
CG9224	16.25 ±1.59	17.70 ±0.82	166.50 ±34.02	163.20 ±12.11	57.45 ±16.50	16.99 ±1.50	5.10 ±0.28	17.00 ±5.55	603.45 ±32.36	556.50 ±17.69	287.50 ±16.77
CG9320	19.94 ±1.32	21.00 ±0.69 **	123.06 ±13.10	187.11 ±9.35	29.34 ±3.22	14.48 ±0.84	5.24 ±0.29	13.33 ±3.13	669.11 ±66.78	516.94 ±17.47	302.22 ±17.24
CG9353	20.60 ±1.09	16.50 ±0.93	93.67 ±7.15	166.20 ±11.15	25.86 ±3.01	16.66 ±1.53	6.26 ±0.45	13.50 ±1.80	947.57 ±116.35	463.17 ±21.95	262.17 ±20.38
CG9852	18.64 ±1.57	16.73 ±1.30	135.45 ±19.09	142.18 ±16.81	32.81 ±4.85	20.66 ±2.78	5.50 ±0.26	24.55 ±4.17	738.91 ±52.32	463.18 ±29.02	298.41 ±18.76
w + +	16.61 ±1.29	14.07 ±0.79	159.29 ±26.02	164.50 ±12.49	38.28 ±5.97	16.82 ±2.06	5.81 ±0.18	23.75 ±3.72	871.32 ±61.63	467.50 ±22.86	233.57 ±24.07
w + + (old)	20.50 ±1.35	17.93 ±1.14	95.67 ±12.44	135.00 ±16.22	20.08 ±2.51	14.66 ±1.33	5.40 ±0.27	33.67 ±9.28	915.77 ±87.83	364.83 ±28.83	262.83 ±25.29
CG17304	17.18 ±1.44	18.59 ±1.24	126.14 ±11.32	192.14 ±11.47	41.11 ±6.44	16.07 ±1.64	5.19 ±0.19	27.95 ±9.13	646.95 ±56.13	527.27 ±23.09	297.50 ±28.55
CG30115	21.93 ±1.26	19.23 ±0.93 *	93.67 ±9.07	203.83 ±10.44	23.88 ±2.85	13.53 ±1.01	5.33 ±0.22	27.00 ±4.88	823.37 ±67.36	439.00 ±22.79	253.67 ±17.38
CG7878	22.15 ±1.12	17.15 ±0.97	72.69 ±6.84 *	174.96 ±14.33	20.37 ±1.63	12.19 ±0.77	5.11 ±0.23	27.50 ±7.64	831.92 ±66.91	428.85 ±27.21	216.15 ±19.78

As can be seen in both Table 8-10 and Table 8-11, a number of gene knockdown lines showed significant differences in sleep compared to controls. Lines in which gene knockdown caused lethality were repeated without the *UAS>Dicer2* enhancer (data not shown).

8.3.8 The results of *UAS>Dicer2* enhanced knockdown experiments are only partially reproducible

Experimental lines showing significant effects in the more robust, *UAS>Dicer2* enhanced knockdown model were subject to repeated investigation to check the reproducibility of the original results (except in cases in which the UAS construct had been lost). Repeat experiments were assessed over three full days of LD rather than two days in an effort to improve the reliability of the results:

Table 8-12 *UAS-Dicer2; c309>GAL4* repeats in fertilised females. Values shown are genotype means ± standard error of the mean recorded over three full days. Genotypes were analysed relative to the no knockdown, congenic control line (*UAS-Dicer2; c309>GAL4 x w; +; +*) using a conservative *post hoc* Dunnett's test on values for individual flies (n=283). Significance levels are shown using asterisks: * denotes p<0.05, ** denotes p<0.01, *** denotes p<0.001

Gene	n	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Waking Activity/Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/12hrs)	Total Sleep in Light (Mins/12hrs)
CG11208	36	16 ±1.12	14 ±0.87*	168 ±17.40	167 ±9.56	5.2 ±0.24	25 ±7.79	754 ±55.78	512 ±18.95	221 ±18.12**
CG32475	30	21 ±1.21	16 ±0.88***	134 ±16.76	175 ±10.61	4.2 ±0.13	19 ±2.55	614 ±29.27	491 ±16.36	226 ±12.16**
CG3397	29	15 ±1.34	14 ±0.91**	182 ±24.05	176 ±9.41	5.0 ±0.19	19 ±2.04	714 ±44.42	542 ±19.27	194 ±17.06
CG3500	22	11 ±1.26**	10 ±1.00	221 ±25.49*	169 ±10.44	4.7 ±0.19	19 ±2.46	605 ±41.10	620 ±14.42**	178 ±23.19
CG4686	31	16 ±1.29	14 ±0.98*	163 ±25.94	175 ±8.96	7.0 ±0.76**	23 ±1.85	1,079 ±113.22*	486 ±21.16	181 ±17.95
CG4974	22	13 ±1.60	15 ±0.80**	217 ±30.08*	162 ±7.08*	5.7 ±0.47	18 ±1.60	807 ±106.76	566 ±23.25	188 ±18.49
CG5532	29	17 ±1.25	12 ±0.93	171 ±23.10	184 ±7.24	4.6 ±0.15	21 ±2.12	717 ±36.28	532 ±19.74	125 ±12.87
CG8680	27	17 ±1.36	14 ±0.94*	200 ±26.91	166 ±7.33	4.1 ±0.16	19 ±3.51	611 ±35.07	541 ±16.57	151 ±15.92
CG8768	26	14 ±1.52	13 ±1.00	197 ±29.15	173 ±8.86	4.9 ±0.26	19 ±1.47	713 ±54.10	558 ±20.00	161 ±14.52
CG9320	21	16 ±1.44	15 ±1.07**	166 ±25.11	177 ±11.11	4.3 ±0.21	25 ±2.48	606 ±47.95	539 ±19.40	212 ±19.25*
w; +; +	10	18 ±1.78	9 ±1.39	133 ±21.28	195 ±16.25	5.1 ±0.26	21 ±2.44	818 ±77.11	511 ±28.38	131 ±23.36

These results in Table 8-12 suggest that the genes *CG11208*, *CG32475* and *CG9320* have a general effects on sleep during the day, *CG3500* affects sleep during the night and *CG4974* affects the length a sleep bout can be maintained. *CG4686* has an effect on activity levels, but not directly on sleep.

Table 8-13 *UAS>Dicer2, timGAL4* knockdown repeats in fertilised females. Values shown are genotype means ± standard error of the mean recorded over three full days. Genotypes were analysed relative to the no knockdown, congenic control line (*UAS-Dicer2, tim>GAL4 x w; +; +*) using a conservative *post hoc* Dunnett's test on values for individual flies (n=292). Significance levels are shown using asterisks: * denotes p<0.05, ** denotes p<0.01, *** denotes p<0.001.

Genotype	n	Bouts of Sleep (Dark)	Bouts of Sleep (Light)	Longest Sleep Duration (Min)	Longest Sleep Start (Bin)	Mean Waking Activity/Bin (/Day)	Night Offset (Min)	Total Activity (/Day)	Total Sleep in Dark (Mins/12hrs)	Total Sleep in Light (Mins/12hrs)
CG11208	37	21 ±1.01	13 ±0.80**	107 ±14.14	189 ±11.11	4.8 ±0.19***	31 ±5.90	837 ±63.79**	427 ±22.12	167 ±18.49
CG32475	21	20 ±1.13	16 ±0.92	82 ±10.04	170 ±16.97	5.7 ±0.22	28 ±5.82	994 ±76.74	366 ±29.28	222 ±24.44
CG3397	29	18 ±1.21	13 ±1.29**	75 ±10.60*	169 ±13.91	8.3 ±0.71*	49 ±9.81	1,727 ±204.82**	297 ±25.47***	152 ±21.53
CG4686	23	20 ±1.20	13 ±1.00**	88 ±13.20	167 ±16.46	5.6 ±0.20	42 ±8.10	1,057 ±68.09	348 ±27.01	169 ±20.94
CG4974	14	16 ±1.97	15 ±1.95	145 ±24.45	181 ±16.62	6.6 ±1.11	30 ±9.38	925 ±173.74	512 ±33.73*	240 ±36.80
CG5532	27	16 ±1.33	10 ±0.93***	163 ±20.08	165 ±12.48	5.8 ±0.26	33 ±10.72	889 ±76.84	508 ±27.52**	184 ±26.87
CG7722	16	16 ±1.99	14 ±1.23	183 ±28.10	183 ±13.97	5.0 ±0.24*	16 ±2.09	697 ±62.05**	546 ±26.92***	208 ±24.17
CG8680	22	15 ±1.20	15 ±0.96	176 ±27.70	165 ±13.24	5.4 ±0.23	38 ±5.25	757 ±59.44**	495 ±21.68*	252 ±23.46*
CG8768	33	21 ±1.18	17 ±0.88	126 ±16.33	141 ±13.33*	5.7 ±0.22	21 ±3.30	799 ±59.14**	438 ±19.50	313 ±21.65***
CG9224	14	17 ±1.51	15 ±1.60	155 ±25.43	176 ±14.82	5.1 ±0.34	18 ±2.26	713 ±61.65**	556 ±21.58***	195 ±25.31
CG9320	23	19 ±1.34	17 ±1.24	124 ±19.53	177 ±14.51	5.3 ±0.29	25 ±4.49	812 ±77.27*	448 ±26.13	244 ±22.85
w; +; +	33	18 ±1.27	17 ±1.03	127 ±21.69	174 ±11.53	6.8 ±0.66	31 ±10.80	1,239 ±175.62	417 ±28.72	190 ±18.17

Comparison of the repeat experiments (Table 8-12 and Table 8-13) with the original knockdown experiments (Table 8-10 and Table 8-11) shows that the reproducibility of the originally significant results is relatively low. Furthermore, the pattern of significant results shows only limited reproducibility:

Table 8-14 Table comparing the significant results from the preliminary screen with significant results in the repeat screen. Metrics for which a line differs significantly from controls are indicated in the table by abbreviations: **TS** = total sleep, **BS** = Bouts of sleep, **NO** = Night offset, **TA** = total activity, **MWA** = mean waking activity, **MBL** = mean bout length, **LSS** = longest sleep start, **LSD** = longest sleep duration. Suffixes '**-L**' or '**-D**' refer to 'light' and 'dark' respectively. Significance levels calculated using a conservative *post hoc* Dunnett test; * denotes P<0.05, ** denotes P<0.01, *** denotes P<0.001. '**Brain level**' refers to whether a gene is judged to be up or down regulated in the brain as assessed by the FlyAtlas project. Results which are found to be significant in both the preliminary screen and the repeat are indicated (**blue highlights**).

Gene	*Brain level*	*c309>GAL4; UAS>Dicer2*	*c309>GAL4; UAS>Dicer2* (2)	*tim>GAL4; UAS>Dicer2*	*tim>GAL4; UAS>Dicer2* (2)
CG11208	Down	TS-D*	BS-L*, TS-L**	NO***,***TA****, TS-D***	BS-L**, MWA***, ***TA****
CG5532	Down	MWA**, MBL-L*	none	MBL-L**, MBL-D**	BS-L***, TS-D**
CG8768	Up	MWA*, TA**, TS-D*	none	***LSD****, MWA*, TS-D*	***LSD****, TA**, TS-L***
CG32475	Up	TSD*	BS-L***, TS-L**	TA*, TSL-L*	none
CG3397	Up	MBL-L*, TSL-L*	BS-L**	none	BS-L**, MWA*, TA**, TS-D***
CG4686	Down	BD-D*, ***TA****, TS-D*	BS-L*, MWA**, ***TA****	none	BS-L**
CG4974	Up	BS-D*, ***LSD****, MBL-D***, MBL-L***, TA**, TSD***	BS-L**, ***LSD****, LSS*	***TS-D****	***TS-D****
CG8680	Down	TA*, TS-L*	BS-L*	none	TA**, TS-D*, TS-L*
CG9320	-	MBL-L**, TA**, ***TS-L****	BS-L**, ***TS-L****	BS-L**	TA*

It is clear that the effects of gene knockdown are specific to each GAL4 driver. Additionally, results are not very replicable; only in the case of *CG9320*, *CG4974*, and *CG8768* does gene knockdown in the repeat experiment have a similar effect on sleep as in the preliminary screen (blue highlights). Although *CG11208* and *CG4686* knockdown has similar effects in both primary and secondary screens, in both cases this predominantly affects total activity, therefore knockdown of these genes may have a more general effect on fly arousal (as in the hyperactive *CG4686*), or general locomotor activity and health (as in the hypoactive *CG11208*) rather than on sleep itself.

Taken together, although ~50% of the genes subject to secondary analysis were confirmed to have a replicable effect on a sleep metric, in general the reproducibility of individual results is lower than that found by Wu *et al.* in their sleep screen (Wu et al., 2008a). This may result in part from Wu *et al.*'s use of P element insertion mutants, thereby circumventing the problems of partial gene knockdown observed when using RNAi for screening.

Higher resolution analysis of the sleep data 'by hour' suggested that the lack of reproducibility might in part be due to the changes in sleep between genotypes increasing over time:

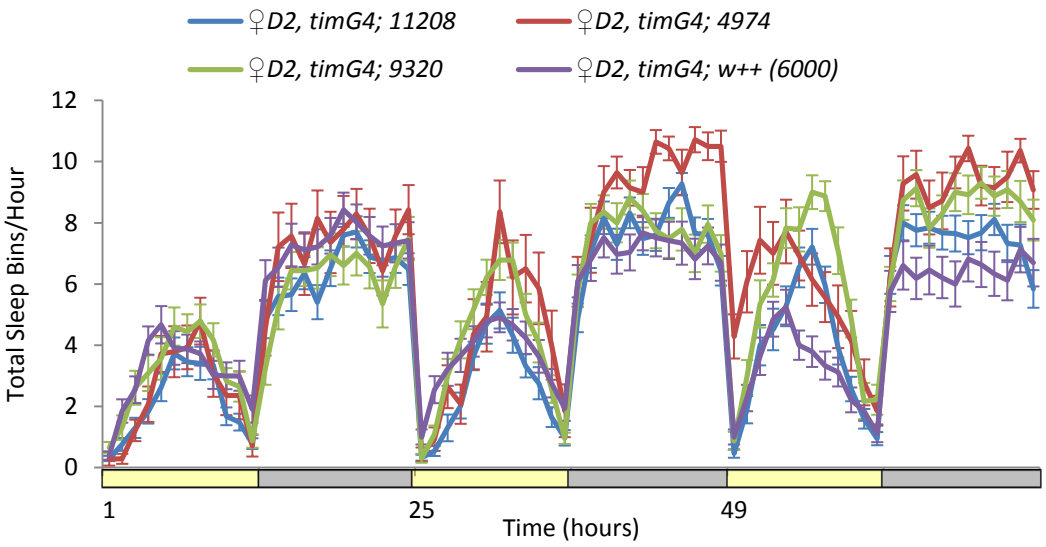


Figure 8-14 Differences in total sleep between knockdown lines and D2, timG4, w++ (6000) controls are accentuated over time. **D2** = UAS>dicer2, **timG4** = tim>GAL4,and the 'CG' prefix to gene numbers is omitted.

This suggested that analysing only the data on the third day of LD conditions, rather than averaging over all three days, might prove a more successful strategy. Sleep differences may accumulate as flies acclimatise to the environment of their activity tube, and recover from any long term effects of the anaesthesia used to load them into tubes.

Table 8-14 also makes clear that not all genes selected for further analysis should be expected to have an effect on sleep; examination of the FlyAtlas database of fly gene expression showed that only a subset of genes selected for secondary screening are likely to be expressed in the brain (e.g. *CG8680*, *CG5532*), yet even those genes which are found to be expressed in the brain do not necessarily have replicable effects on sleep when knocked down (e.g. *CG3397*).

These results suggest that *CG9320*, *CG4974*, and *CG8768*'s role in sleep should be further assessed using in depth sleep deprivation experiments, whilst *CG11208* and *CG4686*'s role in activity and arousal might be better assessed by testing these lines' aggression and longevity phenotypes. Table 8-15 shows that all these genes have classical *P* element insertion mutants available, which may facilitate further study.

Table 8-15 Alleles and likely function of genes identified by this sleep screen.

Gene	Symbol	Classical alleles	Biological process	Molecular function
CG9320	*ns4*	*ns4[c06563], ns4[c06659]*	-	GTP binding
CG4974	*dally*	*dally[f01097] dally[f01984], many others*	*decapentaplegic* receptor signaling pathway ; development	*dpp* binding
CG8768	*CG8768*	*CG8768 [MB06845]*	cellular metabolic process	coenzyme binding, catalytic activity
CG11208	*CG11208*	*CG11208 [KG04289]*	-	oxalyl-CoA decarboxylase activity
CG4686	*CG4686*	*CG4686 [PL00083]*	-	-

8.4 Discussion

In this chapter, a novel tool for the analysis of *Drosophila* sleep was created in the hope that sleep behaviour might represent a sensitive, easily measured behaviour appropriate for testing the models for determining gene function developed in the previous chapter.

8.4.1 The BeFly! sleep analysis module represents a significant technical advance

The BeFly! sleep analysis module developed in this chapter (the use of which is described in detail within appendix D4) represents a significant advance on the alternative sleep analysis packages published to date, both in terms of usability and the quality of the output. The results generated using this software reveal correlations between different sleep metrics similar to those published by other groups (Harbison and Sehgal, 2008), and the overall success of screening (when using the *UAS>Dicer2* enhancer) is of a similar order to that found by other groups screening for sleep mutants (Wu et al., 2008a).

By analysing sleep both 'by hour' and 'by day', the software is flexible enough for use not only in high resolution sleep studies, but also for longer term longevity experiments in which the decline in mean waking activity is a good correlate for the general health of a fly (see manual for a detailed example). The author is currently collaborating with a number of research groups studying both short term and long term neurodegeneration and longevity to further develop the flexibility of the tools described in this chapter.

One criticism that might be levelled at the software in its current form is that graphical output presents genotype mean results rather than medians. As not all sleep metrics are normally distributed (Figure 8-4), median values may be more appropriate representations, however this is not currently possible when using Excel's Pivot Tables to present results. As Pivot Tables are the key to rapid, dynamic examination of results, future versions of the sleep analysis module will retain pivot table functionality, but will implement an option to transform

sleep metrics measuring time, thereby normalising the distribution. Due to skew effects, the 'longest sleep duration' and 'time till first sleep' metrics were not widely used during the screening described in this chapter.

Another criticism might include the somewhat forced distinction between 'by hour' and 'by day' analysis. This distinction however is not only appropriate, but essential to the tool: as sleep bouts may extend over a period of many hours, metrics such as bout number and sleep duration can only be properly assessed over longer periods of time. The higher resolution analysis 'by hour' is particularly useful for visualising changes in activity; indeed, visualising total sleep rather than total activity per hour produces considerably more robust and clearer results (Figure 8-7), suggesting such analysis should be adopted more widely in the circadian field.

One obstacle to high throughput sleep analysis is the existence of outliers. Although transformation of results may minimise the effect of outliers on further analysis, the current implementation of sleep analysis in BeFly! is primarily dependent upon a number of techniques to automatically identify and exclude aberrant flies. The most important distinction in a sleep analysis is that of appropriately discriminating between dead flies and sleeping flies in order to accurately measure sleep. As the BeFly! package can be used for both sleep and longevity analysis, the procedures by which the analysis module handles such dead flies can be varied. Nevertheless, automatically identifying aberrant flies is prone to error, and therefore the author recommends manual inspection of individual sleep patterns before creating group averages (see manual for further discussion of these points). As manual analysis limits the speed and objectivity of high throughput analysis, the procedures for identifying aberrant flies are continually being developed.

8.4.2 The identification of novel sleep genes; further validation for the transheterozygote approach to determining gene function

The proof of concept screening work in this chapter has revealed that knockdown of *CG9320*, *CG4974*, and *CG8768* produces a reproducible effect on sleep, whilst knockdown of *CG11208* and *CG4686* elicits changes in total activity. Although *CG4974 (dally)'s* effects on sleep might be a pleiotropic effect of its important roles in development (Table 8-15), *CG9320* and *CG8768* are likely to represent *bona fide* sleep mutants. Whilst there are no known functions for *CG8768, CG9320* (gene name *ns4*), is part of a family of nucleostemin GTPases which have been shown to act in serotonergic neurons to regulate insulin signalling and control body size (Kaplan et al., 2008), an interesting result given the demonstration in the previous chapter that *Akh* (also involved in insulin/glucagon-like sugar homeostasis) may play an important role in mediating circadian activity. Clearly the role of these genes should now be investigated in greater depth by performing sleep deprivation experiments, in addition to further high resolution analysis of daily sleep patterns using both RNAi knockdown and classical insertion mutants.

The utility of the transheterozygote diallel crossing scheme in the discovery of genes involved in behaviour has therefore been confirmed for both sleep and circadian behaviours. Given that both sleep circadian behaviours are cyclical in nature, one might have expected that correlating phenotypic scores with gene expression recorded only at a single point during the day (ZT3) would not be sufficient to uncover gene function; our method might therefore benefit by repeating microarray experiments using samples taken at ZT15 (i.e. antiphasic to the original samples).

One observation of note is that gene knockdown often affected daytime sleep, but only rarely affected nighttime sleep. Whilst this might reasonably be assumed to be an artefact of our microarray samples being collected during the light phase (for example see

Figure 8-9), it is more likely that this reflects the ease with which daytime 'sleep' mutants can be recovered (R. Greenspan, pers. comm.).

8.4.3 A dataset of lasting value in which to test future analyses

The data generated in this chapter represents a valuable resource with which to test strategies for identifying sleep mutants; as shown in Figure 8-9, a number of low sleeping lines can be identified by eye (such as *CG30404* knockdown using the *c687>GAL4* driver) which have not yet been followed up. This is in part due to the difficulty inherent in properly assessing the significance of a result when the control genotype represents an outlier.

Indeed, there is still some discussion in the field as to the most appropriate statistical tests to be used when analysing unevenly distributed sleep data. A further statistical complication is that even during short activity experiments, flies may become stuck in their food source and die, and therefore not all samples will be of the same size. Although this study used a conservative *post hoc* Dunnett's test after ANOVA to analyse the significance of results (in which STATISTICA corrects for differences in sample size), alternatives such as the Unequal N HSD test might also be empirically tested (though preliminary studies suggested limited difference between the two tests). If sample sizes are allowed to vary, then it might be appropriate to consider more flies of the control genotype, which should make the identification of genotypes showing significant sleep effects simpler. Additionally, the use of more complex nonparametric statistical models might also be more appropriate for the analysis of skewed data distributions such as those found in sleep studies (Wu et al., 2008a).

A further problem observed in this study is that sleep patterns may change over time; although the transheterozygote study suggested that sleep phenotypes remained stable over two days (Figure 8-6), this was not always the case, especially when using RNAi lines (e.g. Figure 8-14) in which the effect of gene knockdown may be cumulative. This suggests that for

RNAi experiments days might more appropriately be analysed independently (rather than averaging over time).

The activity data generated by the knockdown experiments will therefore prove to be an important test bed for development of methods appropriate not only for the study of sleep, but may also allow screening for circadian phenotypes as has been demonstrated in the literature (Wu et al., 2008a) as was shown in this chapter by the apparent identification of a number of arrhythmic lines (Figure 8-10).

8.4.4 Balancing breadth and depth when screening to optimise reproducibility

A key issue in any screen is balancing screen breadth (the number of candidate genes tested) with screen depth (the number of repeats per genotype) given limited experimental resources. Sleep metrics are known to show considerable inter-individual variation (Andretic and Shaw, 2005), however successful screening strategies have assayed sleep in as few as 4 individuals (Wu et al., 2008a), lending credence to the use of 16 flies in preliminary screening during this study.

However, as the results show, reproducing sleep results obtained using 16 fly samples has proven to be very difficult. The approach taken in the early screening experiment arguably focussed on breadth at the expense of depth; the use of six different GAL4 drivers and independent testing of male and female progeny generated a very large dataset, however the utility of this dataset was compromised by the 'high sleeping' phenotype of the control lines, biasing the screen towards detection of short sleeping lines. Although these limitations were addressed in later screening efforts, the reproducibility of results remained low (see Table 8-14), suggesting that screening depth might be more important than breadth when considering sleep.

One might argue that, given the speed with which sleep screening can be performed (experiments lasting only 3 days rather than a minimum of 10 in circadian screens), the limited

return on screening effort is an inevitable yet acceptable by-product of biasing the screen for breadth rather than depth. Further development of the statistical analyses used to analyse sleep screens may help resolve this issue.

8.4.5 Implications for networks

As discussed, the key issue is the inherent variability of sleep. The high resolution transheterozygote sleep data shows considerable variation even in inbred lines carrying phenotypically mild alleles as transheterozygotes (Figure 8-6). The extent of this variation is perhaps less surprising given the recent demonstration that one of the alleles used to generate transheterozygotes (*Csp*) has itself been shown to play a role in the response to sleep deprivation (Gilestro et al., 2009). However, even lines not carrying *Csp* showed extreme phenotypes; whilst *nal/Pen* females sleep only 10 minutes per hour during their midday siesta, *Itp/nmo* females sleep four times as long.

If even subtle network disruption is capable of generating such extreme variation, what are the implications for accurately recovering sleep mutants using more complex, heterogeneous backgrounds such as those evident in the natural populations used for QTL analysis? This question forms the basis of the final discussion chapter.

8.5 Conclusion/further work

This chapter has demonstrated that the transheterozygote screening strategy can be applied to *Drosophila* sleep behaviour in much the same way as circadian behaviour, successfully identifying new sleep gene candidates. However, in the context of this thesis the analysis of sleep was performed in the hope that sleep metrics might be used to test the various models with which to correlate gene expression and phenotype developed in the previous chapter, and in this context the screening has not proven a success. The sheer variability of sleep metrics makes reproducing results particularly difficult, and as such

inappropriate for testing theoretical models. The final chapter discusses alternative approaches to those shown in this chapter.

9 Discussion

9.1 The purpose of this study

The causal relationship between networks of interacting genes and the diversity of phenotypes observable in even the simplest of organisms remains something of a biological 'black box'. As new technologies to determine the expression of genes are becoming increasingly reliable and affordable, realising the scientific and therapeutic potential of these advances depends upon determining the function of the many non-annotated genes in the genome.

The central theme of this thesis has been the determination of gene function in *Drosophila* using both forward and reverse genetics approaches. Throughout this work the focus has been on genes affecting behaviour, primarily the rhythmic outputs of the well characterised circadian clock system, but also considering a number of other behaviours of varying complexity (Sokolowski, 2001) including sleep.

Behaviour is the primary mechanism by which an organism reconciles its genetic makeup with the environment it finds itself in, and is therefore not only subject to rapid evolution, but can itself influence the pace and trajectory of evolution (de Belle, 2002). Individuals show a wide range of behaviours and behavioural strategies, each of which requires the coordinated regulation of many discrete inputs and outputs forming independent biobehavioural modules (Tully, 1996). In addition to the role of behaviour in evolution and ecology, many human diseases and disorders are primarily behavioural in nature (e.g. schizophrenia and depression - Mackay, 2008), whilst behavioural uncoordination is a symptom of many neurodegenerative disorders (e.g. Huntington's disease - Morton et al., 2005). Understanding the basis of behaviour is therefore a topic of considerable significance and broad interest.

Since the identification of the *per* gene (Konopka and Benzer, 1971) it has been clear that a range of complex behaviours can be affected by mutations in a single gene, a finding that has been repeated in other fields (e.g. the characterisation of fruitless - Baker et al., 2001). However, the focus in this work has primarily been on the behavioural phenotypes arising as a result of the misregulation of gene function - as might occur in disease. The demonstration that many mutations have pleiotropic effects on behaviour (Hall, 1994) makes the study of behaviour particularly sensitive to changes in the genetic background, and behavioural studies are therefore particularly suited to model organisms in which the genetic background can easily be controlled.

9.2 The value of automated analysis software

In part because of their sensitivity to genetic background effects, behavioural phenotypes show considerable inter-individual and inter-genotype variation, and therefore the analysis of such phenotypes is often labour intensive, and consequently highly subjective. The first part of this thesis describes the development of a suite of software tools which partially automate such analyses, making behaviour considerably more amenable to high throughput study.

The development of the BeFly! package's accessible, yet powerful and easily scalable tools has been a significant success, representing as it does a significant improvement on the *status quo*, and as such has been adopted by groups spread from California to Scotland, garnering a growing list of acknowledgements in the literature despite not yet being published. Whilst many of the analyses within the BeFly! package are not entirely novel, its ease of use and versatility has allowed its application to fields as diverse as neurodegeneration and tidal rhythmicity, vindicating the decision to write the software within the Excel environment familiar to the majority of scientific users (despite the software running somewhat slower as a result, compared to implementations in dedicated languages such as Python).

In addition to its use in high throughput behavioural screening, the BeFly! package allows more detailed and – crucially - more objective analysis of activity data, both in terms of circadian periodicity and sleep (as discussed in detail in chapters 3 and 4). Of the other tools available in the package, perhaps the most broadly applicable is the cross-correlation tool, which in addition to its primary circadian use in the calculation of phase shifts has other uses in 'best fit' scenarios. The versatility of this tool is demonstrated by the author's recent contribution to a proteomic analysis in the murine system (Deery *et al.* submitted), in which the expression levels of a number candidate cycling genes was cross-correlated to a sinusoidal waveform, the maximum positive or negative correlation coefficient allowing the phase of the data to be calculated. The significance of the correlation was assessed using a Monte Carlo approach (as used in chapter 7), in which the experimental data was randomised and the cross-correlation procedure repeated 1000 times to generate a distribution of possible correlations, the 95th percentile of which being taken to represent the 5% significance level.

Although comprehensive, the BeFly! package will become more widely adopted by other circadian groups if it implements all the different analytical tools currently within the field. The most notable omission is the lack of a procedure to automatically measure the height of the largest autocorrelation 'peak', a metric widely used in the field as an indication of rhythm 'quality' (Dowse, 2007). Although 'peak finding' mathematics can be relatively complex in noisy data series, a simple implementation would facilitate the adoption of the BeFly! package by other groups, offering the community the chance to standardise the use and development of a single analysis package (as was the case in the late '90s with the – now defunct - Brandeis Rhythms package).

9.3 The difficulties of identifying gene function

Serendipitous recovery of lines showing novel phenotypes, whilst hardly a high throughput technique for determining gene function, has played an important role in forward

genetics since the recovery of the first *white* eye colour mutation (Morgan, 1911). In this work, the fortuitous identification of the *Party on* strain stimulated the development of novel behavioural analyses in an effort to dissect the nature of this mutation. Behavioural analysis and preliminary meiotic mapping strongly suggest that the *Party on* mutation is not a novel allele of any known circadian gene on the 2nd chromosome, however the susceptibility of the LL rhythmic phenotype (as with many behaviours) to genetic modifiers precluded more accurate mapping during the course of this work (see discussion in chapter 4 for a further mapping strategy).

Despite the use of in depth analyses, including wavelet techniques not previously applied to activity data, characterising the *Party on* mutation proved difficult; the results generated in this work cannot discriminate between *Po* playing a role in intracellular processes such as PER/TIM degradation, or acting as part of the intercellular clock network. Experiments using *Po* served to emphasise the importance of using genetically homogenous populations in behavioural studies; different *Drosophila* isolates have periods ranging from 23.8 to 24.5 hours, therefore it was unclear whether the DD period of the *Party on* isolate (24.3 hours at 25°C) was 'long' or 'normal'. This distinction is important, as a 'long' period would suggest that the circadian clock mechanism was in some way compromised in DD as well as LL conditions. The later observation that the period of *Party on* flies is temperature sensitive in both DD and LL suggested that that the DD period of 24.3 hours was indeed the result of underlying problems with the oscillator.

9.4 Identifying gene function using genome-wide association analysis

Genome-wide association (GWA) studies represent a powerful approach with which to identify genes involved in quantitatively measurable phenotypes. A number of groundbreaking studies focussing on the genetics of common human diseases has elucidated many of the principals of network structure that underpin variation in disease susceptibility,

principals which are more broadly applicable to phenotypic variation in general. These findings include the demonstration that disease genes are generally loci with only small effects on the variance in disease susceptibility (Wellcome Trust Case Control Consortium, 2007), are nonessential and show no tendency to encode hub proteins in networks - indeed are frequently localized in the functional periphery networks (Goh et al., 2007). Taken together, these findings suggest that efforts to determine gene function using traditional screening methods will fail to saturate the genome (see further discussion see chapter 7).

In an effort to circumvent the limitations of traditional screening, we employed a novel approach to identify genes associated with behaviour, using a controlled genetic background and a diallel crossing scheme to successfully identify novel circadian (chapter 7) and sleep related (chapter 8) genes with a false positive rate similar to that using other techniques (see previous discussion). This approach also allowed some elucidation of the genetic architecture underlying various behavioural traits and the patterns of epistasis among them.

Whilst the discussion in chapter 8 details a number of technical limitations inherent to our statistical analysis and RNAi screening approach, an issue not considered in the discussion of our correlative approach to identifying gene function is that of non-causative correlations between genes and phenotypes:

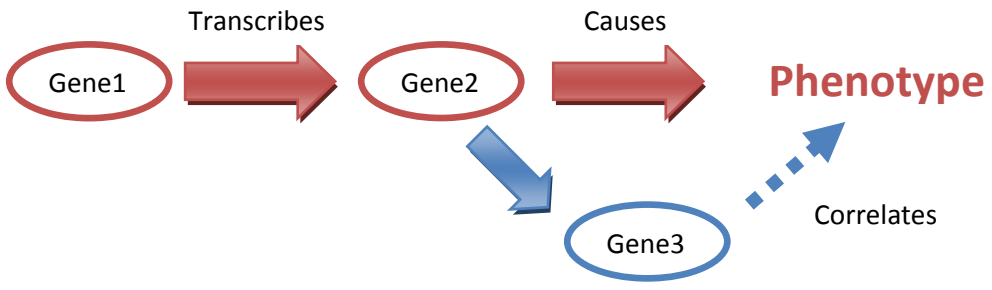


Figure 9-1 Causative and correlative relationships between gene expression and phenotype. Gene1 increases the transcription of **Gene2**, which in turn upregulates **Gene3** in addition to affecting the organismal phenotype. The three transcripts form a coexpression network; whilst each gene may correlate with changes in the phenotype of interest, the association between the phenotype and levels of Gene3 is correlative (dashed arrow) rather than causative as in the case of genes 1 and 2 (solid arrow).

As is clear in the figure above, not all genes that correlate with a phenotype have a causative relationship with that phenotype. As such, knocking down the expression of such genes will not affect the phenotype, as was observed in many cases in our RNAi follow-up screening (see Table 7-8). The efficiency of screening for gene function might therefore be significantly increased by taking into account biological connectivity information to discriminate between causal and correlative (or 'consequential') interactions (Rockman, 2008).

9.5 Gene function in the context of networks and modules

As has been emphasised throughout this work, the regulatory networks underlying phenotypes are complex and highly context dependent, as is evident from the varying patterns of epistasis in our diallel scheme e.g. Table 7-6 (Mackay and Anholt, 2006). As gene networks are not static, network information cannot be retrieved from a database but must be calculated from the expression data generated in each experiment.

In essence, networks represent groups of interconnected transcripts as illustrated in Figure 9-1, and these groups are termed 'modules'. Genes within a given module should show higher correlations with each other than to the rest of the transcriptome (though such correlations should be thought of as statistical relationships rather than physical interactions between genes). The techniques used to build and delineate modules are a subject of intense research, Mackay listing the use of Bayesian techniques (Zhu et al., 2004), partial correlation analysis (Aten et al., 2008) and empirical Bayes procedures (Chen et al., 2007) in her recent review of quantitative genetics techniques (Mackay et al., 2009).

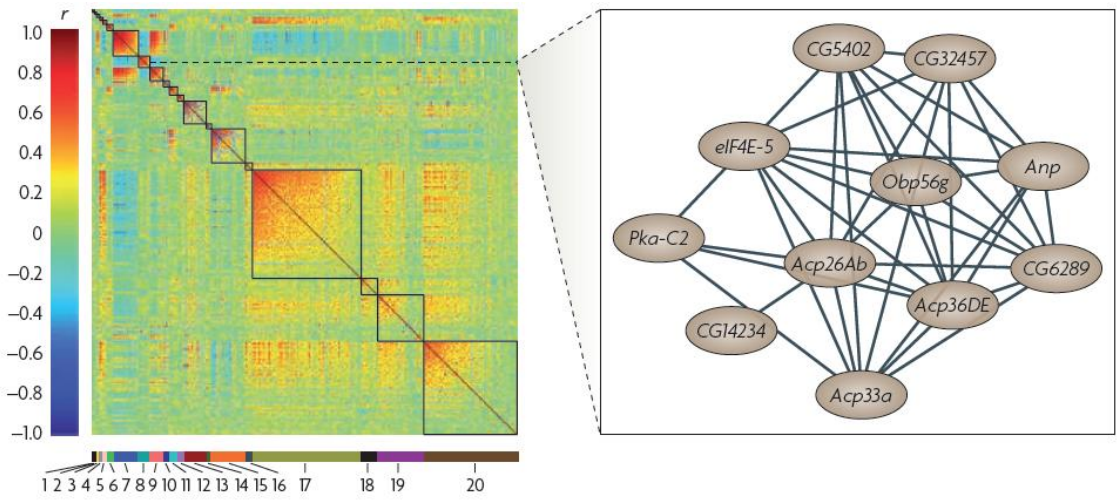


Figure 9-2 Identification of genetic modules using expression data. Left panel: by identifying groups of genes showing significant coexpression, a number of distinct genetic modules can be identified in microarray data. **Right panel**: a genetic module, in which edges connecting nodes represent significant correlations in expression. Figure and caption adapted from Mackay (Mackay et al., 2009).

Modules are particularly useful for screening experiments as they have been shown to represent biologically meaningful groupings, enriched for similar ontology categories, KEGG pathways, protein–protein interactions, tissue-specific expression patterns and transcription factor-binding sites (see Mackay et al., 2009 and references therein). As the genes within a module are likely to play similar functions, screens can be performed more efficiently by verifying only one or two genes within each module, other tests being redundant. Additionally, many module forming algorithms can be 'seeded' using genes known to play a role in a process, allowing existing data to be placed into a modular framework (Li and Horvath, 2007).

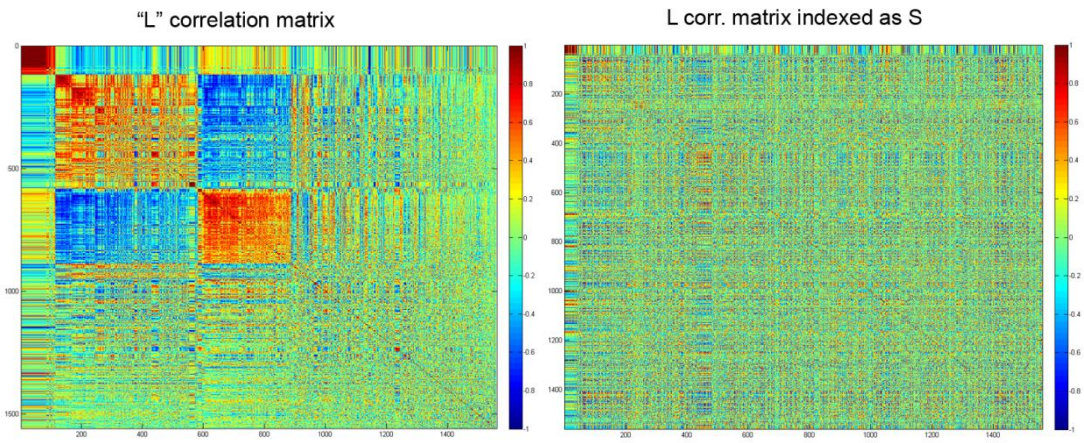


Figure 9-3 Preliminary module analysis of transheterozygote period data. Correlations between transcripts colour coded such that red = +1 and blue = -1. Expression data for transheterozygous lines showing outlier 'long' periods were analysed using the MTOM module forming software (Li and Horvath, 2007) (**left panel**) to generate a number of distinct modules. Examining the same modules in transheterozygote lines showing short periods (**right panel**) reveals that modules are highly dependent on background, as the correlations evident in the left panel are no longer evident. Figure created by Ralph Greenspan.

The concept of modular networks is becoming increasingly important in quantitative genetics. Data generated in this study, by virtue of its emphasis on controlled genetic backgrounds and subtle network disruption, will prove a valuable comparison to QTL studies in which the degree of network disruption is much more significant. It is unclear as to which method will prove most successful in elucidating gene function.

9.6 Utilising existing datasets to inform gene function screens

Although networks are sensitive to genetic context, existing data may yet be informative when attempting to elucidate network structure and gene function. Indeed, in this study considerable use was made of gene ontology predictions to refine lists of candidate genes in an effort to identify those genes most likely to play a role controlling behaviour.

To this end, specialist microarray software packages such as GENESPRING (http://www.chem.agilent.com/) have been developed to interpret microarray data in its biological context. Pathway information, described using standardised annotation such as

BioPAX, (http://www.biopax.org/), includes information gleaned from physical or genetic interaction data, as well as from earlier microarray studies. Analysing microarray data using a knowledge of pathways and modules facilitates the identification of real interactions (e.g. if several components of a pathway are upregulated, one can confidently conclude that pathway plays a major role in mediating the effect of interest).

Gene ontology information can also be mapped onto microarray data, being used to identify significantly enriched classes of genes or pathways (Gene Set Enrichment Analysis, GSEA). However, by taking account of gene ontology it is also possible to identify 'missing' genes in a pathway; for example the transheterozygote period and gene expression data might be used to identify novel components of the ubiquitination pathway correlating with period. Such a gene might well be a good candidate for the *Party on* mutation (as discussed in chapter 4). Ontology information can also be used to visualise the spatial pattern of interactions identified by microarray studies occurring within the cell:

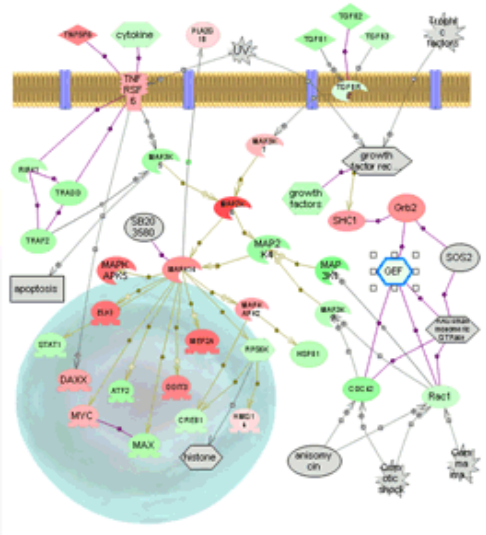


Figure 9-4 Example of genetic interactions from microarray data mapped onto subcellular localisation data using the GENESPRING software.

Currently the author is importing the transheterozygote gene expression data generated using the custom microarray design employed by the FlyChip group into GENESPRING with the aim of performing such analyses. This experiment is limited by a

number of bugs identified by the author in GENESPRING's data import procedures, and should be subject to a fix by tech support in the near future.

9.7 Higher order network interactions

One of the outstanding features of our transheterozygote approach is the typing of the same lines for many different behaviours. Whilst this work was being performed, a similar analysis was performed by the Mackay group using 40 inbred lines and a number of 'ecologically relevant' phenotypes, some of which overlap with this study (Ayroles et al., 2009). By identifying regulatory networks (or modules) contributing to each phenotype, Ayroles *et al.* were able to show a degree of module overlap between certain behaviours:

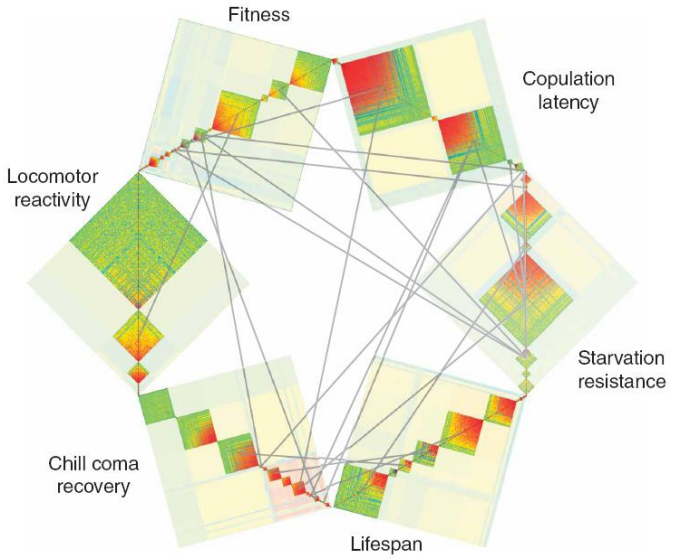


Figure 9-5 Pleiotropy between phenotypic modules. Grey lines connect modules with a significant overlap of greater than four genes between gene lists, as determined by Fisher's exact tests. Figure and caption from Ayroles *et al.* (Ayroles et al., 2009).

The advantage of our study is that, by performing a diallel cross, not only can we identify such overlapping regulatory modules, but by considering the patterns of epistatic relationships revealed by the phenotypes (GCA/SCA scores) we will be able to determine the magnitude of the effect of such interactions.

9.8 Conclusion

Biology as a whole is moving into an era in which high throughput studies are the norm; rapid technological advances such as the availability of whole genome RNAi libraries and the falling cost of microarray experiments have revolutionised the screening of gene function within the past four years of this project. Indeed, the software and methods developed within this thesis are at the cutting edge of such high throughput behavioural analyses, as evidenced by their adoption by a number of groups worldwide and their use in identifying a number of novel circadian and sleep associated genes, notably *Akh*.

Not only has this thesis provided results acting as a proof of concept for our novel gene screening strategy, but the comprehensive collection of genotypic and phenotypic data generated promises to remain a valuable resource in the coming years with which to evaluate emerging techniques, developed both in this thesis and by other groups, promising new ways of identifying gene function. Accomplishing this daunting task remains the key to unlocking to potential of the genomic era.

10 Appendices

10.1 Digital appendices

Appendix	Description
D1	BeFly! manual.
D2	The spreadsheet used for the bioinformatics study.
D3	Summary of transheterozygote gene expression and phenotypic data.
D4	Sleep analysis manual
D5	Results of gene knockdown on sleep metrics

10.2 Appendices to chapter 4

As the data suggest that *Po* might lie near *curved* (recombination position 2-75.5), further mapping might be performed using the following strategy:

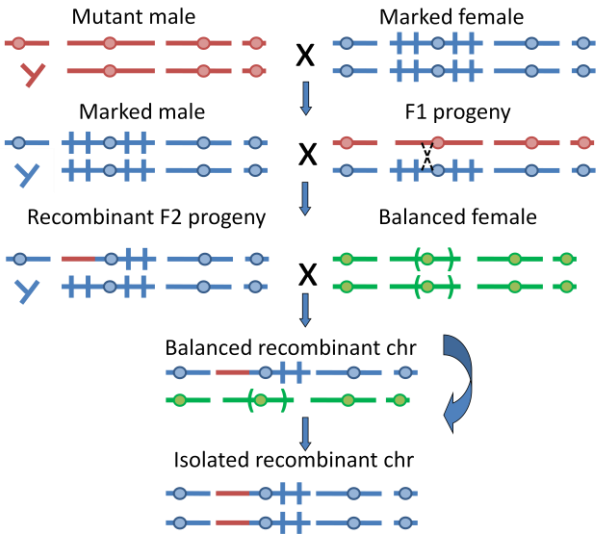


Figure 10-1 Improved *Po* meiotic mapping strategy. In this scheme, individual recombinant chromosomes are isolated and tested *en masse* in the F4 generation, precluding behavioural plasticity or variable penetrance disrupting the mapping of *Po*.

10.3 Appendices to chapter 5

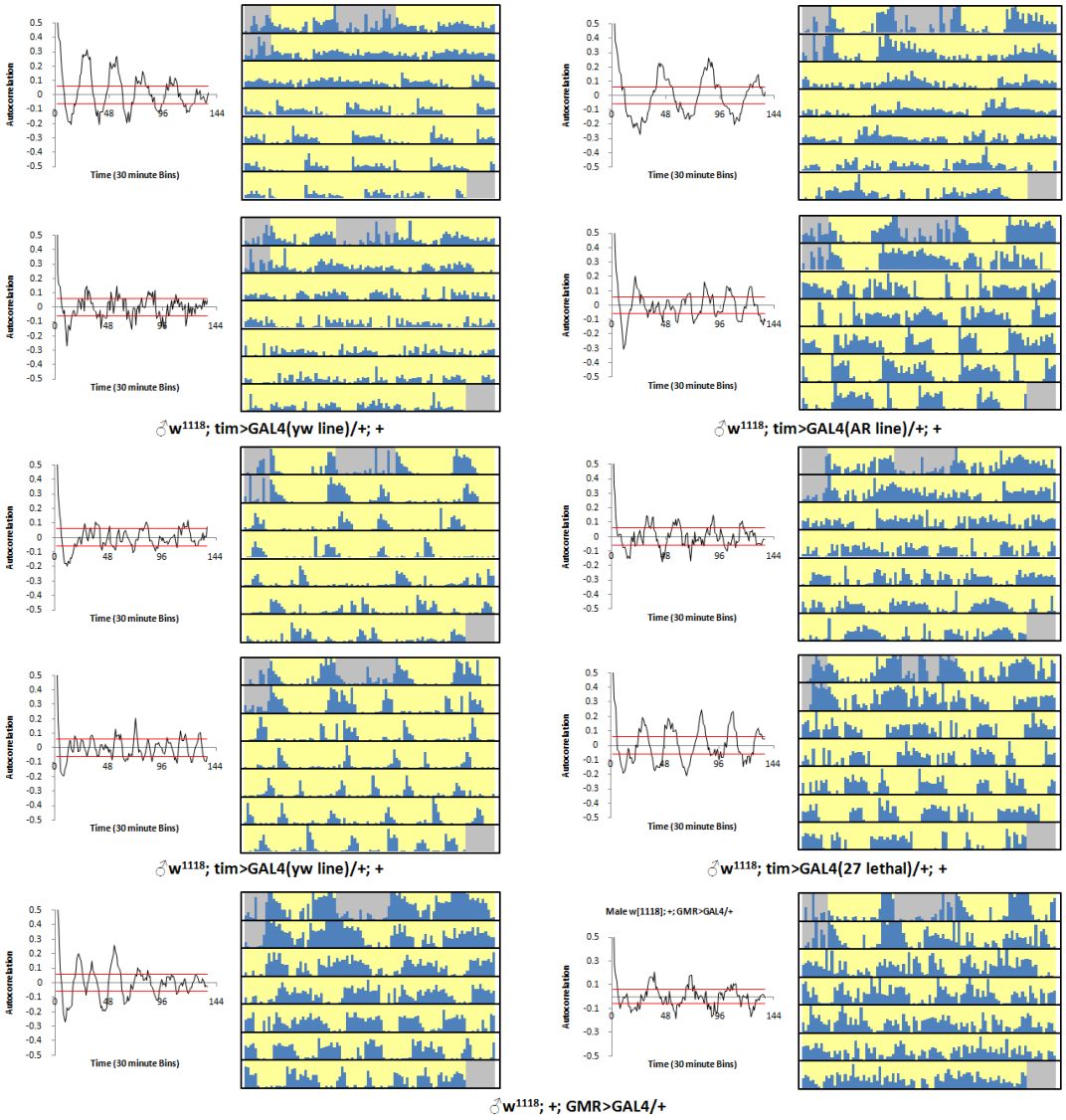


Figure 10-2 Examples of individual activity records for GAL4 lines showing LL rhythmicity in a *per*$^{+}$ background. All genotypes show ultradian rhythmicities of different period and quality, in part due to differences in the length and distribution of activity bouts between genotypes.

10.4 Appendices to chapter 6

Exelixis technology transfer notice Stocks sourced from Exelixis may not be redistributed to any third party outside of the recipient's institution. Recipients may take these stocks with them when they move to a new institution, provided they will be conducting non-commercial research at the new institution. If a stock is no longer available from the Bloomington Stock Center, an investigator may redistribute the stock to other not-for-profit investigators if

recipients are notified in writing of the license requirements and redistribution restrictions. Commercial work using these stocks requires a licence, obtainable from info@exelixis.com. By providing these stocks, Exelixis grants no rights to any patents or patent applications. Exelixis makes no representations regarding the stocks and extends no warranties of any kind. Exelixis shall not be liable for damage associated with the use of these stocks.

DrosDel technology transfer notice The DrosDel Consortium is the source of several stocks, provided on an "as is" basis. Publications that use the RS elements or deletions must use the predefined Df()EDx numbers when naming deletions constructed with the DrosDel kit. This is to avoid labs duplicating efforts in making deletions, and for FlyBase tracking. All constructed deletions should be submitted to FlyBase for curation and be made freely available to the fly community.

10.5 Appendices to chapter 8

Table 10-1 Genes correlating with sleep, as calculated by Bruno van Swinderen in a preliminary study showing gene whose expression correlates with total sleep (p<0.01). This list was further refined by EG to identify genes with ontologies that might suggest a role in sleep regulation, possibly due to involvement in signalling, neural development or protein kinase activities. Genes selected for further study are listed here. **FBgn** = FlyBase gene number, **Sym** = gene symbol, **CG** = gene CG number, **TS** = correlation value with total sleep, **V ID** = Vienna transformant strain ID, **GO** = gene ontology from InterPro database.

FBgn	Sym	CG	TS	V ID	GO Biological Process
FBgn0039915	*CG1732*	*CG1732*	0.683	13359	neurotransmitter transport
FBgn0033837	*CG17034*	*CG17034*	0.675	8136	phospholipid transport
FBgn0034443	*Cer*	*CG10460*	0.586	22752	long-term memory
FBgn0033544	*CG7220*	*CG7220*	0.569	34199	proteolysis
FBgn0037551	*CG7891*	*CG7891*	0.547	26085	small GTPase mediated signal transduction
FBgn0039889	*CG2219*	*CG2219*	0.536	41691	small GTPase mediated signal transduction
FBgn0004055	*Uzip*	*CG3533*	0.528	4066	axonogenesis
FBgn0033544	*CG7220*	*CG7220*	0.485	34199	proteolysis
FBgn0034913	*Usnp*	*CG11173*	0.470	18172	vesicle-mediated transport
FBgn0032644	*CG5131*	*CG5131*	0.469	42025	Proteolysis
FBgn0039821	*CG15556*	*CG15556*	0.462	1791	G-protein coupled receptor protein signaling pathway
FBgn0002526	*LanA*	*CG10236*	0.460	18873	central nervous system development ; axon guidance
FBgn0052091	*CG32091*	*CG32091*	0.458	5222	Transport
FBgn0027356	*Amph*	*CG8604*	0.457	9264	endocytosis
FBgn0033765	*Nemy*	*CG8772*	0.439	7192	locomotory behavior ; memory
FBgn0037846	*CG6574*	*CG6574*	0.431	40903	transport
FBgn0034691	*Synj*	*CG6562*	0.422	46070	dephosphorylation ; synaptic vesicle endocytosi
FBgn0039907	*Lgs*	*CG2041*	0.416	5694	*Wnt* receptor signaling pathway
FBgn0034367	*sec6*	*CG5341*	0.414	22077	neurotransmitter secretion
FBgn0041626	*Or19a*	*CG18859*	0.412	49231	sensory perception of smell
FBgn0034453	*Hpo*	*CG11228*	0.407	7823	protein amino acid phosphorylation ; negative regulation of organ growth

FBgn	Sym	CG	TS	V ID	GO Biological Process
FBgn0023407	*B4*	*CG9239*	-0.389	6295	circadian rhythm
FBgn0036518	*RhoGAP71E*	*CG32149*	-0.445	41861	signal transduction
FBgn0039024	*CG4721*	*CG4721*	-0.452	51619	proteolysis
FBgn0033996	*CG11807*	*CG11807*	-0.459	38564	cell communication
FBgn0047178	*CG32147*	*CG32147*	-0.462	34419	proteolysis
FBgn0002781	*mod(mdg4)*	*CG32491*	-0.471	52268	regulation of chromatin assembly or disassembly ; male meiosis I
FBgn0004868	*Gdi*	*CG4422*	-0.474	26537	neurotransmitter secretion ; regulation of GTPase activity
FBgn0001218	*Hsc70-3*	*CG4147*	-0.475	14882	sleep; response to heat
FBgn0036505	*CG7945*	*CG7945*	-0.501	35296	protein folding
FBgn0041604	*Dlp*	*CG32146*	-0.503	10299	*Wnt* and *Smo* signaling pathways ; detection of light stimulus involved in visual perception
FBgn0025743	*Mbt*	*CG18582*	-0.518	46044	mushroom body development ; activation of MAPK activity ;
FBgn0039628	*CG11841*	*CG11841*	-0.530	14849	proteolysis
FBgn0000422	*Ddc*	*CG10697*	-0.538	3330	dopamine biosynthetic process from tyrosine
FBgn0033663	*ERp60*	*CG8983*	-0.541	51674	protein folding
FBgn0052296	*Mrtf*	*CG32296*	-0.574	34503	positive regulation of transcription, DNA-dependent
FBgn0030755	*CG9906*	*CG9906*	-0.583	5597	protein folding
FBgn0030377	*CG1924*	*CG1924*	-0.590	42397	protein folding
FBgn0003460	*So*	*CG11121*	-0.603	8950	circadian rhythm ; Bolwig's organ morphogenesis
FBgn0034722	*Rtf1*	*CG10955*	-0.616	27341	Notch signaling pathway

Table 10-2 Detailed sleep tables Genes correlating with sleep, as calculated by author Table 10-3 Genes selected for further analysis. **FBGN** = FlyBasegene number, **Symbol** = gene symbol, **Transgene ID** = Vienna Drosophila RNAi Centre transgene ID, **TS** = Total Sleep, **BL** = Bout Length, **BN** = Bout number, **GO** = gene ontology. **W** = wakefulness related, **S** = Sleep related, **P** = Pack identified as differentially expressed in sleep. **1.5** = high stringency dataset, **1.2** = lower stringency dataset.

FBGN	Symbol	Transgene ID	Chr	Existing study	TS p<0.0001	BL p<0.0001	BN p<0.001	GO Biological Process
FBgn0015037	*CG10842*	3331	2	W1.5	-ve			-
FBgn0000024	*CG17907*	3968	3		+ve	+ve		acetylcholine catabolic process
FBgn0052475	*CG32475*	4071	3		-ve			determination of adult life span ;response to stress
FBgn0013764	*CG3924*	30454	2				+ve	axon guidance ; nervous system development ;
FBgn0025454	*CG8453*	4616	3	P	-ve			response to insecticide ;
FBgn0034488	*CG11208*	4720	3		-ve			-
FBgn0038267	*CG17304*	5878	1				+ve	-
FBgn0039005	*CG6969*	6831	2		-ve			response to oxidative stress ;
FBgn0039040	*CG13833*	6845	3		-ve			
FBgn0030341	*CG1967*	12196	3				+ve	post-Golgi vesicle-mediated transport ;
FBgn0010228	*CG17921*	12773	2		-ve	-ve		regulation of transcription, D
FBgn0034935	*CG13565*	12876	3				+ve	-
FBgn0033382	*CG8058*	13314	2		-ve	-ve		-
FBgn0034756	*CG4373*	13846	1		-ve			-
FBgn0011577	*CG4974*	14136	3		+ve			*dpp*, *Wnt* and *tgf* beta signalling ; nervous system development
FBgn0039203	*CG13618*	14714	3		-ve			-
FBgn0039073	*CG4408*	15235	3	W1.5			+ve	proteolysis
FBgn0039205	*CG13623*	17334	3		-ve			iron-sulfur cluster assembly
FBgn0035402	*CG12082*	17567	2				+ve	protein deubiquitination
FBgn0002645	*CG1483*	18403	3		+ve	+ve		microtubule-based process
FBgn0034849	*CG3500*	19703	3		-ve	-ve		-
FBgn0035344	*CG16761*	22917	2	W1.2	-ve			-

FBGN	Symbol	Transgene ID	Chr	Existing study	TS p<0.0001	BL p<0.0001	BN p<0.001	GO Biological Process
FBgn0031684	*CG8680*	23467	3				-ve	-
FBgn0032882	*CG9320*	24141	3		-ve			-
FBgn0030208	*CG2890*	25445	3				+ve	regulation of protein amino acid dephosphorylation
FBgn0041094	*CG7590*	25506	2		-ve			negative regulation of signal transduction and growth
FBgn0033574	*CG7722*	25534	3		+ve			-
FBgn0038206	*CG3172*	25817	2			-ve		actin filament depolymerization
FBgn0034902	*CG5532*	30347	3		-ve			
FBgn0035583	*CG13704*	30356	3	W1.2	-ve			-
FBgn0035715	*CG10103*	31174	2		-ve	-ve		-
FBgn0034688	*CG11474*	31507	2		-ve			-
FBgn0039666	*CG11501*	33115	2		+ve			-
FBgn0032013	*CG7851*	33157	3		-ve			-
FBgn0037975	*CG3397*	34604	3		-ve			potassium ion transport
FBgn0038686	*CG5555*	35013	3			+ve		-
FBgn0037549	*CG7878*	35288	1	W1.2	-ve			-
FBgn0034579	*CG9353*	35447	3		-ve			-
FBgn0026379	*CG5671*	35731	3			+ve		Development, ; insulin receptor signaling pathway ; starvation ;
FBgn0033769	*CG8768*	36022	2		-ve			cellular metabolic process
FBgn0020766	*CG13348*	36554	3		+ve	+ve		tRNA processing
FBgn0036286	*CG10616*	36605	3		+ve			
FBgn0027535	*CG15110*	37186	3		+ve			Wnt, smo and Dpp receptor signaling pathways
FBgn0034724	*CG3624*	37280	2	W1.2	-ve			-
FBgn0003463	*CG9224*	37405	3				-ve	regulator of *torso*, *BMP* and *TGF beta* signaling pathways
FBgn0052412	*CG32412*	38277	2		-ve			proteolysis
FBgn0029973	*CG11386*	38494	2		+ve			-
FBgn0045862	*CG12878*	38722	3		-ve			bicoid mRNA localization

FBGN	Symbol	Transgene ID	Chr	Existing study	TS p<0.0001	BL p<0.0001	BN p<0.001	GO Biological Process
FBgn0035534	*CG15016*	38843	2		-ve			translation
FBgn0042135	*CG18812*	39224	3		+ve			-
FBgn0038736	*CG4583*	39561	3		+ve			protein amino acid phosphorylation
FBgn0050115	*CG30115*	39952	1	S1.2		-ve		regulation of *Rho* protein signal transduction
FBgn0034140	*CG8317*	40704	2	W1.2	+ve			-
FBgn0028386	*CG5067*	40867	3				+ve	negative regulation of transcription ; eye morphogenesis ;
FBgn0034659	*CG4021*	41974	2		+ve	+ve		-
FBgn0033372	*CG13742*	43315	2				+ve	-
FBgn0037141	*CG7143*	44721	3				+ve	-
FBgn0039537	*CG5590*	45462	3		-ve			-
FBgn0038739	*CG4686*	46372	3			-ve		-
FBgn0035359	*CG1143*	47644	3		-ve	-ve		-
FBgn0010340	*CG9852*	48716	3		-ve			multicellular organismal development
FBgn0034761	*CG4250*	50384	3		-ve			-

11 Bibliography

Adams, M.D., S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, R.A. George, S.E. Lewis, S. Richards, M. Ashburner, S.N. Henderson, G.G. Sutton, J.R. Wortman, M.D. Yandell, Q. Zhang, L.X. Chen, R.C. Brandon, Y.H. Rogers, R.G. Blazej, M. Champe, B.D. Pfeiffer, K.H. Wan, C. Doyle, E.G. Baxter, G. Helt, C.R. Nelson, G.L. Gabor, J.F. Abril, A. Agbayani, H.J. An, C. Andrews-Pfannkoch, D. Baldwin, R.M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E.M. Beasley, K.Y. Beeson, P.V. Benos, B.P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M.R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K.C. Burtis, D.A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J.M. Cherry, S. Cawley, C. Dahlke, L.B. Davenport, P. Davies, B. de Pablos, A. Delcher, Z. Deng, A.D. Mays, I. Dew, S.M. Dietz, K. Dodson, L.E. Doup, M. Downes, S. Dugan-Rocha, B.C. Dunkov, P. Dunn, K.J. Durbin, C.C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A.E. Gabrielian, N.S. Garg, W.M. Gelbart, K. Glasser, A. Glodek, F. Gong, J.H. Gorrell, Z. Gu, P. Guan, M. Harris, N.L. Harris, D. Harvey, T.J. Heiman, J.R. Hernandez, J. Houck, D. Hostin, K.A. Houston, T.J. Howland, M.H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G.H. Karpen, Z. Ke, J.A. Kennison, K.A. Ketchum, B.E. Kimmel, C.D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A.A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T.C. McIntosh, M.P. McLeod, D. McPherson, G. Merkulov, N.V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S.M. Mount, M. Moy, B. Murphy, L. Murphy, D.M. Muzny, D.L. Nelson, D.R. Nelson, K.A. Nelson, K. Nixon, D.R. Nusskern, J.M. Pacleb, M. Palazzolo, G.S. Pittman, S. Pan, J. Pollard, V. Puri, M.G. Reese, K. Reinert, K. Remington, R.D. Saunders, F. Scheeler, H. Shen, B.C. Shue, I. Siden-Kiamos, M. Simpson, M.P. Skupski, T. Smith, E. Spier, A.C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A.H. Wang, X. Wang, Z.Y. Wang, D.A. Wassarman, G.M. Weinstock, J. Weissenbach, S.M. Williams, WoodageT, K.C. Worley, D. Wu, S. Yang, Q.A. Yao, J. Ye, R.F. Yeh, J.S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X.H. Zheng, F.N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H.O. Smith, R.A. Gibbs, E.W. Myers, G.M. Rubin, and J.C. Venter. 2000. The genome sequence of Drosophila melanogaster. *Science.* 287:2185-2195.

Agosto, J., J.C. Choi, K.M. Parisky, G. Stilwell, M. Rosbash, and L.C. Griffith. 2008. Modulation of GABAA receptor desensitization uncouples sleep onset and maintenance in Drosophila. *Nat.Neurosci.* 11:354-359.

Akten, B., E. Jauch, G.K. Genova, E.Y. Kim, I. Edery, T. Raabe, and F.R. Jackson. 2003. A role for CK2 in the Drosophila circadian oscillator. *Nat.Neurosci.* 6:251-257.

Allada, R., N.E. White, W.V. So, J.C. Hall, and M. Rosbash. 1998. A mutant Drosophila homolog of mammalian Clock disrupts circadian rhythms and transcription of period and timeless. *Cell.* 93:791-804.

Allan, D.W., N. Ashby, and C.C. Hodge. 1997. The Science of Timekeeping Hewlett Packard Application Note 1289.

Andretic, R., and P.J. Shaw. 2005. Essentials of sleep recordings in Drosophila: moving beyond sleep time. *Methods Enzymol.* 393:759-772.

Andretic, R., B. van Swinderen, and R.J. Greenspan. 2005. Dopaminergic modulation of arousal in Drosophila. *Curr.Biol.* 15:1165-1175.

Armstrong, J.D., M.J. Texada, R. Munjaal, D.A. Baker, and K.M. Beckingham. 2006. Gravitaxis in Drosophila melanogaster: a forward genetic screen. *Genes Brain Behav.* 5:222-239.

Aschoff, J. 1965. Circadian Clocks. North Holland Press, Amsterdam. 479 pp.

Aserinsky, E., and N. Kleitman. 1953. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science.* 118:273-274.

Ashburner, M., K.G. Golic, and S.R. Hawley . 2005. Drosophilia: A Laboratory Handbook. Cold Spring Harbor Laboratory Press, New York. 1409 pp.

Aten, J.E., T.F. Fuller, A.J. Lusis, and S. Horvath. 2008. Using genetic markers to orient the edges in quantitative trait networks: the NEO software. *BMC Syst.Biol.* 2:34.

Ayroles, J.F., M.A. Carbone, E.A. Stone, K.W. Jordan, R.F. Lyman, M.M. Magwire, S.M. Rollmann, L.H. Duncan, F. Lawrence, R.R. Anholt, and T.F. Mackay. 2009. Systems genetics of complex traits in Drosophila melanogaster. *Nat.Genet.* 41:299-307.

Bachleitner, W., L. Kempinger, C. Wulbeck, D. Rieger, and C. Helfrich-Forster. 2007. Moonlight shifts the endogenous clock of Drosophila melanogaster. *Proc.Natl.Acad.Sci.U.S.A.* 104:3538-3543.

Bae, K., C. Lee, P.E. Hardin, and I. Edery. 2000. dCLOCK is present in limiting amounts and likely mediates daily interactions between the dCLOCK-CYC transcription factor and the PER-TIM complex. *J.Neurosci.* 20:1746-1753.

Baggerman, G., A. Cerstiaens, A. De Loof, and L. Schoofs. 2002. Peptidomics of the larval Drosophila melanogaster central nervous system. *J.Biol.Chem.* 277:40368-40374.

Baier, A., B. Wittek, and B. Brembs. 2002. Drosophila as a new model organism for the neurobiology of aggression? *J.Exp.Biol.* 205:1233-1240.

Baker, B.S., B.J. Taylor, and J.C. Hall. 2001. Are complex behaviors specified by dedicated regulatory genes? Reasoning from Drosophila. *Cell.* 105:13-24.

Bammler, T., R.P. Beyer, S. Bhattacharya, G.A. Boorman, A. Boyles, B.U. Bradford, R.E. Bumgarner, P.R. Bushel, K. Chaturvedi, D. Choi, M.L. Cunningham, S. Deng, H.K. Dressman, R.D. Fannin, F.M. Farin, J.H. Freedman, R.C. Fry, A. Harper, M.C. Humble, P. Hurban, T.J. Kavanagh, W.K. Kaufmann, K.F. Kerr, L. Jing, J.A. Lapidus, M.R. Lasarev, J. Li, Y.J. Li, E.K. Lobenhofer, X. Lu, R.L. Malek, S. Milton, S.R. Nagalla, J.P. O'malley, V.S. Palmer, P. Pattee, R.S. Paules, C.M. Perou, K. Phillips, L.X. Qin, Y. Qiu, S.D. Quigley, M. Rodland, I. Rusyn, L.D. Samson, D.A. Schwartz, Y. Shi, J.L. Shin, S.O. Sieber, S. Slifer, M.C. Speer, P.S. Spencer, D.I. Sproles, J.A. Swenberg, W.A. Suk, R.C. Sullivan, R. Tian, R.W. Tennant, S.A. Todd, C.J. Tucker, B. Van Houten, B.K. Weis, S. Xuan, H. Zarbl, and Members of the Toxicogenomics Research Consortium. 2005. Standardizing global gene expression analysis between laboratories and across platforms. *Nat.Methods.* 2:351-356.

Barabasi, A.L., and Z.N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat.Rev.Genet.* 5:101-113.

Belden, W.J., and J.C. Dunlap. 2008. SIRT1 is a circadian deacetylase for core clock components. *Cell.* 134:212-214.

Belenky, G., N.J. Wesensten, D.R. Thorne, M.L. Thomas, H.C. Sing, D.P. Redmond, M.B. Russo, and T.J. Balkin. 2003. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *J.Sleep Res.* 12:1-12.

Bellen, H.J., S. Kooyer, D. D'Evelyn, and J. Pearlman. 1992. The Drosophila couch potato protein is expressed in nuclei of peripheral neuronal precursors and shows homology to RNA-binding proteins. *Genes Dev.* 6:2125-2136.

Bellen, H.J., R.W. Levis, G. Liao, Y. He, J.W. Carlson, G. Tsang, M. Evans-Holm, P.R. Hiesinger, K.L. Schulze, G.M. Rubin, R.A. Hoskins, and A.C. Spradling. 2004. The BDGP gene disruption project: single transposon insertions associated with 40% of Drosophila genes. *Genetics.* 167:761-781.

Bell-Pedersen, D., V.M. Cassone, D.J. Earnest, S.S. Golden, P.E. Hardin, T.L. Thomas, and M.J. Zoran. 2005. Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat.Rev.Genet.* 6:544-556.

Benington, J.H., and H.C. Heller. 1995. Restoration of brain energy metabolism as the function of sleep. *Prog.Neurobiol.* 45:347-360.

Benito, J., H. Zheng, and P.E. Hardin. 2007. PDP1epsilon functions downstream of the circadian oscillator to mediate behavioral rhythms. *J.Neurosci.* 27:2539-2547.

Bennett, M., M.F. Schatz, H. Rockwood, and K. Wiesenfeld. 2002. Huygens's clocks. *Proc. R.Soc. Lon. Series A.* 458:563-579.

Berger, R.J., and N.H. Phillips. 1994. Constant light suppresses sleep and circadian rhythms in pigeons without consequent sleep rebound in darkness. *Am.J.Physiol.* 267:R945-52.

Blanchardon, E., B. Grima, A. Klarsfeld, E. Chelot, P.E. Hardin, T. Preat, and F. Rouyer. 2001. Defining the role of Drosophila lateral neurons in the control of circadian rhythms in motor activity and eclosion by targeted genetic ablation and PERIOD protein overexpression. *Eur.J.Neurosci.* 13:871-888.

Blau, J., F. Blanchard, B. Collins, D. Dahdal, A. Knowles, D. Mizrak, and M. Ruben. 2007. What is there left to learn about the Drosophila clock? *Cold Spring Harb.Symp.Quant.Biol.* 72:243-250.

Blau, J., and M.W. Young. 1999. Cycling vrille expression is required for a functional Drosophila clock. *Cell.* 99:661-671.

Block, G.D., M. Geusz, S.B. Khalsa, S. Michel, and D. Whitmore. 1996. Circadian rhythm generation, expression and entrainment in a molluscan model system. *Prog.Brain Res.* 111:93-102.

Boothroyd, C.E., H. Wijnen, F. Naef, L. Saez, and M.W. Young. 2007. Integration of light and temperature in the regulation of circadian gene expression in Drosophila. *PLoS Genet.* 3:e54.

Boutros, M., A.A. Kiger, S. Armknecht, K. Kerr, M. Hild, B. Koch, S.A. Haas, R. Paro, N. Perrimon, and Heidelberg Fly Array Consortium. 2004. Genome-wide RNAi analysis of growth and viability in Drosophila cells. *Science.* 303:832-835.

Brand, A.H., and N. Perrimon. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development.* 118:401-415.

Bushey, D., R. Huber, G. Tononi, and C. Cirelli. 2007. Drosophila Hyperkinetic mutants have reduced sleep and impaired memory. *J.Neurosci.* 27:5384-5393.

Byrne, A.B., M.T. Weirauch, V. Wong, M. Koeva, S.J. Dixon, J.M. Stuart, and P.J. Roy. 2007. A global analysis of genetic interactions in Caenorhabditis elegans. *J.Biol.* 6:8.

Campbell, S.S., and I. Tobler. 1984. Animal sleep: a review of sleep duration across phylogeny. *Neurosci.Biobehav.Rev.* 8:269-300.

Campesan, S., Y. Dubrova, J.C. Hall, and C.P. Kyriacou. 2001. The nonA gene in Drosophila conveys species-specific behavioral characteristics. *Genetics.* 158:1535-1543.

Ceriani, M.F., T.K. Darlington, D. Staknis, P. Mas, A.A. Petti, C.J. Weitz, and S.A. Kay. 1999. Light-dependent sequestration of TIMELESS by CRYPTOCHROME. *Science.* 285:553-556.

Ceriani, M.F., J.B. Hogenesch, M. Yanovsky, S. Panda, M. Straume, and S.A. Kay. 2002. Genome-wide expression analysis in Drosophila reveals genes controlling circadian behavior. *J.Neurosci.* 22:9305-9319.

Chan, R.C., A. Chan, M. Jeon, T.F. Wu, D. Pasqualone, A.E. Rougvie, and B.J. Meyer. 2003. Chromosome cohesion is regulated by a clock gene paralogue TIM-1. *Nature.* 423:1002-1009.

Chang, D.C. 2006. Neural circuits underlying circadian behavior in Drosophila melanogaster. *Behav.Processes.* 71:211-225.

Chen, D., A. Ahlford, F. Schnorrer, I. Kalchhauser, M. Fellner, E. Viragh, I. Kiss, A.C. Syvanen, and B.J. Dickson. 2008. High-resolution, high-throughput SNP mapping in Drosophila melanogaster. *Nat.Methods.* 5:323-329.

Chen, L.S., F. Emmert-Streib, and J.D. Storey. 2007. Harnessing naturally randomized transcription to infer regulatory relationships among genes. *Genome Biol.* 8:R219.

Cheng, Y., and P.E. Hardin. 1998. Drosophila photoreceptors contain an autonomous circadian oscillator that can function without period mRNA cycling. *J.Neurosci.* 18:741-750.

Chintapalli, V.R., J. Wang, and J.A. Dow. 2007. Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nat.Genet.* 39:715-720.

Cirelli, C., and D. Bushey. 2008. Sleep and wakefulness in Drosophila melanogaster. *Ann.N.Y.Acad.Sci.* 1129:323-329.

Cirelli, C., D. Bushey, S. Hill, R. Huber, R. Kreber, B. Ganetzky, and G. Tononi. 2005a. Reduced sleep in Drosophila Shaker mutants. *Nature.* 434:1087-1092.

Cirelli, C., T.M. LaVaute, and G. Tononi. 2005b. Sleep and wakefulness modulate gene expression in Drosophila. *J.Neurochem.* 94:1411-1419.

Cirelli, C., and G. Tononi. 2008. Is sleep essential? *PLoS Biol.* 6:e216.

Cirelli, C., and G. Tononi. 2005. Total sleep deprivation. *In* Sleep Deprivation: Basic Science, Physiology, and Behavior. Marcel Dekker, New York, NY. 63-79.

Claridge-Chang, A., H. Wijnen, F. Naef, C. Boothroyd, N. Rajewsky, and M.W. Young. 2001. Circadian regulation of gene expression systems in the Drosophila head. *Neuron.* 32:657-671.

Clark, A.C., M.L. del Campo, and J. Ewer. 2004. Neuroendocrine control of larval ecdysis behavior in Drosophila: complex regulation by partially redundant neuropeptides. *J.Neurosci.* 24:4283-4292.

Cohen, B.A., R.D. Mitra, J.D. Hughes, and G.M. Church. 2000. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat.Genet.* 26:183-186.

Collins, B., E.O. Mazzoni, R. Stanewsky, and J. Blau. 2006. Drosophila CRYPTOCHROME is a circadian transcriptional repressor. *Curr.Biol.* 16:441-449.

Collins, B.H., S. Dissel, E. Gaten, E. Rosato, and C.P. Kyriacou. 2005. Disruption of Cryptochrome partially restores circadian rhythmicity to the arrhythmic period mutant of Drosophila. *Proc.Natl.Acad.Sci.U.S.A.* 102:19021-19026.

Cook, T., F. Pichaud, R. Sonneville, D. Papatsenko, and C. Desplan. 2003. Distinction between color photoreceptor cell fates is controlled by Prospero in Drosophila. *Dev.Cell.* 4:853-864.

Crane, P.C. 2001. Applications of the DFT/CLEAN technique to solar time series. *Sol. Phys.* 203:381-408.

Cyran, S.A., A.M. Buchsbaum, K.L. Reddy, M.C. Lin, N.R. Glossop, P.E. Hardin, M.W. Young, R.V. Storti, and J. Blau. 2003. vrille, Pdp1, and dClock form a second feedback loop in the Drosophila circadian clock. *Cell.* 112:329-341.

Darlington, T.K., L.C. Lyons, P.E. Hardin, and S.A. Kay. 2000. The period E-box is sufficient to drive circadian oscillation of transcription in vivo. *J.Biol.Rhythms.* 15:462-471.

Darlington, T.K., K. Wager-Smith, M.F. Ceriani, D. Staknis, N. Gekakis, T.D. Steeves, C.J. Weitz, J.S. Takahashi, and S.A. Kay. 1998. Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. *Science.* 280:1599-1603.

Davis, R.L., and D.L. Turner. 2001. Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning. *Oncogene.* 20:8342-8357.

de Belle, J.S. 2002. Unifying the genetics of behavior. *Nat.Genet.* 31:329-330.

Delaunay, F., and V. Laudet. 2002. Circadian clock and microarrays: mammalian genome gets rhythm. *Trends Genet.* 18:595-597.

Dietzl, G., D. Chen, F. Schnorrer, K.C. Su, Y. Barinova, M. Fellner, B. Gasser, K. Kinsey, S. Oppel, S. Scheiblauer, A. Couto, V. Marra, K. Keleman, and B.J. Dickson. 2007. A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature.* 448:151-156.

Dilda, C.L., and T.F. Mackay. 2002. The genetic architecture of Drosophila sensory bristle number. *Genetics.* 162:1655-1674.

Dissel, S., V. Codd, R. Fedic, K.J. Garner, R. Costa, C.P. Kyriacou, and E. Rosato. 2004. A constitutively active cryptochrome in Drosophila melanogaster. *Nat.Neurosci.* 7:834-840.

Doi, M., J. Hirayama, and P. Sassone-Corsi. 2006. Circadian regulator CLOCK is a histone acetyltransferase. *Cell.* 125:497-508.

Dolezelova, E., D. Dolezel, and J.C. Hall. 2007. Rhythm defects caused by newly engineered null mutations in Drosophila's cryptochrome gene. *Genetics.* 177:329-345.

Donlea, J.M., N. Ramanan, and P.J. Shaw. 2009. Use-dependent plasticity in clock neurons regulates sleep need in Drosophila. *Science.* 324:105-108.

Dowse, H.B. 2008. Mid-range Ultradian Rhythms in *Drosophila* and the Circadian Clock Problem. *In* Ultradian Rhythms from Molecules to Mind: A New Vision of Life. D. Lloyd and E. Rossi, editors. Spring Science + Business Media B.V. 175-199.

Dowse, H.B. 2007. Statistical analysis of biological rhythm data. *Methods Mol.Biol.* 362:29-45.

Dowse, H.B., M.S. Dushay, J.C. Hall, and J.M. Ringo. 1989. High-resolution analysis of locomotor activity rhythms in disconnected, a visual-system mutant of Drosophila melanogaster. *Behav.Genet.* 19:529-542.

Dowse, H.B., J.C. Hall, and J.M. Ringo. 1987. Circadian and ultradian rhythms in period mutants of Drosophila melanogaster. *Behav.Genet.* 17:19-35.

Dowse, H.B., and J.M. Ringo. 1987. Further evidence that the circadian clock in Drosophila is a population of coupled ultradian oscillators. *J.Biol.Rhythms.* 2:65-76.

Drosophila 12 Genomes Consortium, A.G. Clark, M.B. Eisen, D.R. Smith, C.M. Bergman, B. Oliver, T.A. Markow, T.C. Kaufman, M. Kellis, W. Gelbart, V.N. Iyer, D.A. Pollard, T.B. Sackton, A.M. Larracuente, N.D. Singh, J.P. Abad, D.N. Abt, B. Adryan, M. Aguade, H. Akashi, W.W. Anderson, C.F. Aquadro, D.H. Ardell, R. Arguello, C.G. Artieri, D.A. Barbash, D. Barker, P. Barsanti, P. Batterham, S. Batzoglou, D. Begun, A. Bhutkar, E. Blanco, S.A. Bosak, R.K. Bradley, A.D. Brand, M.R. Brent, A.N. Brooks, R.H. Brown, R.K. Butlin, C. Caggese, B.R. Calvi, A. Bernardo de Carvalho, A. Caspi, S. Castrezana, S.E. Celniker, J.L. Chang, C. Chapple, S. Chatterji, A. Chinwalla, A. Civetta, S.W. Clifton, J.M. Comeron, J.C. Costello, J.A. Coyne, J. Daub, R.G. David, A.L. Delcher, K. Delehaunty, C.B. Do, H. Ebling, K. Edwards, T. Eickbush, J.D. Evans, A. Filipski, S. Findeiss, E. Freyhult, L. Fulton, R. Fulton, A.C. Garcia, A. Gardiner, D.A. Garfield, B.E. Garvin, G. Gibson, D. Gilbert, S. Gnerre, J. Godfrey, R. Good, V. Gotea, B. Gravely, A.J. Greenberg, S. Griffiths-Jones, S. Gross, R. Guigo, E.A. Gustafson, W. Haerty, M.W. Hahn, D.L. Halligan, A.L. Halpern, G.M. Halter, M.V. Han, A. Heger, L. Hillier, A.S. Hinrichs, I. Holmes, R.A. Hoskins, M.J. Hubisz, D. Hultmark, M.A. Huntley, D.B. Jaffe, S. Jagadeeshan, W.R. Jeck, J. Johnson, C.D. Jones, W.C. Jordan, G.H. Karpen, E. Kataoka, P.D. Keightley, P. Kheradpour, E.F. Kirkness, L.B. Koerich, K. Kristiansen, D. Kudrna, R.J. Kulathinal, S. Kumar, R. Kwok, E. Lander, C.H. Langley, R. Lapoint,

B.P. Lazzaro, S.J. Lee, L. Levesque, R. Li, C.F. Lin, M.F. Lin, K. Lindblad-Toh, A. Llopart, M. Long, L. Low, E. Lozovsky, J. Lu, M. Luo, C.A. Machado, W. Makalowski, M. Marzo, M. Matsuda, L. Matzkin, B. McAllister, C.S. McBride, B. McKernan, K. McKernan, M. Mendez-Lago, P. Minx, M.U. Mollenhauer, K. Montooth, S.M. Mount, X. Mu, E. Myers, B. Negre, S. Newfeld, R. Nielsen, M.A. Noor, P. O'Grady, L. Pachter, M. Papaceit, M.J. Parisi, M. Parisi, L. Parts, J.S. Pedersen, G. Pesole, A.M. Phillippy, C.P. Ponting, M. Pop, D. Porcelli, J.R. Powell, S. Prohaska, K. Pruitt, M. Puig, H. Quesneville, K.R. Ram, D. Rand, M.D. Rasmussen, L.K. Reed, R. Reenan, A. Reily, K.A. Remington, T.T. Rieger, M.G. Ritchie, C. Robin, Y.H. Rogers, C. Rohde, J. Rozas, M.J. Rubenfield, A. Ruiz, S. Russo, S.L. Salzberg, A. Sanchez-Gracia, D.J. Saranga, H. Sato, S.W. Schaeffer, M.C. Schatz, T. Schlenke, R. Schwartz, C. Segarra, R.S. Singh, L. Sirot, M. Sirota, N.B. Sisneros, C.D. Smith, T.F. Smith, J. Spieth, D.E. Stage, A. Stark, W. Stephan, R.L. Strausberg, S. Strempel, D. Sturgill, G. Sutton, G.G. Sutton, W. Tao, S. Teichmann, Y.N. Tobari, Y. Tomimura, J.M. Tsolas, V.L. Valente, E. Venter, J.C. Venter, S. Vicario, F.G. Vieira, A.J. Vilella, A. Villasante, B. Walenz, J. Wang, M. Wasserman, T. Watts, D. Wilson, R.K. Wilson, R.A. Wing, M.F. Wolfner, A. Wong, G.K. Wong, C.I. Wu, G. Wu, D. Yamamoto, H.P. Yang, S.P. Yang, J.A. Yorke, K. Yoshida, E. Zdobnov, P. Zhang, Y. Zhang, A.V. Zimin, J. Baldwin, A. Abdouelleil, J. Abdulkadir, A. Abebe, B. Abera, J. Abreu, S.C. Acer, L. Aftuck, A. Alexander, P. An, E. Anderson, S. Anderson, H. Arachi, M. Azer, P. Bachantsang, A. Barry, T. Bayul, A. Berlin, D. Bessette, T. Bloom, J. Blye, L. Boguslavskiy, C. Bonnet, B. Boukhgalter, I. Bourzgui, A. Brown, P. Cahill, S. Channer, Y. Cheshatsang, L. Chuda, M. Citroen, A. Collymore, P. Cooke, M. Costello, K. D'Aco, R. Daza, G. De Haan, S. DeGray, C. DeMaso, N. Dhargay, K. Dooley, E. Dooley, M. Doricent, P. Dorje, K. Dorjee, A. Dupes, R. Elong, J. Falk, A. Farina, S. Faro, D. Ferguson, S. Fisher, C.D. Foley, A. Franke, D. Friedrich, L. Gadbois, G. Gearin, C.R. Gearin, G. Giannoukos, T. Goode, J. Graham, E. Grandbois, S. Grewal, K. Gyaltsen, N. Hafez, B. Hagos, J. Hall, C. Henson, A. Hollinger, T. Honan, M.D. Huard, L. Hughes, B. Hurhula, M.E. Husby, A. Kamat, B. Kanga, S. Kashin, D. Khazanovich, P. Kisner, K. Lance, M. Lara, W. Lee, N. Lennon, F. Letendre, R. LeVine, A. Lipovsky, X. Liu, J. Liu, S. Liu, T. Lokyitsang, Y. Lokyitsang, R. Lubonja, A. Lui, P. MacDonald, V. Magnisalis, K. Maru, C. Matthews, W. McCusker, S. McDonough, T. Mehta, J. Meldrim, L. Meneus, O. Mihai, A. Mihalev, T. Mihova, R. Mittelman, V. Mlenga, A. Montmayeur, L. Mulrain, A. Navidi, J. Naylor, T. Negash, T. Nguyen, N. Nguyen, R. Nicol, C. Norbu, N. Norbu, N. Novod, B. O'Neill, S. Osman, E. Markiewicz, O.L. Oyono, C. Patti, P. Phunkhang, F. Pierre, M. Priest, S. Raghuraman, F. Rege, R. Reyes, C. Rise, P. Rogov, K. Ross, E. Ryan, S. Settipalli, T. Shea, N. Sherpa, L. Shi, D. Shih, T. Sparrow, J. Spaulding, J. Stalker, N. Stange-Thomann, S. Stavropoulos, C. Stone, C. Strader, S. Tesfaye, T. Thomson, Y. Thoulutsang, D. Thoulutsang, K. Topham, I. Topping, T. Tsamla, H. Vassiliev, A. Vo, T. Wangchuk, T. Wangdi, M. Weiand, J. Wilkinson, A. Wilson, S. Yadav, G. Young, Q. Yu, L. Zembek, D. Zhong, A. Zimmer, Z. Zwirko, D.B. Jaffe, P. Alvarez, W. Brockman, J. Butler, C. Chin, S. Gnerre, M. Grabherr, M. Kleber, E. Mauceli, and I. MacCallum. 2007. Evolution of genes and genomes on the Drosophila phylogeny. *Nature.* 450:203-218.

Dubruille, R., and P. Emery. 2008. A plastic clock: how circadian rhythms respond to environmental cues in Drosophila. *Mol.Neurobiol.* 38:129-145.

Duffield, G.E. 2003. DNA microarray analyses of circadian timing: the genomic basis of biological time. *J.Neuroendocrinol.* 15:991-1002.

Duffy, J.B. 2002. GAL4 system in Drosophila: a fly geneticist's Swiss army knife. *Genesis.* 34:1-15.

Dulcis, D., R.B. Levine, and J. Ewer. 2005. Role of the neuropeptide CCAP in Drosophila cardiac function. *J.Neurobiol.* 64:259-274.

Dworkin, I. 2005. Canlization, Crptic Variation, and Developmental Buffering: A Critical Examination and Analytical Perspective. *In* Variation. B. Hallgrímsson and B.K. Hall, editors. Academic Press. 131-158.

Edery, I. 2007. A blend of two circadian clocks, seasoned to perfection. *Cell.* 129:21-23.

Edery, I., L.J. Zwiebel, M.E. Dembinska, and M. Rosbash. 1994. Temporal phosphorylation of the Drosophila period protein. *Proc.Natl.Acad.Sci.U.S.A.* 91:2260-2264.

Elkins, T., B. Ganetzky, and C.F. Wu. 1986. A Drosophila mutation that eliminates a calcium-dependent potassium current. *Proc.Natl.Acad.Sci.U.S.A.* 83:8415-8419.

Etter, P.D., and M. Ramaswami. 2002. The ups and downs of daily life: profiling circadian gene expression in Drosophila. *Bioessays.* 24:494-498.

Ewer, J., B. Frisch, M.J. Hamblen-Coyle, M. Rosbash, and J.C. Hall. 1992. Expression of the period clock gene within different cell types in the brain of Drosophila adults and mosaic analysis of these cells' influence on circadian behavioral rhythms. *J.Neurosci.* 12:3321-3349.

Falconer, D.S., and T.F. Mackay. 1995. Introduction to Quantitative Genetics. Longman, London. 480 pp.

Fang, Y., S. Sathyanarayanan, and A. Sehgal. 2007. Post-translational regulation of the Drosophila circadian clock requires protein phosphatase 1 (PP1). *Genes Dev.* 21:1506-1518.

Fedorowicz, G.M., J.D. Fry, R.R. Anholt, and T.F. Mackay. 1998. Epistatic interactions between smell-impaired loci in Drosophila melanogaster. *Genetics.* 148:1885-1891.

Fire, A., S. Xu, M.K. Montgomery, S.A. Kostas, S.E. Driver, and C.C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature.* 391:806-811.

Fisher, R.A. 1918. The correlations between relatives on the supposesition of Mendelian inheritance. *Trans. R. Soc. Edinb.* 52:399-433.

Foltenyi, K., R. Andretic, J.W. Newport, and R.J. Greenspan. 2007a. Neurohormonal and neuromodulatory control of sleep in Drosophila. *Cold Spring Harb.Symp.Quant.Biol.* 72:565-571.

Foltenyi, K., R.J. Greenspan, and J.W. Newport. 2007b. Activation of EGFR and ERK by rhomboid signaling regulates the consolidation and maintenance of sleep in Drosophila. *Nat.Neurosci.* 10:1160-1167.

Freeman, M. 1996. Reiterative use of the EGF receptor triggers differentiation of all cell types in the Drosophila eye. *Cell.* 87:651-660.

Friedman, A., and N. Perrimon. 2007. Genetic screening for signal transduction in the era of network biology. *Cell.* 128:225-231.

Friedman, A., and N. Perrimon. 2004. Genome-wide high-throughput screens in functional genomics. *Curr.Opin.Genet.Dev.* 14:470-476.

Frisch, B., P.E. Hardin, M.J. Hamblen-Coyle, M. Rosbash, and J.C. Hall. 1994. A promoterless period gene mediates behavioral rhythmicity and cyclical per expression in a restricted subset of the Drosophila nervous system. *Neuron.* 12:555-570.

Ganguly-Fitzgerald, I., J. Donlea, and P.J. Shaw. 2006. Waking experience affects sleep need in Drosophila. *Science.* 313:1775-1781.

Gegear, R.J., A. Casselman, S. Waddell, and S.M. Reppert. 2008. Cryptochrome mediates light-dependent magnetosensitivity in Drosophila. *Nature.* 454:1014-1018.

George, H., and R. Terracol. 1997. The vrille gene of Drosophila is a maternal enhancer of decapentaplegic and encodes a new member of the bZIP family of transcription factors. *Genetics.* 146:1345-1363.

Gilestro, G.F., and C. Cirelli. 2009. pySolo: a complete suite for sleep analysis in Drosophila. *Bioinformatics.* 25:1466-1467.

Gilestro, G.F., G. Tononi, and C. Cirelli. 2009. Widespread changes in synaptic markers as a function of sleep and wakefulness in Drosophila. *Science.* 324:109-112.

Glaser, F.T., and R. Stanewsky. 2005. Temperature synchronization of the Drosophila circadian clock. *Curr.Biol.* 15:1352-1363.

Glossop, N.R., J.H. Houl, H. Zheng, F.S. Ng, S.M. Dudek, and P.E. Hardin. 2003. VRILLE feeds back to control circadian transcription of Clock in the Drosophila circadian oscillator. *Neuron.* 37:249-261.

Glossop, N.R., L.C. Lyons, and P.E. Hardin. 1999. Interlocked feedback loops within the Drosophila circadian oscillator. *Science.* 286:766-768.

Goh, K.I., M.E. Cusick, D. Valle, B. Childs, M. Vidal, and A.L. Barabasi. 2007. The human disease network. *Proc.Natl.Acad.Sci.U.S.A.* 104:8685-8690.

Golic, K.G., and M.M. Golic. 1996. Engineering the Drosophila genome: chromosome rearrangements by design. *Genetics.* 144:1693-1711.

Gorska-Andrzejak, J., P.M. Salvaterra, I.A. Meinertzhagen, W. Krzeptowski, A. Gorlich, and E. Pyza. 2009. Cyclical expression of Na+/K+-ATPase in the visual system of Drosophila melanogaster. *J.Insect Physiol.* 55:459-468.

Greenspan, R.J. 1997. Fly Pushing: The Theory and Practice of Drosophila Genetics. Cold Spring Harbor Laboratory Press, New York. 155 pp.

Greenspan, R.J. 2001. The flexible genome. *Nat.Rev.Genet.* 2:383-387.

Greenspan, R.J. 1997. A kinder, gentler genetic analysis of behavior: dissection gives way to modulation. *Curr.Opin.Neurobiol.* 7:805-811.

Griffing, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust J Biol Sci.* 9:463-493.

Griffith, L.C., and M. Rosbash. 2008. Sleep: hitting the reset button. *Nat.Neurosci.* 11:123-124.

Grima, B., E. Chelot, R. Xia, and F. Rouyer. 2004. Morning and evening peaks of activity rely on different clock neurons of the Drosophila brain. *Nature.* 431:869-873.

Grima, B., A. Lamouroux, E. Chelot, C. Papin, B. Limbourg-Bouchon, and F. Rouyer. 2002. The F-box protein slimb controls the levels of clock proteins period and timeless. *Nature.* 420:178-182.

Hall, J.C. 1994. Pleiotropy of behavioural genes. *In* Flexibility and Contraint in Behavioral Systems. R.J. Greenspan and C.P. Kyriacou, editors. John Wiley & Suns, Berlin/New York. 15-28.

Hamasaka, Y., and D.R. Nassel. 2006. Mapping of serotonin, dopamine, and histamine in relation to different clock neurons in the brain of Drosophila. *J.Comp.Neurol.* 494:314-330.

Hamasaka, Y., D. Rieger, M.L. Parmentier, Y. Grau, C. Helfrich-Forster, and D.R. Nassel. 2007. Glutamate and its metabotropic receptor in Drosophila clock neuron circuits. *J.Comp.Neurol.* 505:32-45.

Hanai, S., Y. Hamasaka, and N. Ishida. 2008. Circadian entrainment to red light in Drosophila: requirement of Rhodopsin 1 and Rhodopsin 6. *Neuroreport.* 19:1441-1444.

Harbison, S.T., and A. Sehgal. 2008. Quantitative genetic analysis of sleep in Drosophila melanogaster. *Genetics.* 178:2341-2360.

Hartse, K.M. 1994. Sleep in insects and nonmammalian vertebrates. *In* Principals and Practice of Sleep Medicine. M.H. Kryger, T. Roth and W.C. Dement, editors. Saunders, W.B.;, Philadelphia. 95-104.

Hastings, M.H., A.B. Reddy, and E.S. Maywood. 2003. A clockwork web: circadian timing in brain and periphery, in health and disease. *Nat.Rev.Neurosci.* 4:649-661.

Haywood, A.F.M., L.D. Saunders, and B.E. Staveley. 2002. Dopa decarboxylase(Ddc)-GAL4 dramatically reduces life span. *Dros. Inf. Serv.* 85.

Helfrich, C. 1986. Role of the optic lobes in the regulation of the locomotor activity rhythm of Drosophila melanogaster: behavioral analysis of neural mutants. *J.Neurogenet.* 3:321-343.

Helfrich-Forster, C. 2000. Differential control of morning and evening components in the activity rhythm of Drosophila melanogaster--sex-specific differences suggest a different quality of activity. *J.Biol.Rhythms.* 15:135-154.

Helfrich-Forster, C. 1995. The period clock gene is expressed in central nervous system neurons which also produce a neuropeptide that reveals the projections of circadian pacemaker cells within the brain of Drosophila melanogaster. *Proc.Natl.Acad.Sci.U.S.A.* 92:612-616.

Helfrich-Forster, C., O.T. Shafer, C. Wulbeck, E. Grieshaber, D. Rieger, and P. Taghert. 2007. Development and morphology of the clock-gene-expressing lateral neurons of Drosophila melanogaster. *J.Comp.Neurol.* 500:47-70.

Helfrich-Forster, C., C. Winter, A. Hofbauer, J.C. Hall, and R. Stanewsky. 2001. The circadian clock of fruit flies is blind after elimination of all known photoreceptors. *Neuron.* 30:249-261.

Hendricks, J.C., S.M. Finn, K.A. Panckeri, J. Chavkin, J.A. Williams, A. Sehgal, and A.I. Pack. 2000. Rest in Drosophila is a sleep-like state. *Neuron.* 25:129-138.

Hendricks, J.C., J.A. Williams, K. Panckeri, D. Kirk, M. Tello, J.C. Yin, and A. Sehgal. 2001. A non-circadian role for cAMP signaling and CREB activity in Drosophila rest homeostasis. *Nat.Neurosci.* 4:1108-1115.

Heward, J.A., D.A. Baker, T.C. Lukins, and J.D. Armstrong. 2005. FlyTracker: Real-Time Analysis Of Insect Courtship. *Proceedings of Measuring Behavior 2005, 5th International Conference on Methods and Techniques in Behavioral Research.*

Hirsch, J., and L. Erlenmeyer-Kimling. 1961. Sign of taxis as a property of the genotype. *Science.* 134:835-836.

Högbom, J. 1974. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astrophys. J. Suppl. Ser.* 15:417.

Houl, J.H., W. Yu, S.M. Dudek, and P.E. Hardin. 2006. Drosophila CLOCK is constitutively expressed in circadian oscillator and non-oscillator cells. *J.Biol.Rhythms.* 21:93-103.

Huang, Z.J., K.D. Curtin, and M. Rosbash. 1995. PER protein interactions and temperature compensation of a circadian clock in Drosophila. *Science.* 267:1169-1172.

Huber, R., S.L. Hill, C. Holladay, M. Biesiadecki, G. Tononi, and C. Cirelli. 2004. Sleep homeostasis in Drosophila melanogaster. *Sleep.* 27:628-639.

Hung, H.C., S.A. Kay, and F. Weber. 2009. HSP90, a capacitor of behavioral variation. *J.Biol.Rhythms.* 24:183-192.

Hyun, S., Y. Lee, S.T. Hong, S. Bang, D. Paik, J. Kang, J. Shin, J. Lee, K. Jeon, S. Hwang, E. Bae, and J. Kim. 2005. Drosophila GPCR Han is a receptor for the circadian clock neuropeptide PDF. *Neuron.* 48:267-278.

Iijima-Ando, K., and J.C. Yin. 2005. Transgenic cAMP response element reporter flies for monitoring circadian rhythms. *Methods Enzymol.* 393:302-315.

Inoue, S., K. Honda, and Y. Komoda. 1995. Sleep as neuronal detoxification and restitution. *Behav.Brain Res.* 69:91-96.

Ito, C., S.G. Goto, S. Shiga, K. Tomioka, and H. Numata. 2008. Peripheral circadian clock for the cuticle deposition rhythm in Drosophila melanogaster. *Proc.Natl.Acad.Sci.U.S.A.* 105:8446-8451.

Ivanchenko, M., R. Stanewsky, and J.M. Giebultowicz. 2001. Circadian photoreception in Drosophila: functions of cryptochrome in peripheral and central clocks. *J.Biol.Rhythms.* 16:205-215.

Joiner, W.J., A. Crocker, B.H. White, and A. Sehgal. 2006. Sleep in Drosophila is regulated by adult mushroom bodies. *Nature.* 441:757-760.

Kadener, S., D. Stoleru, M. McDonald, P. Nawathean, and M. Rosbash. 2007. Clockwork Orange is a transcriptional repressor and a new Drosophila circadian pacemaker component. *Genes Dev.* 21:1675-1686.

Kageyama, H., T. Nishiwaki, M. Nakajima, H. Iwasaki, T. Oyama, and T. Kondo. 2006. Cyanobacterial circadian pacemaker: Kai protein complex dynamics in the KaiC phosphorylation cycle in vitro. *Mol.Cell.* 23:161-171.

Kaneko, M., and J.C. Hall. 2000. Neuroanatomy of cells expressing clock genes in Drosophila: transgenic manipulation of the period and timeless genes to mark the perikarya of circadian pacemaker neurons and their projections. *J.Comp.Neurol.* 422:66-94.

Kaneko, M., C. Helfrich-Forster, and J.C. Hall. 1997. Spatial and temporal expression of the period and timeless genes in the developing nervous system of Drosophila: newly identified pacemaker candidates and novel features of clock gene product cycling. *J.Neurosci.* 17:6745-6760.

Kantermann, T., M. Juda, M. Merrow, and T. Roenneberg. 2007. The human circadian clock's seasonal adjustment is disrupted by daylight saving time. *Curr.Biol.* 17:1996-2000.

Kaplan, D.D., G. Zimmermann, K. Suyama, T. Meyer, and M.P. Scott. 2008. A nucleostemin family GTPase, NS3, acts in serotonergic neurons to regulate insulin signaling and control body size. *Genes Dev.* 22:1877-1893.

Kavanau, J.L. 1997. Memory, sleep and the evolution of mechanisms of synaptic efficacy maintenance. *Neuroscience.* 79:7-44.

Keegan, K.P., S. Pradhan, J.P. Wang, and R. Allada. 2007. Meta-analysis of Drosophila circadian microarray studies identifies a novel set of rhythmically expressed genes. *PLoS Comput.Biol.* 3:e208.

Kempinger, L., R. Dittmann, D. Rieger, and C. Helfrich-Forster. 2009. The nocturnal activity of fruit flies exposed to artificial moonlight is partly caused by direct light effects on the activity level that bypass the endogenous clock. *Chronobiol.Int.* 26:151-166.

Kim, E.Y., K. Bae, F.S. Ng, N.R. Glossop, P.E. Hardin, and I. Edery. 2002. Drosophila CLOCK protein is under posttranscriptional control and influences light-induced activity. *Neuron.* 34:69-81.

Klarsfeld, A., S. Malpel, C. Michard-Vanhee, M. Picot, E. Chelot, and F. Rouyer. 2004. Novel features of cryptochrome-mediated photoreception in the brain circadian clock of Drosophila. *J.Neurosci.* 24:1468-1477.

Kloss, B., J.L. Price, L. Saez, J. Blau, A. Rothenfluh, C.S. Wesley, and M.W. Young. 1998. The Drosophila clock gene double-time encodes a protein closely related to human casein kinase Iepsilon. *Cell.* 94:97-107.

Knowles, A., K. Koh, J.T. Wu, C.T. Chien, D.A. Chamovitz, and J. Blau. 2009. The COP9 signalosome is required for light-dependent timeless degradation and Drosophila clock resetting. *J.Neurosci.* 29:1152-1162.

Ko, H.W., J. Jiang, and I. Edery. 2002. Role for Slimb in the degradation of Drosophila Period protein phosphorylated by Doubletime. *Nature.* 420:673-678.

Koh, K., J.M. Evans, J.C. Hendricks, and A. Sehgal. 2006. A Drosophila model for age-associated changes in sleep:wake cycles. *Proc.Natl.Acad.Sci.U.S.A.* 103:13843-13847.

Koh, K., W.J. Joiner, M.N. Wu, Z. Yue, C.J. Smith, and A. Sehgal. 2008. Identification of SLEEPLESS, a sleep-promoting factor. *Science.* 321:372-376.

Koh, K., X. Zheng, and A. Sehgal. 2006. JETLAG resets the Drosophila circadian clock by promoting light-induced degradation of TIMELESS. *Science.* 312:1809-1812.

Konopka, R.J., and S. Benzer. 1971. Clock mutants of Drosophila melanogaster. *Proc.Natl.Acad.Sci.U.S.A.* 68:2112-2116.

Konopka, R.J., C. Pittendrigh, and D. Orr. 1989. Reciprocal behaviour associated with altered homeostasis and photosensitivity of Drosophila clock mutants. *J.Neurogenet.* 6:1-10.

Konopka, R.J., R.F. Smith, and D. Orr. 1991. Characterization of Andante, a new Drosophila clock mutant, and its interactions with other clock mutants. *J.Neurogenet.* 7:103-114.

Korman, M., J. Doyon, J. Doljansky, J. Carrier, Y. Dagan, and A. Karni. 2007. Daytime sleep condenses the time course of motor memory consolidation. *Nat.Neurosci.* 10:1206-1213.

Kramer, J.M., and B.E. Staveley. 2003. GAL4 causes developmental defects and apoptosis when expressed in the developing eye of Drosophila melanogaster. *Genet.Mol.Res.* 2:43-47.

Krishnan, B., J.D. Levine, M.K. Lynch, H.B. Dowse, P. Funes, J.C. Hall, P.E. Hardin, and S.E. Dryer. 2001. A new role for cryptochrome in a Drosophila circadian oscillator. *Nature.* 411:313-317.

Kula, E., E.S. Levitan, E. Pyza, and M. Rosbash. 2006. PDF cycling in the dorsal protocerebrum of the Drosophila brain is not necessary for circadian clock function. *J.Biol.Rhythms.* 21:104-117.

Kume, K., S. Kume, S.K. Park, J. Hirsh, and F.R. Jackson. 2005. Dopamine is a regulator of arousal in the fruit fly. *J.Neurosci.* 25:7377-7384.

Kyriacou, C.P., and J.C. Hall. 1989. Circadian rhythm mutations in Drosophila melanogaster affect short-term fluctuations in the male's courtship song. *PNAS.* 77:6729-6733.

Kyriacou, C.P., M. Oldroyd, J. Wood, M. Sharp, and M. Hill. 1990. Clock mutations alter developmental timing in Drosophila. *Heredity.* 64 (Pt 3):395-401.

Kyriacou, C.P., A.A. Peixoto, F. Sandrelli, R. Costa, and E. Tauber. 2008. Clines in clock genes: fine-tuning circadian rhythms to the environment. *Trends Genet.* 24:124-132.

Landskron, J., K.F. Chen, E. Wolf, and R. Stanewsky. 2009. A role for the PERIOD:PERIOD homodimer in the Drosophila circadian clock. *PLoS Biol.* 7:e3.

Larkin, J.E., T. Yokogawa, H.C. Heller, P. Franken, and N.F. Ruby. 2004. Homeostatic regulation of sleep in arrhythmic Siberian hamsters. *Am.J.Physiol.Regul.Integr.Comp.Physiol.* 287:R104-11.

Lear, B.C., J.M. Lin, J.R. Keath, J.J. McGill, I.M. Raman, and R. Allada. 2005a. The ion channel narrow abdomen is critical for neural output of the Drosophila circadian pacemaker. *Neuron.* 48:965-976.

Lear, B.C., C.E. Merrill, J.M. Lin, A. Schroeder, L. Zhang, and R. Allada. 2005b. A G protein-coupled receptor, groom-of-PDF, is required for PDF neuron action in circadian behavior. *Neuron.* 48:221-227.

Lee, C., K. Bae, and I. Edery. 1999. PER and TIM inhibit the DNA binding activity of a Drosophila CLOCK-CYC/dBMAL1 heterodimer without disrupting formation of the heterodimer: a basis for circadian transcription. *Mol.Cell.Biol.* 19:5316-5325.

Lee, C., K. Bae, and I. Edery. 1998. The Drosophila CLOCK protein undergoes daily rhythms in abundance, phosphorylation, and interactions with the PER-TIM complex. *Neuron.* 21:857-867.

Lee, G., J.H. Bahn, and J.H. Park. 2006. Sex- and clock-controlled expression of the neuropeptide F gene in Drosophila. *Proc.Natl.Acad.Sci.U.S.A.* 103:12580-12585.

Lee, G., and J.H. Park. 2004. Hemolymph sugar homeostasis and starvation-induced hyperactivity affected by genetic manipulations of the adipokinetic hormone-encoding gene in Drosophila melanogaster. *Genetics.* 167:311-323.

Leise, T.L., and E.E. Moin. 2007. A mathematical model of the Drosophila circadian clock with emphasis on posttranslational mechanisms. *J.Theor.Biol.* 248:48-63.

Levine, J.D., P. Funes, H.B. Dowse, and J.C. Hall. 2002a. Resetting the circadian clock by social experience in Drosophila melanogaster. *Science.* 298:2010-2012.

Levine, J.D., P. Funes, H.B. Dowse, and J.C. Hall. 2002b. Signal analysis of behavioral and molecular cycles. *BMC Neurosci.* 3:1.

Li, A., and S. Horvath. 2007. Network neighborhood analysis with the multi-node topological overlap measure. *Bioinformatics.* 23:222-231.

Li, H., S. Chaney, I.J. Roberts, M. Forte, and J. Hirsh. 2000. Ectopic G-protein expression in dopamine and serotonin neurons blocks cocaine sensitization in Drosophila melanogaster. *Curr.Biol.* 10:211-214.

Lin, J.M., V.L. Kilman, K. Keegan, B. Paddock, M. Emery-Le, M. Rosbash, and R. Allada. 2002. A role for casein kinase 2alpha in the Drosophila circadian clock. *Nature.* 420:816-820.

Lin, L., J. Faraco, R. Li, H. Kadotani, W. Rogers, X. Lin, X. Qiu, P.J. de Jong, S. Nishino, and E. Mignot. 1999. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell.* 98:365-376.

Lin, S.C., M.H. Lin, P. Horvath, K.L. Reddy, and R.V. Storti. 1997. PDP1, a novel Drosophila PAR domain bZIP transcription factor expressed in developing mesoderm, endoderm and ectoderm, is a transcriptional regulator of somatic muscle genes. *Development.* 124:4685-4696.

Lin, Y., M. Han, B. Shimada, L. Wang, T.M. Gibler, A. Amarakone, T.A. Awad, G.D. Stormo, R.N. Van Gelder, and P.H. Taghert. 2002. Influence of the period-dependent circadian clock on diurnal, circadian, and aperiodic gene expression in Drosophila melanogaster. *Proc.Natl.Acad.Sci.U.S.A.* 99:9562-9567.

Lin, Y., G.D. Stormo, and P.H. Taghert. 2004. The neuropeptide pigment-dispersing factor coordinates pacemaker interactions in the Drosophila circadian system. *J.Neurosci.* 24:7951-7957.

Lyons, L.C., T.K. Darlington, H. Hao, J. Houl, S.A. Kay, and P.E. Hardin. 2000. Specific sequences outside the E-box are required for proper per expression and behavioral rescue. *J.Biol.Rhythms.* 15:472-482.

Mackay, T.F. 2008. The genetic architecture of complex behaviors: lessons from Drosophila. *Genetica.* 136:295-302.

Mackay, T.F., and R.R. Anholt. 2006. Of flies and man: Drosophila as a model for human complex traits. *Annu.Rev.Genomics Hum.Genet.* 7:339-367.

Mackay, T.F., E.A. Stone, and J.F. Ayroles. 2009. The genetics of quantitative traits: challenges and prospects. *Nat.Rev.Genet.* 10:565-577.

Majercak, J., D. Kalderon, and I. Edery. 1997. Drosophila melanogaster deficient in protein kinase A manifests behavior-specific arrhythmia but normal clock function. *Mol.Cell.Biol.* 17:5915-5922.

Majercak, J., D. Sidote, P.E. Hardin, and I. Edery. 1999. How a circadian clock adapts to seasonal decreases in temperature and day length. *Neuron.* 24:219-230.

Martinek, S., S. Inonog, A.S. Manoukian, and M.W. Young. 2001. A role for the segment polarity gene shaggy/GSK-3 in the Drosophila circadian clock. *Cell.* 105:769-779.

Martinek, S., and M.W. Young. 2000. Specific genetic interference with behavioral rhythms in Drosophila by expression of inverted repeats. *Genetics.* 156:1717-1725.

Matsumoto, A., K. Tomioka, Y. Chiba, and T. Tanimura. 1999. timrit Lengthens circadian period in a temperature-dependent manner through suppression of PERIOD protein cycling and nuclear localization. *Mol.Cell.Biol.* 19:4343-4354.

Matsumoto, A., M. Ukai-Tadenuma, R.G. Yamada, J. Houl, K.D. Uno, T. Kasukawa, B. Dauwalder, T.Q. Itoh, K. Takahashi, R. Ueda, P.E. Hardin, T. Tanimura, and H.R. Ueda. 2007. A functional genomics strategy reveals clockwork orange as a transcriptional regulator in the Drosophila circadian clock. *Genes Dev.* 21:1687-1700.

McBride, S.M., G. Giuliani, C. Choi, P. Krause, D. Correale, K. Watson, G. Baker, and K.K. Siwicki. 1999. Mushroom body ablation impairs short-term memory and long-term memory of courtship conditioning in Drosophila melanogaster. *Neuron.* 24:967-977.

McDonald, M.J., and M. Rosbash. 2001. Microarray analysis and organization of circadian gene expression in Drosophila. *Cell.* 107:567-578.

McGuire, S.E., P.T. Le, and R.L. Davis. 2001. The role of Drosophila mushroom body signaling in olfactory memory. *Science.* 293:1330-1333.

McGuire, T.R. 1981. Selection for central excitatory state (CES) in the blow fly Phormia regina. *Behav.Genet.* 11:331-338.

Mealey-Ferrara, M.L., A.G. Montalvo, and J.C. Hall. 2003. Effects of combining a cryptochrome mutation with other visual-system variants on entrainment of locomotor and adult-emergence rhythms in Drosophila. *J.Neurogenet.* 17:171-221.

Meng, Q.J., L. Logunova, E.S. Maywood, M. Gallego, J. Lebiecki, T.M. Brown, M. Sladek, A.S. Semikhodskii, N.R. Glossop, H.D. Piggins, J.E. Chesham, D.A. Bechtold, S.H. Yoo, J.S. Takahashi, D.M. Virshup, R.P. Boot-Handford, M.H. Hastings, and A.S. Loudon. 2008. Setting clock speed in mammals: the CK1 epsilon tau mutation in mice accelerates circadian pacemakers by selectively destabilizing PERIOD proteins. *Neuron.* 58:78-88.

Mertens, I., A. Vandingenen, E.C. Johnson, O.T. Shafer, W. Li, J.S. Trigg, A. De Loof, L. Schoofs, and P.H. Taghert. 2005. PDF receptor signaling in Drosophila contributes to both circadian and geotactic behaviors. *Neuron.* 48:213-219.

Meyer, P., L. Saez, and M.W. Young. 2006. PER-TIM interactions in living Drosophila cells: an interval timer for the circadian clock. *Science.* 311:226-229.

Morgan, T.H. 1911. The Origin of Five Mutations in Eye Color in Drosophila and their Modes of Inheritance. *Science.* 33:534-537.

Morton, A.J., N.I. Wood, M.H. Hastings, C. Hurelbrink, R.A. Barker, and E.S. Maywood. 2005. Disintegration of the sleep-wake cycle and circadian timing in Huntington's disease. *J.Neurosci.* 25:157-163.

Murad, A., M. Emery-Le, and P. Emery. 2007. A subset of dorsal neurons modulates circadian behavior and light responses in Drosophila. *Neuron.* 53:689-701.

Murray, G., N.B. Allen, and J. Trinder. 2003. Seasonality and circadian phase delay: prospective evidence that winter lowering of mood is associated with a shift towards Eveningness. *J.Affect.Disord.* 76:15-22.

Myers, E.M., J. Yu, and A. Sehgal. 2003. Circadian control of eclosion: interaction between a central and peripheral clock in Drosophila melanogaster. *Curr.Biol.* 13:526-533.

Myers, M.P., K. Wager-Smith, C.S. Wesley, M.W. Young, and A. Sehgal. 1995. Positional cloning and sequence analysis of the Drosophila clock gene, timeless. *Science.* 270:805-808.

Nagoshi, E., K. Sugino, E. Kula, E. Okazaki, T. Tachibana, S. Nelson, and M. Rosbash. 2010. Dissecting differential gene expression within the circadian neuronal circuit of Drosophila. *Nat.Neurosci.* 13:60-68.

Naidoo, N., V. Casiano, J. Cater, J. Zimmerman, and A.I. Pack. 2007. A role for the molecular chaperone protein BiP/GRP78 in Drosophila sleep homeostasis. *Sleep.* 30:557-565.

Naidoo, N., W. Giang, R.J. Galante, and A.I. Pack. 2005. Sleep deprivation induces the unfolded protein response in mouse cerebral cortex. *J.Neurochem.* 92:1150-1157.

Naidoo, N., W. Song, M. Hunter-Ensor, and A. Sehgal. 1999. A role for the proteasome in the light response of the timeless clock protein. *Science.* 285:1737-1741.

Nawathean, P., and M. Rosbash. 2004. The doubletime and CKII kinases collaborate to potentiate Drosophila PER transcriptional repressor activity. *Mol.Cell.* 13:213-223.

Newby, L.M., and F.R. Jackson. 1993. A new biological rhythm mutant of Drosophila melanogaster that identifies a gene with an essential embryonic function. *Genetics.* 135:1077-1090.

Newby, L.M., and F.R. Jackson. 1991. Drosophila ebony mutants have altered circadian activity rhythms but normal eclosion rhythms. *J.Neurogenet.* 7:85-101.

Nichols, C.D. 2007. 5-HT2 receptors in Drosophila are expressed in the brain and modulate aspects of circadian behaviors. *Dev.Neurobiol.* 67:752-763.

Nitabach, M.N., J. Blau, and T.C. Holmes. 2002. Electrical silencing of Drosophila pacemaker neurons stops the free-running circadian clock. *Cell.* 109:485-495.

Nitabach, M.N., V. Sheeba, D.A. Vera, J. Blau, and T.C. Holmes. 2005. Membrane electrical excitability is necessary for the free-running larval Drosophila circadian clock. *J.Neurobiol.* 62:1-13.

Nitabach, M.N., and P.H. Taghert. 2008. Organization of the Drosophila circadian control circuit. *Curr.Biol.* 18:R84-93.

Nitabach, M.N., Y. Wu, V. Sheeba, W.C. Lemon, J. Strumbos, P.K. Zelensky, B.H. White, and T.C. Holmes. 2006. Electrical hyperexcitation of lateral ventral pacemaker neurons desynchronizes downstream circadian oscillators in the fly circadian circuit and induces multiple behavioral periods. *J.Neurosci.* 26:479-489.

Nitz, D.A., B. van Swinderen, G. Tononi, and R.J. Greenspan. 2002. Electrophysiological correlates of rest and activity in Drosophila melanogaster. *Curr.Biol.* 12:1934-1940.

Noyes, B.E., F.N. Katz, and M.H. Schaffer. 1995. Identification and expression of the Drosophila adipokinetic hormone gene. *Mol.Cell.Endocrinol.* 109:133-141.

Ouyang, Y., C.R. Andersson, T. Kondo, S.S. Golden, and C.H. Johnson. 1998. Resonating circadian clocks enhance fitness in cyanobacteria. *Proc.Natl.Acad.Sci.U.S.A.* 95:8660-8664.

Pallier, P.N., E.S. Maywood, Z. Zheng, J.E. Chesham, A.N. Inyushkin, R. Dyball, M.H. Hastings, and A.J. Morton. 2007. Pharmacological imposition of sleep slows cognitive decline and reverses dysregulation of circadian gene expression in a transgenic mouse model of Huntington's disease. *J.Neurosci.* 27:7869-7878.

Papatsenko, D., G. Sheng, and C. Desplan. 1997. A new rhodopsin in R8 photoreceptors of Drosophila: evidence for coordinate expression with Rh3 in R7 cells. *Development.* 124:1665-1673.

Paranjpe, D.A., D. Anitha, A. Joshi, and V.K. Sharma. 2004. Multi-oscillatory control of eclosion and oviposition rhythms in Drosophila melanogaster: evidence from limits of entrainment studies. *Chronobiol.Int.* 21:539-552.

Parisky, K.M., J. Agosto, S.R. Pulver, Y. Shang, E. Kuklin, J.J. Hodge, K. Kang, X. Liu, P.A. Garrity, M. Rosbash, and L.C. Griffith. 2008. PDF cells are a GABA-responsive wake-promoting component of the Drosophila sleep circuit. *Neuron.* 60:672-682.

Park, J.H., and J.C. Hall. 1998. Isolation and chronobiological analysis of a neuropeptide pigment-dispersing factor gene in Drosophila melanogaster. *J.Biol.Rhythms.* 13:219-228.

Park, J.H., C. Helfrich-Forster, G. Lee, L. Liu, M. Rosbash, and J.C. Hall. 2000. Differential regulation of circadian pacemaker output by separate clock genes in Drosophila. *Proc.Natl.Acad.Sci.U.S.A.* 97:3608-3613.

Park, J.H., A.J. Schroeder, C. Helfrich-Forster, F.R. Jackson, and J. Ewer. 2003. Targeted ablation of CCAP neuropeptide-containing neurons of Drosophila causes specific defects in execution and circadian timing of ecdysis behavior. *Development.* 130:2645-2656.

Parks, A.L., K.R. Cook, M. Belvin, N.A. Dompe, R. Fawcett, K. Huppert, L.R. Tan, C.G. Winter, K.P. Bogart, J.E. Deal, M.E. Deal-Herr, D. Grant, M. Marcinko, W.Y. Miyazaki, S. Robertson, K.J. Shaw, M. Tabios, V. Vysotskaia, L. Zhao, R.S. Andrade, K.A. Edgar, E. Howie, K. Killpack, B. Milash, A. Norton, D. Thao, K. Whittaker, M.A. Winner, L. Friedman, J. Margolis, M.A. Singer, C. Kopczynski, D. Curtis, T.C. Kaufman, G.D. Plowman, G. Duyk, and H.L. Francis-Lang. 2004. Systematic generation of high-resolution deletion coverage of the Drosophila melanogaster genome. *Nat.Genet.* 36:288-292.

Pavlidis, T. 1971. Populations of biochemical oscillators as circadian clocks. *J.Theor.Biol.* 33:319-338.

Pegoraro, M., and E. Tauber. 2008. The role of microRNAs (miRNA) in circadian rhythmicity. *J.Genet.* 87:505-511.

Peschel, N., K.F. Chen, G. Szabo, and R. Stanewsky. 2009. Light-Dependent Interactions between the Drosophila Circadian Clock Factors Cryptochrome, Jetlag, and Timeless. *Curr.Biol.* 19:241-247.

Peschel, N., S. Veleri, and R. Stanewsky. 2006. Veela defines a molecular link between Cryptochrome and Timeless in the light-input pathway to Drosophila's circadian clock. *Proc.Natl.Acad.Sci.U.S.A.* 103:17313-17318.

Petri, B., and M. Stengl. 1997. Pigment-dispersing hormone shifts the phase of the circadian pacemaker of the cockroach Leucophaea maderae. *J.Neurosci.* 17:4087-4093.

Pieron, H. 1913. Le Probleme Physiologique Du Sommeil. Masson, Paris.

Pitman, J.L., J.J. McGill, K.P. Keegan, and R. Allada. 2006. A dynamic role for the mushroom bodies in promoting sleep in Drosophila. *Nature.* 441:753-756.

Pittendrigh, C.S. 1965. Biological clocks, the functions, ancient and modern, of biological oscillations. *Science in the Sixties, Proceedings of the 1965 Cloudcroft Symposium.*:96-111.

Pittendrigh, C.S., and S. Daan. 1976. A functional analysis of circadian pacemakers in nocturnal rodents. V. Pacemaker structure: A clock for all seasons. *J Comp Physiol.* 106:333-55.

Power, J.M., J.M. Ringo, and H.B. Dowse. 1995. The effects of period mutations and light on the activity rhythms of Drosophila melanogaster. *J.Biol.Rhythms.* 10:267-280.

Price, J.L., J. Blau, A. Rothenfluh, M. Abodeely, B. Kloss, and M.W. Young. 1998. double-time is a novel Drosophila clock gene that regulates PERIOD protein accumulation. *Cell.* 94:83-95.

Price, T.S., J.E. Baggs, A.M. Curtis, G.A. Fitzgerald, and J.B. Hogenesch. 2008. WAVECLOCK: wavelet analysis of circadian oscillation. *Bioinformatics.* 24:2794-2795.

Raizen, D.M., J.E. Zimmerman, M.H. Maycock, U.D. Ta, Y.J. You, M.V. Sundaram, and A.I. Pack. 2008. Lethargus is a Caenorhabditis elegans sleep-like state. *Nature.* 451:569-572.

Rao, K.R., J.P. Riehm, C.A. Zahnow, L.H. Kleinholz, G.E. Tarr, L. Johnson, S. Norton, M. Landau, O.J. Semmes, R.M. Sattelberg, W.H. Jorenby, and M.F. Hintz. 1985. Characterization of a pigment-dispersing hormone in eyestalks of the fiddler crab Uca pugilator. *Proc.Natl.Acad.Sci.U.S.A.* 82:5319-5322.

Rattenborg, N.C., B.H. Mandt, W.H. Obermeyer, P.J. Winsauer, R. Huber, M. Wikelski, and R.M. Benca. 2004. Migratory sleeplessness in the white-crowned sparrow (Zonotrichia leucophrys gambelii). *PLoS Biol.* 2:E212.

Ray, V.M., and H.B. Dowse. 2005. Mutations in and deletions of the Ca2+ channel-encoding gene cacophony, which affect courtship song in Drosophila, have novel effects on heartbeating. *J.Neurogenet.* 19:39-56.

Rechtschaffen, A., M.A. Gilliland, B.M. Bergmann, and J.B. Winter. 1983. Physiological correlates of prolonged sleep deprivation in rats. *Science.* 221:182-184.

Reddy, A.B., N.A. Karp, E.S. Maywood, E.A. Sage, M. Deery, J.S. O'Neill, G.K. Wong, J. Chesham, M. Odell, K.S. Lilley, C.P. Kyriacou, and M.H. Hastings. 2006. Circadian orchestration of the hepatic proteome. *Curr.Biol.* 16:1107-1115.

Reimund, E. 1994. The free radical flux theory of sleep. *Med.Hypotheses.* 43:231-233.

Renn, S.C., J.H. Park, M. Rosbash, J.C. Hall, and P.H. Taghert. 1999. A pdf neuropeptide gene mutation and ablation of PDF neurons each cause severe abnormalities of behavioral circadian rhythms in Drosophila. *Cell.* 99:791-802.

Ricker, J.P., and J. Hirsch. 1985. Evolution of an instinct under long-term divergent selection for geotaxis in domesticated populations of Drosophila melanogaster. *J.Comp.Psychol.* 99:380-390.

Rieger, D., O.T. Shafer, K. Tomioka, and C. Helfrich-Forster. 2006. Functional analysis of circadian pacemaker neurons in Drosophila melanogaster. *J.Neurosci.* 26:2531-2543.

Rieger, D., R. Stanewsky, and C. Helfrich-Forster. 2003. Cryptochrome, compound eyes, Hofbauer-Buchner eyelets, and ocelli play different roles in the entrainment and masking pathway of the locomotor activity rhythm in the fruit fly Drosophila melanogaster. *J.Biol.Rhythms.* 18:377-391.

Ripperger, J.A., and U. Schibler. 2006. Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions. *Nat.Genet.* 38:369-374.

Roberts, D.H., J. Lehar, and J.W. Dreher. 1987. Time series with CLEAN. Deviation of a spectrum. *J. Astronomy.* 93:968-989.

Rockman, M.V. 2008. Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature.* 456:738-744.

Roenneberg, T., S. Daan, and M. Merrow. 2003. The art of entrainment. *J.Biol.Rhythms.* 18:183-194.

Rosato, E., and C.P. Kyriacou. 2006. Analysis of locomotor activity rhythms in Drosophila. *Nat.Protocols.* 1:559-568.

Rosbash, M. 2009. The implications of multiple circadian clock origins. *PLoS Biol.* 7:e62.

Rosekind, M.R. 2005. Underestimating the societal costs of impaired alertness: safety, health and productivity risks. *Sleep Med.* 6 Suppl 1:S21-5.

Roth, T. 2005. Prevalence, associated risks, and treatment patterns of insomnia. *J.Clin.Psychiatry.* 66 Suppl 9:10-3; quiz 42-3.

Rutila, J.E., O. Maltseva, and M. Rosbash. 1998a. The timSL mutant affects a restricted portion of the Drosophila melanogaster circadian cycle. *J.Biol.Rhythms.* 13:380-392.

Rutila, J.E., V. Suri, M. Le, W.V. So, M. Rosbash, and J.C. Hall. 1998b. CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of Drosophila period and timeless. *Cell.* 93:805-814.

Saifullah, A.S., and T.L. Page. 2009. Circadian regulation of olfactory receptor neurons in the cockroach antenna. *J.Biol.Rhythms.* 24:144-152.

Sandrelli, F., E. Tauber, M. Pegoraro, G. Mazzotta, P. Cisotto, J. Landskron, R. Stanewsky, A. Piccin, E. Rosato, M. Zordan, R. Costa, and C.P. Kyriacou. 2007. A molecular basis for natural selection at the timeless locus in Drosophila melanogaster. *Science.* 316:1898-1900.

Sarov-Blat, L., W.V. So, L. Liu, and M. Rosbash. 2000. The Drosophila takeout gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell.* 101:647-656.

Sathyanarayanan, S., X. Zheng, S. Kumar, C.H. Chen, D. Chen, B. Hay, and A. Sehgal. 2008. Identification of novel genes involved in light-dependent CRY degradation through a genome-wide RNAi screen. *Genes Dev.* 22:1522-1533.

Sathyanarayanan, S., X. Zheng, R. Xiao, and A. Sehgal. 2004. Posttranslational regulation of Drosophila PERIOD protein by protein phosphatase 2A. *Cell.* 116:603-615.

Sawyer, L.A., J.M. Hennessy, A.A. Peixoto, E. Rosato, H. Parkinson, R. Costa, and C.P. Kyriacou. 1997. Natural variation in a Drosophila clock gene and temperature compensation. *Science.* 278:2117-2120.

Sehadova, H., F.T. Glaser, C. Gentile, A. Simoni, A. Giesecke, J.T. Albert, and R. Stanewsky. 2009. Temperature entrainment of Drosophila's circadian clock involves the gene nocte and signaling from peripheral sensory tissues to the brain. *Neuron.* 64:251-266.

Seugnet, L., J. Boero, L. Gottschalk, S.P. Duntley, and P.J. Shaw. 2006. Identification of a biomarker for sleep drive in flies and humans. *Proc.Natl.Acad.Sci.U.S.A.* 103:19913-19918.

Seugnet, L., Y. Suzuki, L. Vine, L. Gottschalk, and P.J. Shaw. 2008. D1 receptor activation in the mushroom bodies rescues sleep-loss-induced learning impairments in Drosophila. *Curr.Biol.* 18:1110-1117.

Shafer, O.T., D.J. Kim, R. Dunbar-Yaffe, V.O. Nikolaev, M.J. Lohse, and P.H. Taghert. 2008. Widespread receptivity to neuropeptide PDF throughout the neuronal circadian clock network of Drosophila revealed by real-time cyclic AMP imaging. *Neuron.* 58:223-237.

Shafer, O.T., M. Rosbash, and J.W. Truman. 2002. Sequential nuclear accumulation of the clock proteins period and timeless in the pacemaker neurons of Drosophila melanogaster. *J.Neurosci.* 22:5946-5954.

Shang, Y., L.C. Griffith, and M. Rosbash. 2008. Light-arousal and circadian photoreception circuits intersect at the large PDF cells of the Drosophila brain. *Proc.Natl.Acad.Sci.U.S.A.* 105:19587-19594.

Shaw, P.J., C. Cirelli, R.J. Greenspan, and G. Tononi. 2000. Correlates of sleep and waking in Drosophila melanogaster. *Science.* 287:1834-1837.

Shaw, P.J., G. Tononi, R.J. Greenspan, and D.F. Robinson. 2002. Stress response genes protect against lethal effects of sleep deprivation in Drosophila. *Nature.* 417:287-291.

Sheeba, V., K.J. Fogle, M. Kaneko, S. Rashid, Y.T. Chou, V.K. Sharma, and T.C. Holmes. 2008a. Large ventral lateral neurons modulate arousal and sleep in Drosophila. *Curr.Biol.* 18:1537-1545.

Sheeba, V., H. Gu, V.K. Sharma, D.K. O'Dowd, and T.C. Holmes. 2008b. Circadian- and light-dependent regulation of resting membrane potential and spontaneous action potential firing of Drosophila circadian pacemaker neurons. *J.Neurophysiol.* 99:976-988.

Sheldon, B.L., and M.K. Milton. 1972. Studies on the scutellar bristles of Drosophila melanogaster. II. Long-term selection for high bristle number in the Oregon RC strain and correlated responses in abdominal chaetae. *Genetics.* 71:567-595.

Shimizu, H., M. Shimoda, T. Yamaguchi, K.H. Seong, T. Okamura, and S. Ishii. 2008. Drosophila ATF-2 regulates sleep and locomotor activity in pacemaker neurons. *Mol.Cell.Biol.* 28:6278-6289.

So, W.V., L. Sarov-Blat, C.K. Kotarski, M.J. McDonald, R. Allada, and M. Rosbash. 2000. takeout, a novel Drosophila gene under circadian clock transcriptional regulation. *Mol.Cell.Biol.* 20:6935-6944.

Sofola, O., V. Sundram, F. Ng, Y. Kleyner, J. Morales, J. Botas, F.R. Jackson, and D.L. Nelson. 2008. The Drosophila FMRP and LARK RNA-binding proteins function together to regulate eye development and circadian behavior. *J.Neurosci.* 28:10200-10205.

Sokolowski, M.B. 2001. Drosophila: genetics meets behaviour. *Nat.Rev.Genet.* 2:879-890.

Spaggiari, M.C. 2008. Sleep medicine in occupational health. *G.Ital.Med.Lav.Ergon.* 30:276-279.

Stanewsky, R., M. Kaneko, P. Emery, B. Beretta, K. Wager-Smith, S.A. Kay, M. Rosbash, and J.C. Hall. 1998. The cryb mutation identifies cryptochrome as a circadian photoreceptor in Drosophila. *Cell.* 95:681-692.

Stark, C., B.J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, and M. Tyers. 2006. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34:D535-9.

Stelling, J., U. Sauer, Z. Szallasi, F.J. Doyle 3rd, and J. Doyle. 2004. Robustness of cellular functions. *Cell.* 118:675-685.

Stoleru, D., P. Nawathean, M.P. Fernandez, J.S. Menet, M.F. Ceriani, and M. Rosbash. 2007. The Drosophila circadian network is a seasonal timer. *Cell.* 129:207-219.

Stoleru, D., Y. Peng, J. Agosto, and M. Rosbash. 2004. Coupled oscillators control morning and evening locomotor behaviour of Drosophila. *Nature.* 431:862-868.

Stoleru, D., Y. Peng, P. Nawathean, and M. Rosbash. 2005. A resetting signal between Drosophila pacemakers synchronizes morning and evening activity. *Nature.* 438:238-242.

Suh, J., and F.R. Jackson. 2007. Drosophila ebony activity is required in glia for the circadian regulation of locomotor activity. *Neuron.* 55:435-447.

Suri, V., Z. Qian, J.C. Hall, and M. Rosbash. 1998. Evidence that the TIM light response is relevant to light-induced phase shifts in Drosophila melanogaster. *Neuron.* 21:225-234.

Tartar, J.L., C.P. Ward, J.T. McKenna, M. Thakkar, E. Arrigoni, R.W. McCarley, R.E. Brown, and R.E. Strecker. 2006. Hippocampal synaptic plasticity and spatial learning are impaired in a rat model of sleep fragmentation. *Eur.J.Neurosci.* 23:2739-2748.

Tauber, E., K.S. Last, P.J. Olive, and C.P. Kyriacou. 2004. Clock gene evolution and functional divergence. *J.Biol.Rhythms.* 19:445-458.

Tauber, E., M. Zordan, F. Sandrelli, M. Pegoraro, N. Osterwalder, C. Breda, A. Daga, A. Selmin, K. Monger, C. Benna, E. Rosato, C.P. Kyriacou, and R. Costa. 2007. Natural selection favors a newly derived timeless allele in Drosophila melanogaster. *Science.* 316:1895-1898.

Taylor, P., and P.E. Hardin. 2008. Rhythmic E-box binding by CLK-CYC controls daily cycles in per and tim transcription and chromatin modifications. *Mol.Cell.Biol.* 28:4642-4652.

Toma, D.P., K.P. White, J. Hirsch, and R.J. Greenspan. 2002. Identification of genes involved in Drosophila melanogaster geotaxis, a complex behavioral trait. *Nat.Genet.* 31:349-353.

Tomioka, K., K. Yamada, S. Yokoyama, and Y. Chiba. 1991. Mutual interactions between optic lobe circadian pacemakers in the cricket *Gryllus bimaculatus*. *J Comp Physiol.* 169:291-298.

Tononi, G., and C. Cirelli. 2006. Sleep function and synaptic homeostasis. *Sleep Med.Rev.* 10:49-62.

Torrence, C., and G.P. Compo. 1998. A practical guide to wavelet analysis. *Bull. Amer. Meteorol. Soc.* 79:61-78.

Tully, T. 1996. Discovery of genes involved with learning and memory: an experimental synthesis of Hirschian and Benzerian perspectives. *Proc.Natl.Acad.Sci.U.S.A.* 93:13460-13467.

Turrigiano, G.G., and S.B. Nelson. 2000. Hebb and homeostasis in neuronal plasticity. *Curr.Opin.Neurobiol.* 10:358-364.

Ueda, H.R., S. Hayashi, W. Chen, M. Sano, M. Machida, Y. Shigeyoshi, M. Iino, and S. Hashimoto. 2005. System-level identification of transcriptional circuits underlying mammalian circadian clocks. *Nat.Genet.* 37:187-192.

Ueda, H.R., A. Matsumoto, M. Kawamura, M. Iino, T. Tanimura, and S. Hashimoto. 2002. Genome-wide transcriptional orchestration of circadian rhythms in Drosophila. *J.Biol.Chem.* 277:14048-14052.

Van der Horst, D.J., W.J. Van Marrewijk, and J.H. Diederen. 2001. Adipokinetic hormones of insect: release, signal transduction, and responses. *Int.Rev.Cytol.* 211:179-240.

van Swinderen, B., and K.A. Flores. 2006. Attention-like processes underlying optomotor performance in a Drosophila choice maze. *J.Neurobiol.* 67:129-145.

van Swinderen, B., and R.J. Greenspan. 2005. Flexibility in a gene network affecting a simple behavior in Drosophila melanogaster. *Genetics.* 169:2151-2163.

Veleri, S., C. Brandes, C. Helfrich-Forster, J.C. Hall, and R. Stanewsky. 2003. A self-sustaining, light-entrainable circadian oscillator in the Drosophila brain. *Curr.Biol.* 13:1758-1767.

Venken, K.J., and H.J. Bellen. 2007. Transgenesis upgrades for Drosophila melanogaster. *Development.* 134:3571-3584.

Venken, K.J., and H.J. Bellen. 2005. Emerging technologies for gene manipulation in Drosophila melanogaster. *Nat.Rev.Genet.* 6:167-178.

Vosshall, L.B., J.L. Price, A. Sehgal, L. Saez, and M.W. Young. 1994. Block in nuclear localization of period protein by a second clock mutation, timeless. *Science.* 263:1606-1609.

Vyazovskiy, V.V., C. Cirelli, M. Pfister-Genskow, U. Faraguna, and G. Tononi. 2008. Molecular and electrophysiological evidence for net synaptic potentiation in wake and depression in sleep. *Nat.Neurosci.* 11:200-208.

Wang, H., R. Kingsland, H. Zhao, Y. Wang, W. Pan, X. Dong, J. Guo, and F. Huang. 1995. Time of symptom onset of eight common medical emergencies. *J.Emerg.Med.* 13:461-469.

Waterhouse, J., B. Edwards, A. Nevill, S. Carvalho, G. Atkinson, P. Buckley, T. Reilly, R. Godfrey, and R. Ramsay. 2002. Identifying some determinants of "jet lag" and its symptoms: a study of athletes and other travellers. *Br.J.Sports Med.* 36:54-60.

Wayne, M.L., M. Telonis-Scott, L.M. Bono, L. Harshman, A. Kopp, S.V. Nuzhdin, and L.M. McIntyre. 2007. Simpler mode of inheritance of transcriptional variation in male Drosophila melanogaster. *Proc.Natl.Acad.Sci.U.S.A.* 104:18577-18582.

Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* 447:661-678.

Wheeler, D.A., M.J. Hamblen-Coyle, M.S. Dushay, and J.C. Hall. 1993. Behavior in light-dark cycles of Drosophila mutants that are arrhythmic, blind, or both. *J.Biol.Rhythms.* 8:67-94.

Wijnen, H., F. Naef, C. Boothroyd, A. Claridge-Chang, and M.W. Young. 2006. Control of daily transcript oscillations in Drosophila by light and the circadian clock. *PLoS Genet.* 2:e39.

Wilcockson, D., and L. Zhang. 2008. Circatidal clocks. *Curr.Biol.* 18:R753-R755.

Winfree, A.T. 1974. Suppressing drosophila circadian rhythm with dim light. *Science.* 183:970-972.

Woelfle, M.A., Y. Ouyang, K. Phanvijhitsiri, and C.H. Johnson. 2004. The adaptive value of circadian clocks: an experimental assessment in cyanobacteria. *Curr.Biol.* 14:1481-1486.

Wu, M.N., K. Koh, Z. Yue, W.J. Joiner, and A. Sehgal. 2008a. A genetic screen for sleep and circadian mutants reveals mechanisms underlying regulation of sleep in Drosophila. *Sleep.* 31:465-472.

Wu, Y., G. Cao, and M.N. Nitabach. 2008b. Electrical silencing of PDF neurons advances the phase of non-PDF clock neurons in Drosophila. *J.Biol.Rhythms.* 23:117-128.

Yang, M., J.E. Lee, R.W. Padgett, and I. Edery. 2008. Circadian regulation of a limited set of conserved microRNAs in Drosophila. *BMC Genomics.* 9:83.

Yang, Z., and A. Sehgal. 2001. Role of molecular oscillations in generating behavioral rhythms in Drosophila. *Neuron.* 29:453-467.

Yoshii, T., K. Fujii, and K. Tomioka. 2007. Induction of Drosophila behavioral and molecular circadian rhythms by temperature steps in constant light. *J.Biol.Rhythms.* 22:103-114.

Yoshii, T., Y. Funada, T. Ibuki-Ishibashi, A. Matsumoto, T. Tanimura, and K. Tomioka. 2004. Drosophila cryb mutation reveals two circadian clocks that drive locomotor rhythm and have different responsiveness to light. *J.Insect Physiol.* 50:479-488.

Yoshii, T., Y. Heshiki, T. Ibuki-Ishibashi, A. Matsumoto, T. Tanimura, and K. Tomioka. 2005. Temperature cycles drive Drosophila circadian oscillation in constant light that otherwise induces behavioural arrhythmicity. *Eur.J.Neurosci.* 22:1176-1184.

Yoshii, T., M. Sakamoto, and K. Tomioka. 2002. A temperature-dependent timing mechanism is involved in the circadian system that drives locomotor rhythms in the fruit fly Drosophila melanogaster. *Zoolog Sci.* 19:841-850.

Young, M.W., and B.H. Judd. 1978. Nonessential Sequences, Genes, and the Polytene Chromosome Bands of DROSOPHILA MELANOGASTER. *Genetics.* 88:723-742.

Yu, W., H. Zheng, J.H. Houl, B. Dauwalder, and P.E. Hardin. 2006. PER-dependent rhythms in CLK phosphorylation and E-box binding regulate circadian transcription. *Genes Dev.* 20:723-733.

Yu, W., H. Zheng, J.L. Price, and P.E. Hardin. 2009. DOUBLETIME plays a non-catalytic role to mediate CLOCK phosphorylation and repress CLOCK-dependent transcription within the Drosophila circadian clock. *Mol.Cell.Biol.*

Yuan, Q., W.J. Joiner, and A. Sehgal. 2006. A sleep-promoting role for the Drosophila serotonin receptor 1A. *Curr.Biol.* 16:1051-1062.

Yuan, Q., F. Lin, X. Zheng, and A. Sehgal. 2005. Serotonin modulates circadian entrainment in Drosophila. *Neuron.* 47:115-127.

Zeeberg, B.R., J. Riss, D.W. Kane, K.J. Bussey, E. Uchio, W.M. Linehan, J.C. Barrett, and J.N. Weinstein. 2004. Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics. *BMC Bioinformatics.* 5:80.

Zeng, H., Z. Qian, M.P. Myers, and M. Rosbash. 1996. A light-entrainment mechanism for the Drosophila circadian clock. *Nature.* 380:129-135.

Zepelin, H., and A. Rechtschaffen. 1974. Mammalian sleep, longevity, and energy metabolism. *Brain Behav.Evol.* 10:425-470.

Zhao, J., V.L. Kilman, K.P. Keegan, Y. Peng, P. Emery, M. Rosbash, and R. Allada. 2003. Drosophila clock can generate ectopic circadian clocks. *Cell.* 113:755-766.

Zhao, W., E. Serpedin, and E.R. Dougherty. 2009. Identifying genes involved in cyclic processes by combining gene expression analysis and prior knowledge. *EURASIP J.Bioinform Syst.Biol.2009:683463 Epub.*:683463.

Zheng, X., K. Koh, M. Sowcik, C.J. Smith, D. Chen, M.N. Wu, and A. Sehgal. 2009. An isoform-specific mutant reveals a role of PDP1 epsilon in the circadian oscillator. *J.Neurosci.* 29:10920-10927.

Zhu, J., P.Y. Lum, J. Lamb, D. GuhaThakurta, S.W. Edwards, R. Thieringer, J.P. Berger, M.S. Wu, J. Thompson, A.B. Sachs, and E.E. Schadt. 2004. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenet.Genome Res.* 105:363-374.

Zimmerman, J.E., D.M. Raizen, M.H. Maycock, G. Maislin, and A.I. Pack. 2008. A video method to study Drosophila sleep. *Sleep.* 31:1587-1598.

Zimmerman, J.E., W. Rizzo, K.R. Shockley, D.M. Raizen, N. Naidoo, M. Mackiewicz, G.A. Churchill, and A.I. Pack. 2006. Multiple mechanisms limit the duration of wakefulness in Drosophila brain. *Physiol.Genomics.* 27:337-350.

Zimmerman, W.F., C.S. Pittendrigh, and T. Pavlidis. 1968. Temperature compensation of the circadian oscillation in drosophila pseudoobscura and its entrainment by temperature cycles. *J.Insect Physiol.* 14:669-684.

Zordan, M.A., C. Benna, and G. Mazzotta. 2007. Monitoring and analyzing Drosophila circadian locomotor activity. *Methods Mol.Biol.* 362:67-81.