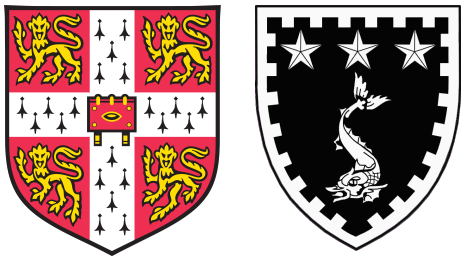

University of Cambridge

Murray Edwards College

MODELS OF NEURODEGENERATION USING COMPUTATIONAL APPROACHES

by

ELEONORA KHABIROVA

This dissertation is submitted for the degree of

Doctor of Philosophy

July 2016

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University of similar institution except as declared in the Preface and specified in the text.

Посвящается самой лучшей маме на свете,
а также всем моим мишкам

To all the fallen flies that sacrifice
their lives in the name of Science

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. Damian Crowther and Dr. Florian Hollfelder for the endless support and always inspiring discussions. In particular, I wish to thank sincerely Dr. Crowther for reading my thesis even during the holiday season and always helping me every step of the way.

As most of the experimental work was a result of the collaborative effort from multiple departments, I would like first to thank Giorgio Favrin from Biochemistry department for helpful discussions and ideas. I would like to thank Kofan Chen for his patience and willingness to explain me the concepts of circadian biology and showing me how to perform the experiments. Also I would like to thank our collaborator in LMB John O'Neill for lending equipment and his helpful suggestions.

SUMMARY

Alzheimer's disease (AD), as one of the most common neurodegenerative diseases, is characterized by the loss of neuronal dysfunction and death resulting in progressive cognitive impairment. The main histopathological hallmark of AD is the accumulation and deposition of misfolded Aβ peptide as amyloid plaques, however the precise role of Aβ toxicity in the disease pathogenesis is still unclear. Moreover, at early stages of the disease the important clinical features of the disorder, in addition to memory loss, are the disruptions of circadian rhythms and spatial disorientation.

In the present work I first studied the role of Aβ toxicity by comparing the findings of genome-wide association studies in sporadic AD with the results of an RNAi screen in a transgenic *C. elegans* model of Aβ toxicity. The initial finding was that none of the human orthologues of these worm genes are associated with risk for sporadic Alzheimer's disease, indicating that Aβ toxicity in the worm model may not be equivalent to sporadic AD. Nevertheless, comparing the first degree physical interactors (+1 interactome) of the GWAS and worm screen-derived gene products have uncovered 4 worm genes that have a +1 interactome overlap with the GWAS genes that is larger than one would expect by chance. Three of these genes form a chaperonin complex and the fourth gene codes for actin, a major substrate of the same chaperonin.

Next I have evaluated the circadian disruptions in AD by developing a new system to simultaneously monitor the oscillations of the peripheral molecular clock and behavioural rhythms in single *Drosophila*. Experiments were undertaken on wild-type and Aβ-expressing flies. The results indicate the robustness of the peripheral clock is not correlated with the robustness of the circadian sleep and locomotor behaviours, indicating that the molecular clock does not directly drive behaviour. This is despite period length correlations that indicate that the underlying molecular mechanisms that generate both molecular and behavioural rhythms are the same. Rhythmicity in Aβ-expressing flies is worse than in controls. I further investigated the mechanism of spatial orientation in *Drosophila*. It was established that in the absence of visual stimuli the flies use self-motion cues to orientate themselves within the tubes and that in a *Drosophila* model of Aβ toxicity this control function is disrupted.

LIST OF PAPERS

Material in this thesis is published or being prepared for publication in the following papers:

1. "**The TRiC/CCT Chaperone Is Implicated in Alzheimer's Disease Based on Patient GWAS and an RNAi Screen in Aβ-Expressing *Caenorhabditis elegans***" Khabirova E, Moloney A, Marciniak SJ, Williams J, Lomas DA, et al. (2014) *PLoS ONE* 9(7): e102985. doi: 10.1371/journal.pone.0102985

2. **"Simultaneous Monitoring of Behavior & Peripheral Clocks in *Drosophila* Reveals Unstructured Sleep in an Alzheimer's Model"**
   Eleonora Khabirova, Ko-Fan Chen, John S O'Neill, Damian C Crowther (2015). *bioRxiv,* doi: http://dx.doi.org/10.1101/013730. / epub

CONTENT

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Neurodegenerative disease is a general term for a range of disorders related to the progressive loss of function or structure of neurons in the human brain resulting in cognitive decline. They are a major threat to public health, with Alzheimer's disease (AD) being the most common disorder in adults. Other examples of such disorders include Parkinson's and Huntington's diseases (Forman, Trojanowski et al. 2004, Chiti and Dobson 2006). AD is thought to be caused by the self-association of at least two proteins, the amyloid beta peptide (Aβ) and tau, resulting in a spectrum of aggregates that may have features of amyloid. Furthermore, the aggregation of misfolded proteins characterizes other neurodegenerative diseases, meaning that they might have similar mechanism of pathogenesis. This chapter will give an overview of neurodegenerative diseases, focusing on Alzheimer's disease and amyloid beta toxicity.

1.1 NEURODEGENERATIVE DISEASES

Neurodegenerative diseases can be hereditary or sporadic, or in the case of prion disease they may rarely be infectious. They are characterized by progressive nervous system dysfunction, often associated with neuronal loss in the central nervous system, caused by accumulation of intracellular or extracellular protein aggregates (Forman, Trojanowski et al. 2004).

1.1.1 MECHANISMS OF NEURODEGENERATION

Recent studies of familial forms of these neurodegenerative diseases have led to the development of model systems, providing new insights into disease pathogenesis. One of the conclusions of such studies was that the misfolding, in case of neurodegenerative diseases, is caused by genetic mutations, most of which are located in completely unrelated genes. In many cases, these mutations lead to the overproduction of misfolded proteins, which have high aggregation propensity resulting in the deposition of protein aggregates in extra- and intracellular space (Forman, Trojanowski et al. 2004). For example, AD is characterized by the

extracellular aggregation and deposition of amyloid-beta (Aβ) peptide, particularly isoforms that are longer than normal - for example the $A\beta_{42}$ isoform, with a length of 42 amino acids, is more aggregation prone than the $A\beta_{40}$ isoform that is 40 amino acids residues in length. Similarly, aggregation prone variants of alpha-synuclein are believed to be linked to Parkinson's disease (PD) pathogenesis, while mutations in microtubule-associated protein tau are thought to be a cause of frontotemporal dementia (Soto 2003). Mutations identified in the familial cases can also provide insights into the pathogenesis of sporadic forms as the aggregates of the same proteins are found in the majority of the cases. In particular, all mutations in amyloid precursor protein, presenilin-1 and presenilin-2, as identified in the cases of familiar form of AD, lead to the overproduction of aggregation-prone isoforms of Aβ peptides, resulting in plaque deposition, which is the major pathological hallmark of the disease in both familial and sporadic forms (Weggen and Beher 2012). In the same way, mutations in the alpha-synuclein gene SNCA cause the accumulation of alpha-synuclein proteins in the brain, a major neuropathological feature of familial and sporadic PD (Bertram and Tanzi 2005). Therefore, the study of these mutations can lead to further understanding of how and why some normally soluble proteins convert into insoluble, fibrillar aggregates.

Familial forms of neurodegenerative diseases only account to the small percentage of total number of cases, specifically 5-10% cases of AD, ~10% of PD (Bertram and Tanzi 2005). In contrast, the vast majority of the cases occur as sporadic forms, most likely resulting from a combination of genetic and environmental factors that are not yet fully understood (Tillement, Lecanu et al. 2010).

As mentioned before, the deposits of protein aggregates in the nervous system are a characteristic of most adult-onset neurodegenerative diseases and therefore this suggests that they share a common pathogenic mechanism. One explanation of this phenomenon could be age-related degradation of proteostasis network, a protein quality control system, leading to the progressive accumulation of protein damage and formation of misfolded protein aggregates (Douglas and Dillin 2010). Furthermore, it has been suggested that accumulation of misfolded proteins associated with many of these diseases may be age-related and not necessarily a disease-related process (Lindner and Demarez 2009). Aging is thought to be a stochastic process combining the genetic and environmental factors and leading to the progressive physiological

decline (Hindle 2010). One of the main molecular phenotypes associated with aging is the non-linear accumulation of molecular and cellular damage which consequently leads to the time-dependent accumulation of protein aggregates (Lindner and Demarez 2009). Moreover, frailty of old age, or this age-related accumulation of health deficits, is characterized by the combined effect of many health related factors and leads to the increased likelihood of death over time. The mortality rate grows exponentially with age, which however slows down at very advanced age (Partridge and Mangel 1999).

In health however the correct conformation of the protein products, including neuronal proteins, is achieved by the protein quality system with the help of chaperone proteins, which inhibit formation of aggregates, and protein-clearance machinery that are in charge of degrading the damaged proteins by proteasomes and autophagy (Labbadia and Morimoto 2014). With age, the damage inflicted on these control mechanism progressively increases and consequently this leads to the accumulation of misfolded proteins and their aggregation (Hartl and Hayer-Hartl 2009). Moreover, although most of the proteins associated with neurodegeneration are expressed systemically, the resulting toxicity has been identified only in the neuronal cells. This might be due to the increased susceptibility of the brain structure, especially the neuronal cells, to the impairment of the protein quality system (Soto 2003).

As global life expectancy is progressively increasing, the prevalence and incidence of most neurodegenerative diseases will also rise dramatically in the elderly. The pathogenesis of many of these diseases is still unknown and they pose a threat to the economic and social wellbeing of current and future generations (Forman, Trojanowski et al. 2004). Therefore, it is important to understand the mechanisms of toxicity in order to develop the therapeutic solutions to treat and prevent these disorders.

1.1.2 NEURODEGENERATIVE DISEASE CLASSIFICATIONS

Several neurodegenerative diseases are classified as “proteopathies”, where protein misfolding causes toxicity. In these cases, DNA mutations in the encoding genes have been associated with the hereditary forms of neurodegenerative diseases. For example, in many different diseases the mutated gene has a repeat of CAG nucleotide

triplet that encodes the glutamine amino acid. These disorders with many repeats of glutamine are known as polyglutamine (polyQ) diseases, one example of which is Huntington's disease (MacDonald, Ambrose et al. 1993). As a result, extra glutamine residues have toxic effects, such as impaired protein folding and aggregation that deposits itself as intranuclear inclusions (Figure 1.1).

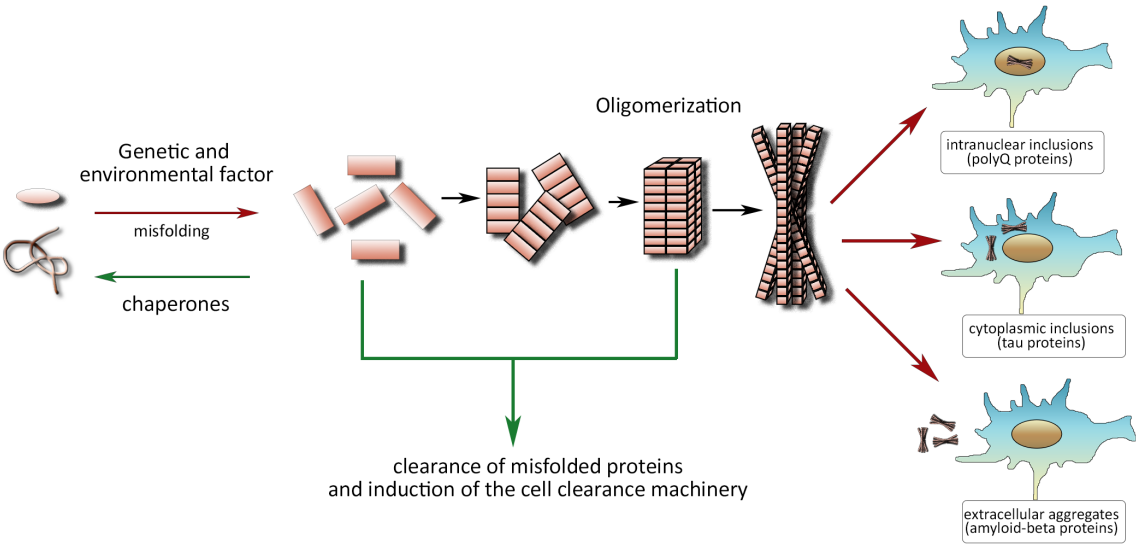

Figure 1.1 Overview of protein misfolding and aggregation. Genetic and environmental factors can accelerate the process of protein misfolding and aggregation, while cellular quality-control systems, such as molecular chaperones, regulate the correct folding or refolding of the proteins. If the native state of proteins is unattainable, misfolded proteins are selectively removed through proteolysis or autophagy. However, failure to eliminate misfolded proteins leads to the aggregation and accumulation of toxic aggregates which can deposit themselves in the nucleus (for example polyQ deposits in Huntington's disease), cytoplasm (alpha-synuclein in Parkinson's disease) or in the extracellular space (Aβ in Alzheimer's disease).

In some of the cases the proteins aggregate and deposit in the intracytoplasmic space. One example of such a disorder is Parkinson's disease that is caused by the mutation in alpha-synuclein gene. The disease is characterized, in most cases, by aggregation of alpha-synuclein that forms insoluble fibrils in dopaminergic neurons of Lewy bodies (Shulman, De Jager et al. 2011). On the other hand, mutated protein can also be secreted extracellularly, for example amyloid-beta in Alzheimer's disease (Querfurth and LaFerla 2010). Some evidence suggests that the inclusion bodies, consisting of abnormal aggregated proteins, are not the main cause of toxicity and whether soluble monomers, oligomers or larger aggregates have the most toxic effect in many of these conditions is uncertain. Moreover, it has been shown that in some

circumstances inclusions can serve a protective role (Soto 2003). However, in general the tendency of mutated protein to aggregate into oligomeric amyloid species is correlated with toxicity (Forman, Trojanowski et al. 2004). Alzheimer's disease, as one of the most common neurodegenerative disease, is an excellent candidate disorder to study the mechanism of protein aggregation and its role in disease pathogenesis as it characterized by the presence of both intra and extracellular protein deposits.

1.2 ALZHEIMER'S DISEASE

Alzheimer's disease (AD) accounts for more than 60% of dementia cases and affects approximately 46 million people worldwide (Alzheimer's 2014). AD is characterized by the progressive dysfunction and then loss of neurons and synapses especially in the cerebral cortex and by the presence of two neuropathological lesions: extracellular plaques, mainly composed of beta-amyloid peptide, and intracellular tangles, made up largely of cytoskeletal protein tau (Querfurth and LaFerla 2010).

1.2.1 TAU PROTEIN

Tau is a highly soluble microtubule-associated protein (MAP). One of the main functions of tau protein is to bind to the microtubule and promote their assembly and stability. Other functions include tubulin polymerization, maintaining neuronal integrity, axonal transport and axonal polarity. These proteins are mostly found in neurons of central nervous system as compared to non-neuronal cells. However, tau is not present in dendrites, being confined normally to the axons where it provides microtubule stabilization (Ballatore, Lee et al. 2007). In adult human brain tau proteins are produced as a result of alternative splicing from a single gene labelled as a microtubule-associated tau protein (MAPT) and located on chromosome 17q21, composed of 16 exons. There are six tau isoforms in human brain tissue that are classified according to the number of binding domains. Each of the tau isoforms has certain physiological functions and appears to be differentially expressed during development of the brain (Goedert, Ghetti et al. 2000).

The binding of tau to tubulin is regulated by its phosphorylation state. The longest isoform of tau protein has 79 potential Serine or Threonine phosphorylation sites. Notable, phosphorylation of tau is also developmentally regulated, when it is substantially higher in foetal brain than in adult brain. Phosphorylation of tau is regulated by coordinated action of a number of kinases and phosphatases (Ingram and Spillantini 2002). Under pathological conditions, such as in case of AD, the equilibrium of tau binding to tubulin is disrupted, resulting in microtubule disorganization and abnormal increase of unbound tau molecules (Figure 1.2). Consequently, hyperphosphorylation of tau is associated with misfolding and conformational changes in the normal structure of tau and leads to the aggregation into fibrillar structures, such as straight or paired helical filaments (SFs and PHFs respectively), inside the affected neurons. The PHFs is further self-assembly and forms neurofibrillary tangles (NFTs), which is one of the major hallmarks of AD (Brandt, Hundelt et al. 2005). At the cellular level, abnormal phosphorylation of tau introduces alterations in the binding abilities to tubulin and microtubule assembling, causing the interruption of axonal transport and consequently synapse dysfunction (Roy, Zhang et al. 2005).

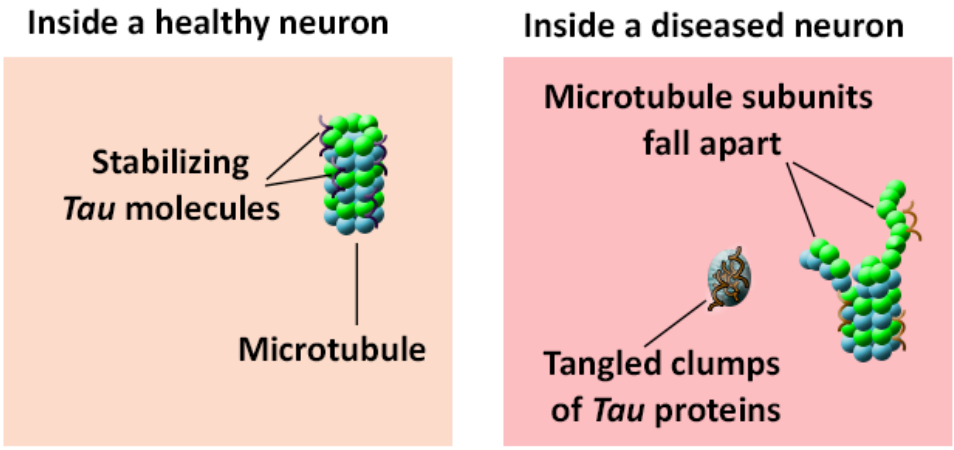

Figure 1.2 Microtubule disintegration as a result of hyperphosphorylated tau. In healthy neurons tau binds to microtubules and stabilizes them, while in diseased neurons tau, as a result of hyperphosphorylation, detaches from the microtubules and aggregates into neurofibrillary tangles.

Previous findings show that the tau-mediated toxicity appears to be caused by the mutations in tau genes. In fact, the abundance of NTFs consisted mainly of hyperphosphorylated tau, correlates well with familial forms of frontotemporal

dementia and AD. Even though the presence of neurofibrillary tangles is associated with neuronal loss in AD, their role in pathology of AD is not yet clear (Goedert, Ghetti et al. 2000, Ingram and Spillantini 2002).

1.2.2 AMYLOID-BETA

Familial form of AD is associated with the mutation in one of the genes encoding amyloid precursor protein (APP) and Presenilins 1 and 2. Many disease-linked mutations have been identified and each results in either an increased production of total Aβ protein or else a change in the balance of Aβ isoforms resulting in more aggregation-prone species predominating. Aβ denotes the peptides that consist of 36-43 amino acids and are generated as a cleavage product from the membrane protein APP. The gene coding the APP is located on chromosome 21 which may explain why individuals with Down's syndrome, who have three copies of that chromosome, have a high incidence of AD (Querfurth and LaFerla 2010). This suggests that the overexpression of APP can be a cause of AD.

As a transmembrane protein, APP is metabolized by sequential proteolysis by enzymes including α–, β– and γ-secretase (Hardy 2006). APP can undergo proteolytic processing by two alternative pathways (LaFerla, Green et al. 2007), one of which leads to the generation of Aβ peptide (amyloidogenic) and the second one does not (non-amyloidogenic) as shown in Figure 1.3.

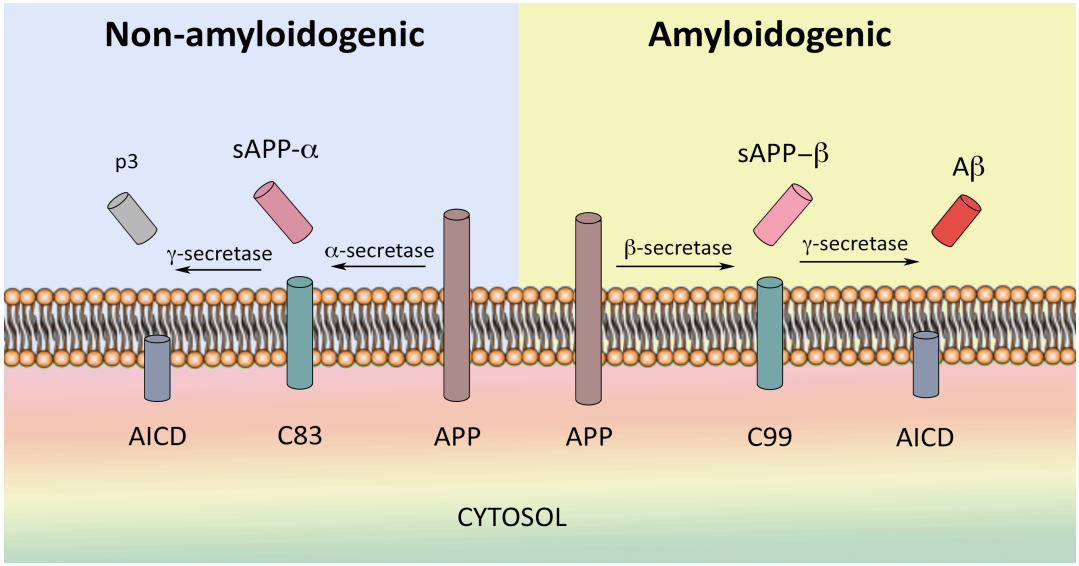

Figure 1.3 Sequential cleavage of amyloid precursor protein. APP is processed by α– or β-secretases. The processing by α-secretase precludes Aβ formation, whereas processing by β-secretase and sequentially by γ-secretase lead to the production of Aβ peptide. Depending on the site of cleavage Aβ can be 36-43 amino acids long.

Crucial steps in APP processing happen at the cell surface, where APP can be transported to the cell surface or to endosomal compartment. On the cell surface APP undergoes non-amyloidogenic processing, where it is cleaved directly by α-secretase and then γ-secretase, a process that involves the alternative cleavage of Aβ into shorter fragments and therefore precludes amyloid-beta formation. Amyloidogenic processing of APP occurs in the endosomal compartments and involves sequential enzymatic actions of β- and γ-secretases (De Strooper, Vassar et al. 2010). Extracellular cleavage by β-secretase creates soluble extracellular fragments (sAPP-β) and membrane-bound fragments labelled as C99. Following the release of sAPP-β ectodomain, γ-secretase cleaves the transmembrane region of APP and generates amyloid-beta (Aβ) fragments of 30-51 amino acids residues in length (LaFerla, Green et al. 2007).

Monomers of $A\beta_{40}$ and $A\beta_{42}$ are the most common, with the longer isoform being more toxic due to the higher level of aggregation. It is yet unclear why some surface APP is cleaved directly by α-secretase while the others being internalized into endosomes to produce Aβ fragments (De Strooper, Vassar et al. 2010).

1.2.3 GENETICS OF ALZHEIMER'S DISEASE

The cause of Alzheimer's disease is still unclear, except for around 5-10% of the cases where genetic mutations have been identified. Familial cases of AD (FAD), based on the reviews of twins and families studies, are characterized by the early onset (before the age of 65) and inherited in dominant autosomal manner. While most cases of AD do not show dominant inheritance and are termed sporadic, pathological and clinical studies of these forms of AD suggest that these two types have similar pathologically and clinically features (Hardy 2006, Querfurth and LaFerla 2010). Therefore, the genetic study of FAD can shed the light on the mechanism of the pathogenesis of this disease.

Genetic analysis of FAD has revealed that mutations in several genes disrupt the equilibrium of Aβ production. These identified genes encode for amyloid precursor protein (APP) and presenilin 1 and 2 (PSEN1 and PSEN2 accordingly). Most of the mutations within APP cluster around the γ-secretase cleavage site, β-secretase cleavage site or cause the changes to the sequence of Aβ peptides. The

mutations around the β-secretase cleavage site increase the production of Aβ, while the mutations in γ-secretase cleavage site promote the production of longer, less soluble and more toxic $A\beta_{42}$ (LaFerla, Green et al. 2007, Querfurth and LaFerla 2010). Furthermore, the mutation in the APP gene related to the Aβ sequence accelerates the self-aggregation of Aβ into amyloid fibrils and is highly toxic, as described in the case of Arctic mutations (Chiti and Dobson 2006).

On the other hand, sporadic cases of AD, which do not exhibit autosomal dominant inheritance, have a complex aetiology and multiple studies suggest that they can be caused by a combination of environmental and genetic factors (Sleegers, Lambert et al. 2010). Currently, several potential risk genes have been identified. The most important risk gene is ApoE, discovered in 1993 by genetic linkage analysis; it is linked to a threefold increase in the risk of AD. Further analysis showed that only one of three alleles (namely the E4 allele) was a risk factor for this disease (Bu 2009). However, the exact mechanism behind such an association is still unclear. A number of studies suggest a role of ApoE in amyloid binding and deposition (Bu 2009, Bettens, Sleegers et al. 2010, Querfurth and LaFerla 2010). Some evidence suggests that while under the normal conditions ApoE enhances proteolytic clearance of Aβ; the E4 allele isoform is not as effective at promoting the breakdown as other alleles and thus increases susceptibility to the disease in individuals with that gene variant (Fuentealba, Liu et al. 2010). Furthermore, recent genome-wide association studies (GWAS) have suggested a number of new genetic risk factors associated with AD (Bertram and Tanzi 2005, Lambert, Heath et al. 2009). These studies are based on the assumption that if one type of genetic variant (or alleles) is more frequent in people with disease, then that particular single-nucleotide polymorphism is associated with disease and individuals carrying that polymorphism are at increased risk of disease. Interestingly, the genes involved in FAD (APP, PSEN1 and PSEN2) are absent from the list of the candidate genes associated with sporadic form of AD as identified by GWAS studies, while all of the studies confirmed the association of ApoE4 allele with the disease. On the one hand, this finding might suggest the lack of sufficient power to detect the alleles with low frequency but high penetrance, but on the other hand it might suggest that sporadic and familial forms of AD have different mechanism of pathogenesis (Gandhi and Wood 2010).

1.2.4 Aβ MEDIATED TOXICITY

According to the amyloid hypothesis, accumulation and aggregation of Aβ peptide in the brain is the primary neuropathological event that triggers other pathological aspects of AD including synaptic dysfunction, inflammation, neurofibrillary tangles and consequently neuronal death. Consequently, the amyloid cascade hypothesis also proposes that the changes in tau and formation of neurofibrillary tangles results from the imbalance between Aβ formation and Aβ clearance (Hardy and Selkoe 2002).

The support for this hypothesis comes from various experimental observations. For example, transgenic mice overexpressing both mutant APP and mutant tau are characterized by the accelerated neuronal loss and show higher level of neurofibrillary tangle formation (as compared to transgenic mice model for tau alone), while the structure and amount of amyloid plaques remain on the same level (Lewis, Dickson et al. 2001). Moreover, it has been shown that in triple transgenic mice with APP, PS1 and tau mutations the Aβ deposition is developed prior to the NFTs (Oddo, Caccamo et al. 2003). This finding imply that the APP processing occurs upstream of tau and this assumption has been supported by the recent experimental results that demonstrate tau-dependent Aβ toxicity in mouse neuronal culture. Finally, APP transgenic mouse model with reduced level of endogenous tau show less cognitive impairment (Roberson, Scearce-Levie et al. 2007).

1.3 SYNAPTIC DYSFUNCTION

Recent studies suggest that toxic Aβ aggregates bind to the neurons and change their structure so that they disrupt synaptic communications. Moreover, persistent synaptic dysfunction may account for the evident neuronal loss on later stages of disease (Shankar and Walsh 2009). Similarly, it has been proposed that Aβ peptide also plays a crucial role in synaptic physiology and regulation of synaptic vesicle release (O'Brien and Wong 2011). Several lines of evidence have shown that Aβ impairs synaptic plasticity by inhibiting long-term potentiation (LTP), responsible for the learning and memory formation in synapses, and at the same time facilitating long-term depression (LTD), which reduces the efficacy of neuronal synapses (Walsh,

Klyubin et al. 2002). Aβ aggregates disrupt the release of presynaptic neurotransmitters and postsynaptic glutamatergic transmission by inducing the endocytosis of surface receptors, such as the N-methyl-D-aspartate (NDMA) receptor (Snyder, Nong et al. 2005). Similar effects have been observed during normal aging, when the balance between LTP and LTD is altered. However Aβ seems to be able to trigger such synaptic deficits even earlier (Querfurth and LaFerla 2010). In addition, experimental studies have shown that Aβ binds to neurotropic receptors essential for cognitive processing, such as brain derived neurotrophic factor (BDNF), impair the maintenance of LTP and consequently reduce the neurotrophic support (Dechant and Barde 2002).

Exposure to Aβ during disease progression inhibits key mitochondrial enzymes, causing the damage to ATP production, oxygen consumption and electron transport. In Alzheimer's disease, as well as during normal aging, mitochondrial DNA (mtDNA) undergoes high levels of oxidative stress (Querfurth and LaFerla 2010).

1.4 COGNITIVE DYSFUNCTIONS IN AD

Several meta-analyses of risk factors for Alzheimer's disease have reported that lifestyle factors can reduce the risk of developing the disease, specifically higher levels of education, social engagement, regular physical activity and cognitive stimulation can lower the risk of cognitive decline and AD (Buschert, Bokde et al. 2010). By comparison, cerebrovascular and cardiovascular factors, which include smoking, obesity, alcohol intake, high blood pressure and diabetes, increase the risks of developing AD (Citron 2010).

While the risk factors described for AD are well characterized they are unfortunately so wide-spread within the population that they are not sufficient to assist with early diagnosis of disease. The timely diagnosis of AD therefore rests on detecting the early clinical features. For many patients memory problems are the initial complaints however they may well already have deficits in other domains, in particular sleep/wake biology and spatial orientation.

1.4.1 MEMORY PROBLEMS

Alzheimer's disease is clinically characterized by cognitive and functional dysfunctions manifested by the memory problems and impairment in other cognitive domains such as language, perceptual thinking, orientation, attention and functional abilities (Burns and Iliffe 2009). At the early stage of the disease, the loss of short-term memory is the most noticeable deterioration, which presents as difficulties in acquiring new memories or remembering recent facts. Depending on the progression of the disease, the impairment of memory and other intellectual functions, including long-term and implicit memory, become more prominent and can be accompanied by psychotic episodes, when patient have delusional symptoms, exhibit unpremeditated aggression or fails to recognize close relatives (Mayeux 2010).

1.4.2 SLEEP PROBLEMS

In addition, Alzheimer's disease is accompanied by non-cognitive disturbances. As discussed previously, progression of AD leads to the loss of brain tissue and impaired mental abilities that may also disrupt the sleep/wake cycle. As a consequence, as a disease progress patients with AD lose their ability to maintain a full night's sleep and have more fragmented sleep (Hatfield, Herbert et al. 2004, Musiek, Xiong et al. 2015, Peter-Derex, Yammine et al. 2015). Biological and physiological processes, including sleeping, body temperature, heart rate and others characteristics, are controlled by circadian rhythms that are driven by circadian clocks. In mammals central circadian clock is located in suprachiasmatic nucleus (SCN) in hypothalamus. In AD the normal functioning of circadian clocks are supposedly disrupted as a result of the progressive degeneration and loss of neuronal connections in hypothalamus, which includes the damage to SCN, however no conclusive evidence has been presented so far proving that the SCN is particularly affected by AD pathology (Volicer, Harper et al. 2001). On the other hand, the disturbance of sleep pattern may aggravate the cognitive impairment due to the impairment of sleep-dependent memory consolidation process (Peter-Derex, Yammine et al. 2015). Overall, circadian rhythms disturbances experienced by the patients with AD are more severe that those seen in healthy elderly individuals, characterized by the higher percentage of nocturnal activity, reduced amplitude of circadian rhythm and delayed body temperature rhythms (Volicer, Harper et al. 2001).

Recent anatomical studies have shown that there is convective flow between cerebrospinal fluid and the lymphatic vessels, called glymphatic system, which can help to remove toxic metabolites (Iliff, Wang et al. 2012). Moreover, it has been suggested that glymphatic clearance is regulated by sleep-wake cycle and this system is strongly stimulated by sleep. Interestingly, during sleep the Aβ clearance activity in mice is twice as fast as during waking hours (Xie, Kang et al. 2013). According to the clinical studies, the quality and duration of sleep may be early indicative symptom or AD and predict the onset of this disease (Mendelsohn and Larrick 2013).

1.4.3 SPATIAL DISORIENTATION

Among the various behavioural and cognitive dysfunction, experienced by patients with AD, is spatial disorientation, or tendency to get lost in a familiar environment. The orientation involves the ability to extract the selective spatial information and consequently generate the mental representation of the environment. Failure to do so could result in the situation when patients with AD are unable to find their way from one room to another even in their own houses (Henderson, Mack et al. 1989, Monacelli, Cushman et al. 2003). Therefore, the spatial navigation impairment, that usually occurs at the early stage of AD, together with the disruption of sleep-wake cycle can interfere with the independent living of people with AD and is particularly burdensome to caregivers.

1.5 ANIMAL MODELS OF AD

In humans, the study of neuropathological changes that occur during AD progression is limited because the data can only be acquired *post mortem* (Spires and Hyman 2005). For this reason animal models of the disease can facilitate our understanding of underlying pathogenic mechanisms and provide new insights into the disease (Gotz and Ittner 2008). Even though the role of particular cellular pathways can be studied *in vitro* (Cleary, Walsh et al. 2005), the use of animal models of the disease is preferred as it provides the opportunity to study complex biological interactions within the whole living organism (Spires and Hyman 2005). As a result, animal models of AD have proven to be an excellent tool to study the disease-related progression of Aβ and tau pathology, to examine synaptic abnormalities and

behavioural dysfunctions over time (Gotz and Ittner 2008). Eventually, the animal models can be used to develop and test a new therapeutic agents and validate diagnostic tests (Crowther, Kinghorn et al. 2004).

Transgenic models of AD have been developed in several organisms including vertebrates (specifically mouse, rats and monkeys) and invertebrates (fruit flies and worms) (Gotz and Ittner 2008).

1.5.1 MOUSE MODELS OF AD

The mouse has been the mammalian model system of choice for genetic research and allowed to study the role of Aβ toxicity in the amyloid cascade hypothesis (Hall and Roberson 2012). Transgenic mice with mutations in APP, a gene associated with familial AD, have developed Alzheimer-like neuropathology, including the induced Aβ deposition, memory deficits and in some cases the neurofibrillary tangles (Games, Adams et al. 1995, Lewis, Dickson et al. 2001). Similarly, transgenic presenilin mutant mice are characterized by the increased levels of $A\beta_{42}$ but they do not develop AD pathology (Huang, Yee et al. 2003). To enhance Aβ pathology, double transgenic models have been developed, expressing both APP and presenilin (Hall and Roberson 2012). They are characterized by the increased levels of cognitive dysfunctions occurred before the appearance of Aβ deposits, suggesting the causal role of Aβ toxicity in disease pathogenesis. In addition, transgenic mouse models of tau were developed to emulate the tau-related pathologies. Tau transgenic mouse model has been shown to be characterized by the enhanced neurofibrillary pathology, but in absence of Aβ deposits (Lee, Kenyon et al. 2005). At present, more complex mouse models have been developed to recapitulate the complete set of AD-like features. In these triple transgenic mouse models three mutant genes were simultaneously expressed: of APP, tau and presenilin 1. These triple transgenic mice showed the progressive development of Aβ and tau pathology, but interestingly the Aβ deposition developed prior to the tau tangles (Oddo, Caccamo et al. 2003) confirming the amyloid cascade hypothesis.

1.5.2 INVERTEBRATE MODELS OF AD

Invertebrate models of AD have emerged as a powerful tool to study the disease-related neurodegeneration. Both the nematode *Caenorhabditis elegans* (*C. elegans*) and the fruit fly, *Drosophila melanogaster* (*Drosophila*), are the most commonly used invertebrate models organisms. Invertebrate models offer many advantages over vertebrates, including their short lifespan, comprehensively studied genomics, variety of representing phenotypes and availability of genetic tools to express the human genes of interest. In particular, *C. elegans* was the first multicellular organisms of which the genome was fully sequenced (Hillier, Coulson et al. 2005). Moreover, the basic forms of learning and memory, as well as circadian rhythms, can be evaluated in *Drosophila* providing the functionally relevant neuroanatomy for human disorders. Generally, the fundamental aspects of cell biology are highly conserved in both humans and fruit flies (Crowther, Kinghorn et al. 2004, Spires and Hyman 2005, Crowther, Page et al. 2006, Gotz and Ittner 2008).

Most of the genes involved in familial AD have homologues in *Drosophila* and *C. elegans*. For example, the orthologue of APP has been identified in *Drosophila* (amyloid precursor protein-like protein APPL) as well as in *C. elegans* (amyloid precursor-like protein 1 APL-1) (Rosen, Martin-Morris et al. 1989, Daigle and Li 1993). Both of these orthologues resemble the human counterpart throughout their length, but Aβ expressing region lacks the similarity to human APP. Despite this dissimilarity, overexpression of APPL in *Drosophila* leads to the axonal transport defects and increased neuronal loss in the developmental stage, suggesting the role of APP in axonal transport (Spires and Hyman 2005). In a similar manner, it has been reported that, in *C. elegans*, expression of Aβ in bodywall muscle results in the progressive paralysis and decreased lifespan (Dostal and Link 2010). The double transgenic models of AD in *Drosophila* with the coexpression of APPL and tau have been shown to induce the neuronal dysfunction and disrupt axonal transport (Torroja, Chu et al. 1999). In *C. elegans*, the overexpression of tau results in neurodegenerative changes and behavioural and synaptic dysfunctions (Kraemer, Zhang et al. 2003). In addition, the expression of human $A\beta_{42}$ peptide have been reported to lead to accumulation of toxic amyloid aggregates and neurodegeneration, while the *Drosophila* model of less aggregation prone variant of $A\beta_{40}$ developed the age-dependent defects but no neurodegeneration (Crowther, Kinghorn et al. 2004,

Crowther, Page et al. 2006, Lu and Vogel 2009). The presence of the neurodegeneration, which can be evaluated histologically for $A\beta_{42}$, can be observed and quantified by longevity assays and behavioural tests (Crowther, Kinghorn et al. 2004, Crowther, Page et al. 2006).

CHAPTER 2: PROJECT AIMS

Neurodegenerative diseases are currently incurable disorders associated with progressive loss of neuronal cells and supposedly caused by the accumulation of misfolded proteins. Alzheimer's disease, one of the most common types of neurodegeneration, is assumed to be a result of disrupted balance between production and clearance of amyloid beta peptide. Elevated levels of Aβ aggregates in the brain tissue have been implicated in both familial and sporadic AD. Excessive amount of extracellular Aβ tend to self-aggregates into insoluble fibrils, which not only induce the oxidative stress to neuronal cells but also impairs synaptic function, leading to the progressive memory loss and cognitive deterioration associated with AD.

The study of underlying pathogenic mechanisms of the disease in patients with AD is limited; nevertheless the animal models of AD can be used to investigate the molecular and behavioural dysfunctions associated with the disease. The invertebrate models, such as *C. elegans* and *Drosophila*, provide the opportunities to investigate the disease-related neurodegeneration and behavioural characteristics, as well as to perform the necessary genetic manipulations.

Better understanding of disease pathogenesis and finding effective treatment for Alzheimer's disease relies on the discovering the potential risk genes identified by genome-wide association studies and further studies of additional mechanisms of neurotoxicity, in particular the circadian rhythms disruption and underlying mechanisms of spatial disorientation, commonly associated with AD.

Therefore, in this thesis three main research questions were addressed:

1. Can we use an animal model of Aβ toxicity to provide further insights into the human genetics by cross validating worm modifier genes with GWAS candidate genes? Taking into account the amyloid cascade hypothesis, that suggests the central role of Aβ toxicity in AD pathology (Hardy and Selkoe 2002), and on the other hand the absence of genes associated with familial AD from the GWAS findings brings into question the importance of Aβ in disease pathogenesis. In chapter 3, I evaluate the role of Aβ in AD by comparing the findings of GWAS in sporadic AD with the results of knockdown screen in a *C. elegans* model system that exclusively reports Aβ toxicity.

2. To what extent are the circadian disruptions associated with AD governed by dysfunction of the central clock? Several lines of evidence suggested that circadian disruptions contribute to the pathology of AD (Coogan, Schutova et al. 2013, Chen, Possidente et al. 2014), however it is still unclear whether the neurodegeneration damages the central clock itself or the communication pathways leading to the peripheral tissues. In chapter 4, I will describe the method developed to simultaneously monitor oscillations of peripheral clocks and behavioural rhythms in *Drosophila* model of AD.

3. How does the fruit fly orientate itself in space and how is this disrupted in a model of AD? It has been previously shown that most animals have periodic positional preferences, namely they prefer to sleep and move in the certain locations (Campbell and Tobler 1984), indicative of highly developed mechanisms of spatial orientation, however the underlying control of such behaviour is yet to be determined. Moreover, in patients with AD, the similar ability to navigate within the familiar environment is disrupted at the early stage of the disease. In chapter 5, I will investigate the control of spatial orientation in fruit flies and note its disruption in a *Drosophila* model of AD.

CHAPTER 3: USING A WORM MODEL OF Aβ TOXICITY TO DERIVE MORE INFORMATION FROM HUMAN GWAS IN ALZHEIMER'S DISEASE

Finding potential candidate genes associated with an increased risk of a disease can contribute to improving our understanding of its pathogenesis; ultimately, leading to the discovery of possible pathological pathways involved in its development. This chapter describes methods to compare the findings of genome-wide association studies in sporadic AD with the results of a genome-scale screen in a model organism. The transgenic *C. elegans* model system used in this study allows the detection of RNAi constructs that suppress the Aβ-induced paralysis phenotype. By inferring which worm genes are knocked down by each RNAi construct we can gain information about which worm genes are essential for Aβ toxicity. The interactome network of the modifier genes was then compared to the interactome network for human AD GWAS hits and the degree of similarity was assessed.

3.1 INTRODUCTION

According to a growing body of evidence, aging is the greatest risk factor in developing Alzheimer's disease (AD). Genetics, however, plays an important role in the likelihood of developing AD as genetic inheritance has proven to be a major aetiological risk factor; indeed the heritability of so-called sporadic Alzheimer's disease is estimated to be 58-79% (Gatz, Reynolds et al. 2006). The vast majority of AD is sporadic and exhibits heterogeneous, multifactorial inheritance. The most common assumption is that sporadic AD is likely to be regulated by a set of low-penetrance risk genes across a variety of undetermined loci. These genetic variants are believed to be involved in a number of pathways, many of which are likely to alter the production, aggregation and clearance of Aβ. Other pathways that have been implicated include inflammatory pathways, particularly innate immunity. Although a comprehensive genetic characterization of AD has yet to be completed, there is a substantial evidence that the combination of these risk genes have a strong impact on the susceptibility to sporadic AD (Bertram and Tanzi 2009). Under these

circumstances, recent progress in GWAS has furthered our understanding of the genetic influence on the aetiology of sporadic AD.

3.1.1 THE PRINCIPLES OF GENOME WIDE ASSOCIATION STUDIES

Genome-wide association studies (GWAS) simultaneously analyse a large number of genetic variants across the entire genome in unbiased hypothesis-free manner. GWASs are based on the assumption that particular genetic variants that occur in diseased individuals more frequently than in healthy controls contribute to the heritability of the disease. The sequence variants may have a direct effect on gene function but more commonly they are simply a marker for some other nearby pathogenic variant that itself may not be included in the screen. The most common type of DNA sequence variation within a population is single-nucleotide polymorphisms (SNP), therefore, in GWAS the markers for genetic variants are read using an SNP array.

GWAS usually compares SNPs of two large groups of individuals: people diagnosed with disease and similar people without disease. Each of the SNPs is then thoroughly examined to determine if the allele frequency is significantly different between these two groups. The result is usually reported as an odds ratio: a ratio of odds of having disease for individual with a specific allele against the odds of having disease for individual without the same allele. The significance of the result is measured in terms of p-value that is normally calculated using a chi-squared test. The final results of genetic association studies are normally presented in a form of Manhattan plot (Figure 3.1); it shows the negative logarithm of p-values against the corresponding genomic locations. Each point on the Manhattan plot represents the individual SNP with its genomic positions plotted along the X-axis and significance level (negative logarithm of p-value) along the Y-axis. The SNPs with the most significant association will stand out on the plot as the smallest p-values will be plotted as a greatest number on Y-axis (due to the negative logarithm conversion). However, as GWAS compares thousands, sometimes, millions of SNPs, the analysis is prone to both false positives and false-negatives resulting in false associations. Therefore, the p-value threshold for GWAS is adjusted for multiple comparisons using the Bonferroni correction method. Note that although this stringent analysis reduces false positives, it can also increase the number of false negatives because in a

multigenic sporadic disorder most genetic associations are likely to have low odds ratios (Pearson and Manolio 2008).

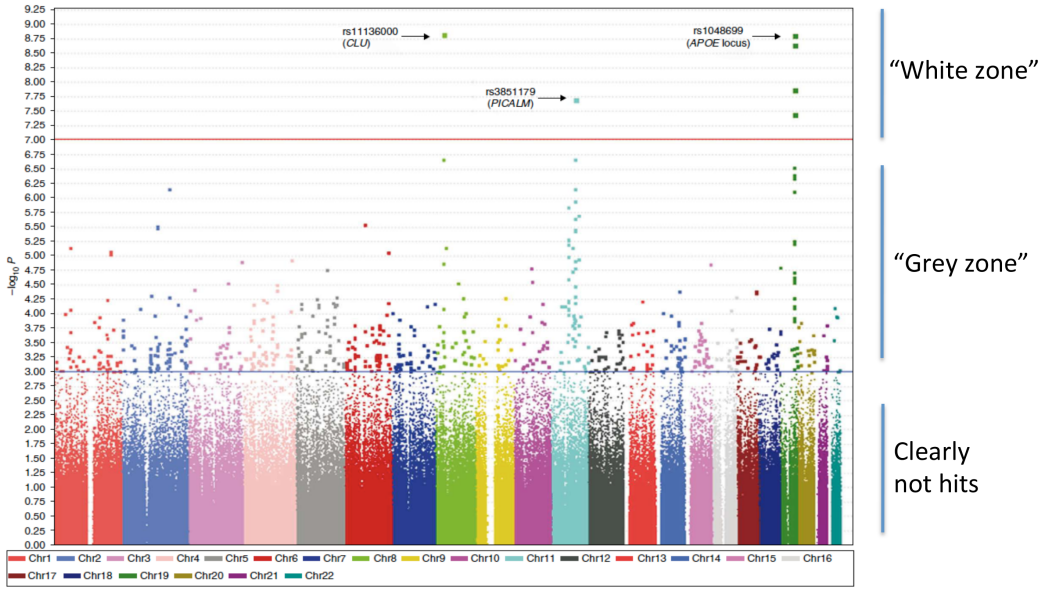

Figure 3.1 Manhattan plot (adapted from (Harold, Abraham et al. 2009). The significance of GWAS findings is displayed as a collection of points, each represented by chromosomal position on x-axis and $-\log_{10}$ GWAS p-value on y-axis.

To make more use of the GWAS data we have I have used complementary data from a worm genetic modifier screen along with further analysis to reveal some previously unidentified associations.

3.1.2 GWAS IN AD

The various genome-wide association studies of AD completed to date have so far proved to be less consistent (Bertram and Tanzi 2009), with the exception that all studies coherently reported the strong association of ApoE4 locus with disease. The first generation of high-resolution, collaborative GWAS in AD was published in 2009 and reported four genes (ApoE, CLU, CR1 and PICALM) with significant association with AD (Harold, Abraham et al. 2009, Lambert, Heath et al. 2009). Finding of three previously undiscovered genes demonstrates the potential of GWAS to identify new genetic variants that may play a role in the pathogenesis of sporadic AD. Additionally, a growing body of evidence has suggested that these genes may be involved in amyloid cascade and alter the clearance, trafficking or production of Aβ peptide. The second generation of GWAS in AD identified a further 17 risk genes;

therefore, at the time of this research there are a total 21 of risk genes associated with AD (Hollingworth, Harold et al. 2011, Lambert, Ibrahim-Verbaas et al. 2013).

It is important to note that the genes that are known to be involved in FAD were absent from the GWAS findings; this may undermine the role of Aβ in pathogenesis of sporadic AD. The absence of these genes may be due to the fact that GWAS lacks sufficient sensitivity to detect rare variants, which might have low frequency of occurrence but in turn have higher penetrance. Obviously, one way to overcome this issue is to increase the size of samples by recruiting more subjects or by using higher resolution SNPs array. However, in this work a different approach is used and is described below.

In order to evaluate the importance of Aβ in the aetiology of sporadic AD, we intend to compare the findings of GWAS in AD with the results of a genome-wide knockdown screen in an animal model organism that reports solely Aβ toxicity, such as the nematode worm *Caenorhabditis elegans*. There are many advantages of using *C. elegans* as a model system of AD. For example, transgenic worms can have easily observable phenotypes caused by Aβ toxicity. Like mammals, they have complex biochemical pathways, most of which are well conserved. Worms are inexpensive to maintain and have a short lifespan, thereby facilitating large-scale laboratory studies. *C. elegans* was the first eukaryote with a nervous system to have its genome sequenced (Consortium 1998) and is the only organism for which we have a complete nervous system wiring diagram (White, Southgate et al. 1986). Moreover, the genetic characteristics of *C. elegans* have been well studied (Hillier, Coulson et al. 2005) and the utility of RNA interference in *C. elegans* discovered in 1998 (Fire, Xu et al. 1998) has been extensively applied affording the opportunity to silence expression of target genes (Kamath and Ahringer 2003).

Additionally, with purpose of identifying new potential risk genes, we propose to include the GWAS findings of borderline significance (GWAS "grey zone", p-values threshold below 10^{-5}) together with the genes with conventional GWAS significance (p-value$<10^{-7}$, GWAS "white zone"). Although this extended list of candidate genes will contain false positive results, it might additionally include previously unidentified risk genes that may contribute to the pathogenesis of sporadic AD. We hypothesize that if Aβ toxicity is indeed involved in sporadic AD, then we expect to see a statistically significant overlap between the list of genes identified by

GWAS and the list of genes derived from the genome-wide knockdown screen. However, if the leading cause of sporadic AD were distinct from the accumulation of Aβ peptide, we would expect the overlap between those two lists to occur due to chance.

3.2 MATERIALS AND METHODS

The effect of Aβ toxicity on neurodegeneration in AD was studied using *C. elegans* model organism, which can be genetically modified to express human Aβ peptide.

3.2.1 DESIGN OF THE EXPERIMENT (WORM SCREEN)

In order to study the aggregation of Aβ, *C. elegans* transgenic lines have been generated expressing human β-amyloid (Aβ). In the *C. elegans* transgenic line CL4176 Aβ is expressed in muscle cells using the myosin promoter (Drake, Link et al. 2003). The identification of the pathways that are involved in the mechanism of Aβ toxicity in *C. elegans* can be achieved by employing genome-wide RNA-mediated interference (RNAi) as part of a genetic modifier screen.

3.2.1.1 *C. ELEGANS* STRAINS

The *C. elegans* strain CL4176 [*smg-1*(cc546ts) dvIs27(pAF29 Pmyo-3::Aβ42+pRF4)], expressing Aβ in the body-wall muscle, was used as a model of Aβ toxicity (Link 1995). Briefly, in order to create the temperature-dependent induction of Aβ, mRNA-surveillance system of the *C. elegans* (the *smg* system) was used to generate transgenic constructs. At non-permissive temperatures (23 ^{o}C) this system is inactive and allows the induction of transgene mRNA for human $A\beta_{42}$. The inducible Aβ-expressing transgene was introduced to the strain CL4176 by inserting an Aβ minigene fragment into expression vector pPD118.60. The non-Aβ expressing strain CL802 [*smg-1*(cc546ts) *rol-6*(su1006)] was used as a background strain. Both CL4176 and matched transgenic control CL802 were cultured at permissive

temperature for *smg* mutation of 16 ^{o}C (Link, Taft et al. 2003) on solid peptone nematode growth media (NGM) as described in (Brenner 1974).

3.2.1.2 RNA INTERFERENCE (RNAi) SCREEN FOR *C. ELEGANS*

In the genome-wide RNAi screen, the worms were fed bacteria *Escherichia coli (E. coli)* carrying RNAi-expressing plasmids that target a particular gene of interest; typically the specific destruction of the target gene's mRNA results in reduced expression of the gene product.

The genome-wide RNAi modifier screen in *C. elegans* was performed by Dr. Aileen Moloney, working in the laboratory of Professor David Sattelle in the MRC Functional Genomics Unit in The University of Oxford. The screen was undertaken using the RNAi library generated in *E. coli* and obtained from Source BioScience (formerly GeneService, Nottingham, UK). The library was first generated by Dr. Julie Ahringer laboratory in Cambridge (Kamath and Ahringer 2003).

Briefly, the construction of bacterial RNAi library first involves the PCR amplification of genomic DNA fragments from the target genes, then cloning those PCR products into the L4440 RNAi feeding vector and consequently transforming them into bacterial strain HT115 (Fraser, Kamath et al. 2000). Separate clones of the HT115 *E.coli* strain generated in this way were cultured overnight at 37 ^{o}C in 500 μl LB medium with 50 μg/ml ampicillin. Similarly, control *E.coli* containing empty plasmids (without an RNAi insert) were cultured simultaneously and under the same conditions. Those bacterial strains were transferred drop-wise onto solid NGM agar plates, containing 25 μg/ml carbencillin and 1 mM IPTG, and incubated overnight at room temperature. Worm embryos collected from synchronized worm cultures (Strange, Christensen et al. 2007) were transferred onto the NGM agar plates with either RNAi-expressing bacteria or control bacteria, and left for 48 h until they reach the third larval stage. The expression of Aβ in CL4176 *C. elegans* strain was induced by upshifting the surrounding temperature to 23 ^{o}C for 24 h and subsequently the worms were observed under the microscope and scored for the paralysis phenotype.

3.2.1.3 PRIMARY AND SECONDARY SCREEN TO DETECT RNAI MODIFIERS

After the induction of Aβ expression by temperature uplift, the effect of RNAi-expressing clone on the severity of resulting paralysis phenotype in *C. elegans* strains was evaluated in triplicate and compared with triplicate wells containing worms fed control bacteria with empty plasmid. Additionally, as a positive control for RNAi activity, each plate included a well with RNAi-expressing bacteria that targeted the *unc-22* gene, a regulator of muscle contraction; *unc-22* RNAi normally resulted in uncoordinated twitching behaviour. Therefore, any plate with worms that were fed *unc-22* RNAi bacteria, but nevertheless failed to twitch, was discarded and similar experiment repeated again.

The locomotor behaviour of worms was analysed in triplicate at two distinct time points. The first experiment, performed 22 hours post-induction, was aimed at identifying the enhancers of Aβ toxicity by checking if worms, assessed in each triplicate for each RNAi clone, were still paralysed under the conditions when they would otherwise be moving normally, for example when Aβ-expressing worms were fed control bacterial cultures. Moreover, only Aβ-expressing worms have displayed such a phenotypic modification, and none of the worms from control strain CL802, without the expression of Aβ, have demonstrated a similar paralysis. Therefore, clones that induce a non-specific paralysis phenotype in worms were not considered. At the later time point, at 32 hours, the Aβ-expressing worms were expected to be paralysed so any RNAi clones that protected the locomotor function of worms, in triplicate wells, were defined as suppressors of Aβ toxicity.

Secondary screen have been repeated three times for each modifier from the primary screen, each repeat was undertaken in triplicate wells. Hourly observations of locomotor behaviour were used to generate the paralysis curves, where at each time point the percentage of worms that have not yet been paralysed in each plate is plotted against time from temperature upshift initiation. These paralysis curves were later analysed using the Kaplan-Meier estimator, a non-parametric maximum likelihood estimator of the probability that a particular worm in a given population of worms will have not be paralysed after a particular time point. The statistical significance of the estimation was calculated using Mantel-Cox Log Rank test based on p-value cut-off of 0.05.

3.2.2 LIST OF GWAS FINDINGS IN AD

The list of GWAS findings in AD was obtained from the Database of Genotype and Phenotypes (dbGaP) of the National Center for Biotechnology Information (Mailman, Feolo et al. 2007); dbGaP is a public repository of phenotype, genotype and association data as identified by the published studies. The data were accessed in June 2013 using the keywords "Alzheimer disease" for the disease and the results were displayed only with p-values $<10^{-5}$ (http://www.ncbi.nlm.nih.gov/projects/gapplusprev/sgap_plus.htm).

3.2.3 HUMAN ORTHOLOGUES OF WORM GENES

In order to compare the list of genes from the worm screen with GWAS findings, the first step of the analysis was to assign a human orthologue to each worm gene that modified the paralysis phenotype in the screen. In general, orthologues are the genes in a different organism that have evolved from a common ancestral gene by a specification event and they normally maintain the same function (Koonin 2005). As mentioned before, the worm genes are highly conserved on molecular and cellular levels with mammalian genes.

The human orthologues of worm screen hits were identified using EnsemblCompara (http://www.ensembl.org), a phylogeny based orthology prediction method that collects the genomic information from external biological resources (Flicek, Ahmed et al. 2013). Only results with more than 25% of identity with respect to worm genes sequence were chosen (Lechner, Findeiss et al. 2011). Due to the definition of orthology in principal, evolutionary development of a certain gene can lead to the gene in one organism having multiple orthologues in another organism. So in a case when there was more than one human orthologue for a given worm gene, only the gene with highest percentage identity was chosen. However, in nearly 15% of cases there were multiple human orthologues with the same percent identity for a particular worm gene; in these cases a single orthologue was selected using MetaPhylogenyBasedOrthologs (MetaPhOrs) (Pryszcz, Huerta-Cepas et al. 2011) based on the consistency score. MetaPhOrs is a global public repository of phylogeny-based orthology predictions that calculates a consistency score based on the combined seven popular homology prediction sources of phylogenetic information. In MetaPhOrs, consistency score is a ratio of number of methods

confirming given relationship between particular protein pair and ranges from 0 to 1. This score is equal to 0 if none of the databases predict a particular orthology relationship and approaches 1 if predicted orthology is consistent among all the available data sources. Therefore, as a rule preferred orthologue out of multiple human orthologues for a given worm gene was chosen based on the highest consistency score.

3.2.4 GENE ONTOLOGY ENRICHMENT ANALYSIS

Gene ontology (GO), which consists of species-neutral terms representing gene product properties, is routinely used to describe gene product function or location. In order to interpret the functional profile of the worm screen hits, gene ontology enrichment analysis was implemented using the clueGO plugin for Cytoscape that analyses and visualizes a functionally organized biological term network for the list of genes (Bindea, Mlecnik et al. 2009).

3.2.5 GENERATION OF GENE NETWORK

The straighforward comparison between list of genes in GWAS white+grey zone and list of human orthologues of worm screen hits have demonstrated the lack of similarity, although we have discovered that 6 of the human orthologues of worm screen hits interact directly with genes in GWAS list. The comparison was performed using the BioGrid database (http://thebiogrid.org, v. 3.2.96) that contains the genetic and protein interaction data compiled from various studies and manually curated (Stark, Breitkreutz et al. 2006). The lists of genes in GWAS white+grey zone and the human orthologues of worm screen hits were analysed using data from the BioGrid database (v.3.2.96); specifically the primary lists of genes have been supplemented with genes encoding the first-degree physical interactors (+1 interactome) of the protein products of genes.

3.2.6 STATISTICAL SIGNIFICANCE OF OVERLAP

The statistical significance of the overlap between GWAS +1 interactomes and worm +1 interactome was determined by generating random lists of genes and comparing the experimental overlap with randomly generated. To do so, for each

worm screen gene we have randomly generated many +1 interactomes of the same size as the interactome of the gene being tested (Figure 3.2).

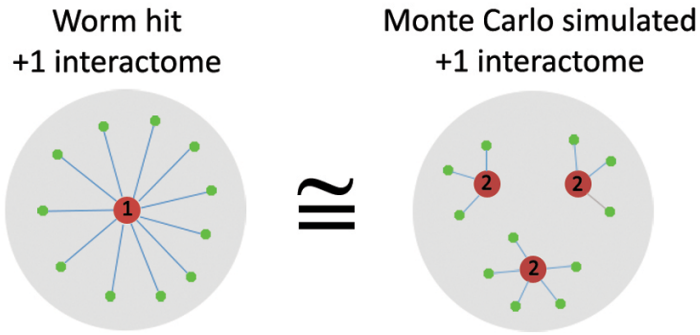

Figure 3.2 Generation of equivalent random +1 interactomes (Khabirova, Moloney et al. 2014). Random +1 interactome (right) was generated in a way that it contains the same number of genes (in this example, N=11). Human orthologue of worm-screen hits is labeled with "1" (left). Random human genes with a worm orthologue are labeled with "2" (right). The +1 interactors are represented by small green dots.

Furthermore, in order to obtain a statistically significant result we have generated 1000 random worm +1 interactomes of the equivalent size for each gene in the worm screen list. Then, for each gene in the worm list the p-value was calculated based on the number of times that the overlap of the GWAS +1 interactome with a random worm +1 interactome was bigger or smaller than the overlap with an interactome of the appropriate size for that particular gene. To estimate the significance of our findings, we have also repeated this analysis 100 times for 60 randomly generated worm genes (Figure 3.3).

Overall, the computational algorithm to estimate the statistical significance of the overlap between +1 interactome for any particular worm screen hit "*g*" and GWAS +1 interactome is depicted on Figure 3.3 and follows the steps below:

Algorithm 1: Estimating the statistical significance of the +1 interactome overlap

1. For each human orthologue of the hits from the worm screen, "*g*", the size of its +1 interactome "*N"* was evaluated.
2. Using a Monte Carlo approach, we have randomly generated the +1 interactome of size "*N*" (Figure 3.2). This goal was accomplished by repeatedly concatenating the +1 interactome lists of randomly-chosen human orthologues of worm genes until the accumulated random interactome reached "*N*" entries. If the +1 interactome of the first randomly selected worm gene exceed the size requirements, then only the first "*N*" genes in its +1 interactome were considered.

3. The overlap (number of identical entries in both lists) between +1 interactomes of randomly generated worm list and GWAS list was quantified.
4. Step 2 and 3 repeated N = 1000 times, producing 1000 Monte Carlo overlap scores "$MC_{i,g}$", where $i \in 1,2\ldots N$.
5. The random overlap score derived from this Monte Carlo approach "$MC_{i,g}$" is compared with the experimental overlap score "$Exp.score_g$" between +1 interactomes of worm modifier gene g and GWAS. The number of times that $MC_{i,g} \geq Exp.score_g$ is counted as "$n_{right,g}$" (Figure 3.4, panel A). Likewise, number of times when $MC_{i,g} \leq Exp.score_g$ is counted as "$n_{left,g}$" (Figure 3.4, panel B).

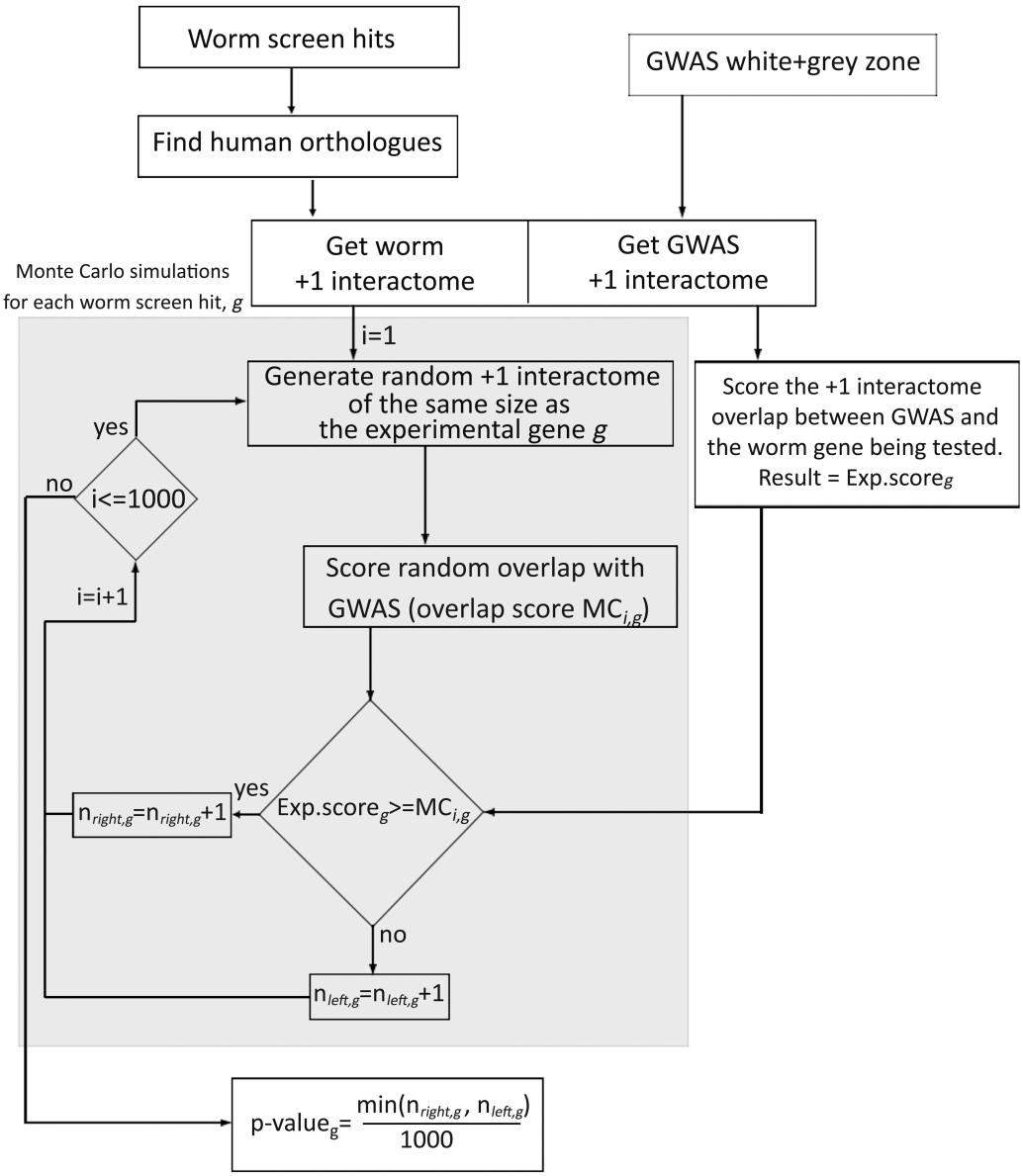

Figure 3.3 Computational pipeline for determining the significance of the overlap between the +1 interactomes of worm modifier genes and GWAS white+grey zone genes (Khabirova, Moloney et al. 2014).

6. The p-value of the +1 interactome overlap for experimental gene "*g*" was calculated as a ratio of either "$n_{right,g}$" or "$n_{left,g}$", whichever is smaller, to the total number of simulations *N*, thus simulating the one-tailed test.

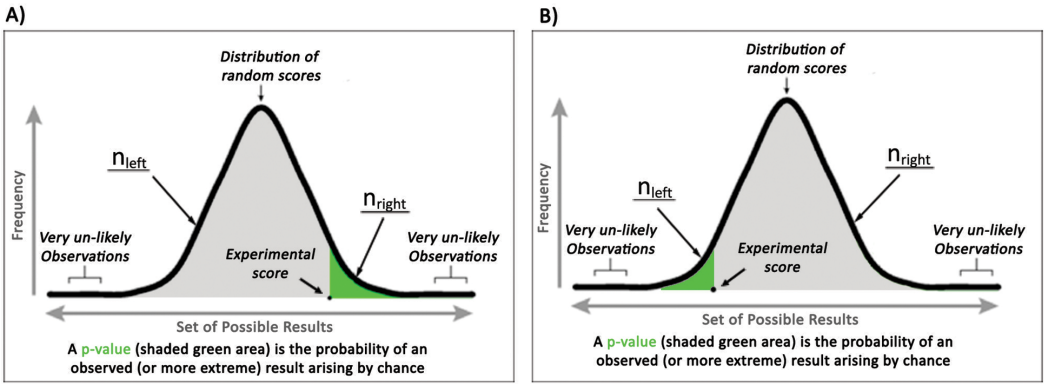

Figure 3.4 Estimating the significance of highly overlapping/poorly overlapping +1 interactomes (Khabirova, Moloney et al. 2014). The +1 interactome of human orthologues of worm genes can overlap more than expected by chance (panel A, contributing to n_{right}) or less than expected (panel B, contributing to n_{left}). The p-value of overlap can be derived by generating 1000 random +1 interactomes.

3.3 RESULTS

The expression of human Aβ in *C. elegans*, some but not all of which is secreted out of the muscle cells, leads to the aggregation of Aβ in muscles and results in progressive paralysis. Similarly, several studies have shown that Aβ is also deposited intramuscularly in Inclusion Body Myositis, a progressive human muscular disorder (Dostal and Link 2010). Moreover, the line that we used was designed to exhibit temperature-inducible expression of Aβ allowing toxic peptide expression to be initiated at specified times. The age-specific and synchronised expression of Aβ through temperature uplift results in a reproducible paralysis phenotype. Overall, exhaustive primary and secondary screens have been implemented to identify RNAi clones that significantly suppress or induce the paralysis phenotype in *C. elegans*.

3.3.1 RNAI WORM SCREEN

The identification of the pathways that are involved in the mechanism of Aβ toxicity in *C. elegans* can be achieved by employing genome-wide RNA-mediated

interference (RNAi) as part of a genetic modifier screen. In the genome-wide RNAi screen, the worms are fed bacteria (*Escherichia coli)* that carry RNAi-expressing plasmids which target a particular gene of interest; typically the specific destruction of the target gene's mRNA results in reduced expression of the gene products. Genes that modify the severity of the worm locomotor phenotype when knocked-down were considered "hits" in this screen.

It was determined that 7970 human protein-encoding genes had orthologues in *C. elegans*. The transcription level of each of those genes was systematically knock down by using the Ahringer *C. elegans* RNAi feeding library (Kamath and Ahringer 2003). After the induction of Aβ expression by temperature uplift, the resulting paralysis phenotype was evaluated in the presence of each genetic knockdown in triplicate and at two distinct time points. Identical *C. elegans* transgenic strains that did not express Aβ expression were used as controls.

It has been identified from the Kaplan Meier estimator plot that induction of Aβ expression in worms of 48 h old (in their third larval stage) results in 50% of the worm population being paralysed 30 h post induction. Thorough primary and secondary screens were used to identify the RNAi clones that either significantly suppress or enhance the Aβ-induced paralysis phenotype (p-value<0.05, 4 biological replicates). The modifier clones were subsequently positively identified by DNA sequencing. Out of the initial 7970 worm genes, 78 have been identified as significantly suppressing the Aβ-induced paralysis and only three genes specifically enhanced the Aβ-induced paralysis (Appendix A.1).

Furthermore, when we analysed the data looking for gene ontology enrichment (using the clueGO plugin for Cytoscape) we found a number of labels that were over-represented in the list of the worm screen hits, such as ATP synthesis, purine metabolism, protein binding and chaperone activity, and components of the translational machinery (Appendix A.2).

3.3.2 PREPARING THE LIST OF GENES FOR COMPARISON

Using the pipeline of identifying the human orthologues for worm genes described in 3.2.3, corresponding human orthologues were found for 61 out of 78 worm modifier genes.

The list of GWAS findings in AD with p-value<10^{-5} was obtained from dbGaP as described in 3.2.2; it consists of 63 genes (Appendix A.3) and will be referred below as "GWAS white+grey zone".

3.3.3 GENE INTERCONNECTIVITY (NETWORK ANALYSIS)

Given the list of 61 human orthologues of worm screen hits and comparing it with 63 genes in GWAS white+grey zone, we expect significant overlap between these two gene lists as the worm screen reports exclusively on Aβ toxicity and GWAS findings account for the genes involved in the pathogenesis of sporadic AD. However, we found no overlap between these two gene lists. As a matter of fact, the gene ontology annotations linked to the GWAS white+grey zone genes and to the human orthologues of worm modifier genes do not show any significant resemblance (gene enrichment analysis was performed using WEB-based GEne SeT AnaLysis Toolkit (Zhang, Kirov et al. 2005)).

However, if we consider the hypothesis that the GWAS finding in AD and worm screen hits might be related to each other on a higher level and act in the same pathways, we can gain some valuable insights into the role of Aβ in sporadic AD. Indeed, when we compared the direct physical interactors of genes in GWAS white+grey zone and human orthologues of worm screen hits, we have discovered that 6 of the human orthogues interact directly with genes in GWAS list. The comparison was performed using the BioGrid database (v. 3.2.96) that contain the genetic and protein interaction data compiled from various studies and manually curated (Stark, Breitkreutz et al. 2006).

In order to find the genes in the worm modifier list that interact with the genes in GWAS list in a statistically significant manner, we have combined the candidate gene lists with the corresponding genes encoding the first-degree physical interactors (+1 interactome) using BioGrid database as described in 3.2.4. Whereas not every gene was annotated in BioGrid database, 60 out of 61 worm genes and 52 out of 63 of GWAS gene lists were present. As a result, GWAS +1 interactome (Figure 3.5) contains genes in GWAS white+grey zone and all the genes (Figure 3.5, circles) with which they physically interact (Figure 3.5, blue lines). Accordingly, the worm +1 interactome (Figure 3.6) is composed of human orthologues of worm modifier screen

hits and their direct interactors. The gene networks were constructed and visualized using Cytoscape (Shannon, Markiel et al. 2003).

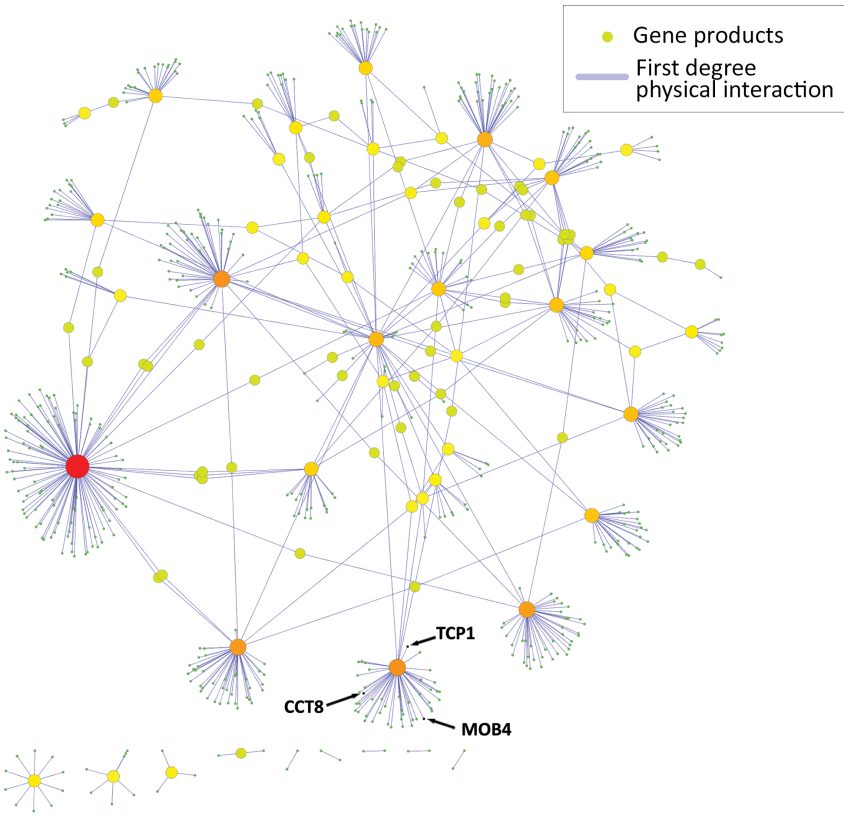

Figure 3.5 GWAS genes +1 interactome (Khabirova, Moloney et al. 2014)). The network contains the first-degree physical interactors (circles) of 52 GWAS genes with intermediate and high significance ($p<10^5$). There were in total 703 interactions (blue lines) in the +1 interactome of 52 genes in the GWAS white+grey zones. Circle size and colour is proportional to the number of physical interactions associated with each gene product, where green colour represents fewer interactions whereas red colour indicates higher interconnectivity.

It is clear that the two +1 interactomes (Figure 3.5 & 3.6) are different in term of relative size; genes in worm +1 interactome have many more interactive genes as compared to GWAS +1 interactome. In fact, on average each human orthologue of worm screen hits have 53 interactors while genes in GWAS list have less than 14.

Whether or not the human orthologues of worm screen hits were significantly more highly interconnected than one would expect by chance was determined by comparing the size of the observed +1 interactomes of these genes with the simulated +1 interactomes of a randomly generated list of genes.

Our approach was to rank all the genes in BioGrid database according to their degree of node (number of interactors). We then assessed whether the human

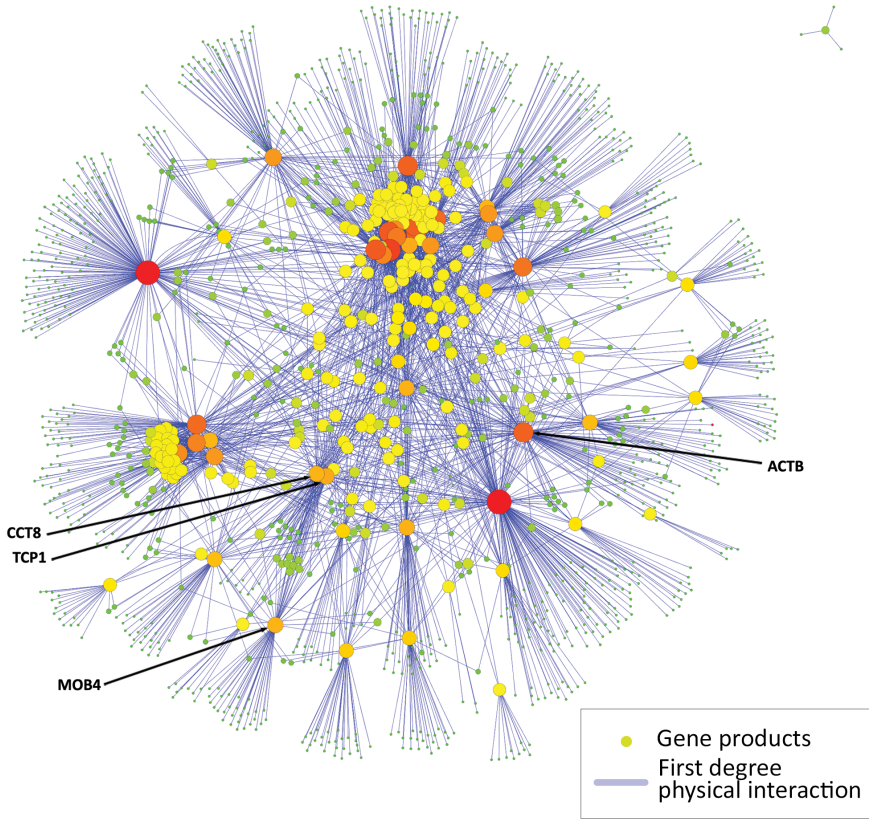

Figure 3.6 Worm genes +1 interactome (adapted from (Khabirova, Moloney et al 2014)). The +1 interaction network was constructed for the human orthologues of the 60 worm RNAi screen hits (each circles represent a gene product). There were in total 3191 interactions (blue lines) in the +1 interactome of the 60 worm-screen genes with human orthologues. Circle size and colour is proportional to the number of physical interactions associated with each gene product, where green color represents fewer interactions whereas red colour indicates higher interconnectivity.

orthologues of worm screen hits tended to partition into the more highly interconnected bins. However, the set of human orthologues of worm genes consisted of 7970 entries out of the 13928 genes in the BioGrid database at the time of the analysis. If we had simply clustered these 13928 genes into bins of equal size, then bias could be introduced if the human orthologues worm genes as a whole were more or less interconnected than the average human gene. We then clustered this ranked list of all human genes into 10 bins in such a way that each bin contains the same number of human orthologues of worm genes (specifically 797 genes in each bin). For instance, the most highly interconnected 952 human genes were assigned to the first bin, of these 797 were human orthologues of worm genes; in the same way the bin boundaries for the all the genes in BioGrid database were defined, resulting in bin boundaries as listed: 1-952, 953-2006, 2007-3175, 3176-4403, 4404-5735, 5736-

7134, 7135-8624, 8625-10200, 10201-12104 and 12105-13928. As a control distribution, we generated 100 lists containing 60 randomly selected human orthologues of worm genes using Monte-Carlo approach (Metropolis and Ulam 1949).

The distribution of gene numbers per bin is shown on Figure 3.7. The y-axis represents the logarithmic change of number of genes that falls into each bin as a percentage and x-axis shows the bin boundaries of gene ranking in a log scale. As anticipated, genes from randomly generated lists fall into each bin with equal frequency (Figure 3.7, triangles). By contrast, the highly interacting property of genes in worm lists create a skewed linear distribution with more than 50% of genes falling into the first bin with the most highly interconnected genes (Figure 3.7, diamonds).

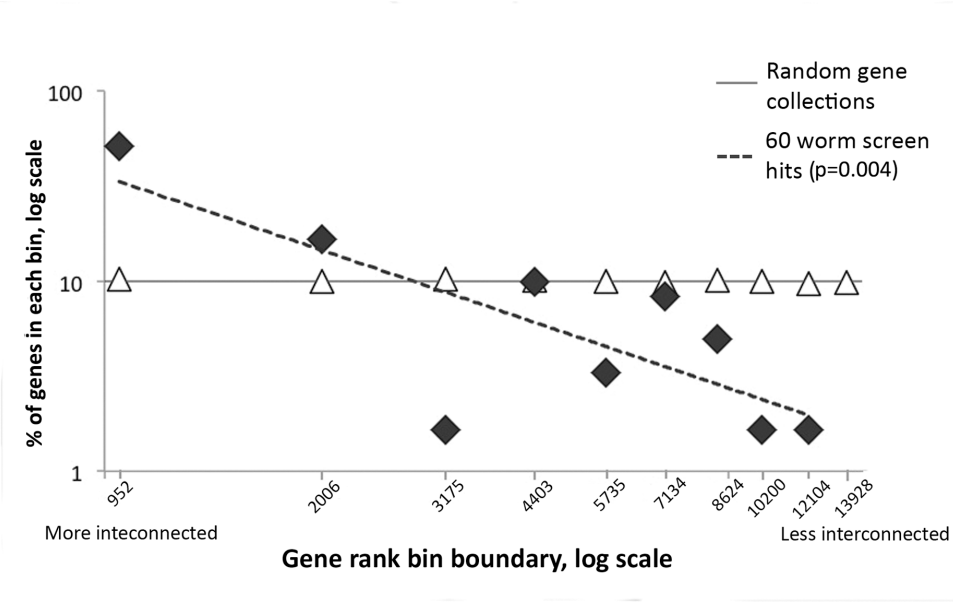

Figure 3.7 Distribution of rankings of worm genes (Khabirova, Moloney et al. 2014). The log of the gene ranking bin boundaries is presented in the x-axis in a decreasing order of gene-product connectedness from left to right. The y-axis represents the log of the fraction of genes that fall within each bin (where 100% is 60 genes). The dashed line represents the linear regression for the worm screen results. The results of the screens are shown as black diamonds and results of random simulations are shown as empty triangles

It is notable that the genes in worm list demonstrate the power law distribution (meaning that relative chance in one quantity results in a proportional relative change in other quantity) so that we can approximately fit a linear regression line (gradient -1.1, R^2 = 0.6) to a log-log plot of bin boundaries of gene ranks vs. frequency of observation. In order to establish if the distribution of worm genes interconnectivity is

significantly different from the random distribution we performed two-tailed Student's t-test (Rice 2006), comparing the gradient of the experimental regression line with the randomly generated data in the Monte Carlo simulations. The resulting p-value of 0.004 confirms that the genes in our worm screen list are indeed more highly interconnected as compared to a similar collection of random genes.

In order to check whether the bias towards the highly interacting genes is a common feature of worm RNAi screens, the equivalent analysis was repeated for three published worm RNAi screens of human neurodegenerative diseases (modelling polyglutamine aggregation (Nollen, Garcia et al. 2004), tau (Kraemer, Burgess et al. 2006) and alpha-synuclein pathology (van Ham, Thijssen et al. 2008) as mentioned in 1.1.2). The resulting distribution of gene interconnectivity in various worm screens is consistent with our results and characterized by the skew towards highly interconnected genes (Figure 3.8 A-C).

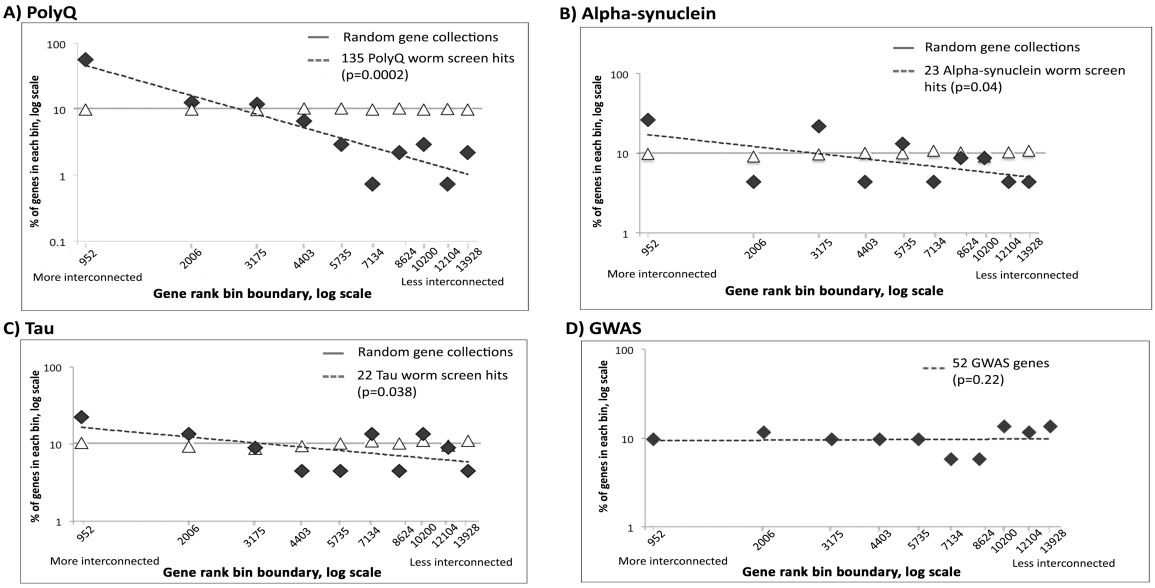

Figure 3.8 Distribution of rankings of worm models of neurodegenerative diseases (Khabirova, Moloney et al. 2014). A) polyglutamine screen; B) α-synuclein screen; C) tau screen; D) GWAS candidate white+grey zone genes for AD. The log of the gene ranking bin boundaries is presented in the x-axis in a decreasing order of gene-product connectedness from left to right. The y-axis represents the log of the fraction of genes within each bin (where 100% is 135, 23, 22 & 52 genes for panels A, B, C & D respectively). The dashed line shows the linear regression for the worm screen results. The results of the screens are shown as black diamonds and results of random simulations are shown as empty triangles.

Using one-tailed Student's t-test we found that the gradient of the fitted regression lines was significantly different from randomly generated data in all three worm screens, producing the p-value=0.0002 for polyQ screen (Figure 3.8, panel A), p-value=0.04 for alpha-synuclein screen (Figure 3.8, panel B) and p-value=0.04 for tau screen (Figure 3.8, panel C). By comparison, the gradient of the regression line fitted to the distribution of interconnectivity of genes in GWAS list was equal to zero, which is equivalent to the randomly generated population of genes (p-value=0.22, Figure 3.8, panel D).

As we have discovered, there exists the potential bias towards the highly interconnected genes in the results of all examined worm modifier screens, including our own. Under these circumstances, the straight forward analysis of the overlap between the worm screen and GWAS +1 interactomes is not possible due to the fact that some of the worm genes have many more interactors as compared to the genes in GWAS list and therefore they are likely to create the overlap simply by chance. For instance, a single gene in the list of human orthologues of worm screen hits, namely HSPA8, accounts for more than 30% of the overlap between worm and GWAS +1 interactomes. Therefore, a different method was developed, as described below, to compare the overlap of worm and GWAS +1 interactome and determine if it is significant or not.

3.3.4 COMPARING WORM AND GWAS INTERACTOMES

In order to compare worm and GWAS interactomes and overcome the bias towards highly interacting genes in the worm list, we have developed the following Monte Carlo method: for each worm screen gene we have randomly generated +1 interactome of the same size as the interactome of the gene being tested as described in 3.2.5. Briefly, the statistical significance (p-value) of the overlap between +1 interactome of each worm screen modifier hit and GWAS +1 interactome was estimated by first generating 1000 random worm +1 interactomes of the same size as the experimental gene. The p-value for each gene in the list of worm screen hits was determined by counting the number of times the overlap between experimental +1 interactome with GWAS +1 interactome was either greater or smaller that the overlap between a random +1 interactome with GWAS +1 interactome. Furthermore, to

evaluate the significance of these findings, the above-mentioned analysis was repeated 100 times for each of the worm gene, resulting in the generation of 100 random lists of 60 worm genes.

As a result, Monte Carlo simulations of 100 random lists of 60 worm genes have shown that on average 3.14 genes (confidence interval ± 0.34) exhibit the +1 interactome overlap with GWAS +1 interactome that is significantly different than expected by chance at the significance level p-value<0.05 (Appendix A.4). This result is consistent with the expected number of genes with non-random overlap with p-value<0.05 occurred by chance when performing multiple comparisons ($60 \text{ genes} \cdot 0.05 = 3 \text{ genes}$).

By contrast, the analysis of experimental worm screen list has demonstrated that 7 worm genes had a significantly non-random +1 interactome overlap with the GWAS network. Examining 100 runs of Monte Carlo simulations, we have discovered that only 5 simulations resulted in 7 or more genes having non-random +1 interactome overlap, giving the p-value=0.05 for our experimental observation.

3.3.5 GENES WITH SIGNIFICANT +1 INTERACTOME OVERLAPS

Looking closely into the 7 experimental worm genes with non-random +1interactome overlap, we have noticed that 3 genes (NOP56, PSMA1 and RPL10L) had a lower +1 interactome overlap with GWAS network that one would expect by chance (Figure 3.9). In comparison, the remaining four genes had the +1 interactome overlap with GWAS +1 interactome that is significantly larger than expected by chance.

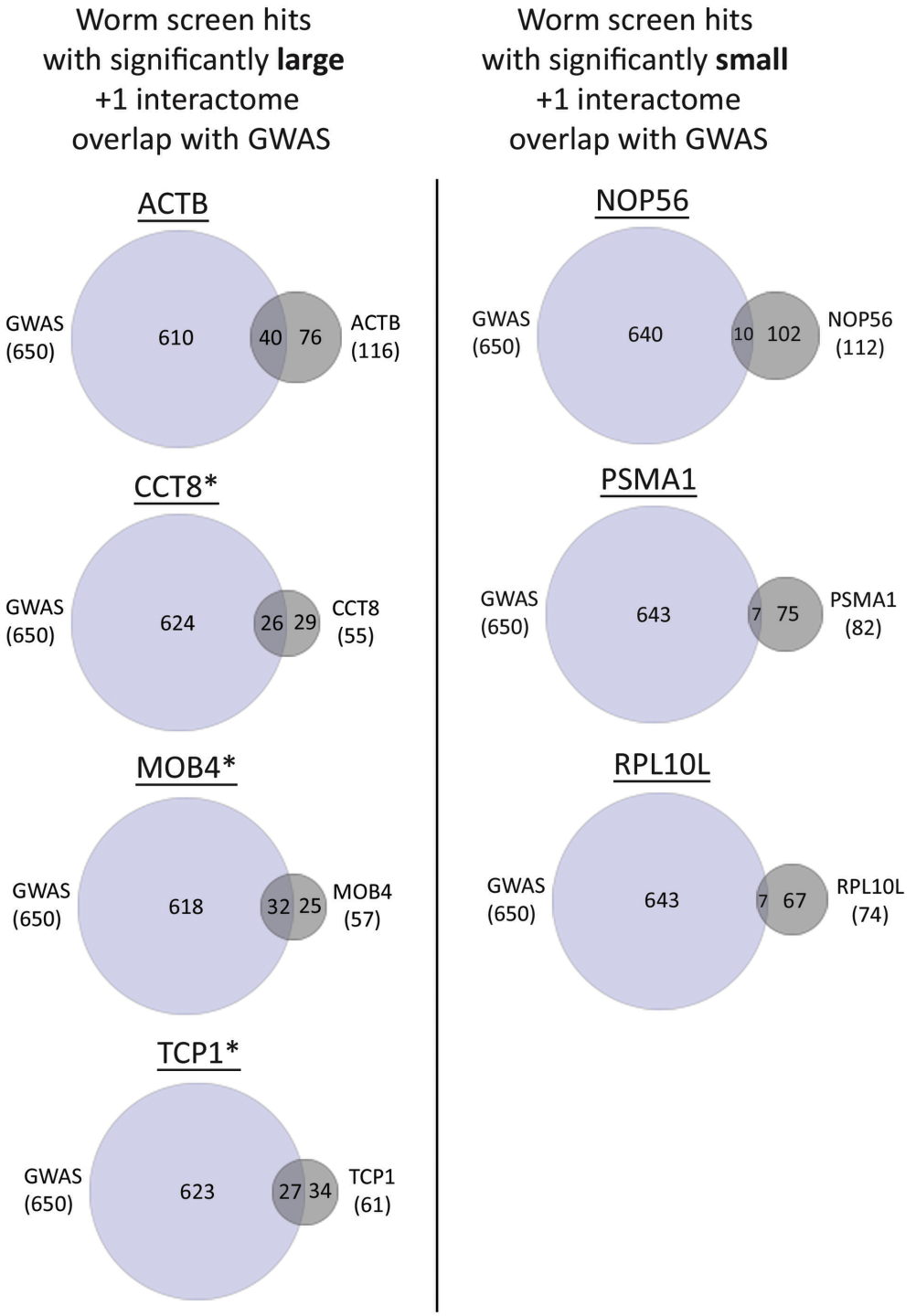

Figure 3.9 Venn diagram indicating the degree of the +1 interactome overlap for GWAS white+grey genes and each of the significant genes from the worm screen (Khabirova, Moloney et al. 2014). Seven genes from worm screen have been identified as having +1 interactome that overlap more or less than one would expect by chance with GWAS +1 interactome. Four of them (left panel) overlap more and three less (right panel) than expected. The area of the circle or overlap is proportional to the number of genes. The human orthologues of worm genes that interact directly with GWAS list are marked with asterisk.

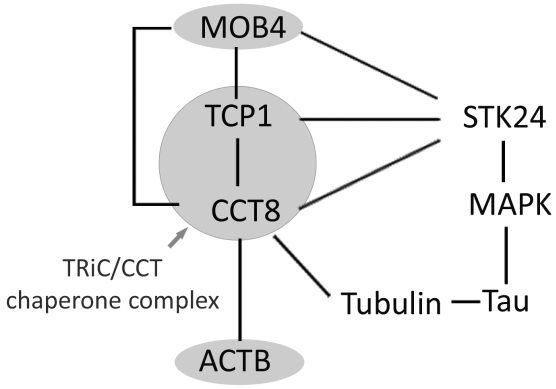

Figure 3.10 The interaction network of 4 significant worm genes (Khabirova, Moloney et al. 2014). Four human orthologues of worm screen hits (MOB4, TCP1, CCT8, ACTB) have +1 interactome that overlap more than expected with +1 interactome of GWAS gene list. Two of those genes form the abundant cytoplasmic TRiC/CCT chaperone complex, third gene (MOB4) interact closely and fourth (ACTB), along with tubulin, is a major substrate of the same chaperone complex. TRiC/CCT interacts with STK24 and PP2a to form a complex that regulates the MAPK pathway. This network of interactions may have a bearing on tau phosphorylation and neuronal regeneration.

Surprisingly, three of those genes (TCP1, CCT8 and MOB4) are constituents of the same chaperone complex TRiC/CCT and fourth gene actin (ACTB) is a major folding substrate of that particular chaperone complex (Goudreault, D'Ambrosio et al. 2009, Cong, Baker et al. 2010). Moreover, all three genes that constitute the chaperone complex have a direct interaction with a single gene STK24 in GWAS grey zone gene list (Goudreault, D'Ambrosio et al. 2009) as shown on Figure 3.10.

3.4 DISCUSSION

Genome-wide association studies have provided some insight into the pathogenic pathways and genes involved in sporadic AD but the exact pathogenesis of this disease is still unknown. However, none of the genes implicated in familial AD are associated with the risk variants according to the number of GWAS in sporadic AD. These observations raise doubts about the involvement of Aβ toxicity, which clearly plays a role in the pathogenesis of familial AD, in a more complex case of sporadic AD. In order to further investigate this issue, we have used an invertebrate

model system reporting exclusivity on Aβ toxicity. This was done by implementing the genome-wide RNAi screen in *C. elegans* and finding the set of worm gene transcripts that modify the Aβ-induced paralysis phenotype.

The goal of current analysis was to compare the results of worm modifier screen with GWAS finding in sporadic AD of borderline significance. The initial straightforward comparison has shown that none of the genes in GWAS white+grey zone is identical to the 61 human orthologues of worm modifier genes. The absence of genes associated with Aβ toxicity may suggest that, at least for predominantly Caucasian population, sporadic AD is governed by different pathogenic mechanism. Nevertheless, we have noticed that human orthologues of 6 worm screen hits presumably had direct interactions with genes in GWAS white+grey zone. So we decided to further investigate in this direction.

The first noticeable property of the worm modifier genes was that they are more likely to have a human orthologue than expected by chance. This high level of conservation may implicate the involvement of core and essential genes in promoting the profound toxic effect of Aβ aggregates.

Furthermore, examining closely the +1 interactome of worm screen hits reveals the highly interconnected nature of these genes. In particular, the distribution of interconnectivity according to the gene ranking has clearly shown the power law relationship as can be seen in Figure 3.7. For the worm screen hits, the gradient of regression line fitted to this distribution was significantly different from the gradient of random distribution, where the frequency of occurrence of a gene is independent of its interconnectivity.

We have additionally examined the genes from other worm modifier screen of neurodegenerative diseases (polyQ, alpha-synuclein and tau) and have observed similar skewed distributions of interconnectivity. One potential explanation of the bias towards highly interacting properties of worm screen hits could be outlined as a "sociological bias" (Chen, Aronow et al. 2009). The main idea of this hypothesis is that the previous worm screens, which identified similar set of genes, have promoted further interest in those genes within the field and which has in turn lead to more interaction data being collected and deposited in BioGrid database. Indeed, we have found that worm screens for Aβ and polyQ have produced 27 common genes, with 20 of them having highly interconnected properties (fall into first bin in plots on Figure

3.7 & 3.8). However, the presence of sociological bias does not discredit our subsequent results as we have carefully compensated for any differences in network connectivity.

On the other hand, the alternative hypothesis could be that the genes reported by the worm screen do indeed have more highly interacting protein products as compared to a random set of human orthologues of worm genes (Goh, Cusick et al. 2007). It is likely that these highly interconnected gene products, also known as protein hubs, have a higher chance to be involved in various biological pathways so that they are more likely to modify the disease-related phenotypes in our worm genome-wide RNAi screen. Interestingly, the genes in GWAS white+grey zone do not have the same highly interacting properties, with their distribution of interconnectivity is being similar to the one of random population of genes. However, taking into account that we have included the genes with borderline significance to GWAS significant findings, this could be explained be the presence of false-positives genes.

In order to evaluate the statistical significance of the overlap between the worm and GWAS +1 interactomes we have developed the algorithm as described in Figure 3.3. In brief, for each gene in the worm screen hits list the significance is determined by observing to what extent the overlap of its +1 interactome with GWAS +1 interactome is greater or lesser than expected by chance. We have used Monte Carlo approach to generate 1000 random +1 interactomes of the same size as the experimental interactome and compute the distribution of overlap scores. As the result, 7 genes out of 60 experimental worm genes had the significantly different overlap from what one would expect by chance.

In particular, three of those genes have +1 interactome overlap with +1 interactome of GWAS white+grey zone genes smaller than expected by chance, indicating that they could be the false positive modifiers of the worm phenotype and probably not involved in the human disorder. On the other hand, the remaining four genes have shown a greater than expected +1 interactome overlap with genes in GWAS white+grey zone. Moreover, two of those genes (TCP1 and CCT8) are both components of the same TRiC/CCT chaperone complex and third gene actin beta (ACTB) closely interact with it. It is noteworthy that two genes, namely TCP1 and ACTB, were identified as modifiers of polyQ toxicity in the RNAi worm screen (Nollen, Garcia et al. 2004) whereas CCT8 and MOB4 have not been observed in any

worm modifier screens modelling neurodegeneration. This might indicate that Aβ and polyQ share common pathological mechanisms, but might also mean that the modifiers of TCP1 and ACTB have non-specific effect in the worm model systems. It is also worth mentioning that TCP1 level are significantly reduced in foetal Down's syndrome (Yoo, Fountoulakis et al. 2001). Besides, TCP1, CCT8 and MOB4 all interact directly with each other and additionally with the kinase STK24, one of the genes in GWAS grey zone, which has been associated with axonal regeneration in rats (Lorber, Howe et al. 2009).

The exact role of TRiC/CCT in the pathogenesis of sporadic AD is still unknown. Nevertheless, the fact that RNAi knockdown in the worm protects against Aβ-induced paralysis may indicate that this particular chaperone complex is somehow mediate Aβ toxicity. However, several studies have shown that the reduction of TRiC/CCT chaperone activity has a protective effect (Guisbert, Czyz et al. 2013) and speculated that the lower levels of misfolded proteins could activate, or maintain, the activity of regenerative pathways.

3.5 CONCLUSION

To summarize, the failure to find a direct overlap between genes implicated by GWAS in AD and worm modifier screen hits may suggest that Aβ toxicity cannot be considered equivalent to AD. However, implementing the +1 interactome network analysis has revealed 7 genes whose +1 interactomes overlap with the GWAS genes cannot be explained by chance. Four of these worm genes have exhibited a much bigger overlap with the GWAS network than expected by chance; three of these are part of cytoplasmic chaperone complex and the fourth is an important substrate of the same chaperone.

CHAPTER 4: STUDY OF CIRCADIAN DISRUPTIONS IN ALZHEIMER'S DISEASE USING *DROSOPHILA* AS A MODEL ORGANISM

It is widely known that Alzheimer's disease (AD) is associated with memory problems, but the fact that AD is characterized by sleep disturbances as well is often omitted. Changes in sleep and behaviour patterns usually present themselves in the earliest clinical stages even before the memory problems start (Musiek 2015). Furthermore, several studies have suggested that circadian dysregulation may contribute to the pathogenesis of AD (Coogan, Schutova et al. 2013, Chen, Possidente et al. 2014). However, examining the reciprocal relationships between aging, neurodegeneration and circadian behaviour has proved to be a complicated problem due to the challenges of performing the experiment at different levels of biological system in a freely moving organism. In this chapter, I will describe novel collaborative method we have developed to capture molecular and behavioural circadian oscillations simultaneously for individual animals and analyse their relationships in terms of rhythm, period and amplitude stability.

4.1 INTRODUCTION

As mentioned before in 1.4.2, patients with AD often have disorganized daytime activity patterns and fragmented sleep. Until quite recently, these sleep and circadian disruptions were considered to be a symptom of the disease, yet, increasingly, the evidence indicates that behavioural and circadian abnormalities precede other symptoms and most likely contribute to the pathogenesis of AD (Musiek, Xiong et al. 2015).

Generally, AD patients experience disturbances in timing and duration of the sleep cycle, mainly in terms of increased night time wakefulness and daytime sleepiness. Moreover, it has been shown that the duration of rapid eye movement (REM) stage of the sleep is significantly decreased in patients with AD as compared to the age-matched healthy people, resulting in the cumulative lack of REM sleep (Prinz, Peskind et al. 1982). In support of this observation, several studies in

transgenic mice have demonstrated that accumulation of Aβ peptides in the brain can induce the disruption of sleep patterns even before the characteristic emergence of amyloid plaques (Musiek, Xiong et al. 2015). In addition, a recent neuroimaging study supports the association between shorter sleep duration and increased Aβ depositions in 70 older adults (Spira, Gamaldo et al. 2013).

4.1.1 CIRCADIAN RHYTHMS IN MAMMALIAN SYSTEMS

Circadian rhythms are endogenous and entrained oscillations of any biological process with a period of around 24 hours. A wide range of biological processes and activities are controlled by circadian rhythms, including locomotor and feeding behaviour and sleep. The endogenous oscillations are driven by molecular and cellular circadian clocks, which in turn receive daily corrective signals from the environment. Such timekeeping mechanisms help maintain homeostasis and synchronize the behaviour with cycles of day and night; providing a competitive advantage to the organism by allowing it to predict changes in the physical and social environments. The sleep/wake cycle is an important example of a circadian rhythm in all mammals including humans.

In mammals and other organisms circadian rhythms are manifest at the full range of biological levels starting with transcriptional and translational regulation of genes and proceeding through specific neuronal circuits that underpin the clock through to the highest levels of behavioural output. At the molecular level, circadian rhythms are governed by internal autonomous circadian oscillators that form negative and positive autoregulatory feedback loops. In mammals, the correct functioning of the oscillator is sustained by the so-called 'master', or central, clock, located in suprachiasmatic nucleus (SCN) of the hypothalamus. In addition, peripheral organs and tissues of the body also contain molecular clocks, which are controlled, by a greater or lesser extent, by the master clock (Mohawk, Green et al. 2012). The master clock receives light signals from the retina and synchronizes the peripheral clocks to the external subjective day. Even though molecular clocks located in the SCN and throughout peripheral tissues are thought to have essentially identical molecular architectures, they have a different degree of intercellular coupling. In particular, neurons in SCN have a higher degree of intercellular coupling and so form a neuronal network that is unaffected by the phase perturbations from internal cues, while

peripheral clocks are receptive to adjustments from the SCN clock. As a result of the tight coupling of neuronal network in the SCN, the master clock is able to faithfully maintain an intrinsic rhythm of about 24 hour (Partch, Green et al. 2014).

4.1.2 CIRCADIAN RHYTHMS IN *DROSOPHILA*

Recent studies have shown that circadian clocks are highly conserved throughout the animal kingdom, from flies to humans, and have allowed investigators to perform genomic analysis of circadian oscillations at the molecular level using animal model systems. In particular, much of the progress in this field has been driven by experimental work in *Drosophila melanogaster* (common fruit flies (Prussing, Voigt et al. 2013)). Moreover, the first clock gene, called the "Period" gene, was discovered in 1971 in a mutagenesis screen in *Drosophila* (Konopka and Benzer 1971). This discovery has further on led to the finding of other genes involved in the circadian regulation.

As a result, the basic molecular mechanisms of biological clocks have been studied comprehensively in *Drosophila* and there are many tools available to analyse their complex behaviour at the molecular and behavioural levels. Additionally, fruit flies have circadian patterns of locomotor behaviour that are easily measurable. For example, they have two peaks of locomotor activity (morning and evening) and it has been shown that these oscillations remain rhythmic even when external entraining cues are removed, such as in constant darkness (Peschel and Helfrich-Forster 2011).

In *Drosophila* the central clock relies on the daily expression rhythms of the clock proteins period (PER) and timeless (TIM) as shown on Figure 4.1.

Briefly, heterodimers of the transcription factors CLOCK and CYCLE (CLK/CYC) bind to the highly conserved Enhancer box (E-box), which acts as a promoter and drive the initial gene transcription of PER and TIM. As PER and TIM accumulate in the cytoplasm, they bind together and form dimers. The PER/TIM dimers translocate into the nucleus and bind to CLC/CYC and thus inhibiting further transcription of PER and TIM. This results in a decline of PER and TIM mRNA levels, until the inhibition is removed and cycle starts again. As such, the levels of TIM and PER are controlled by a delayed negative feedback, which means that the protein products of these genes in fact turn off production of more protein production and therefore generate an oscillation over time.

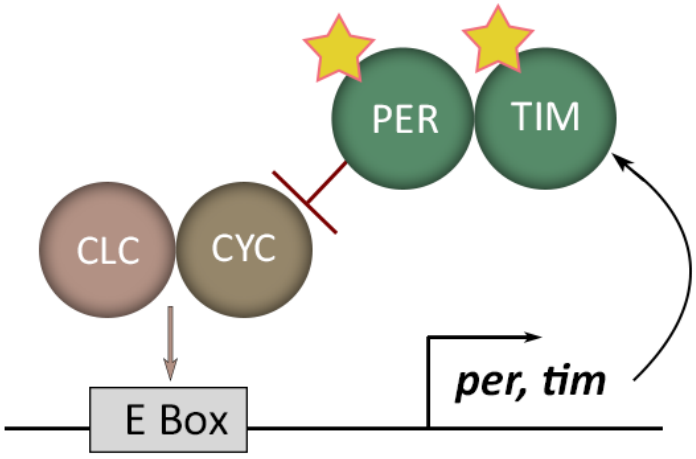

Figure 4.1 Transcriptional-translational negative feedback loop of central clock regulation in *Drosophila*. Hererodimeric transcription factors CLC and CYC activate transcription of PER and TIM by binding to E-box. Stars represent sites of post-transcriptional control.

The oscillations of circadian rhythms can be visualized non-invasively in *Drosophila* by expression of a firefly luciferase reporter construct under the control of the promoter region of a clock protein, such as PER. A particularly effective reporter consists of the PER protein fused to luciferase, known as *XLG-luc2* (Veleri, Brandes et al. 2003). Feeding these flies with the luciferase substrate (luciferin) results in bioluminescence (light emission), with the intensity of emitted light reporting the level of the fusion protein (PER). This bioluminescence signal reports the cycling of the molecular clock in every cell where it is expressed (Brandes, Plautz et al. 1996).

At the neuronal level, the central clock in *Drosophila* is positioned in the small ventral lateral neurons (LNvs) located in the lateral brain. Similar to the SCN, light signals transfers to LNvs from retinal photoreceptors in the compound eyes and, additionally, from extra-retinal photoreceptors within the brain. Interestingly, under constant dark condition, the central clock in the fly brain is able to maintain robust rhythms of gene expression and locomotor activity. Some studies have shown that circadian clocks in peripheral tissues in flies can also sustain robust rhythmicity for at least 7 days in cell culture, implying that fly peripheral clocks are not entirely dependent on the central clock found in the LNvs; in any case they are less closely coupled to the central clock as compared to mammalian systems. In particular, it appears that light can penetrate the fly cuticle in peripheral tissues and entrain peripheral clocks directly, therefore revoking the necessity of central clock to synchronize other oscillators (Glossop and Hardin 2002).

4.1.3 PERIPHERAL AND CENTRAL CLOCK COMMUNICATION

Several lines of evidence have shown that in mammalian system circadian clocks comply with the "master-slave" theory, in which peripheral clocks are tightly coupled to the central clock and synchronized solely by it in a hierarchical order. According to this theory, the central clock in the SCN receives external entrainment cues, mainly light, and projects its rhythm onto cell-autonomous clocks in peripheral tissues. As a result, peripheral clocks are synchronized by the central clock with a constant phase delay. In mammals, the central clock in the SCN is able to entrain the peripheral oscillators with via the endocrine and autonomic nervous systems, both directly and indirectly. For example, it has been shown that in mice hormonal signals, modelled using glucocorticoid agonists that are normally sent by central clock, are capable of shifting the phase of peripheral oscillators (Balsalobre, Brown et al. 2000). A more indirect entrainment mechanism, for instance of the liver, is possible by the SCN controlling rest/activity cycle and consequently the feeding schedule. Therefore, it is possible to uncouple the peripheral clocks from SCN by artificially setting the timing of feeding (Reppert and Weaver 2002). Several studies have confirmed that temporal feeding restriction for several days can lead to the desynchronization of SCN and peripheral circadian oscillations, shifting the phase of clock gene expression in peripheral tissues but not in SCN (Stokkan, Yamazaki et al. 2001, Yoo, Yamazaki et al. 2004)

Opposite to the mammalian system, it has been suggested that in simpler organisms, such as *Drosophila*, circadian rhythms can follow a so-called "orchestra" model, where the central clock is considered as a conductor and each of the peripheral clocks as an orchestra member. In this scenario, each peripheral clock can play its own "instrument", or have its own rhythms, but nevertheless they are guided by the "conductor" to ensure the efficient and coordinated melody. Specifically, as mentioned before, most of the cell-autonomous clocks are light sensitive and can be individually entrained by the environment, meaning that the peripheral clocks can have their own rhythms according to the external or internal stimuli. Yet their oscillations maybe coordinated to varying degrees by the light-dark cues received by the central clock (Richards and Gumz 2012). One example of how a central clock may facilitate a functional, whole-organ rhythm has been illustrated by (Welsh, Yoo et al. 2004), where by using bioluminescence imaging of fibroblast cells they

observed that individual fibroblasts can maintain robust rhythms with undiminished amplitude but failed to synchronize to each other. Similar experiments, but in vivo, have shown that in SCN-lesioned hamsters the rhythmic oscillations were eliminated not only for circadian behaviour, but also for the expression of clock genes in peripheral tissues; this observation suggests that loss of rhythmicity for an organ may not be due to loss of the molecular rhythms within each cell but rather due to the loss of a conductor resulting in cell-cell desynchronisation. In the absence of a central clock the progressive desynchronisation within an organ is inevitable if coupling between cells is weak (Welsh, Yoo et al. 2004). By comparison, the lesions of the SCN in mice did not destroy circadian oscillation in peripheral tissues, but rather caused the loss of phase coupling within different tissues of an animal and between animals (Yoo, Yamazaki et al. 2004). This indicates that coupling between cells within murine organs is stronger.

With reference to AD pathogenesis, circadian rhythms in humans, for example sleep-wake cycle, hormone secretion and blood pressure, are directly controlled by central clock in SCN. The control centre in SCN sends the instructions via various mechanisms, in particular through autonomic nervous system and hormone release, to synchronize peripheral clocks in the body. For instance, the secretion of melatonin, also known as "the hormone of darkness", regulates the sleep onset and several studies have reported that in mice the damage to SCN results in the disruption of circadian rhythms and consequently sleep-wake cycle (Musiek, Xiong et al. 2015). Postmortem studies of the brain of patients with AD imply that the loss of neuronal cells in SCN is linked to circadian abnormalities. However, it is still not clear at which stage of the signal transmission the damage occurs, in particular whether the damage is inflicted on the central clock itself or the communication pathway leading to the periphery. Interestingly, the recent study by (Chen, Possidente et al. 2014) has shown that despite maintaining an intact central clock, Aβ overexpression in transgenic *Drosophila* lead to age-related circadian behavioural disruption. Taking into account that circadian behaviour is driven by the central clock in SCN, this discovery might suggest that the damage caused by Aβ toxicity affect the communication pathways between different tissue clocks.

This observations bring into question the mechanism of coupling between central and peripheral clock, and, consequently, behavioural rhythms. For example, it

is still unclear to what extent the peripheral clocks are being controlled by the central clock. Understanding the pathway by which the central and peripheral clocks communicate with each other and control behaviour can potentially shed the light on the role of circadian disruptions in the complex pathogenesis of Alzheimer's disease.

4.1.4 METHODS TO MEASURE CIRCADIAN CLOCKS AND BEHAVIOUR

At the present, assessing the circadian biology in human subjects with AD is limited to largely non-invasive techniques, but several methods exist to do so in animal models of the disease, with *Drosophila* being arguably the most convenient. The first and most commonly used method to non-invasively measure the oscillations of molecular rhythms *in vivo* was developed by (Brandes, Plautz et al. 1996) with the use of bioluminescence, as mentioned in 4.1.2. In this method, the transgenic firefly luciferase gene was fused with the promoter region of one of the clock genes in *Drosophila.* When the organism is fed the luciferase substrate (luciferin), it emits light according to the rhythmic expression of luciferase and therefore, to the oscillations of clock gene expression. Such measurements are usually undertaken in constant darkness, practically this is because of the low intensity of the bioluminescence. Generally, two methods have been developed to automatically monitor the luciferase oscillations: using a luminometer or a charge-coupled device (CCD) camera.

The first robust and frequently used method to observe the bioluminescence is based on using the Packard Topcount multiplate scintillation counter by Perkin Elmer, in this case functioning as a "luminometer", in which the flies are placed in individual wells in a 96 well plate each of which is prefilled with fly food doped with luciferin. The light readout of each well containing a fly is then measured at the regular intervals by photomultiplier tube-based luminescence counter in which absorption of a photon results in the emission of many electrons. This method is an elegant approach for quantitatively monitoring the circadian rhythms but unfortunately is incapable of studying their circadian behaviour at the same time. This is due to the fact that flies are kept in circular domes and practically immobilized.

Recent developments in imaging technology have opened up new avenues to analyse the circadian bioluminescence with the use of CCD cameras, allowing the highly sensitive imaging with low background noise. Capturing the images with CCD

cameras is based on the photoelectric effect, converting the photons that impact individual spots on a silicon chip into free electrons and channelling them in a controlled fashion to quantify the number of electrons at each picture elements (pixels). The sensitivity of the camera or proportion of incoming photons being detected is expressed as quantum efficiency and depends on wavelength of light, which is 560 nm for firefly luciferase. Occasionally, some of the free electrons are thermally generated in the absence of incoming photons and give rise to dark noise. The amount of dark noise is proportional to the exposure time, so that the long-exposure images will contain more of the dark noise. Another technical issue of CCD cameras is the presence of readout noise due to the uncertainty during the process of quantifying the electronic signal. Recent improvements in CCD technology have led to the development of cameras suitable for low-light luminescence images with lower noise levels. One of the commonly used scientific camera is electron-multiplying CCD camera (EM-CCD), where the electrons are multiplied on-chip prior to readout event, so that the readout noise is negligible in comparison to pre-amplified low-light level signal. Overall, the equipment is comparatively expensive but a correctly chosen camera allows a highly sensitive and low-noise detection of dim signals. However, none of the existing CCD camera based methods reviewed to date have been able to monitor circadian clocks and circadian behaviour at the same time, at least in *Drosophila.*

On the other hand, there are many techniques have been developed to track the fly movement in constant darkness. The easiest and traditionally used method was first introduced in 1971 in (Konopka and Benzer 1971) and called infrared beam split system. The automatic data collection monitors were later on developed by Trikinetics, USA and termed "*Drosophila* activity monitor" (DAM). This method analyses fly movement by quantifying the number of times each individual fly breaks an infrared beam located in the middle of the tube (Figure 4.2). According to this technique, each individual fly is placed in a long capillary tube that contains cotton plug on one end and food on the other end. Taking into account that fly circadian clock is not sensitive to infrared light, an infrared beam is shone from one side of the middle of the tube and detected by a sensor on the other side. When a fly crosses the beam, it disrupts the passage of the infrared light and generates a binary signal that is registered by a computer. The result is usually organized into half an hour bins, in which each bin represented by the number of times fly broke the infrared beam.

However, the main drawback of this technique when applied to locomotor activity and especially sleep analysis is the extremely low spatial resolution as the system only detects the movement when the fly crosses the midline of the tube. Moreover, it has been shown that the DAM system strongly overestimates sleep, particularly in case of single-beam system. Currently, there are a variety of methods used to register the locomotor activity of flies, most of which based on using the infrared camera and red LED illuminators (Donelson, Kim et al. 2012, Gilestro 2012, Oren, Tarrant et al. 2015). Despite the wide range of fly movement video tracking methods to date, none of them are capable of simultaneously monitoring the oscillation of circadian clocks.

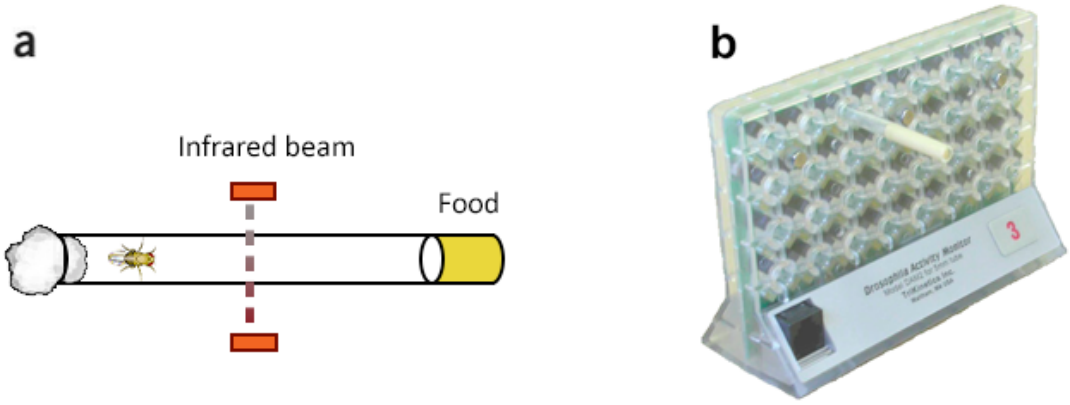

Figure 4.2 The *Drosophila* activity monitor system. a) Infrared beam-split system. The activity of a single fly is estimated by counting number of times a fly breaks the beam of infrared light in the midline of the tube (red dashed line) by passing through it. b) *Drosophila* activity monitor by TriKinetics, USA. The activity monitors have the infrared emitters and receivers and are connected to the computer to record the

Therefore, the current experimental paradigm consists of measuring the oscillation of central/peripheral clocks for one subset of flies and locomotor behaviour for another subset of flies. As a result, assessing the relationship between different oscillations is only possible on the population scale and has large inter-individual variability. To overcome this technical issue, the new method of simultaneously measuring circadian clocks and locomotor behaviour, together with sleep structure, was developed to study the circadian disruptions in *Drosophila* model of AD.

4.2 MATERIALS AND METHODS

Recent advances in bioluminescence imaging has led to new techniques being developed to study how the central clock synchronizes oscillators in peripheral tissues

in mammalian systems (Saini, Liani et al. 2013), but no similar method has ever been developed for the *Drosophila* model. Therefore, the purpose of the project was to develop the system that will allow us to examine both the molecular clocks and the locomotor activity for the same individual organism, namely fruit flies.

4.2.1 EXPERIMENTAL DESIGN

The fruit flies for the first and second datasets were prepared by a postdoctoral fellow in Damian Crowther's group (Department of Genetics, University of Cambridge), Kofan Chen, as part of a collaboration. With his training I subsequently prepared the flies and performed the third run of the experiment. The recording was undertaken using the EM-CCD camera incorporated within a Cairn Alligator system in Laboratory for Molecular Biology, together with the collaborator Dr John O'Neill.

4.2.1.1 FLY STRAINS

A model of the toxicity of the amyloid beta (Aβ) peptide in Alzheimer's disease has been generated in *Drosophila* using the GAL4/UAS system (Figure 4.3) that allows overexpression of human transgenes in a specific tissue. This system has two parts: the GAL4 gene, encoding the yeast transcription activator protein Gal4, and the UAS (Upstream Activation Sequence), an enhancer to which GAL4 binds in order to activate gene transcription. As AD is a neurodegenerative disease that is thought to result from the overproduction and aggregation of the Aβ peptide; consequently the peptide, linked to a secretion signal peptide, was overexpressed using the pan-neural promoter elav (embryonic lethal, abnormal vision), so that the GAL4 expression is driven by the elav promoter and expressed specifically in the neurons of fly brain. The presence of GAL4 in those cells has little or no effect, since GAL4, being a yeast transcriptional activator, is inactive in insect cells in the absence of a cognate UAS region.

In the reported line the Arctic variant of Aβ peptide 1-42 ($A\beta_{42}$), as a model of amyloid toxicity, is placed downstream of a UAS activation domain that consists of GAL4-binding sites. This genetic construct is carried by every cell of the fly, but in the absence of GAL4 nothing happens, because the transgenic promoter is not activated. The transgenes are only activated by crossing flies from the elav-GAL4

driver line to flies with UAS-transgenes. Their offspring will express the human pathogenic transgene $A\beta_{42}$ Arctic in all the neurons of the brain and so exhibit the symptoms of the disease (St Johnston 2002).

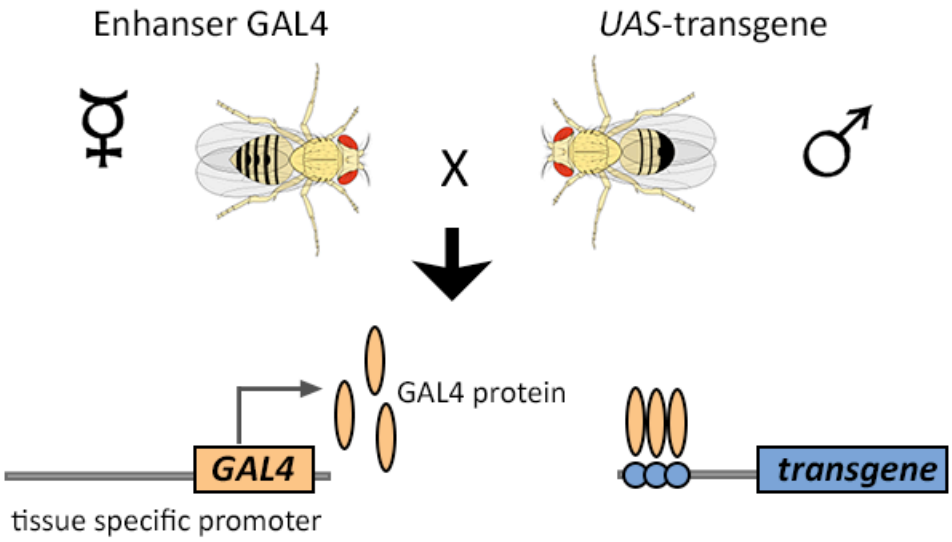

Figure 4.3 GAL4/UAS system. The GAL4 gene have been inserted at certain position in *Drosophila* genome so that GAL4 is expressed under the control of genetic enhancer. The expression of human transgene can be activated by crossing the GAL4 enhancer-trap line to flies with UAS-transgene.

Drosophila mutant strain w^{1118} was used as a background strain. This *Drosophila* mutant strain is characterized by the lack of eye pigment so that the flies have white-eye phenotype. The wild-type strain 51D without any insert was used as a control line.

In order to monitor clock gene expression, the Period-luciferase fusion method was used as described above. Due to the low number of central clock neurons (150 in total), acquiring a bioluminescence signal using the available CCD camera proved impossible. As an alternative approach, both control and UAS-$A\beta_{42}$ Arctic *Drosophila* strains were combined with flies carrying the period-luciferase transgene XLG-luc2, in which the expression of luciferase is driven by *period* promoter region in peripheral tissues. To ensure the correct genotype after crosses and prevent recombination between chromosomes, balancer chromosomes, that have different marker phenotypes, were used. For instance, we used flies with UAS-$A\beta_{42}$ transgene expressed over the balancer CyO on the second chromosome, that produces curly wings in flies. Period-luciferase transgene XLG-luc2 was expressed over balancer TM6B in third chromosome, that produced a humeral (so called "hairy shoulder")

phenotype (Figure 4.4).

As shown in Figure 4.4, there are 4 possible phenotypic groups of offspring as a result of fly crosses, but only the one class (number 4, non-curly, non-humeral, Figure 4.4) was selected to ensure the presence of Aβ and *period*-luciferase expression according to the following criteria: normal wings shape (absence of CyO

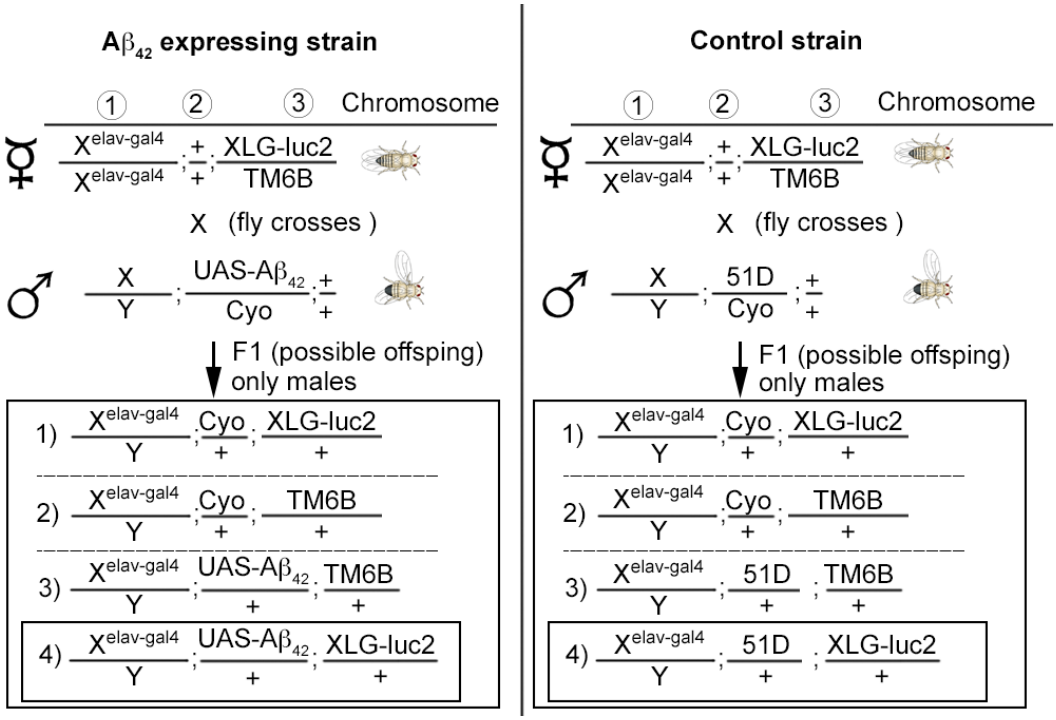

Figure 4.4 Fly crosses. The flies with required genotype were generated by crossing virgin flies with GAL4 and XLG-luc2 expression in the peripheral tissues and male flies with UAS-Aβ$_{42}$ expression. Out of 4 possible genotypes of the offspring, the flies with genotype number 4 were selected for the experiment.

balancer, non-curly wings phenotype) and non-humeral (absence of TM6B balancer, non-hairy shoulders phenotype). The same criteria were applied when generating the control flies that have an empty transgene receptor site on chromosome 2 rather than the Aβ transgene (strain 51D).

4.2.1.2 FLY ARENA

Individual flies were placed in capped glass tube of standard length (5mm x 65 mm) so that they were able to walk horizontally up to 20 times of their body length (Figure 4.6, panel A). The tube was filled with fly food on one end and cotton wool on the other end. The food containing luciferin was prepared by mixing 5% sucrose

with 1% agar and 15 mM (first run) or 50 mM (second and third run) of luciferin. The tubes containing flies were placed in a microscope slide storage tray and separated by the black card divider stripes. One tray could hold up to 48 tubes (Figure 4.5).

4.2.1.3 BIOLUMINESCENCE RECORDING

After flies hatched and collected, they were aged on standard cornmeal food until they reached the required age for the experiment. In order to synchronise the intrinsic circadian rhythms, flies were entrained to a 12:12 light-dark (LD) cycle for 3 days. The tubes, each containing a single fly, were placed in the apparatus and kept in constant darkness in the incubator (Figure 4.5 A). The environment was also kept constant with regard to temperature and humidity over the course of the experiment to ensure that no environmental signals affect the circadian rhythms.

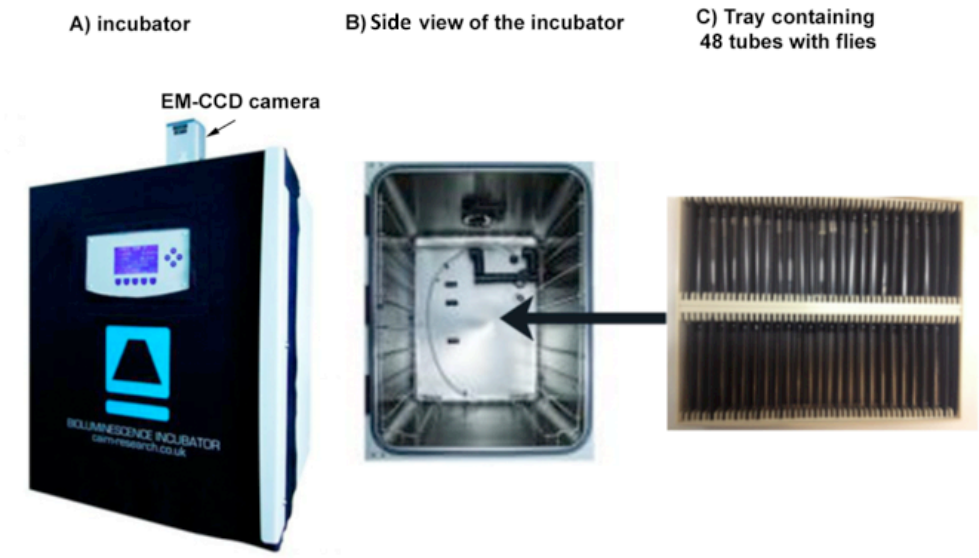

Figure 4.5 Experimental setup. A) Incubator kept the constant temperature and humidity. The EM-CCD camera was incorporated in the top plate of the incubator B) Top view of the incubator. C) 48 cappilary tubes were placed in a storage tray.

The top of the incubator was equipped with cooled electron-multiplying charge-coupled device (EM-CCD) camera that summed the emitted bioluminescence over 5 min exposures (that integrates the photon counts) for a total of seven days (Figure 4.5). In order to compare the performance of the experiment with conventional methods of measurements, flies were additionally placed in 96-well

plates and covered by pierced plastic domes to restrict their movement (Brandes, Plautz et al. 1996), as described in 4.1.4, and placed in the incubator, where the recording was done in parallel with tube-based setup. The circadian behaviour of comparable flies, contemporaneously housed in glass tubes with luciferin-free food, was also recorded using DAM infrared beam split system as described in 4.1.4.

4.2.2 IMAGE ANALYSIS

The images of the flies constrained within a glass capillary tube have been captured using EM CCD camera with 5 min exposure time so that each image is represents a 300 sec integration of photon counts. As a result, each image contains sufficient information to quantify the locomotor behaviour of the fly (with spatial resolution of the order of millimetres and temporal resolution of 300 sec), together with a simultaneous estimation of the level of peripheral clock gene expression (the total number of photons captured in the area of the tube). Taking into account that immobility is opposite to locomotor behaviour, it is possible to measure additionally the periods of inactivity and infer sleep-like events that satisfy standard criteria of sleep (Gilestro 2012).

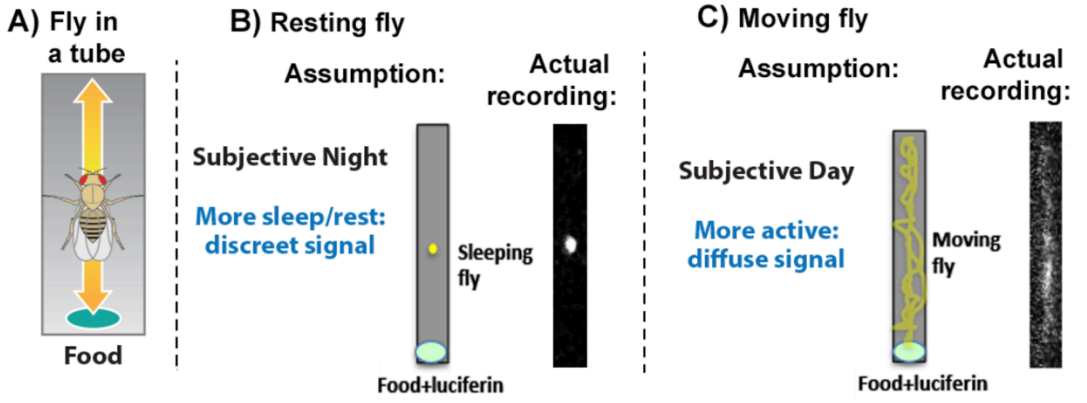

Figure 4.6 Examples of fly behaviour taken during the experiment. A) A single fly was placed in a capillary tube with the food on one end of the tube. B) Resting fly being stationary in one place will produce a bright spot. C) Moving fly will produce a smear of dots as it walks within the tube.

Analysis of the fly's behaviour is one of the crucial parts and is based on the assumption that the images taken with EM-CCD camera translate into 300 seconds of fly movement. Given that each image is 300 seconds of photon integration and the flies emit light, we expect that if the fly were to spend the majority of time in one

place it will produce bright localized spot in the image (Figure 4.6, panel B). By contrast, if the fly were to move continuously then it would produce a smear of dots in the image (Figure 4.6, panel C).

In order to extract some meaningful features from the acquired images, the preprocessing stage is required. This involves several steps of image processing and analysis.

4.2.2.1 NOISE REDUCTION AND IMAGE ROTATION

As the EM-CCD camera is a very sensitive device and acquires images by summing the photon counts over 5 minutes; as such the images contain some level of read out noise. In order to improve signal-to-noise ratio, the first step of noise reduction has been performed. It consists of estimating the background noise level by averaging the brightness of pixel intensities taken at the bottom left 64x64 square for each of the image (Figure 4.7). The average value of the background noise is then subtracted from every pixel of the image.

Image rotation was performed so that all the tubes in the parallel rows are aligned with each other and tube centres can be easily identified. The required rotation angle was calculated based on the manual location of two marker points positioned at the top and bottom right hand side corners of the tray where fly tubes are placed (Figure 4.7) in the illuminated image, acquired at the end of the experiment. The angle is calculated based on the law of cosine and then rotation take place around the centre of the image using the rotation matrix. In the correctly rotated images the centres of fly tubes will be aligned above each other along the x-axis.

The centre of the fly tubes can be found by summing up pixel intensities along y-axis (Figure 4.7). This summation is performed for 200 images to account for the intensity variability and random noise, such as cosmic rays. If the rotation step has been done correctly, the centres of the tubes will lie above each other and produce the peaks in intensity profile. In order to identify the top and bottom boundaries of the tubes similar process has been done, summing up the pixel intensities along x-axis (Figure 4.7). The rectangular boundaries of each tube were defined around the centres, assuming the tubes are 17 pixels wide and 185 pixels long.

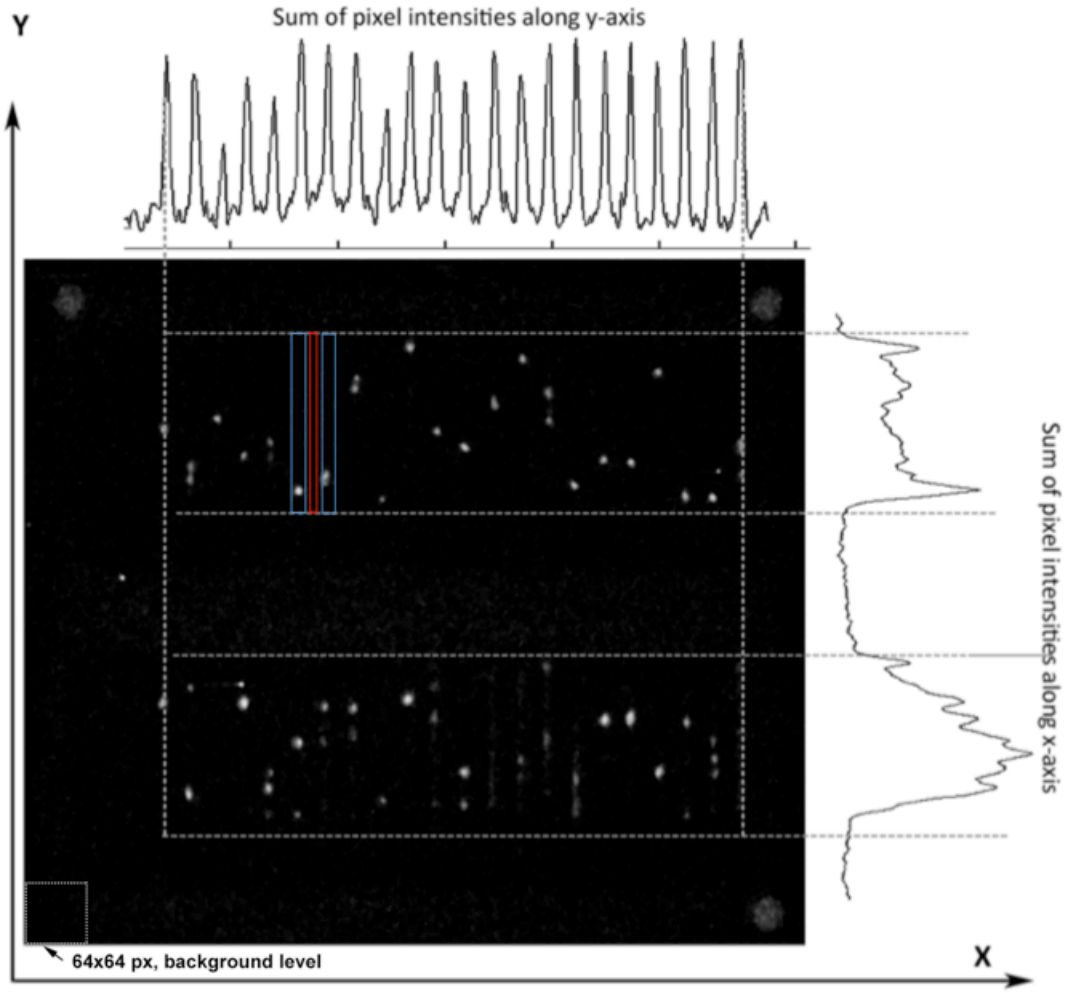

Figure 4.7 Rotating images and finding the boundaries of the tubes. Three circles (grey circles) were attached at the corners of the tray as landmark for image rotation. The boundaries of the tubes within the image were identified by summing up the pixel intensities along x-axis (right plot) and y-axis (top plot).

4.2.2.2 IMAGE ENHANCEMENT

The main core of the analysis is quantification of fly walking behaviour and periods of inactivity from the images, as will be described below. The quantification of images usually implies that the images have same level of background noise and overall background brightness. However, due to the luciferase decay, being reported in previous studies (Brandes, Plautz et al. 1996), the images at the end of 7-day experiment have considerably lower brightness as compared to the images at the beginning. Therefore, in order to reliable identify flies the contrast and brightness in the images were enhanced by using contrast-limited adaptive histogram equalization (Reza 2004). This algorithm increases the contrast between the darks and the lights.

Contrast enhancement was limited inversely according to the number of images (as images were numbered every 5 minutes) meaning that the contrast for initial images was enhanced to a lesser extent than for images at the end. As a result, using this algorithm will transform all of the images so that all of them will have uniform contrast histograms. These enhanced images were only used to allow sensitive feature detection, principally spot detection, but quantitative measurements were based on the raw images, processed only to remove background noise.

Another technical issue is that the limitations of the lens optical system of EM-CCD camera cause the vignetting effect in the images, which means that there is a small systematic variation in pixel brightness between different regions of the image. In order to account for this effect, second step of background subtractions has been performed. In this case background noise level was evaluated on tube-by-tube basis to get the best possible estimate of background pixel brightness in the region next to the tube of interest. The boundaries of the appropriate background rectangle (Figure 4.7, red rectangle) were set as the top and bottom boundaries of the tubes on y-axis and on x-axis it was taken as a 5 pixels wide region in between the pair of tubes (Figure 4.7, tube rectangles shown as blue rectangles).

4.2.3 BIOLUMINESCENCE AND LOCOMOTION ESTIMATION

Based on the images taken with the EM CCD camera, we were able to calculate different parameters characterising fly circadian clock oscillations and circadian behaviour, including sleep.

4.2.3.1 MEASURING MOLECULAR CLOCK OSCILLATIONS

The level of the circadian reporter construct was defined as the sum of noise-corrected signal intensity within each tube rectangle. The oscillation of circadian clocks was estimated by plotting the summed intensity over time. This parameter reflects the cellular circadian rhythms of clock gene expression in peripheral tissues of the fly.

4.2.3.2 ANALYSIS OF ACTIVITY AND SLEEP

In order to convert the array of dots in the image into the timeline of fly movement, we first needed to find the regions where fly was stationary (resting periods). As each tube containing the fly is identified as 185 pixels long and 17 pixels wide, each tube was divided into 4 pixels high bins along long axis (y-axis) to reduce the effect of minor observation errors, resulting in 46 bins of 4 pixels long and 17 pixels wide. If we sum the intensity of light in each bin and plot the profile, we can identify those resting periods (Fig. 4.8). Bins containing a stationary fly are identified as peaks in the intensity profile along the tube; we call them resting peaks. On the contrary, we expect that and actively moving fly will produce smear of dots (Figure 4.6, panel C) and we approximate this profile to a horizontal line across the bins at a particular level above background (Fig. 4.8); we call this the activity level.

Assuming that the shape of the resting peaks follows a normal distribution, the dominant peaks can be identified if the value of the intensity in particular bin is above the mean value plus a standard deviation of the light intensity among all the bins in the enhanced images (Figure 4.8, "Mean+SD" values). The boundaries of resting peaks were assigned to the adjacent non-peak bins or to the value of activity level ("Mean-peak") in case when the value of adjacent bin was below "Mean-peak" value (Figure 4.8, blue shade). Once the enhanced data was used in this way to identify bins containing data from a resting flies thereafter only raw data was used. Specifically, the mean level of bin intensity was calculated (using the raw data), but only considering bins that do not contain a resting fly ("mean-peak", right hand panels in Figure 4.8).

Assuming that there are only two distinct kinds of fly locomotor behaviour, namely *resting* and *walking*, it follows that the 300 sec of exposure time must be partitioned between these two states for each camera image. Using the raw image data, the contribution of each parameter was estimated as an area under the corresponding curve. Total area is defined as:

$$Total\ area\ (300\ sec) = \sum_{N\ peaks} \int resting\ peaks + \int activity\ level \quad (4.1)$$

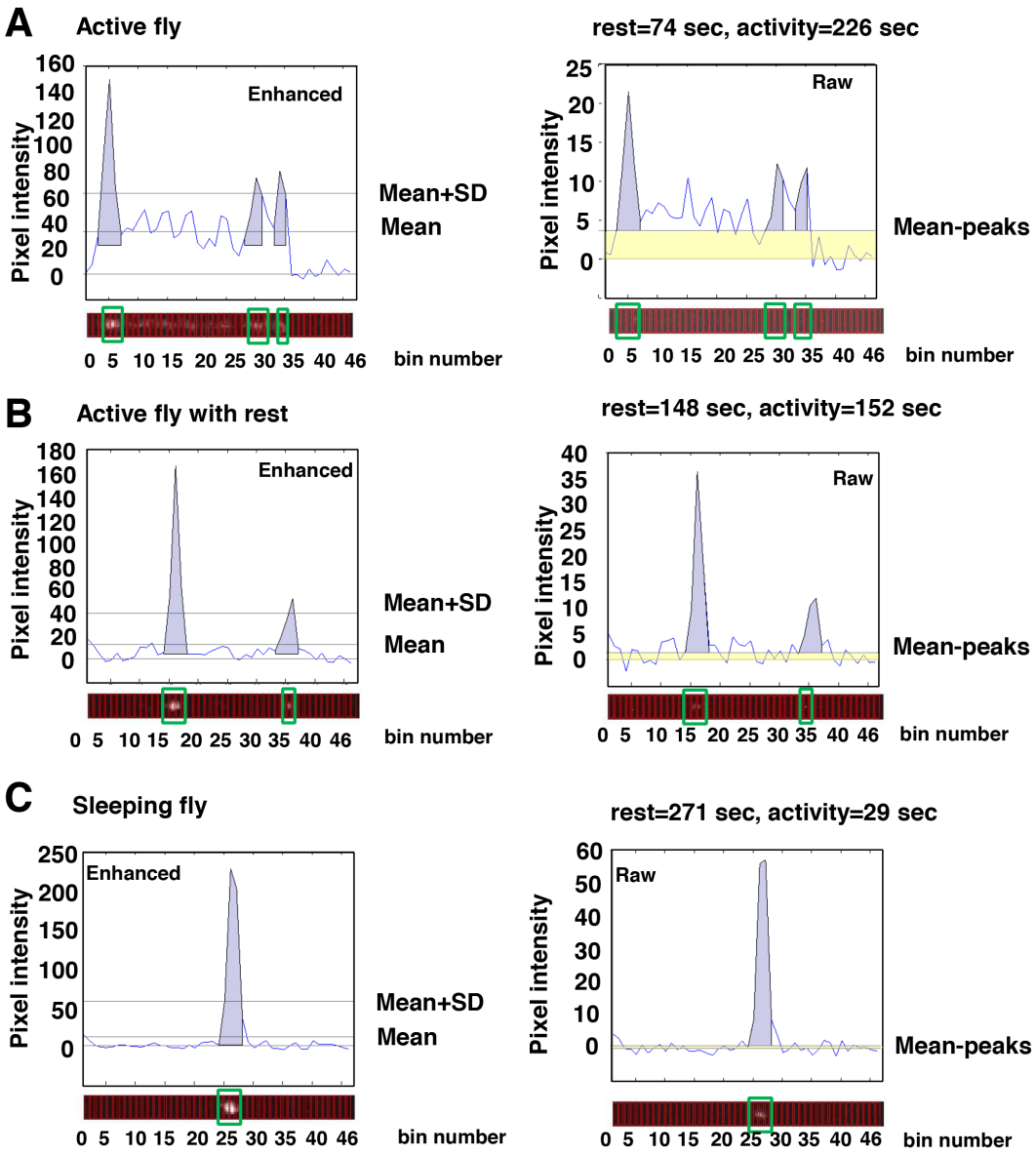

Figure 4.8 Frame-by-frame identification of behavioural status. The intensity pixel profile was binned into 46 segments along the length of the tube for each timeframe. The enhanced images were used to detect the bins containing the peaks of intensity (resting peaks, blue area under the peaks): if intensity value of the bin was above mean and standard deviation of the pixel intensities (Mean+SD) for the tube. Using the raw images, the activity level was estimated (yellow shade). The total of the areas under resting peaks (blue shade) and activity level (yellow shade) is equal to 300 sec. Representative data is shown for (A) an active fly, (B) an active fly with periods of rest and (C) a sleeping fly.

Under this notation, the contribution of activity to the intensity profile was estimated as an area between the activity level ("Mean-peak") and the background. Likewise, the proportion of resting episodes was evaluated as the sum of areas under each resting peaks.

$$Time_{active} = 300\ sec \cdot \frac{\int activity\ level}{Total\ area} \quad (4.2)$$

$$Time_{resting} = 300\ sec \cdot \frac{\sum_{N\ peaks} \int resting\ peaks}{Total\ area} \quad (4.3)$$

Several studies have demonstrated that in *Drosophila*, sleep-like behaviour is characterized by increased arousal threshold and homeostatically regulated rebound following sleep deprivation. In these studies sleep is defined as a period of immobility for at least 5 consecutive minutes (Gilestro 2012). Therefore, with reference to such sleep-like states, the presence of a sleep episode is defined by three criteria: 1) an image frame may be part of a sleep episode only if it contains a single resting peak, meaning that the fly was resting in one position; 2) for subsequent images of the same tube there must be a resting peak at the same position, indicating that the fly did not move between frames; if there are no resting peaks, in a subsequent images of the tube then the sleep episode is assumed to be over; 3) the rest times summed across two consecutive frames must exceed 300 sec. Even though this definition of sleep in *Drosophila* is conservative, it does not overestimate the short rest episodes as sleep episodes. Using this approach, each fly within each image frame was assigned either 1 or 0 in the binary sleep array (1=asleep and 0=awake).

Under these circumstances, each fly in each image frame was designated with one of two possible behavioural states: active (that allows for resting periods <300 sec) or sleeping (Figure 4.8). However, the binary representation of the sleep data provides an insight into the total amount of sleep but not the "quality" or length of sleep episodes. Because sleep fragmentation is a feature of AD and other disorders the data was analysed further and a novel sleep consolidation index was calculated, that measured the length of each sleep episode block (Figure 4.14). For a particular fly, this index was calculated by examining the binary sleep array and estimating the duration of each sleep block. The sleep consolidation array consisted of a value for each frame representing the total duration of the sleep episode of which it was a part. Two examples of the sleep consolidation array could be: 1) an active frame that has a value of zero and 2) each component frame of a five frame sleep event containing the value of five.

As a result, five different parameters were calculated and presented as a time series for each fly: 1) intensity of emitted light (the molecular clock signal), 2) time spent active (locomotor activity), 3) time spent resting (reciprocal to time spent active), 4) sleep binary array (presence of sleep-like episodes), 5) sleep consolidation

array (duration of any sleep episode that the frame is a part of).

4.2.3.3 BINNING OF FINAL TIME SERIES

As mentioned in 4.1.4, the sampling interval of the circadian activity data collected by DAM system is conventionally set to be 30 minutes. Therefore, in order to assess the performance of the newly established method with performance of the widely used DAM system, the temporal resolution of the computed time series of circadian oscillation was reduced from 5 minutes to 30 minutes by binning the consecutive time points. Binned values of sleep binary and time spent active or resting were represented by the sum of constituent timepoints, while for sleep consolidation the bin was assigned with the average consolidation of non-zero entries in the constituent timepoints; for bioluminescence, average light intensity value of the constituent timepoints was set as a binned value.

4.2.4 TIME SERIES ANALYSIS

Within our experimental data, we are mainly interested in robustness of the rhythmicity exhibited by the oscillations of molecular clock and circadian behaviour of the flies in free run (constant darkness, no external cues). Most of the circadian rhythms have a repeating pattern of around 24 hours, so that the analysis of circadian oscillation requires the analysis of time series data. However, it has been previously shown that other than 24 hours periodicities are also present in the circadian rhythms of *Drosophila*. For example, periodicities in the range of 5 to 18 hours, called ultradian rhythms, have been found in the *Drosophila* circadian clock mutants and can occur in either the presence or absence of normal circadian rhythms (Dowse and Ringo 1992). Likewise, long-range periodicities can also arise as a result of behavioural or enzymatic activity that are independent of circadian rhythms. Both of these additional short- or long-range rhythms can interfere with the analysis of 24-hour periodicities and therefore experimental data required a pre-filtering stage.

4.2.4.1 SIGNAL FILTERING

Several studies have previously reported that luciferase decays over time, so that the absolute amplitude of bioluminescence decreases exponentially over the course of experiment (Brandes, Plautz et al. 1996). This type of decreasing trend can interfere with the extraction of periodicity in the circadian range and therefore need to be taken into account before performing the time series analysis. Indeed, the visual inspection of the bioluminescence time series reveals the exponential decay (Figure 4.9, top).

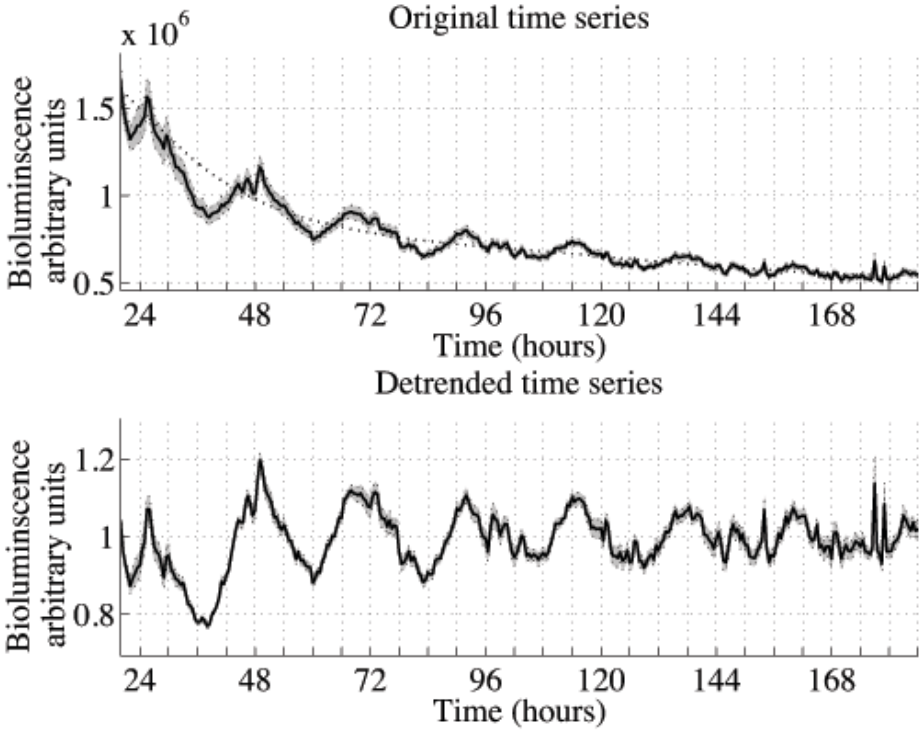

Figure 4.9 Detrending time series data. Trend curve for the bioluminescence decay is indicated as a dashed line and superimposed on the data (top plot). Detrending of the original data was achieved by dividing each point in the time series by the corresponding point on the trend curve

The trend curve was identified by using low-pass filter that only passes the signal with a frequency lower that 1/72, or with periodicity longer than 72 hours, and attenuates signals with higher frequencies (Figure 4.9, dashed line). The bioluminescence signal was then detrended by dividing each data point by the corresponding data point on the trend curve, resulting in a detrended signal with a mean value of 1 and preserving the amplitude of the variation around the trend curve

(Levine, Funes et al. 2002). As a consequence of such filtering, the signal appears to be more robust than in the raw data and the units of measurement are eliminated so that the temporal features can be compared with other signals such as behavioural rhythms.

4.2.4.2 ESTIMATION OF RHYTHMICITY

The robustness of the circadian oscillations can be assessed quantitatively by the well-established autocorrelation method (Levine, Funes et al. 2002), that measures the degree of periodicity of the time series. This method is based on the cross-correlation of the signal with itself at different time delays so that when the lag between two series is equal to 0, the signals are perfectly aligned and the correlation coefficient is 1; however, when the two signals are slipped out of register with one another, the correlation coefficient decreases. For example, the correlation of random signal will rapidly decrease close to zero and remain there. On the other hand, if there is repeated pattern within the signal then peaks and troughs will slip back into register at the lag time close to the pattern period and the correlation coefficient will increase again (Hansen 1995). The autocorrelation coefficient ρ for the observed values $x_1,\ldots,$ x_N and lag h is given by:

$$\rho_h = \frac{\sum_{t=h+1}^{N}(x_t-\overline{x})(x_{t-h}-\overline{x})}{\sum_{t=1}^{N}(x_t-\overline{x})^2} \quad (4.4)$$

Conventionally, the strength of the rhythmicity as a function of regularity of the underlying process can be measured numerically in terms of the rhythmicity index (RI) value as described in (Levine, Funes et al. 2002). This index is assessed as the height of the third peak (equivalent to a lag time of two periods) in the autocorrelation function. When multiple experiments are performed an estimate of the statistical significance of the periodicity, called rhythmicity statistic (RS) value, may be evaluated as a ratio of the RI value to the absolute value of 95% confidence interval at each timepoint. In practice, based on 30 minutes sampling intervals, it is widely accepted that statistical significance is achieved if the RS value is above 1.5.

4.2.4.3 ESTIMATION OF PERIOD

There are many different methods for period estimation, most of which is based on Fourier analysis, also known as spectral analysis. First applied to the analysis of circadian rhythms in early 1950s, Fourier analysis is based on the assumption that any function, regardless of its shape of regularity, can be decomposed into a series of cosine and sine waves of different frequencies as determined by the goodness-of-fit to the data. In Fourier analysis, at a given frequencies 1/N, 2/N…(N-1)/2N, the sum of cosine and sine waves represents the spectral energy of the signal at that frequency. Although Fourier analysis has proven to be efficient method of period estimation, it has some disadvantages when applied in chronobiology. For example, Fourier analysis is only able to estimate what frequencies are present but not how the period of oscillations is changing over time. Shortcomings of the approach include poor frequency resolution, unless the number of cycles is large, and artefactual boosting of long-period signal in noisy time series with uneven sampling. One particular variation of Fourier analysis, namely the Lomb-Scargle (LS) method, can address this issue quite successfully. In addition, this method allows the analysis of non-evenly spaced data (as a result of equipment failure or other problems).

Lomb Scargle method is a variation of Discrete Fourier transform and estimates the frequency spectrum based on the least squares fit of sine and cosine waves to the data. Given N observations and x_i values obtained at times $t_i = 1…N$, first the mean $\overline{x}$ and variance σ^2 of the signal is calculated as follow:

$$\overline{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i \; ; \tag{4.5}$$

$$\sigma^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \overline{x})^2 \tag{4.6}$$

For a given frequency component f, the time offset τ of the angular frequency $\omega = 2\pi f$ is calculated as:

$$\tan(2\omega\tau) = \frac{\sum_{j=1}^{N} \sin(2\omega t_j)}{\sum_{j=1}^{N} \cos(2\omega t_j)} \tag{4.7}$$

The Lomb Scargle periodogram estimates the spectral power of the given frequency in the signal as a function of ω and is defined by:

$$P_{LS}(\omega) = \frac{1}{2\sigma^2} \cdot \left\{ \frac{\left[\sum_{j=1}^{N} (x_j - \overline{x}) \cdot \cos \omega(t_j - \tau)\right]^2}{\sum_{j=1}^{N} \cos^2 \omega(t_j - \tau)} + \frac{\left[\sum_{j=1}^{N} (x_j - \overline{x}) \cdot \sin \omega(t_j - \tau)\right]^2}{\sum_{j=1}^{N} \sin^2 \omega(t_j - \tau)} \right\} \tag{4.8}$$

The periodogram is normalized by the variance of the data to ensure that the power $P_{LS}(\omega)$ is uniformly distributed and provides the probability p that the power P_{LS} at a

given frequency ω is real:

$$p\big(P_{LS}(\omega)\big) = 1 - (1 - e^{-\max(P_{LS})})^N \quad (4.9)$$

A review of circadian period estimation methods by Zielinski and colleagues (such as Enright periodogram, Lomb-Scargle periodogram, Fast Fourier Transform Non-Linear Least Squares, mFourfit, Maximum Entropy Spectral Analysis (MESA) and Spectrum Resampling) has shown that MESA was able to more precisely identify the underlying rhythms as compared to the other methods, however the amount of noise in the artificial data was not comparable with our experimental data (Zielinski, Moore et al. 2014). Another recent review compared the performance of the commonly used method for the analysis of circadian rhythms, including Discrete Fourier transform, Autocorrelation and Enright periodogram, but did not consider MESA or Lomb Scargle periodogram (Mourao, Satin et al. 2014).

Therefore, I undertook a comparison of the four conventionally-used methods (Lomb Scargle periodogram, MESA, Autocorrelation and Enright periodogram) using synthetic signal data contain circadian oscillation with an appropriate level of noise, comparable to our experimental data. The principles of MESA, Autocorrelation and the Enright periodogram are described in (Mourao, Satin et al. 2014)

First, the synthetic signal was generated, consisting of sine waves and added noise. Synthetic signal with one period as a function of time t was generated as follows:

$$y(t) = Asin(\frac{2\pi}{T} \cdot t + \phi), \quad (4.10)$$

where A – amplitude, T – period and ϕ – phase.

In addition, synthetic signal with 2 periods T_1 and T_2 (fast and slow), as an example of ultradian rhythms, was generated according to the following expression:

$$y(t) = A_1 sin(\frac{2\pi}{T_1} \cdot t + \phi_1) + A_2 sin(\frac{2\pi}{T_2} \cdot t + \phi_2) \quad (4.11)$$

The noise was introduced to the synthetic time series to represent experimentally observed errors during the measurements. As a result, the synthetic signal with Gaussian distributed random noise can be generated as:

$$Y(t) = y(t) + \varepsilon \cdot N(0,1), \quad (4.12)$$

where ε represents the strength of the noise and N $(0,1)$ is Gaussian distributed error with mean 0 and standard deviation 1. For example, Figure 4.10, panel A shows the sine wave component of the synthetic signal in blue and 20% noise component in red; the resulting synthetic signal is presented in Figure 4.10, panel B.

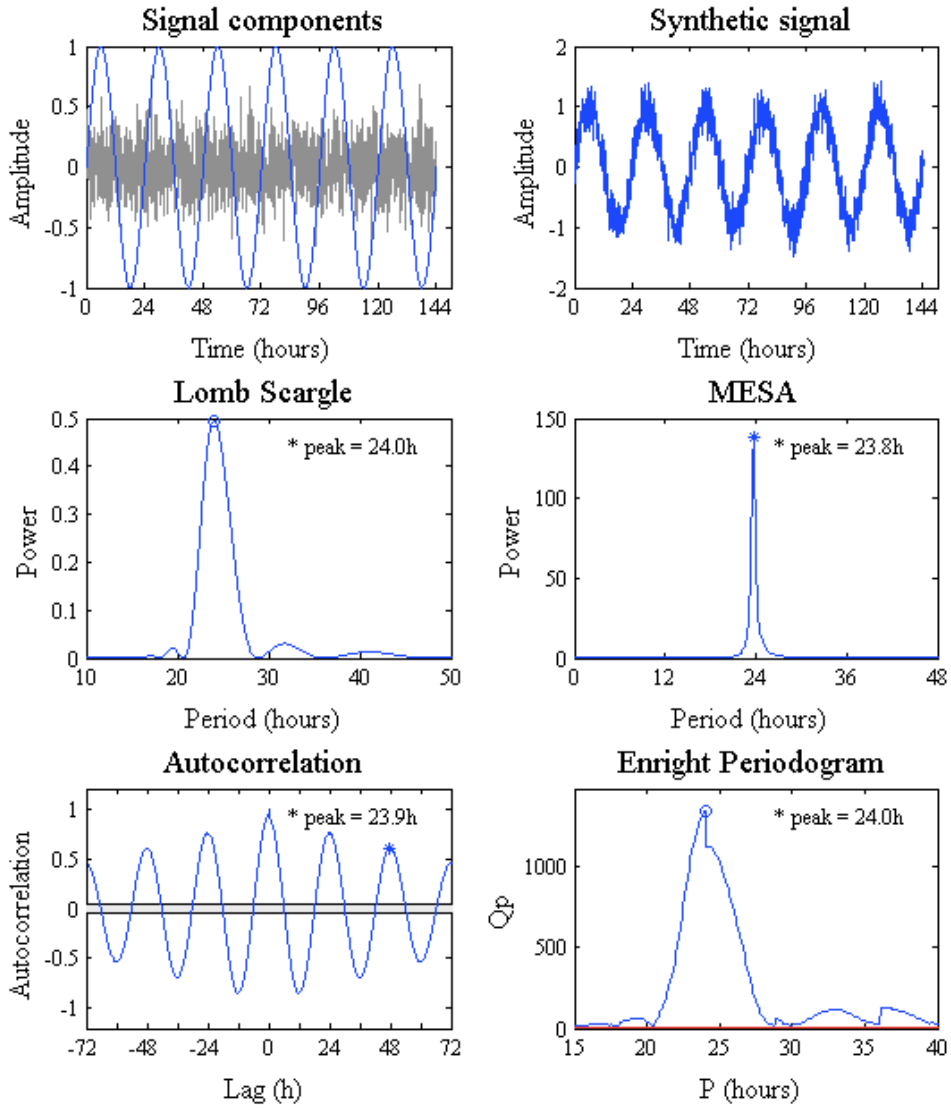

Figure 4.10 Comparison of period estimation methods. A) Synthetic signal was generated as a sine wave with 24 hour period (blue traces) and 20% of the random noise was added (grey traces). The period (value after *peak) of resulting synthetic signal (B) was estimated using 4 different methods (C-F)

As it can be seen from Figure 4.10, only Lomb Scargle and Enright periodograms identified the correct period of 24 hour in the synthetic signal with 20% added noise, while MESA and autocorrelation methods underestimated the period value.

The relative difference (RD), or accuracy of the estimation of a simulated period, for each method was calculated as follows:

$$RD = \frac{|P_m - P_r|}{P_r}, \tag{4.13}$$

where P_m - period identified by a particular method, P_r - real period used to generate the data.

Using the synthetic data with different amount of added noise, the accuracy of each method was evaluated as shown in Figure 4.11; panel A shows the value of RD according to the percentage of added noise and panel B shows the mean value of RD for different methods. As it can be seen, each method estimated the period of synthetic data with high accuracy (<4% error in period estimation), whereas Lomb Scargle and Enright methods produced the best estimates of the period of interest.

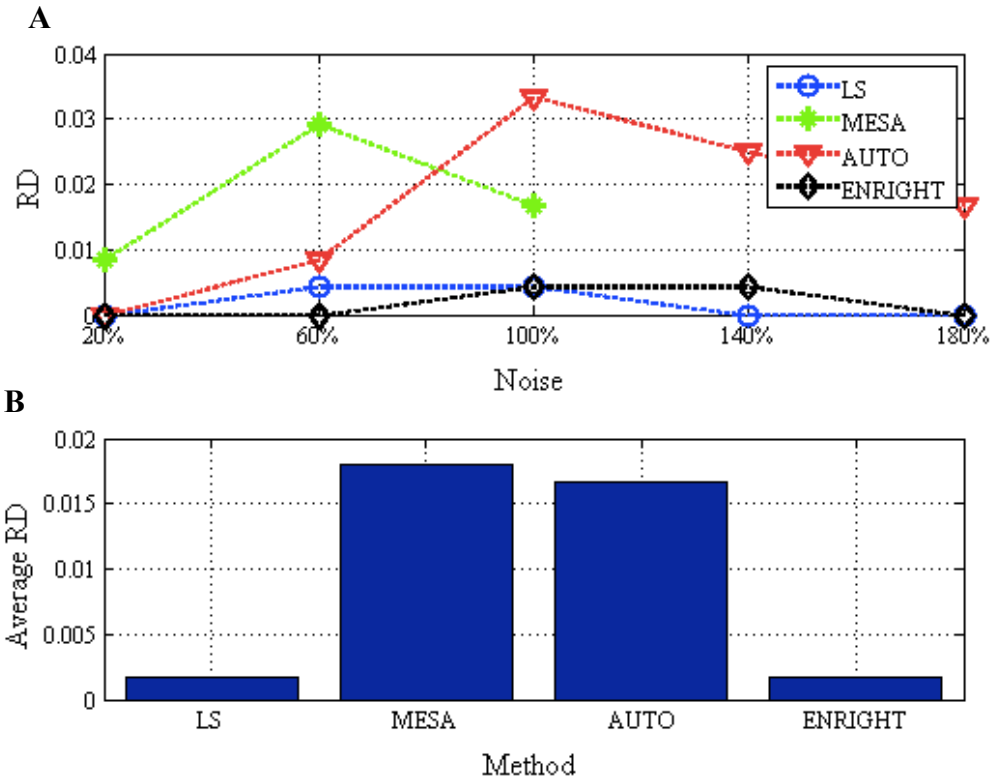

Figure 4.11 Period estimation errors for different methods. A) Period of oscillations in synthetic data with 5 different amounts of added noise was estimated by 4 methods. The values of relative distance (RD), or accuracy of the period estimation, were calculated and plotted on y-axis for each method. B) The bar chart shows the mean RD value of estimating the period in synthetic data with 5 amounts of added noise for each method.

As mentioned before, circadian rhythms in *Drosophila* can have multiple periodicities. To evaluate the accuracy of multiple period estimation of different considered methods, synthetic data with two periods was generated with different amount of noise. For example, Figure 4.12 panel A illustrates the 12- and 24-hour period components (red and blue traces accordingly) and 180% of added noise; the

resulting synthetic signal is shown in panel B.

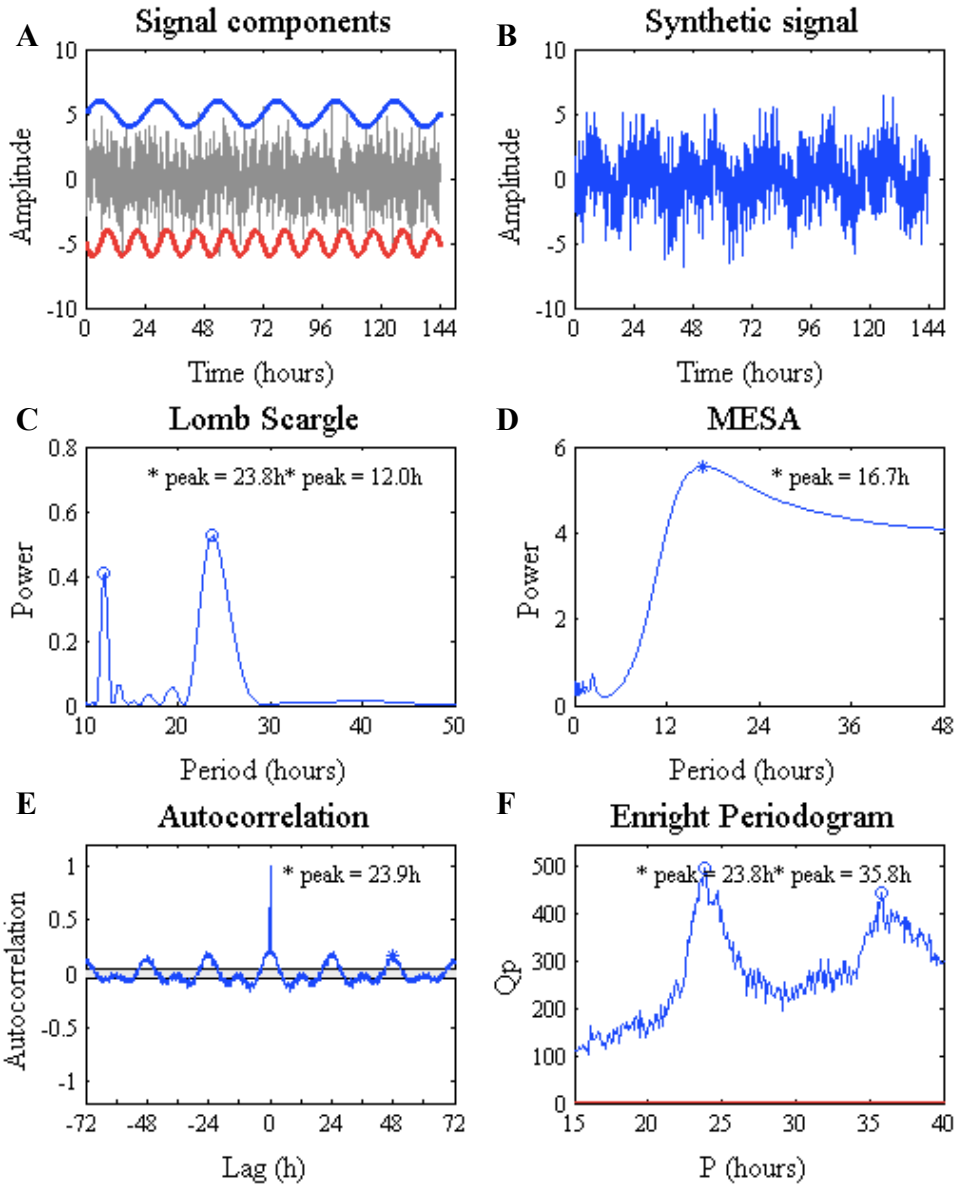

Figure 4.12 Comparison of period estimation methods for oscillations with multiple periods. A) Synthetic signal was generated as combination of a sine wave with 24 hour period (blue traces) and with 12 hour period (red traces) and 20% of the random noise was added (grey traces). The period (value after *peak) of resulting synthetic signal (B) was estimated using 4 different methods (C-F)

As it can be seen, only Lomb Scargle has identified both periods correctly, while MESA found only one period and considerably underestimated 24-hour component. Autocorrelation method similarly identified only one period, whereas although Enright periodogram recognize two periods, the 12-hour component was

misidentified as having a 36 hour period.

The accuracy of the period estimation for synthetic data with two components have (Figure 4.13) demonstrated the supremacy of Lomb Scargle periodogram, especially in the presence of high noise level.

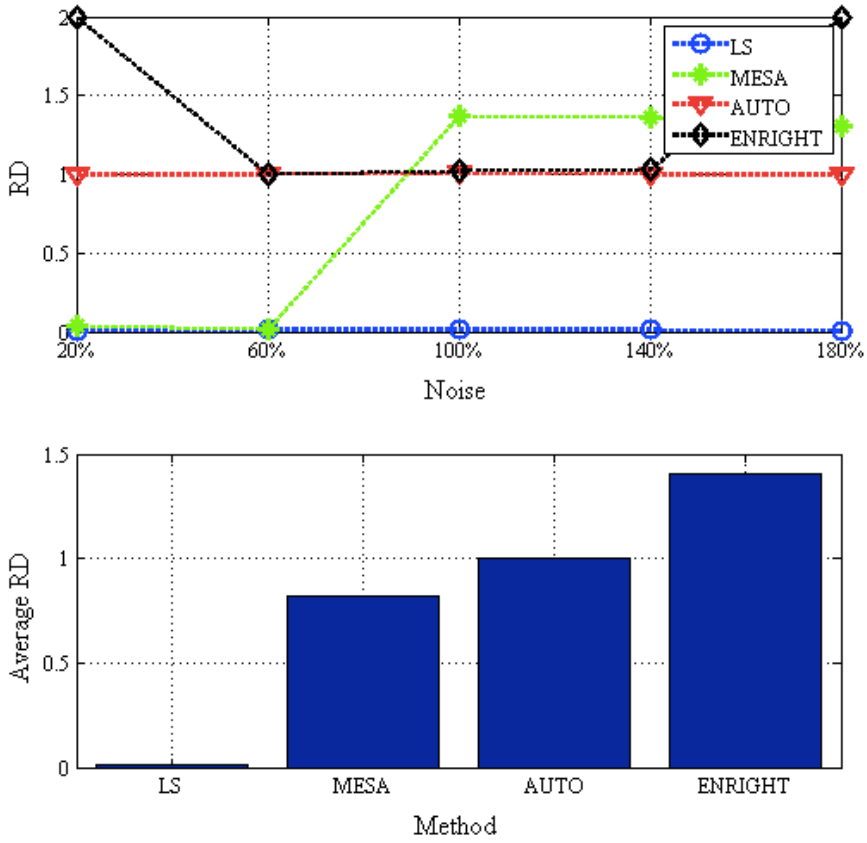

Figure 4.13 Period estimation errors for synthetic signal with two waveforms. Two periods of oscillations in synthetic data with 5 different amounts of added noise were estimated by 4 methods. The values of relative distance (RD), or accuracy of the periods estimation, were calculated and plotted on y-axis for each method (top plot). The bar chart shows the mean RD value of estimating the periods in synthetic data with 5 amounts of added noise for each method (bottom plot).

Therefore, the Lomb Scargle periodogram was chosen as the main method to evaluate the periodicity of circadian oscillations. In order to investigate how the rhythmicity of data changes over the course of the experiment, the Lomb Scargle method was implemented based on a series of overlapping windows of analysis, in which the rhythm variability in periods was detected using 3 days worth of data and 12 hours offset (first window contained the first 72 hours (day 1-3) of experimental data, second window: 12-84 hours (day 1.5-3.5) and so on).

4.3 RESULTS

This project was initiated to test whether it was possible to record concurrent circadian molecular rhythms and behavioural characteristics for the same individual fly. For this purpose, the experimental strain of period-luciferase transgenic flies was crossed with mutant strain of $A\beta_{42}$ expressing flies so that their offspring will have the overexpression of $A\beta_{42}$ peptide in the brain and produce bioluminescence according to the *period* gene expression in peripheral tissues (Figure 4.4). Collected flies from the offspring were monitored using sensitive EM CCD camera with five minutes exposure time so that, when flies are constrained within horizontally positioned capillary tubes, they demonstrate the one-dimensional locomotor behaviour (Figure 4.6). The images were preprocessed to remove the background noise and neutralize the vignetting effect and saved as raw images; additionally, raw images were further enhanced using contrast-limited adaptive histogram equalization (saved as enhanced images).

4.3.1 ANALYSIS OF PARAMETERS

Each image frame consisted of 300 seconds of photon integration that can be directly translated into the timeline of fly behaviour (spatial distribution of photons within the tube) and be used to measure the oscillations of clock gene expression (sum of pixel intensity within the tube, Figure 4.6). The spatial distribution of photons for each fly was presented in forms of two intensity profiles: active fly appeared as a smear of dots and the duration of the movement in each image frame was estimated as rectangle area in the intensity profile (Figure 4.8, yellow shade); in comparison, a resting fly generated a bright, localized dot and produced resting peaks along the intensity profile (Figure 4.8, blue shade). Using the enhanced images, the resting peaks and activity level were identified. Based on the intensity profile in the raw images, the area under resting peaks (Figure 4.8, blue shade) was quantified in terms of time a particular fly spent stationary and the area under the activity level (Figure 4.8, yellow shade) was used to evaluate the time a particular fly spent moving; the sum of the areas under those curves was assumed to be equal to 300 seconds for each fly in each image frame. Moreover, given the sufficient sampling frequency, the

sleep-like episodes were inferred according to the standard criteria (Gilestro 2012). Overall, four different parameters were estimated for every fly at each time frame: 1) the intensity of bioluminescence, 2) the duration of activity & rest, 3) the presence of sleep-like episodes and 4) the sleep consolidation index. The time series of calculated parameters were binned into 30 min bins to compare the performance of this novel technique with the performance of other conventional assays.

Due to the luciferase decay, the bioluminescence raw data was detrended in such a way that the signal have the same mean and amplitude the over the course of the experiment using the method described in 4.2.4.1. The example of the time series for a particular fly from the control strain is presented in Figure 4.14 A, where molecular clock represents the oscillations of the peripheral circadian clocks and locomotor activity is plotted as an average percentage of time the particular fly spent moving around with 100% being 300 seconds. The conventional sleep binary array depicted the presence of sleep episode but do not provide any information about the sleep structure and episode duration. Moreover, most *Drosophila* sleep studies report the total amount of sleep or average length of sleep episodes with the data collected usually for less than three days, which make the analysis of circadian fluctuations unfeasible. In my study I further analysed the sleep data by creating the sleep consolidation array as described in 4.2.3.2, in doing so I revealed the circadian oscillations of sleep episode length (Figure 4.14, C). Clearly visible is the transition of sleep from a consolidated state, with long sleep episodes, during the subjective night (black blocks along the x axis) to the short naps in the daytime.

The close examination of the sleep consolidation plot (Figure 4.14C) showed that the longest and most consolidated sleep episode usually occurred during first 6 hours in the subjective night, whereas shorter discrete episodes of 40-50 minutes long are presented later in the night. Remarkably, I observed sleep episodes of over 3 hours that usually occur in the early part of the night, being observed in at least one night in 52% of control flies. The observed sleep architecture is consistent with the previously reported circadian sleep patterns in *Drosophila* (Gilestro 2012) and resembles the structure of human sleep, when a greater amount of deep sleep occurs particularly in the early night while the proportion of rapid eye movement (REM) sleep increases towards the morning (Wulff, Gatti et al. 2010).

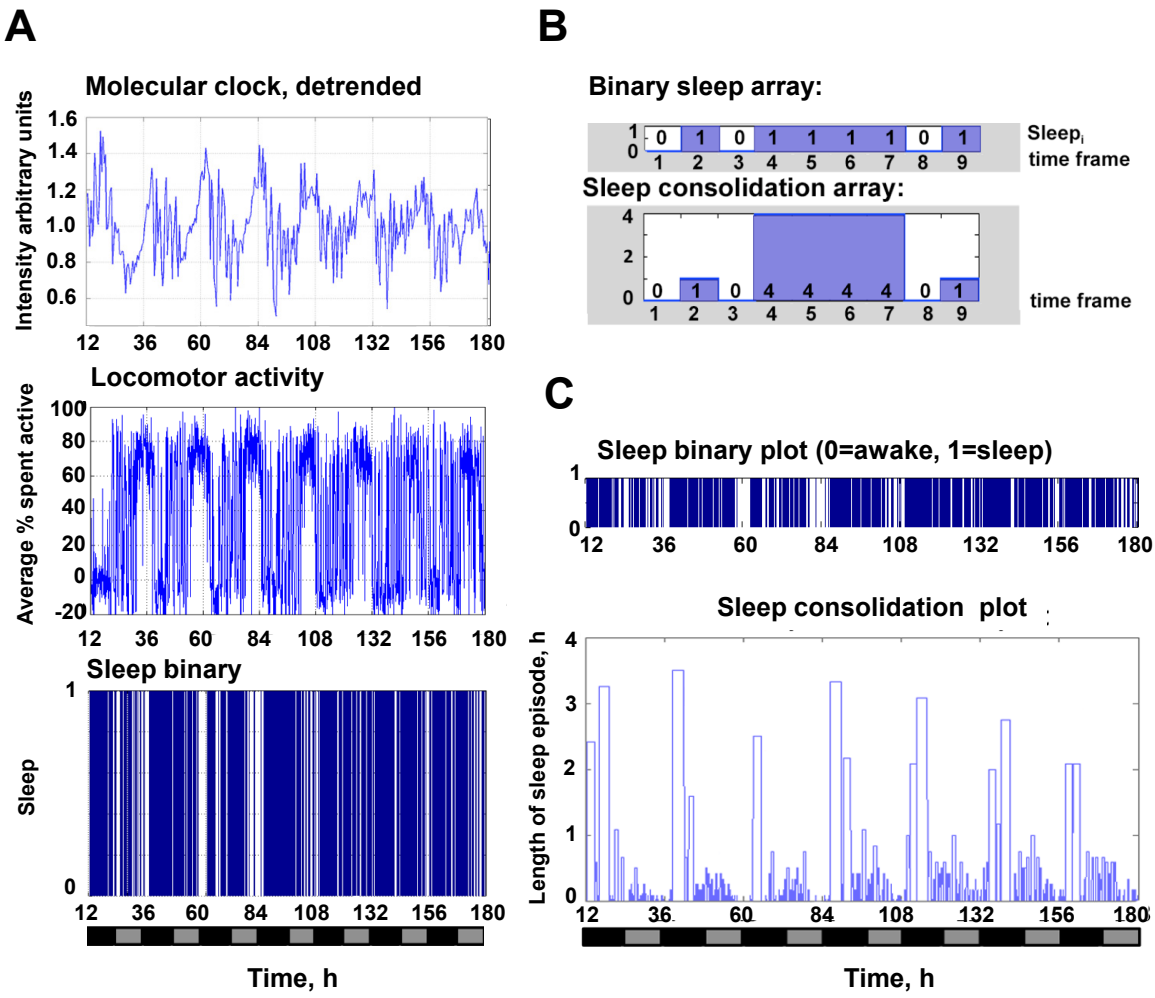

Figure 4.14 Example of calculated parameters for a particular fly. A) The oscillation of the molecular clock is reported by the bioluminesence intensity of a single fly during the 7 days of the experiment in contant dakrness. The percentage of time a particular fly spent active (Locomotor activity) was estimated and the sleep status (Sleep binary) was assigned to each time frame. B) The presence of sleep-like episode (Sleep binary array) was estimated based on three criterias described in 4.2.3.2. Each sleep frame in sleep consolidation array contained the total length of the encompassing sleep episode. C) The structure of the sleep consolidation exhibited circadian oscillations. X-axis in A) and C) represent the circadian time in hours with indicated subjective day (grey) and subjective night (black). The recording started at subjective dusk, when circadian time=12 hr.

4.3.2 COMPARISON OF THE FLYGLOW PERFORMANCE WITH EXISTING METHODS

To evaluate the performance of the developed system, named FLYGLOW, circadian behavioural and molecular oscillations were recorded for 23 day old control period-luciferase transgenic flies, and flies co-expressing both luciferase and $A\beta_{42}$, for seven days in constant darkness and temperature. Previously, it was demonstrated that the $A\beta_{42}$ flies of similar age show progressive behavioural arrhythmia, despite their central molecular clock remaining rhythmic (Chen, Possidente et al. 2014).

Additionally, the locomotor behaviour of flies with the same genotype was recorded using the conventional infrared beam split system (DAM). Visually, the average locomotor rhythm profile of control flies measured by FLYGLOW appear more rhythmic as compared to the one detected by DAM (Figure 4.15). As expected (Chen, Possidente et al. 2014), $A\beta_{42}$-expressing flies exhibit considerably weaker rhythmicity as compared to the control group, however the locomotor traces identified by DAM appeared substantially arrhythmic whereas spatial resolution of FLYGLOW allowed the detection of the retained oscillations.

The robustness of the fly's behavioural rhythmicity was quantified using the autocorrelation rhythmicity index (RS value, (Levine, Funes et al. 2002)), where the oscillations are considered as rhythmic if RS>1.5. Provided the same temporal sampling resolution (30 minutes binning), it was observed that RS values of $A\beta^{42}$-expressing flies detected by both FLYGLOW and DAM are significantly lower as compared to control group (Figure 4.15 B). Comparison of the distribution of RS values for the population of flies recorded by FLYGLOW and DAM revealed an essentially identical pattern (χ^2 test, $p<0.05$), yet FLYGLOW identified more highly rhythmic flies (green bars on the right, Figure 4.15 C), demonstrating higher sensitivity of detecting rhythm variation.

Furthermore, in order to test the possibility that the freedom of movement seen in FLYGLOW could add a systematic artefact we sought to compare the molecular clock signal with the well established luminometer method (described in 4.1.4, Brandes, Plautz et al. 1996) where the flies are restrained under a small plastic dome. A possible source of artefact could be the lower abundance of luciferin substrate in the tubes as compared to domes may result in a false variation in bioluminescence.

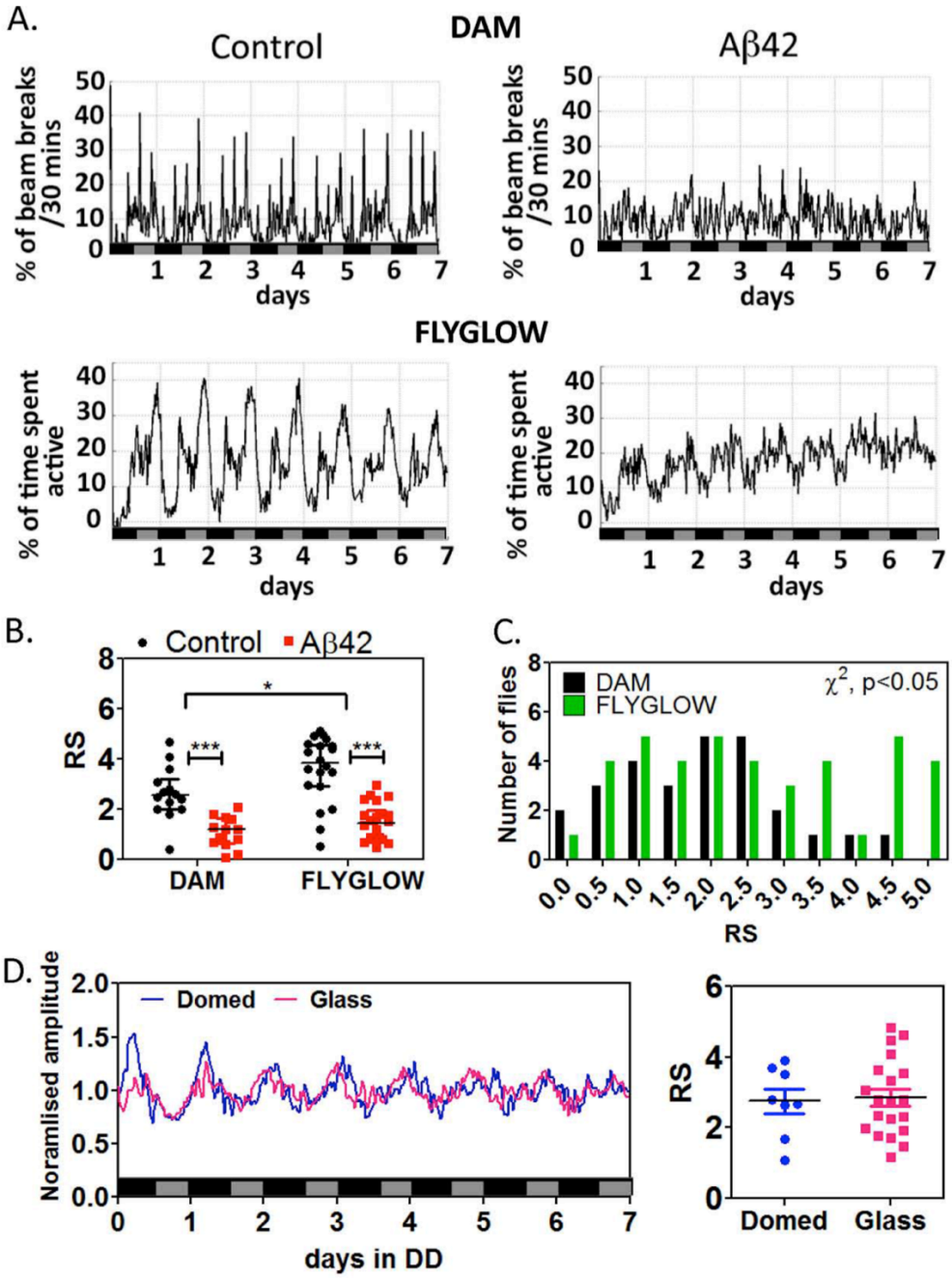

Figure 4.15 Comparison of FLYGLOW with conventional methods. A) The mean locomotor activity for control and $A\beta_{42}$-expressing flies was estimated using DAM (y-axis represents the percentage of beam breaks per 30 mins) and FLYGLOW. X-axis represents circadian time in hours with subjective day (grey bars) and night (black). B) The robustness of locomotor rhythmicity was quantified by calculating RS values. Rhythms were significantly stronger in control flies as compared to $A\beta_{42}$-expressing flies (determined by 2-way ANOVA test). RS values identified by FLYGLOW were not different from that of DAM (2-way ANOVA). C) Frequency distribution of RS values from FLYGLOW is similar to those from DAM. D) Comparing the mean bioluminescence from FLYGLOW and flies observed in domed well plate indicate that molecular clock signals are equivalent (right). No significant difference was found between RS values of molecular clock oscillations between FLYGLOW (blue) and classical approach (red).

After accounting for the luciferase decay, it was observed that both FLYGLOW and the current gold standard housing designs generated essentially identical molecular clock oscillations with similar distribution of RS values (Figure 4.15D). In summary, the above-mentioned comparison indicates that the novel FLYGLOW system is capable of detecting locomotor and molecular oscillations simultaneously with better sensitivity and resolution than existing methods.

4.3.3 SIMULTANEOUS COMPARISON OF CIRCADIAN PARAMETERS

The following analysis considers three time series datasets from control flies expressing only the luciferase reporter construct and also those flies expressing the $A\beta_{42}$ peptide. The dataset are: 1) oscillations of the peripheral molecular clock as reported by the intensity of emitted light, 2) locomotor activity as a percentage of time spent moving and 3) sleep consolidation. For convenience, these parameters are presented as mean traces for a population of flies (Figure 4.16).

As it can be seen in Figure 4.16 A, the detrended time series of molecular clock for control flies from three experiment runs (total number of flies in control group N=60) are characterized by the high amplitude of oscillations that peak in the early part of subjective night (black bars along x axis) and robust circadian oscillations with gradual dampening over the course of 7-day experiment (Veleri, Brandes et al. 2003). Similarly, $A\beta_{42}$-expressing flies (three experiment runs, total number of flies N=45) exhibited the rhythmic oscillations of peripheral molecular clocks, however the deterioration started from day 4 onwards.

The locomotor activity of control flies, as estimated by percentage of time fly spent in activity, showed highly circadian rhythms with two distinguishable peaks of activity; the first usually occurred just before the subjective dusk, when flies spent up to 50% of time moving, and second just before dawn with approximately 40% of time spent on movement (Figure 4.16 A). In contrast, $A\beta_{42}$-expressing flies exhibit low-amplitude and less rhythmic oscillations of locomotor activity that can be described as a state of continuous restlessness during the subjective day and night.

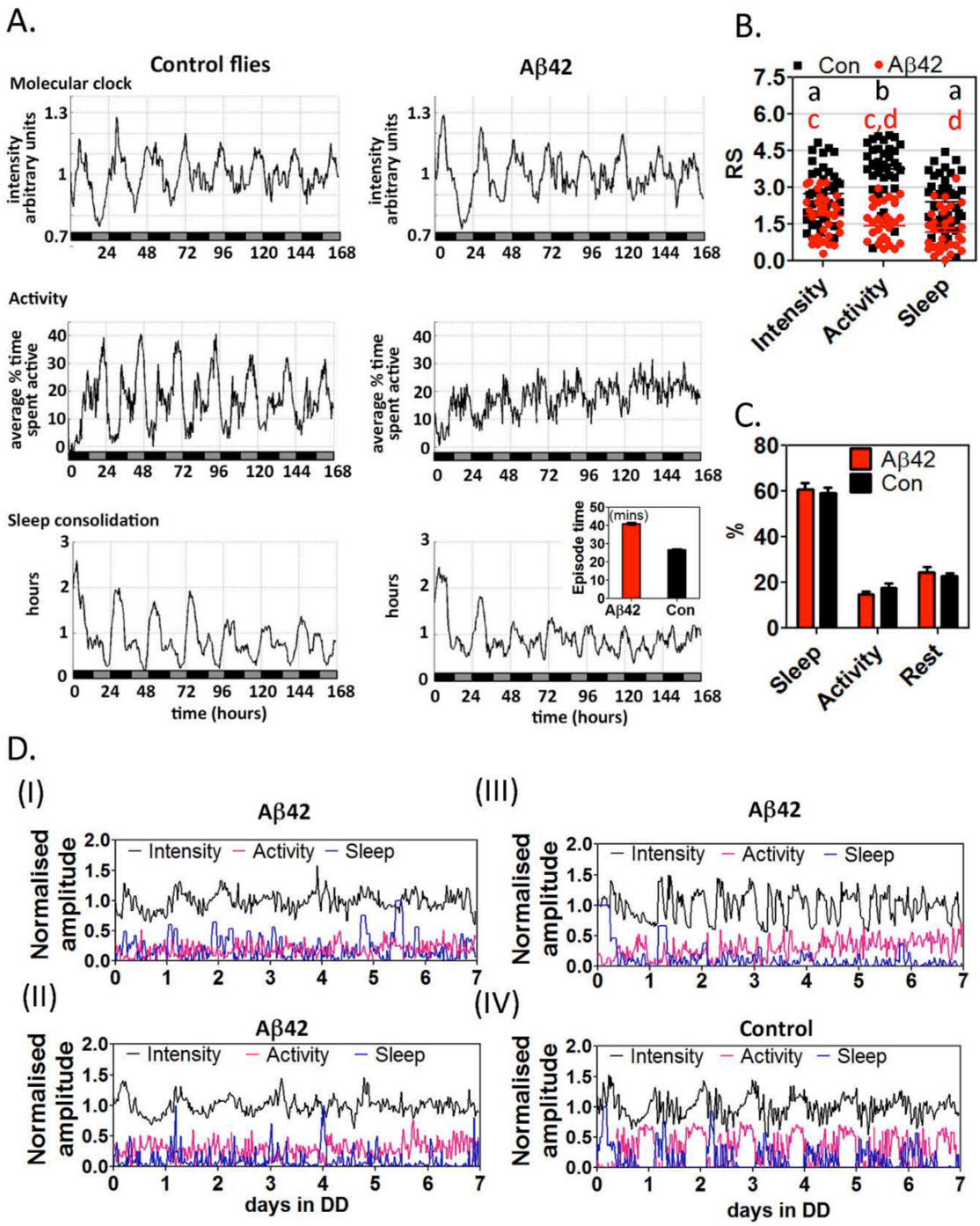

Figure 4.16 Comparing the rhythmicity of control and $A\beta_{42}$-expressing flies. A) Population mean of bioluminescence (Molecular clock), percentage of time spent active (Activity) and length of each sleep episodes (Sleep consolidation) were plotted for control and $A\beta_{42}$-expressing flies. Insert demonstrated that daytime sleep episodes were longer in $A\beta_{42}$-expressing flies (red) as compared to controls (black). B) All 3 rhythms are significantly less robust in $A\beta_{42}$-expressing flies (red dots) as compared to controls (black dots), determined by RS values. The same letter alphabet is indicated for the group with no difference in the RS values (one-way ANOVA). C) The distribution of relative partitioning of time between sleep, activity and rest showed no difference for $A\beta_{42}$-expressing flies (red) and controls (black). D) Panels I and II: $A\beta_{42}$-expressing flies exhibited weak behavioural rhythms (activity in red and sleep consolidation in

As can be expected, the oscillations of sleep consolidation are in anti-phase with oscillations of locomotor activity, where often the longest sleep episodes occur when locomotor activity is at its lowest (Figure 4.16 A). Specifically, in the early part of the subjective night, when control flies spend less than 5% of their time moving, the duration of an average sleep episode in flies is surprisingly high, exceeding 2 hours of undisrupted sleep. This structure of sleep consolidation superficially resembles the inverted human hypnogram (graph representing sleep stages as a function of time), where the deep sleep stage, as analogues to the longest sleep episode in flies, is usually appear earlier in the night. During the morning time of the subjective day, the flies normally take short naps, not longer than 30 minutes, and start to demonstrate significant locomotor activity (20-30% of time spent moving). Peak locomotor activity was observed in the late afternoon coincide with the shortest sleep episodes. Remarkably, the inset in the in Figure 4.16A, that represent duration of daytime sleep episodes, suggests that $A\beta_{42}$-expressing flies (red), similarly to patients with AD, have longer sleep episodes in the subjective day as compared to the control group of flies (black).

The visual interpretation of circadian oscillation assessed at the population level (Figure 4.16 A) has been validated quantitatively by calculating rhythmicity statistic (RS) value for each individual fly. According to the RS values (Figure 4.16 B) $A\beta_{42}$-expressing flies (red dots) are characterized by the significantly reduced robustness of all three rhythms as compared to control group of flies (black dots). Moreover, significantly higher proportion of flies in the control group exhibit rhythmic oscillations (85% for molecular clock, 71% for locomotor activity and sleep consolidation) in comparison with $A\beta_{42}$-expressing (66% for molecular clock, 42% for locomotor activity and 42% sleep consolidation). Importantly, the average proportions of sleep (sum of rests longer than 5 minutes), activity and rest (sum of rests shorter than 5 minutes) was the same for both control and $A\beta_{42}$-expressing flies (Figure 4.16 C). Notwithstanding this, $A\beta_{42}$-expressing flies on average had significantly longer sleep episodes during the daytime. This observation indicates that the differences in sleep are attributed to its temporal organization rather than total amount of sleep.

The simultaneous recording of different circadian oscillation gives an opportunity to investigate the relationships between the individual rhythms using

single-fly data (Figure 4.16 D). For example, a large proportion of $A\beta_{42}$-expressing flies exhibits arrhythmic behaviour but have rhythmic oscillations of molecular clock (panel I and II). Moreover, some of the individual $A\beta_{42}$-expressing flies had disrupted rhythmicity of all three parameters (panel III), whereas the majority of the flies in control group retain the rhythmic oscillations of all three parameters throughout the course of the 7-day experiment (panel IV). In general, these observations might suggest that Aβ expression in the nervous system in the *Drosophila* can potentially results in the progressive deterioration of behavioural and molecular rhythms, but these oscillations decay independently of each other.

4.3.4 SPECTRAL ANALYSIS OF PERIODICITY OF THE OSCILLATIONS

In order to examine more closely the period stability and estimate if the circadian oscillations undergo the period doubling bifurcation, spectral analysis was employed. Conventional methods of spectral analysis, such as MESA and FFT-based periodograms, normally estimate the multiple components of the period for the whole time series and report the result in terms of global periodicities detected over the course of experiment. Therefore, to investigate how the circadian oscillations progressively deteriorate over the course of the experiment, which applies especially to the rhythms exhibited by $A\beta_{42}$-expressing flies, Lomb Scargle periodogram was implemented based on a sliding window principle, as mentioned in 4.2.4.3. The application of this approach to the time series of sleep consolidation (Figure 4.17 A), revealed the most significant differences in period during the course of the experiment. This spectral analysis reveals that control fly retained stable periods of around 24 hours for sleep consolidation (Figure 4.17 B, left) throughout the 7 days of the experiment. In contrast, sleep consolidation in $A\beta_{42}$-expressing fly is characterized by the progressive period deviation from 24 hours (Figure 4.17B, right). At a population level, the average spectral power of the sleep rhythms (black traces in Figure 4.17 C, left) demonstrates that control flies exhibit strong 24-hour period of oscillations, which is consistent among the flies in control group (pink shade in Figure 4.17 C, left). In $A\beta_{42}$-expressing flies, spectral power for 24-hour component is less pronounced and has higher variability between within the population of flies (Figure 4.17 C, right), indicating that $A\beta_{42}$-expressing flies have difficulties in maintaining a regular sleep homeostasis in constant darkness.

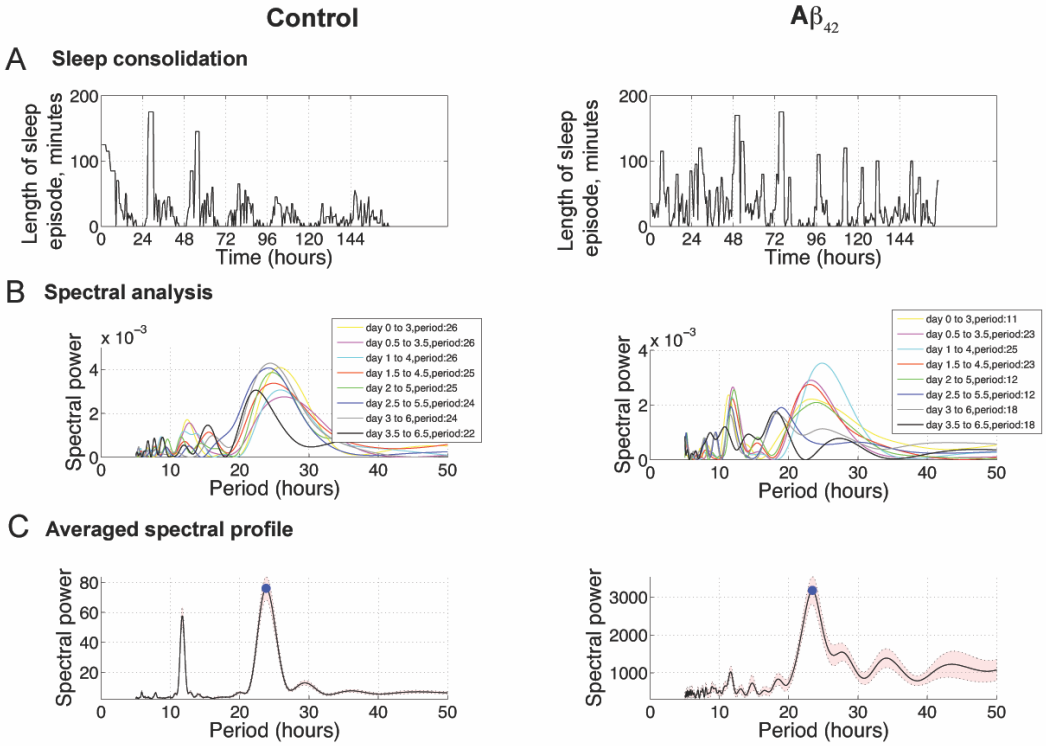

Figure 4.17 Spectral analysis of sleep consolidation time series. A) Sleep consolidation for representative single control or $A\beta_{42}$-expressing fly was plotted as a function of time. B) The spectral analysis of sleep time series using the sliding window approach showed the stable period of oscillations for control fly but in $A\beta_{42}$-expressing fly period progressively deviates from 24 hour component. C) Average spectral profile of control fly (black line) demonstrated the presence of strong 24 hour component with additional ~12 hour component in the sleep time series. $A\beta_{42}$-expressing flies are characterized by the higher variability within the population of flies (pink shade).

4.3.5 COORDINATION BETWEEN MOLECULAR CLOCKS AND CIRCADIAN BEHAVIOURS

An additional benefit of recording multiple circadian rhythms simultaneously with FLYGLOW is that subtle correlations between signals can be detected. Of particular interest are correlations in rhythm robustness between two signals suggesting they may be coupled. Moreover, correlations between the free-running periods of rhythms may indicate that there are underlying mechanistic similarities even if the rhythms are not directly coupled. Moreover, simultaneous measurement of multiple circadian rhythms for the same individual fly provide opportunities to

perform paired statistical tests, which has not been possible previously because different rhythms were observed in different groups of flies using different experimental equipment. With regards to the robustness of rhythm, the general assumption is that if RS values of two rhythms of interest are correlated, then they might be coupled together. For instance, if one rhythm is coupled to another then if one of the rhythms is robust, then it is expected that the other will also be robust, and vice-versa. Moreover, if a master rhythm drives a slave then it is impossible that the slave can be rhythmic if the master is arrhythmic.

Implementing a three-way analysis of RS correlations for individual rhythmic flies in control group have indicated that there is no coupling between molecular clock and either of the behavioural rhythms (black dots in Figure 4.18 A, panel I and II) as no significant correlation was found using Pearson correlation coefficient. However, significant correlation was found for RS values for locomotor behaviour and sleep consolidation (panel III). Taking into account that sleep consolidation is not the simple reciprocal of activity (but rather represents the length of each sleep episode) and so this is not a trivial correlation, it might suggest that both locomotor activity and sleep rhythms are coordinated by the same central clock mechanism, but supposedly through different output pathways. Likewise, there were no robustness correlations observed between behavioural rhythms and peripheral clock oscillation in rhythmic $A\beta_{42}$-expressing flies (red dots in Figure 4.18 A, panel I and II). Furthermore, there the correlation between sleep and locomotor activity rhythms seen in control flies (red dots in Figure 4.18 A, panel III) was not present in $A\beta_{42}$ flies. This dissimilarity might indicate that neuronal coordination is compromised in $A\beta_{42}$-expressing flies and consistent with significantly reduced robustness of behavioural rhythms even in rhythmic $A\beta_{42}$-expressing flies as compared to controls (lower RS values in Figure 4.18 A).

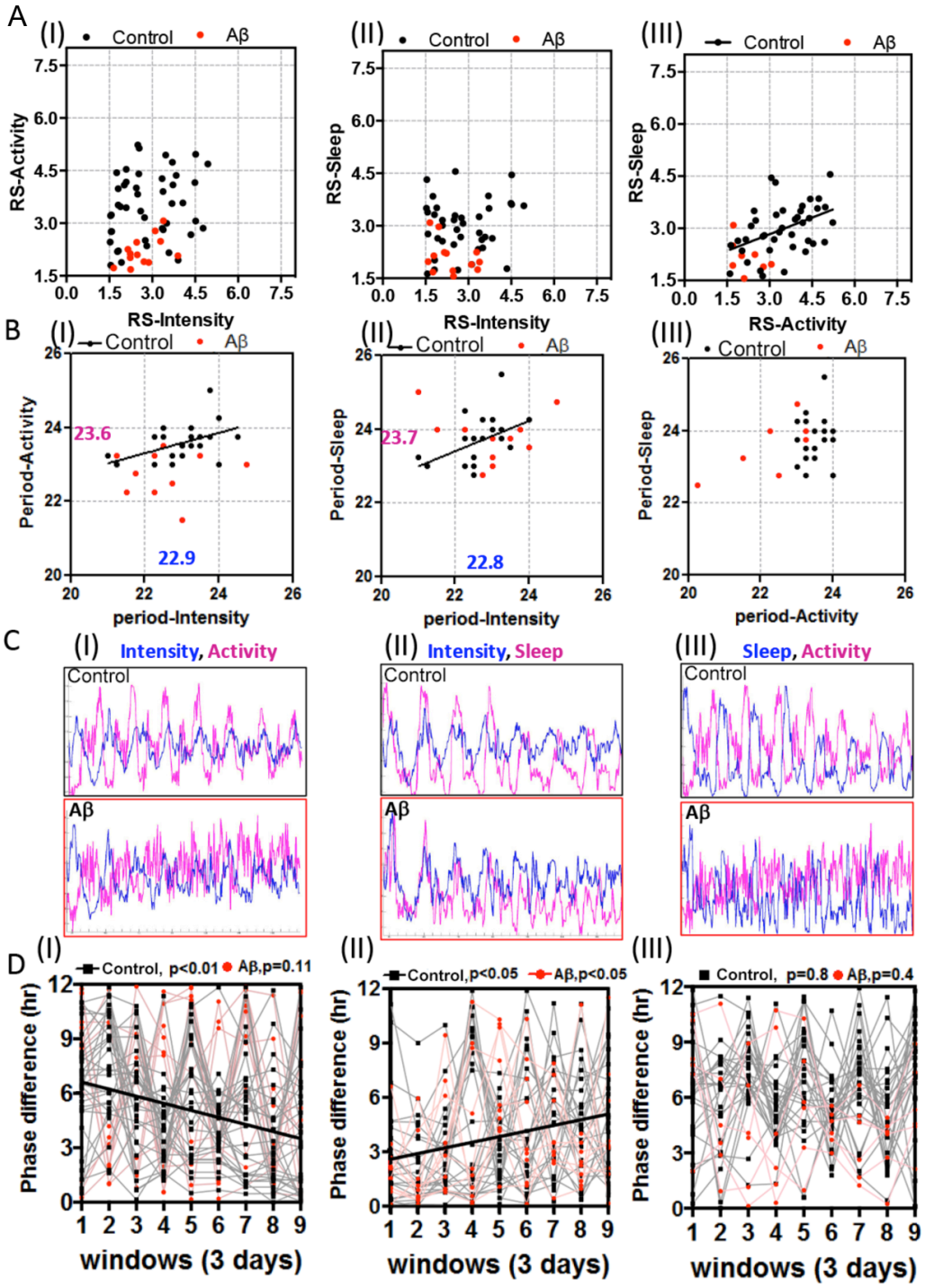

Figure 4.18 Investigating the RS values and periods correlations. A) Paired RS values for various rhythms were plotted for each fly in control (black) or $A\beta_{42}$-expressing population of flies (red). B) Paired periods for various rhythms were plotted for each fly. C) The correspondent pairwise rhythm comparisons were plotted. X-axis represents the circadian time in hours. D) The phase difference between rhythms across the time course was calculated using the sliding window approach and plotted for each fly (the phase differences were plotted between the oscillation of molecular clock (intensity) and locomotor behaviour (activity) in panel I, between molecular clock and sleep consolidation in panel II and locomotor activity and sleep consolidation in panel III.

A similar systematic three-way analysis of rhythm periods in individual control flies have demonstrated the correlation between peripheral molecular clock and both of the behavioural rhythms, even though their RS values are not correlated and the actual periods of the rhythms are different (Figure 4.18 B, panel I and II). These observations indicate that period co-variation is significant for molecular clock vs. locomotor activity and for molecular clock vs. sleep consolidation. The implications are that these pairs of rhythms share common underpinning mechanisms because, taking into account slight variation between individuals, control elements that change period of one rhythms also tend to modify the period of the correlated rhythms. Interestingly, there we could not detect period correlations for sleep consolidation and locomotor activity (panel III). While the underpinning molecular mechanisms are likely to be identical, as deduced from the two period correlations described above, the implication is that sleep rhythms are modified by additional environmental factors that are highly variable between flies, such as the need for homeostatic regulation of total sleep in an environment that may disturb sleep (Wu, Koh et al. 2008, Kunst, Hughes et al. 2014). Similarly to the control flies, the periods of the sleep and activity rhythms do not show any correlation for $A\beta_{42}$-expressing flies. Furthermore, the periods of peripheral clocks did not correlate with either of sleep or activity rhythms (red dots in Figure 4.18 B), indicating that the accumulation of Aβ in the brain damages the connections between the molecular clock and behavioural oscillations even in flies with rhythmic oscillations of molecular clock.

Notably, comparison of the periods of each rhythms using the paired t-test have demonstrated that the periods of sleep and activity oscillations (23.7 and 23.6 hours respectively) are in fact significantly longer than the period of peripheral molecular rhythms (~23 hours) in control population of flies (Figure 4.18 B), suggesting the involvement of non-circadian mechanism in the daily behavioural oscillations. The shorter period of the molecular clock can be visualized by its progressive desynchronisation with respect to the behavioural readouts. Indeed, in Figure 4.18C there are 8 peaks of the molecular clock signal for only 7 peaks of the behavioural rhythms).

In order to explore this difference in period in more detail, the phase of the peaks in both behavioural rhythms and molecular oscillations for each individual fly were compared with each for 3 day windows across the recording time. As can be

seen in Figure 4.18 D, for control flies there was a progressive change in the phase difference between behavioural and molecular oscillations over the course of the experiment (black symbols, decreasing trend between activity and molecular rhythms in Figure 4.18 D-I and increasing differences between molecular and sleep rhythms in Figure 4.18 D-II), while for control flies the phase difference between the two behavioural rhythms exhibited a stable ~6 hours phase delay (black symbols, Figure 4.18 D-III).

On the other hand, no significant phase difference was detected between either of the activity or sleep oscillations and molecular rhythms (red dots, Figure 4.18 D) for $A\beta_{42}$-expressing flies, probably because of the compromised rhythmicity of the oscillations (Figure 4.18 C, red). Visually quantifying the difference in peaks numbers between averaged behavioural and molecular rhythms is more challenging for what seems like chaotic waveforms. In particular, sleep oscillations of $A\beta_{42}$-expressing flies are characterized by the emergence of the shorter wave components (~12 hours) towards the end of the experiment (Figure 4.18 C, panel II), suggesting the disruption of sleep structure and sleep fragmentation.

Overall, systematic correlation analysis has implied that behavioural rhythms and peripheral molecular clock oscillate in the constant darkness with different periods for control flies. However, neuronal damage caused by Aβ expression in fly brain likely lead to the independently compromised rhythmicity of behavioural and molecular oscillations and result in period instability and chaotic phase relationships between rhythms.

4.4 DISCUSSION

Recent studies have shown that circadian rhythms, in addition to their role in the regulation of a variety of biological processes in living organism, also either associated or constitute risk factors for human disease. Several lines of evidence suggest that environmental disruption of circadian rhythms increases the risk of the diseases, while in contrast some studies have demonstrated that some diseases are accompanied by the circadian deficits from the early stage (Reddy and O'Neill 2010, Rakshit, Thomas et al. 2014). Moreover, circadian abnormalities, such as disrupted daily rhythms of sleep and activity, have been observed as one of the symptoms of

several neurodegenerative disorders, including Alzheimer's disease. Generally, until recently the circadian deficits have been conventionally viewed as a consequence of the neurodegeneration. However, emerging literature indicate that disruptions of circadian rhythms may contribute to the neurodegeneration process and influence disease pathology (Musiek 2015). Recent finding that use of sleeping tables (in particular, benzodiazepine) is associated with 50% increase in risk of AD show the involvement of circadian deficits at the early stage of the disease, as a part of dementia prodrome (Steurer 2014). As the disease progresses, patients with AD experience poor quality sleep fragmentation and disruption of normal day/night pattern.

Since circadian clocks are highly conserved, it is a common practice to use animal models to study disease pathogenesis. The circadian mechanisms of one of such model organism, namely *Drosophila melanogaster*, have been comprehensively studied and, when used as model of AD, it successfully mimics key phenotypical aspects of disruption of circadian/sleep regulation associated with human aging and neurodegeneration (Tomioka, Uryu et al. 2012, Chen, Possidente et al. 2014, Rakshit, Thomas et al. 2014). Understanding the mechanisms by which the molecular clock controls behaviour may provide a valuable insight into the disease pathogenesis but to date no appropriate methods have been available, at least for *Drosophila,* to longitudinally monitor multiple behavioural and molecular oscillations in individual animal.

In order to investigate the causal relationship between multiple circadian rhythms, a novel recording system, FLYGLOW, has been developed to simultaneously record peripheral molecular oscillations, sleep and activity rhythms in the same fly. To recognize the oscillations of the molecular clock, a fly strain which expresses period-luciferase transgene XLG-luc2, was studied, with and without crossing in a transgene for Aβ-expression. Using a sensitive EM-CCD camera mounted in the environmentally controlled incubator, images were recorded with the temporal resolution that is compatible with the current working definition of fly sleep (300 sec of immobility) in such a way that the spatial distribution of light signal in a particular tube can be translated into the duration of fly movement or inactivity. This conversion was performed using image analysis techniques (described in 4.2 Materials and Methods), allowing the quantification of multiple circadian parameters,

such as molecular, locomotor and sleep oscillations for the individual organism kept in constant darkness for 7 days. Even though FLYGLOW is not designed to run the experiment in light-dark cycles, it can be used a powerful tool to study non-photo entrainment, using cues such as temperature. FLYGLOW system is an improvement over current alternative methods in *Drosophila*, which can measure only one of the circadian parameters in any given experiment (either circadian behaviour or bioluminescence). Because of the experimental limitations of the current paradigm, comparative rhythm analysis is only possible on a population level, rather than for individual flies, and is unable to assess intra-individual correlations.

Careful examination of the results has shown that, as a population, control flies exhibit complex oscillations of sleep and locomotor behaviour, that remain robust for the duration of the experiment in constant darkness (Figure 4.16). Notably, in the early night the typical duration of sleep episode reaches up to 3 hours of undisrupted sleep. However, in $A\beta_{42}$-expressing flies this long, consolidated sleep was lost early and replaced by fragmented sleep distributed uniformly across the day, with longer daytime naps, as have been confirmed in a recent study (Tabuchi, Lone et al. 2015). Similar disruptions of normal sleep structure and destabilization of sleep mechanism have been observed in patients with AD (Wulff, Gatti et al. 2010, Coogan, Schutova et al. 2013). Remarkably the total amounts of rest, sleep and activity in $A\beta_{42}$-expressing flies are essentially identical to control flies, indicating that the disruption is of sleep architecture rather than the total amount of sleep. In particular, spectral analysis of sleep consolidation time series provided evidence of the progressive period destabilization in $A\beta_{42}$-expressing flies, suggesting the disruption of the control mechanism of a regular sleep homeostasis.

Simultaneously measuring the circadian oscillations using FLYGLOW have revealed that molecular clock in peripheral tissues was more robust than behavioural rhythms in $A\beta_{42}$-expressing flies. Consistent with previous observations that central molecular clock becoming progressively desynchronized from behaviour in $A\beta_{42}$-expressing flies (Chen, Possidente et al. 2014, Long, Blake et al. 2014), the comparisons of rhythm quality for individual $A\beta_{42}$-expressing flies have demonstrated many examples of flies with robust molecular rhythms but arrhythmic behaviour. These recurring findings of the molecular-behavioural asynchrony may imply that behavioural deficits during Aβ related neurodegeneration may be due to the

disruption in outputs of the clocks rather than damage to the peripheral or central clocks themselves (Luo, Chen et al. 2012, Chen, Possidente et al. 2014, Rakshit, Thomas et al. 2014).

The advantage of simultaneous single-fly observations of multiple circadian rhythms using FLYGLOW is the opportunities provided to dissect the causal relationships between circadian oscillations by performing paired correlation analyses of rhythm robustness and period length, which have not been possible before at a population level analysis. As a result of my study, no correlation was found between RS values of the peripheral molecular clock and either sleep or locomotor behavioural rhythms in mid-aged control and $A\beta_{42}$-expressing flies. These observations indicate an absence of inter-coupling between peripheral molecular clock and behavioural rhythms, as have been previously suggested by several studies (Hege, Stanewsky et al. 1997, Giebultowicz, Stanewsky et al. 2000). Performing similar correlation analysis of periodicities of the circadian oscillation has revealed the interdependence of peripheral molecular clock and locomotor behavioural rhythms despite non-correlated RS values, confirming the common belief that the pace of the central molecular clock determines the pace of circadian behaviour. Notably, such correlation was not recognized for $A\beta_{42}$-expressing flies, providing the evidence that Aβ–mediated toxicity damages the connection between molecular clock and circadian behaviours. Under these circumstances, the developed method may be of particular interest to investigate the role of peripheral clock in the age-related behavioural arrhythmia (Luo, Chen et al. 2012, Rakshit, Thomas et al. 2014), which characterized by the higher degree of inter-individual variation in the progressive circadian behavioural disruptions (Koudounas, Green et al. 2012).

On the other hand, the observed correlation between robustness of sleep consolidation, which characterises the daily fluctuations of sleep duration rather than being a simple reciprocal of locomotor activity, and locomotor behaviour indicate a degree of coupling between their rhythms. However, the absence of correlation between the periods of sleep and activity oscillations suggests more complex mechanism of sleep/wake regulation that may respond to other homeostatic and regulatory cues (Bushey, Tononi et al. 2011). In particular, in humans forced desynchrony of 14:14 light:dark cycle lead to the similar dissociation, when circadian rhythms become progressively decoupled from sleep-wake cycle.

Interestingly, it was observed that peripheral and behavioural rhythms free run in the constant darkness with slightly different endogenous periods. The shorter period (22.6 hour) of luciferase oscillations of the XLG-luc2 transgenic flies has been reported in previous studies, but the period of circadian behaviours has not previously been documented (Veleri, Brandes et al. 2003). The difference in periodicity between behavioural and molecular rhythms is visualized graphically by plotting the phase relationships between two oscillations throughout the course of the experiment. My observation confirms an earlier study showing that when the peripheral clock in transplanted Malpighian tubules is out of phase it continues to free run independently of the circadian clock of the host fly (Hege, Stanewsky et al. 1997). Even though FLYGLOW recorded the bioluminescence emitted from the whole adult fly and unlikely obtained signal from this specific excretory organ, it nevertheless agree with the hypothesis that the majority of peripheral clocks in *Drosophila* are autonomously entrained by environmental light cues and are independent of central clock (Plautz, Kaneko et al. 1997).

4.5 CONCLUSION

In this chapter, a novel recording system FLYGLOW was developed to simultaneously monitor circadian behaviours and peripheral molecular clocks, which gives an opportunity to study the causal relationships between multiple circadian rhythms in the freely moving flies. The results of the experiment have demonstrated that the peripheral clock is not directly coupled to circadian behaviour in mid aged flies. The advantage of the simultaneous measurements of circadian rhythms permitted paired statistical comparisons that boost the statistical power of my insights into the aetiology of complex age-related neurodegenerative diseases, such as Alzheimer's disease. Moreover, FLYGLOW can be readily applicable to more broadly based behavioural, genetic and drug-based screens.

CHAPTER 5: STUDY OF POSITIONAL PREFERENCE IN *DROSOPHILA* AND HOW THIS IS DEGRADED IN A MODEL OF ALZHEIMER'S DISEASE

It is commonly observed that patients with Alzheimer's disease, in addition to the progressive memory loss, experience spatial and temporal disorientation as one of the main symptom of neurodegeneration. Spatial orientation is complex cognitive task that relies on the topographic learning and perception of the surrounding and may be impaired in AD, leading to the severe difficulties to navigate in a familiar environment (Giannakopoulos, Gold et al. 2000). This chapter will focus on understanding the spatial orientation in *Drosophila* and how this may be disturbed in a model of AD.

5.1 INTRODUCTION

In most cases of AD the hippocampus, which is responsible for memory formation, is one of the first regions of the brain to be damaged. As the disease progresses, the neurodegeneration spreads to the cortical regions, resulting in a wide range of cognitive dysfunction, one of which is spatial disorientation. Recent discovery of the positioning system within the brain have demonstrated the involvement of hippocampus in the spatial memory, where the "place cells" formed a map within the brain alongside the "grid cells' in the entorhinal cortex, which are responsible for navigation and positioning (O'Keefe and Dostrovsky 1971, Fyhn, Molden et al. 2004).

Moreover, the ability to navigate and orient themselves is an important survival skill for the any motile organisms, from bacteria to flies and vertebrates. In order to succeed, living organisms have to extract relevant cues from the environment, process them and incorporate into internal representation to determine the adequate respond to any situation. For example, simple organisms, such as bacteria *E.coli*, direct their movement according to the chemical stimuli, a process known as chemotaxis. These œbacteria use locomotor organelles called flagella to propel themselves at speeds of 10-20 body lengths per second. In uniform chemical

environment, *E.coli* swims in random walk pattern as a result of episodes of straight runs lasting for about a second and episodes of random flagella rotation. In an attractant or repellent environment, the cell assesses chemical concentration and modulates the probability of tumbling frequency. It is important to note that during a tumbling event *E.coli* is unable to choose the subsequent direction of movement but simply resets its orientation randomly. Thus bacteria have a "run-and-tumble" method of locomotion consisting of two modes; the first is the "run" where the microbe moves in a straight line and the second is the random tumble. As a result, bacteria direct their movement towards the favourable location, up the gradient of a chemoattractant, by controlling the probability of transition between these two modes of motion in a stimulus dependent manner (Berg and Purcell 1977).

It has been found that other living organisms use similar chemotactic strategy to some extent. Several lines of evidence shown that insects may use stereo chemosensation, a mechanism of locating odour sources by comparing the inputs transmitted by each antenna. For example, adult flies demonstrate a wide range of odour driven behaviours. Odour cues regulate not only identification of food sources and social communication between individual flies, but also walking and flight behaviour (Gomez-Marin, Duistermars et al. 2010). In general, the information obtained from the olfactory system is integrated with input from other modalities, such as visual system, and translated into the set of motor commands that direct the fly towards its target. Interestingly, it has been demonstrated that motor response to the visual and olfactory cues presented simultaneously are identical to the linear superposition of stimuli presented in isolation, suggesting that they are governed by separate but parallel sensorimotor pathways (Frye and Dickinson 2004). However, in absence of both visual and olfactory stimuli, such as for example in case of confined space and complete darkness, *Drosophila* can estimate its position relative to the reference point by continuously monitoring its movement and therefore relying on self-motion cues. Moreover, in the absence of both visual and self-motion cues, the representation of the fly position is sustained through persistent activity, a potential equivalent to the short-term memory (Seelig and Jayaraman 2015).

In addition, several reports about fly behaviour have found that *Drosophila* have periodic preference patterns for their position within a tube, specifically that prior to rest, the flies positioned themselves relative to their food (Hendricks, Finn et al. 2000, Donelson, Kim et al. 2012). However, little is known about the rationale

behind the positional preferences, indeed the current literature indicates that flies use presumed odour gradients to position themselves in a tube in the absence of other stimuli. Furthermore it is unknown to what extend the neurodegeneration associated with AD affects spatial orientation in the flies. Therefore, in order to investigate the control function underpinning diurnal orientation of the flies, the previously developed FLYGLOW system was used to observe flies behaviour as described in chapter 4. The flies were confined in what is essentially a one-dimensional space, provided by the capped capillary tubes, and polarized by food at one end and air at the other end of the tube. The experimental design ensured that the flies were maintained in constant darkness, ensuring that no visual cues were present and we reasoned therefore that the fly can estimate its position along the tubes based only on olfactory and self-motion cues. However, the exact mechanism of how fly integrate the feedback from these cues is still unclear. For this reason, two models of fly behaviour in response to the olfactory cue (food smell) in constant darkness were developed. In model 1 it is assumed that flies use olfactory cues (food odour stimuli) to orient themselves in the tubes and therefore their location preference relies on judgment of the relative distance along the tube, namely the percentage of the tube length away from the food. In comparison, model 2 is based on the assumption that flies use self-motion cues to estimate their position and, as a result, the positional preferences are based on the absolute distance from the attractant source, meaning that flies can effectively count their steps away from the food. For example, Donelson and colleagues (Donelson, Kim et al. 2012) suggested that flies have a positional preference to sleep a quarter of the way along the tube close to the food. This observation does not distinguish however between model 1, where the flies estimate their position based on the olfactory gradient of food, or model 2, where spatial awareness is based on the absolute distance away from the food end of the tube. It was only possible to determine which model best accounts for the position preferences by housing the flies in capillary tubes of different lengths and using FLYGLOW to observe their movements.

5.2 MATERIALS AND METHODS

To study the positional preferences in *Drosophila*, the FLYGLOW system (as described in chapter 4) was used to monitor the locomotor activity and determine the presence of sleep-like state of control and $A\beta_{42}$-expressing flies constrained in tubes of different length.

5.2.1 EXPERIMENTAL DESIGN

For this project, the experiment was set up in a same way as FLYGLOW system described in chapter 4. Control and $A\beta_{42}$-expressing strains of flies were generated with the same genotype as described in 4.2.1.1. In brief, to monitor fly movement in constant darkness, *period*-luciferase transgenic fly strain XLG-luc2 was crossed to the UAS-$A\beta_{42}$ or a background control strain. The resulting $A\beta_{42}$-expressing fly strain, elav-gal4;UAS-$A\beta_{42}$/+;XLG-luc2/+, and control strain, elav-gal4; +/+;XLG-luc2/+, were used in the experiment.

In order to examine the fly spatial navigation and positional preference in either sleep or wake state, custom made glass tubes of different length were used. Control and $A\beta_{42}$-expressing flies were placed in capped glass tubes with 5 mm in diameter and 4 different lengths: 1) 3.5 cm, 2) 7 cm (standard length, conventionally used), 3) 10.5 cm and 4) 14 cm. Due to the limited capacity of the incubator, 8 flies of each of the required genotypes were loaded into tubes of each of the four different lengths.

The bioluminescence images of control and $A\beta_{42}$-expressing flies, taken with EM CCD camera, were analysed and five different parameters were calculated for each fly as described in 4.2.3. The original 5 minutes temporal resolution of time series was used for the analysis of fly positional preference (1961 time points taken every 5 minutes = 6.8 days of experiment).

5.2.2 FLY POSITION IN A TUBE

In order to establish the position of a particular fly in a tube, two time series of the circadian parameters, measured with the FLYGLOW system, were used:

bioluminescence data in terms of intensity of emitted light and the sleep binary array, that codes for the presence of a sleep-like state.

As several studies have shown, most animals exhibit environmental preferences for sleeping (Campbell and Tobler 1984), accordingly fly positional preferences were assessed when the fly was in each of two behavioural states: when awake and when asleep. The data can be acquired from the series of captured images, as a time series of bioluminescence distributions of a particular fly along a tube over a 300 sec exposure. A sleep or wake state that was inferred is held in the sleep binary array, calculated as described in chapter 4. The length of tubes was represented as the number of pixels the tube occupies in an image. Knowing the absolute length of tubes, it was estimated that 1 mm corresponds to 3 pixels. As stated in 4.2.3.2, the tube length represented by the image was divided into 4 pixel high bins along long axis to reduce the effect of minor errors, resulting in 27 bins for 3.5 cm long tube, 48 bins for 7 cm tube, 76 bins for 10.5 cm tube and 101 bins for 14 cm tube. The slight variation of the tube lengths was due to the fact that effective length of the tube, proportional to the tube length available to the fly for movement, was limited by the cotton wool on one end and fly food loaded in the tube on the other end.

Considering the data for either wake or sleeping flies separately, the fly position in a tube was visualized using the heat map approach, where individual values of intensity of light are represented as colours. In order to standardize the fluctuation of the intensity across the time points, the intensity values X_i for an individual fly for a given image frame i were normalized as follows:

$$X_{iN} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (5.1)$$

where X_{iN} - normalized intensity of emitted light in the range [0,1], X_i – original intensity value, X_{min}/X_{max} – minimum and maximum values of the intensity at a given image frame. The x-axis in the heat map represents the elapsed time, while the y-axis corresponds to the spatial bin location. The colours in the heat map indicate the probability of the fly being at a given location in terms of spatial distribution of the bioluminescence across the tube, where lighter colour represent the highest probability and black indicate the absence of the fly presence at a given position. The heat map plots consisted of the mean residency probability for the 8 flies of each genotype for each position within each tube and at each time point.

Furthermore, the average residency of every fly along the tube was established. First, a mean values of the intensity profile were extracted for day 1 to day 3 (in accordance with multiple experimental procedures, for example, by (Donelson, Kim et al. 2012, Faville, Kottler et al. 2015), averaging the data across the time points. To get the average positional preference, these values for day 1-3 were summed and normalized as described above, so that values are in a range [0, 1].

5.2.3 MODELLING FLY MOVEMENT

In order to determine the underlying mechanism behind the food attraction and positional preferences, specifically during the awake periods when the fly is mobile, two models of fly location preferences were developed.

5.2.3.1 MODEL 1: OLFACTORY CUES

It is assumed that the odour gradient in a confined space is distributed linearly as a result of gaseous diffusion, with the highest concentration near the food source and lowest near the air end of the tube.

Assuming that the particles diffuse at a constant rate in an approximately one dimensional space with the open end, the equation of the volumetric concentration of particles is given by:

$$-D \cdot \frac{\partial c}{\partial x} = A \tag{5.2}$$

where $\partial c/\partial x$ – change of the particle concentration c along the position x, D - diffusion coefficient, A – constant.

Solving the equation for ∂x gives the linear equation of the concentration gradient:

$$C(x) = -Ax/D + B \tag{5.3}$$

Assuming that the concentration of the particles at position x = 0 (food end): $C(0) = C_1$ and at position x = L (air end): $C(L) = C_2$:

$$C_1 = -A/D \cdot 0 + B, \Rightarrow B = C_1 \tag{5.4}$$

$$C_2 = -A/D \cdot L + C_1, \Rightarrow -\frac{A}{D} = \frac{C_2 - C_1}{L} \tag{5.5}$$

$$C = \left(\frac{C_2 - C_1}{L}\right) \cdot x + C_1 \tag{5.6}$$

Approximating the boundary conditions so that the particle concentration at the food end will be equal to 100% (C_1 = 100 %) and at the air end concentration decreases down to the unknown value y near zero (C_2 = y %), the equation of particle concentration will be as follows:

$$C(x) = \frac{(y-100)}{L} \cdot x + y \qquad (5.7)$$

According to this simplified equation of particle diffusion with the assumption of steady state environment and the constant rate of the concentration flux, concentration of the food smell in the capped capillary tube with open end on one side is a linear function of position, x.

The concentration gradient of the food smell along the tubes of different length decreases linearly from 100% near the food end to some constant values y% near the air end and can be plotted as shown in Figure 5.1.

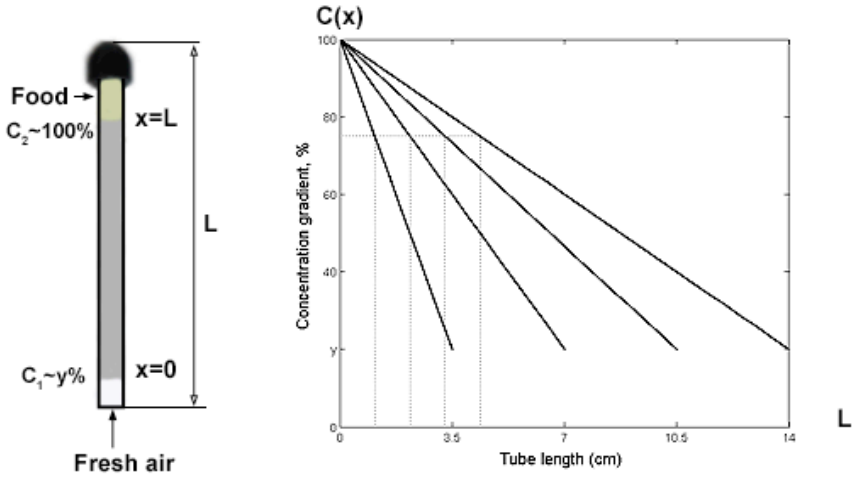

Figure 5.1 Concentration gradient in different lengths tubes. It was assumed that in a capillary tube of length L the concentration of the food smell will diffuse linearly from 100% near the odor source to a relatively small value y% near the air end of the tube (left). The plot on the right showed the distribution of the concentration gradient (y-axis) within the tube of different length (represented in x-axis).

With reference to the finding from (Donelson, Kim et al. 2012) that flies prefer to stay quarter of the tube length away from the food, if the fly behaviour follows model 1, meaning that the fly choose preferred position based on the olfactory cues, then we can expect that flies will stay at the location with 75% of the concentration gradient and in tubes of different length this will result in the 25% of the tube length away from the food (1.1, 2.2, 3.3 and 4.4 cm away from food end for 3.5, 7, 10.5 and 14 cm long tubes respectively, dotted line in Figure 5.1).

5.2.3.2 MODEL 2: SELF-MOTION CUES

According to model 2, it is assumed that the flies have absolute spatial awareness and use self-motion cues instead of olfactory stimuli. Generally, this means that irrespective of tube length, flies will always prefer to stay at certain position away from the food (example is shown in Figure 5.2).

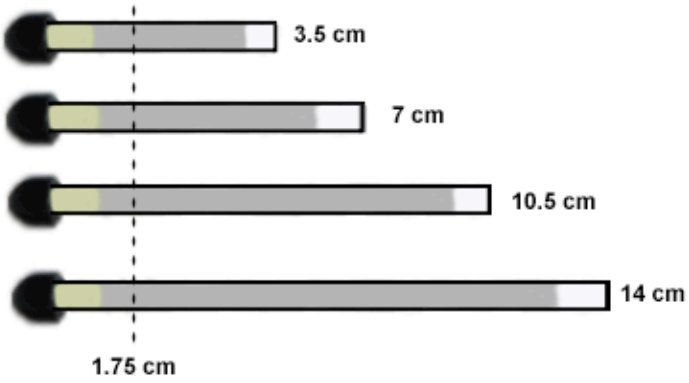

Figure 5.2 Model 2: self-motion cues. Fly stays a certain distance away from the food (for example, 1.75 cm)

Considering the above-mentioned observation made by (Donelson, Kim et al. 2012), the fly preferred position according to model 2 will be quarter of the tube length away from the food, which is 1.75 cm for the standard 7 cm long tube used in the experiment.

5.2.4 SIMULATING FLY MOVEMENT USING RANDOM WALKS

Fruit fly has an ability to navigate in the environment to search for food, mate or shelter which is mainly achieved during walking. As it has been previously observed that without visual and self-motion cues fly walks persistently towards the target, careful analysis of the fly positional behaviour in the tube may be used to infer some attributes of complex locomotor behaviour. As a consequence, understanding the rules underlying fly locomotion and positional preferences is complicated process. Therefore, in this project we propose to develop a stochastic model of fly locomotion, akin to the run-and-tumble model for *E. coli*, and use the experimental data to test it.

The model of fly movement was based on the assumption that in one-dimensional space the flies move in random walk manner and the persistence of the activity was simulated as a probability of changing direction (where the directional persistence probability, DPP=1 the fly will always walk in one direction and where DPP=0 the fly will change direction with every step). As we observed from the experimental data, the positional preference appears as a skewed normal distribution, specifically the DPP was modelled as a log-normal distribution. There was also a consistent peak in residency near the air end of the tube; because this air peak appeared to be a rather stereotyped end effect these data were not included when assessing the goodness of the models.

The probability distribution function (PDF), that results from the DPP, $P(x)$ for the log-normal distribution was calculated for positions along the tube x = [0 ... L] as follows:

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{(\ln x - \mu)}{2\sigma^2}} \quad (5.8)$$

In order to find the best fit of the model to our experimental data, the PDF of DPP was calculated for 7 cm long tube (as it is conventionally used) and the parameters of the distribution were optimised in a sensible range of μ = [1.8 ... 3.4] and σ = [0.02 … 0.2]. As we were more concerned with the positional preferences near the food end of the tube, the spike of the activity near the open end of the tube was not taken into account when evaluating the model and the last two bins of the tube length were excluded in the consequent analysis (so 46 bins instead of 48 bins for 7 cm long tube). To avoid the over-fitting of the model to the experimental data, the set of parameters for the model optimisation was limited to only two (mean and standard deviation for the PDF). In addition, PDF was normalized to unity, so that the probabilities values at each position along the tube will sum up to 1 and can be used to estimate the likelihood of the fly directional persistence.

In our model, locomotion of the fly is governed by a few rules: 1) The fly starts each simulation in the middle of the tube; 2) at each iteration of the model simulation, each equivalent to 1 sec or real time, the probability that the fly will change direction is determined according to the PDF value for its previous position; 3) at the end of either tube ends the fly reverses direction; 4) the fly walks with constant velocity. For each of the 300000 iterations that comprised a simulation experiment (comparable to the 3 day duration of the experiments: 1 step/sec x 60 sec

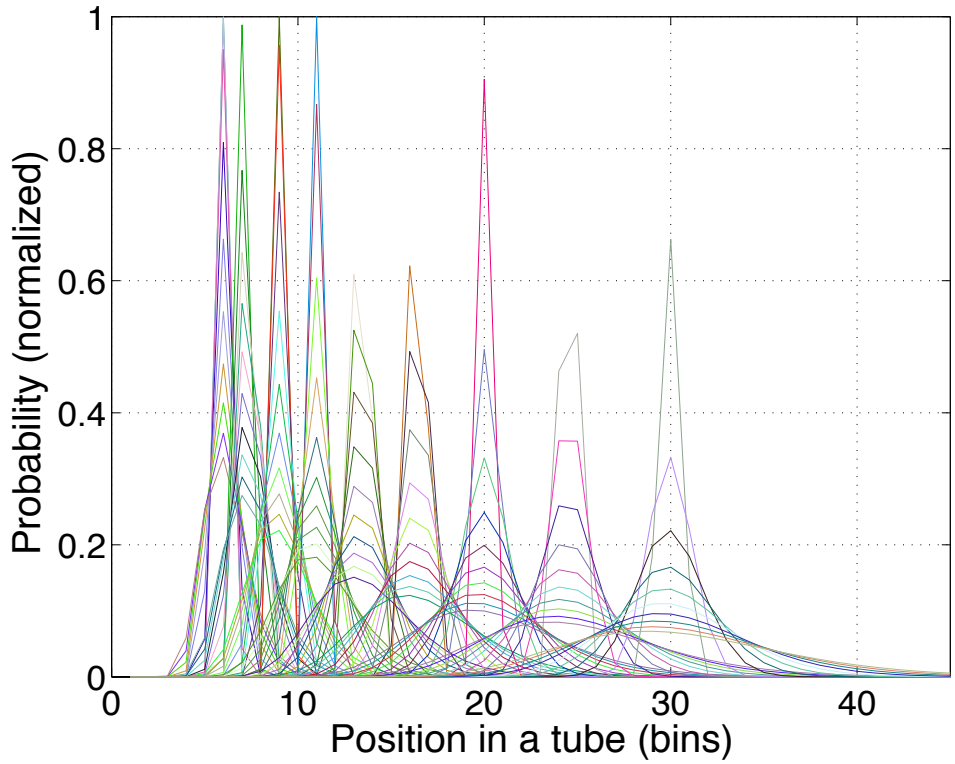

Figure 5.3 Normalized probability distribution function. The PDF was estimated for log-normal distribution with the desired set of parameters.

x 60 minutes x 24 hours x 3 days = 259200 steps) a random number (range 0-1) was drawn from a standard uniform distribution and compared with the DPP value for the previous position of the fly in the tube. The fly reversed its direction only if the drawn random number was higher than the probability derived from DPP. It was therefore necessary to adjust the boundary conditions of the DPP so that the lowest probability should be above 0 (to prevent being trapped by changing direction every iteration), and the highest probability value below 1 (so that there is always at least a small background chance of changing direction). The lower boundary was set to 0.05 and higher boundary to 0.99. The pseudocode describing the algorithms is presented below.

Algorithm 2: Modelling the fly behaviour in constant darkness

Input: tube length L, PDF of experimental data for tube 7 cm long
direction_space = [left, right]; *tube_bounds* = [0, L]; #position boundaries of the tube
for μ=1.8 to 3.4 step 0.2 **do**
    for σ=0.02 to 0.2 step 0.02 **do**
        n = 1 #simulation; *direction(1)*=right; *x(1)*=L/2 #initial position in a tube;
        DPP = lognormal(μ, σ, L)

```
7:          while n<300000 do
8:            random_number = random(0,1) #draw random number between  0 and 1
9:            n = n+1;
10:           if random_number<DPP(x(n-1)) do  #same direction of movement
11:                 direction(n)=direction(n-1);
12:           else #change direction of movement
13:                 direction(n) = difference(direction(n-1), direction_space);
14:           update current position x(n)
15:           if x(n) exceed tube_bounds do #change direction
16:                 direction(n) = difference(direction(n-1), direction_space);
```

In total, 8 simulations for each tube length were performed, replicating the experimental design. The time resolution of each simulation was binned into 5 minutes intervals by summing up the occurrence of the fly at every position along the tube, and the average residency of modelled fly was calculated as described above.

5.2.5 GOODNESS OF MODEL FIT

The accuracy of this model was tested by calculating the mean-squared errors (MSE), that is the difference between the average positional preferences as estimated by the model and average positional preferences given by the experimental data:

$$MSE = \frac{\sum(\hat{Y}_i - Y_i)^2}{N}, \quad (5.9)$$

where Y_i – average positional preference at position i, $\hat{Y}_i$ – predicted positional preference at position i, N – length of the tube.

5.3 RESULTS

To investigate the control system behind the positional preferences in fruit fly, a previously developed technique FLYGLOW, as described in chapter 4, was used to get the estimate of flies' spatial location in capillary tubes of 4 different lengths. Two sets of data produced by FLYGLOW were used in the subsequent analysis: location of the fly along the tube was determined using the spatial distribution of the bioluminescence, and sleep/wake cycle was established with the help of sleep binary array. Furthermore, two models of fly behaviour were to see which better explained the role of olfactory and/or position cues in determining positional preferences in the absence of visual stimuli: model 1 hypothesised that fly uses olfactory stimuli to orient itself in complete darkness, whereas model 2 is based on the assumption that

the fly uses self-motion cues to estimate its position in the tube. Moreover, a stochastic model of fly behaviour in constant darkness was developed in order to simulate the persistence of fly movement towards the food.

5.3.1 SPATIAL LOCATION OF THE FLIES

The bioluminescence intensity emitted by the flies gave an opportunity to identify the position of the fly in the tube, so this information was used to investigate the control mechanism behind the positional preferences in healthy population of flies. Considering the timeframes when the fly was in the awake state, the average position of the population of control flies in the tubes of different lengths was visualized using the heatmap approach (Figure 5.4).

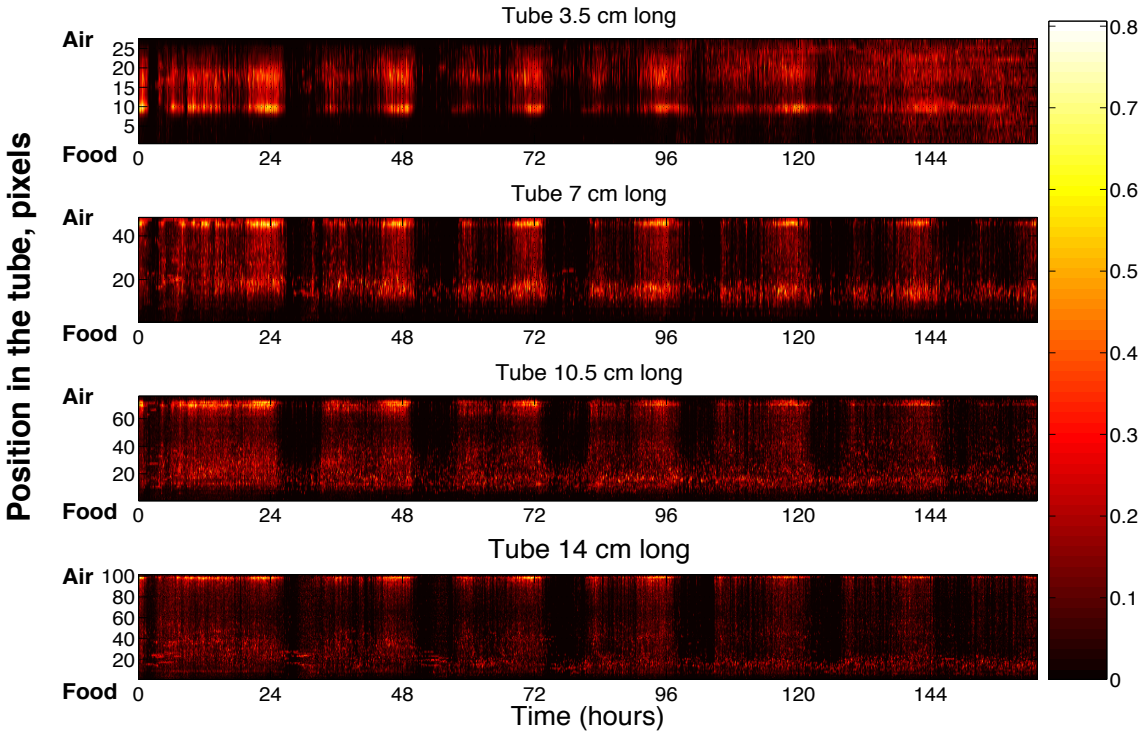

Figure 5.4 Heatmap of mean residency of control flies in the tube during the wake periods. Lighter color represents the higher probability of the residency within the tube. X-axis: time in hours, y-axis: position along the tube as a bin number.

As can be expected from the previous findings of animal positional preferences, the awake fly positions in the tubes distributed unevenly with seemingly two peaks of location preferences around both ends of the tube. Notably, the flies in tube of 7, 10.5 and 14 cm long all show similar pattern of positioning themselves

within the tubes with the clear preference towards the air end of the tube. Moreover, the flies have demonstrated the periodic preference to stay near the food end of the tube, mostly just before falling asleep or in-between the sleep episodes. Consistent with multiple lines of evidence (Donelson, Kim et al. 2012, Faville, Kottler et al. 2015), the flies showed the distinct preference to remain near the food end of the tubes when they enter into the sleep-like state (Figure 5.5).

On the other hand, flies in 3.5 cm long tube have a slightly different positional distribution with rather chaotic movements when they are awake towards the end of 7 days experiment. Indeed, it has been found that all of the flies in 3.5 cm long tubes died on day 5.

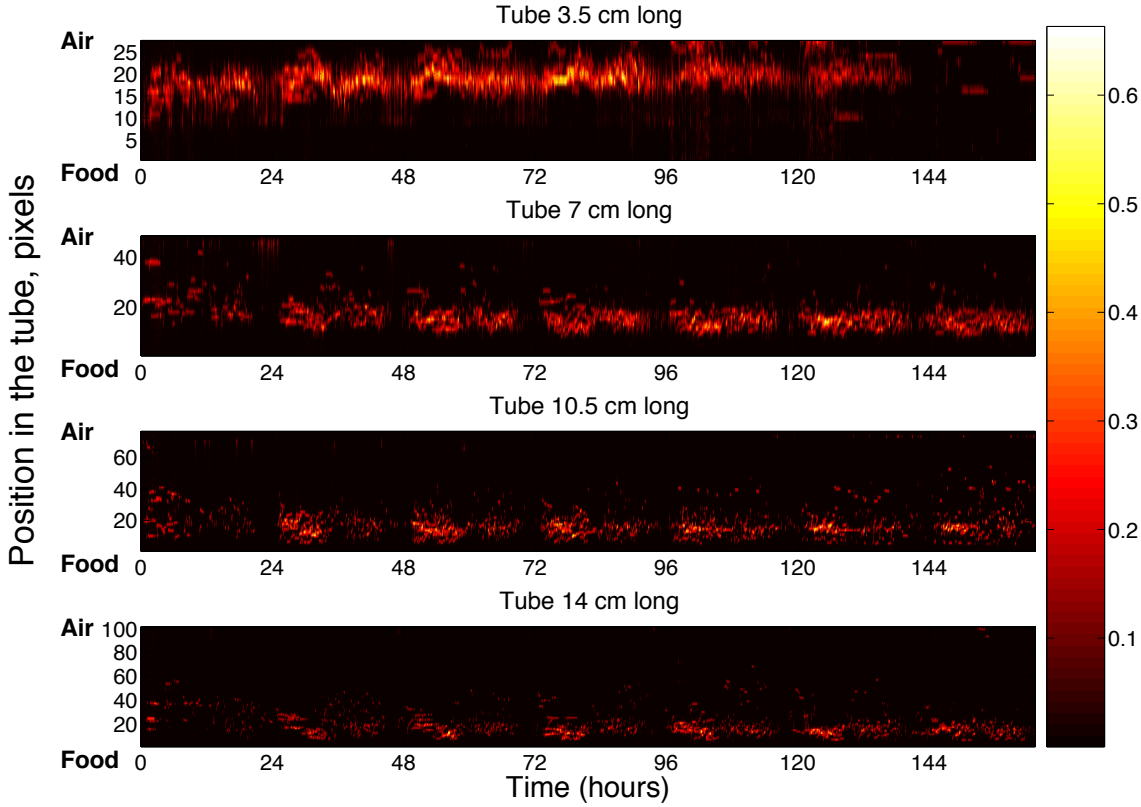

Figure 5.5 Heatmap of mean residency of control flies in the tube during the sleep periods. Lighter color represents the higher probability of the residency within the tube. X-axis: time in hours, y-axis: position along the tube as a bin number.

Interestingly, it may be observed from the relative brightness of the top panel in Figure 5.5 that on average flies in shorter tubes have greater numbers of sleep-like episodes than those in longer tubes, particularly in case of 3.5 cm tubes. In order to determine whether the mean amount of sleep is significantly different in tubes of

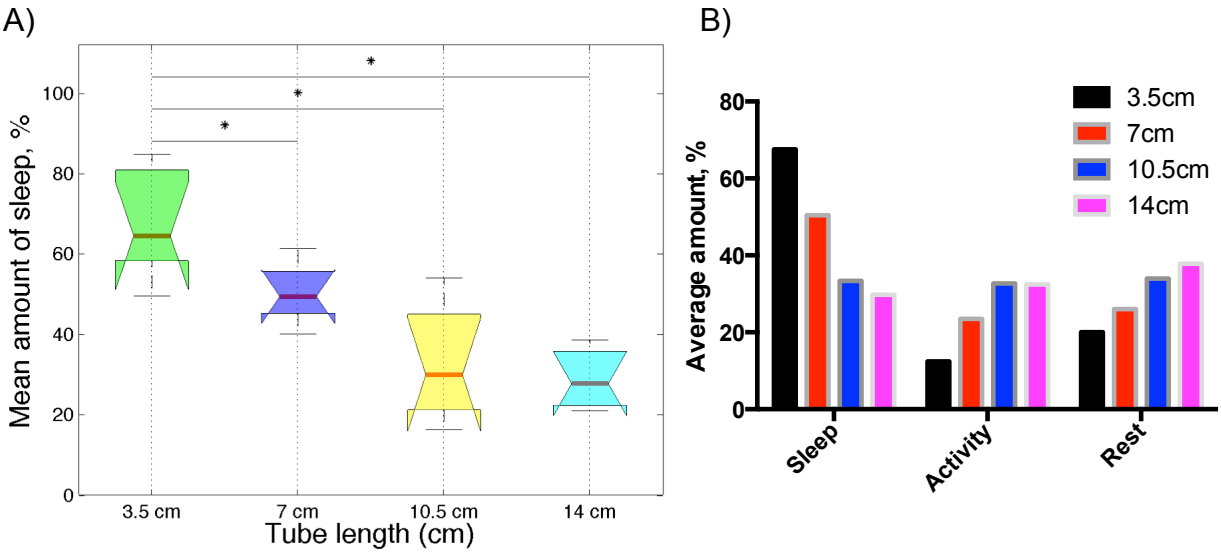

Figure 5.6 Comparing the sleep amount for control flies in tubes of different lengths. A) Average amount of sleep was calculated for the population of flies and presented as a percentage (with 100% means that a fly was sleeping all the time during the 3 days of the experiment). B) Distribution of the proportion of each behavioural parameter in tubes of different length.

different lengths, a one-way analysis of variance (ANOVA) test was performed. For each fly, the mean duration of sleep was determined as the sum of the number of sleep episodes starting from day 2 (the data from first day was omitted in the assumption that flies need to accommodate to the new environment) and until day 4 so that the arrhythmic behaviour of the flies just before their death is excluded. The null hypothesis was that the flies sleep the same amount in tubes of different lengths. Using this analysis, it is apparent that flies in shortest tubes (3.5 cm) sleep for 67% of the time which is significantly longer than flies in other tubes (Figure 5.6 A). Furthermore, the flies in longer than conventionally-used tubes, that is 10.5 and 14 cm tubes, showed significantly lower amount of sleep on average, specifically around 30% of the time during the three day window. Accordingly, the average amount of both activity and rest for flies in 3.5 cm long tubes are significantly different from longer tubes, and the same is true for flies in 7 cm long tubes (Figure 5.6 B). On the other hand, the flies in both 10.5 and 14 cm long tubes did not show any significant difference in either of sleep, activity or rest amounts between two populations.

The visual inspection of the positioning of $A\beta_{42}$-expressing flies within the tubes of different lengths in timeframes when they were identified as being in a sleep state have shown a slightly different pattern of positional preferences as compared to control flies. In particular, the location of $A\beta_{42}$-expressing flies are more scattered

along the tubes, which is most noticeable in 10.5 cm long tubes where the flies fall asleep near either air and food ends of the tubes in the equal proportions (Figure 5.7, third heatmap). Similarly to the control population of flies, it has been found that $A\beta_{42}$-expressing flies in shortest tubes (3.5 cm long) sleep significantly longer than in any other tubes, while $A\beta_{42}$-expressing flies in longer tubes (10.5 and 14 cm) did not showed any difference in the average amount of sleep. Consistent with our previous finding in 4.3.3, the average proportion of sleep, activity and rest of $A\beta_{42}$-expressing flies in tubes of different lengths have the same distribution as the one for the control flies.

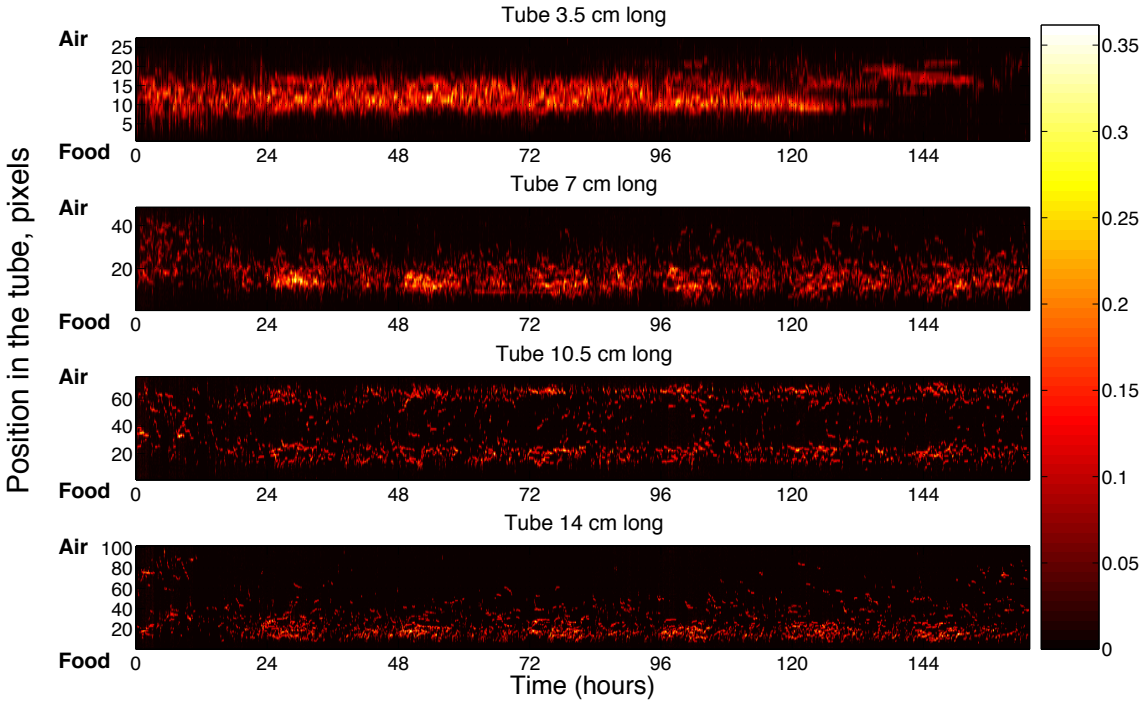

Figure 5.7 Positional preferences of $A\beta_{42}$-expressing flies during sleep. The heatmap shows the mean residency of the flies in the tube during the sleep periods. Lighter color represents the higher probability of the residency within the tube. X-axis: time in hours, y-axis: position along the tube as a bin number.

5.3.2 MODEL OF FLY BEHAVIOUR WHILE AWAKE IN CONSTANT DARKNESS

In order to experimentally test the two models of fly location preference, the average residency along the tubes was established for awake flies in complete

darkness, as shown in Figure 5.8. The averaged positional preferences of the control group of flies during the three days period (black traces in Figure 5.8) in 7, 10.5 and 14 cm long tubes have exhibits the similar distribution with the highest peak of residency near the air end (right hand side of the plots in Figure 5.8) and the second highest point of preferred position for flies was at some distance away from the food. However, the control flies in 3.5 cm long tubes have rather inverted pattern: the peaks of preferred residency is equally distributed between either being some distance away from the food or from the fresh air.

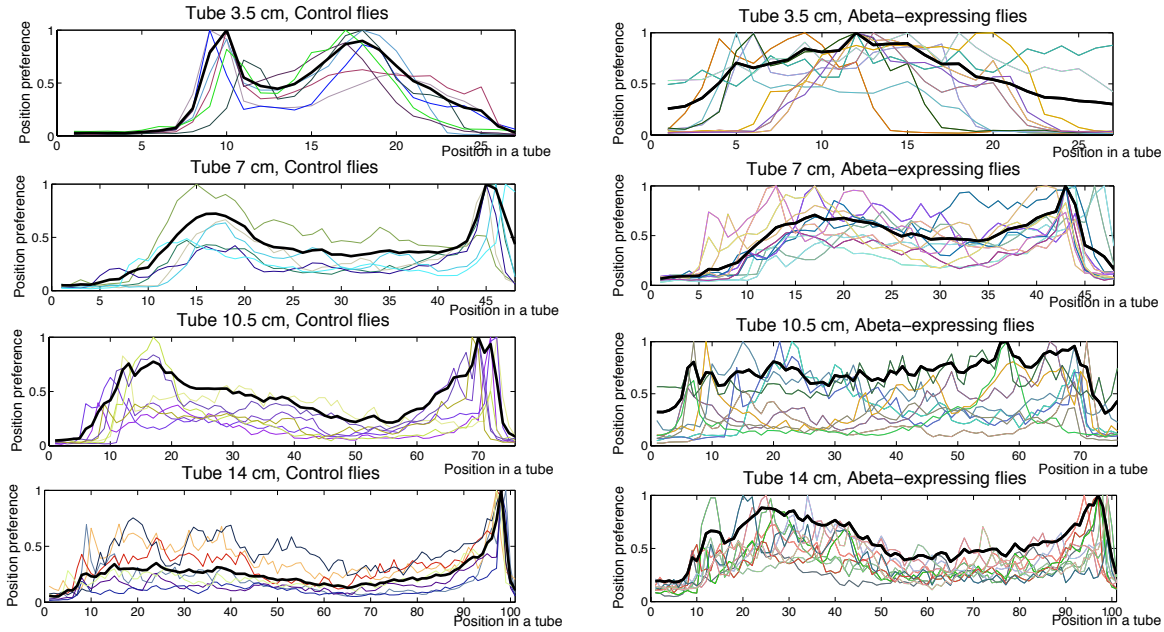

Figure 5.8 The average residency of flies along the tubes of different lengths. Black traces represent the mean value for the population of flies.

The averaged positional preferences of $A\beta_{42}$-expressing flies are characterized by higher degree of variability between individuals (multicolour traces in Figure 5.8), which is particularly pronounced for flies in 10.5 long tubes. In contrast to control flies, $A\beta_{42}$-expressing flies in 3.5 cm long tubes demonstrated the bias to stay in the middle section of the tubes.

The average positional preferences for the control populations of flies were used to test the two models of their response to olfactory and/or spatial cues. When the position data is plotted against the relative distance along the tubes, that is consistent with control by an olfactory gradient, the peak in residency occurs at between 14% and 35% of the way along the tube (Figure 5.9, top panel, inverted triangles). By contrast when the data is plotted as absolute distance from the food, that

is consistent with control by internal motion cues, the peak is observed in a narrower range, 10-18 mm, from the food. This initial analysis indicated that model 2 is more likely to explain the position preferences of the flies, relying on self-motion cues rather than olfactory cues (Figure 5.9). Indeed, apart from flies in 3.5 cm long tubes, control flies have demonstrated the distinct desire to stay on average 16 bins away from the food source as if they have spatial awareness in constant darkness (triangles, Figure 5.9). By contrast, according to model 1 it would be expected than if flies use smell of food to orient themselves within the tube then they will stay at same fraction of the tube length away from the food, which was not the case as shown in Figure 5.9.

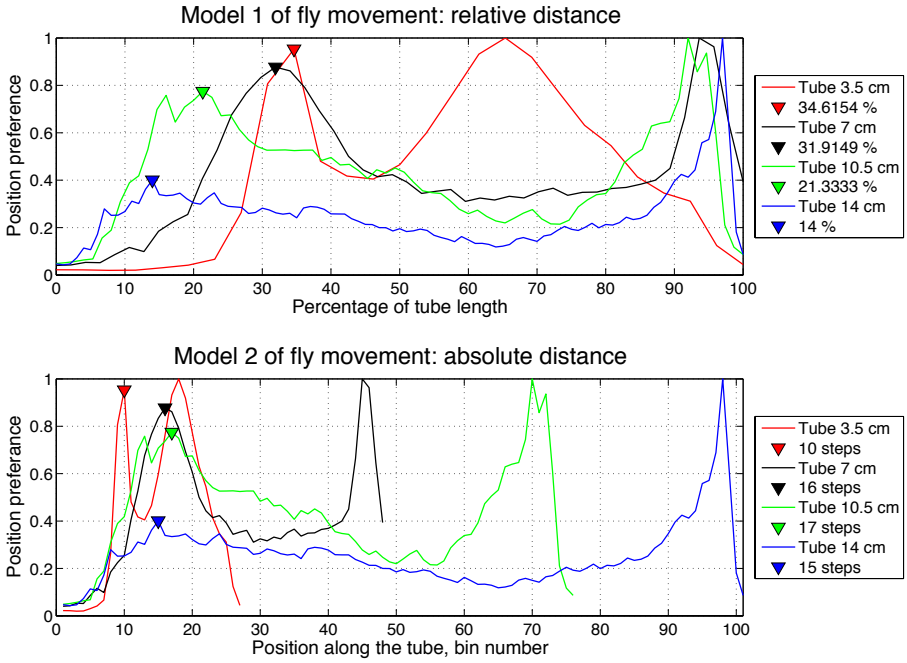

Figure 5.9 Two models of fly response to the cues available in constant darkness. Model 1 (top) assumed that the fly response to the olfactory gradient while model 2 (bottom) hypothesized that the fly use self-motion cues in darkness. Y-axis represents the positional preferences of the flies within the tube, while x-axis represents either a percentage of the tube length (away from food at 0) or bins along the tube.

5.3.3 SIMULATION OF FLY PERSISTENT BEHAVIOUR IN CONSTANT DARKNESS

To understand the underlying control system that governs the flies locomotion behaviour and persistence of movement toward the target, the stochastic model was

developed and tested by using the experimental data from conventionally used 7 cm long tubes. The persistence of fly locomotion was simulated by assigning the directional positional probability to each position along the tube and comparing it with the random number drawn at every iteration of the simulations. The resulting simulations of fly walking behaviour within the tube were averaged into 5 min bins (Figure 5.10, top heatmap) to replicate the experimental conditions and converted into the average positional residency. The best fitting model was found with the following parameters of log-normal distribution: μ=2.8, σ=0.18 as shown in Figure 5.10 (bottom plot, black traces – experimental data, red traces – DPP, green traces – average positional preferences of 8 simulations).

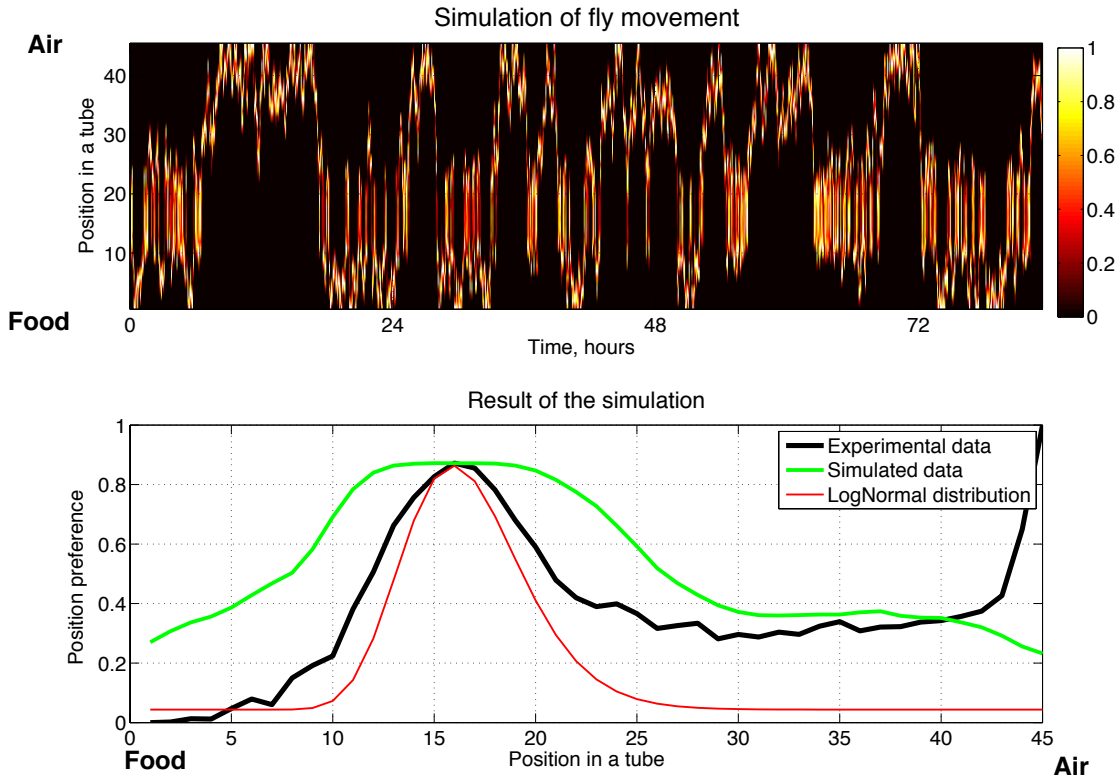

Figure 5.10 Best fitting model of fly persistent behaviour in darkness. Top plot showed the example of the simulated data presented as a heatmap. Bottom plot showed the average residency along the tubes for the control population (black) and superimposed average residency of simulated data

As can be seen, lognormal DPP distribution roughly replicates the distribution of the experimental positional preferences for flies in 7 cm long tube and the same is true for the averaged positional residency data from 8 individual runs of the simulation. Due to the random nature of the simulations, the resulting average

residency along the tube is not a simple conversion of the direction persistence probability but rather a reduction of simulated data dimensionality by averaging along the time frames.

The flies' behaviour was modelled in a similar way for tubes of different lengths using the same parameters as have been previously optimised for the simulation model in 7 cm long tube. Moreover, the two models of spatial orientation in constant darkness in the tubes of different lengths were evaluated by fitting the averaged simulated data to the experimental data captured for the control population of flies as shown in Figure 5.11 (black traces – experimental data, red traces – simulated data). And again model 2 demonstrated the better fit to the experimental

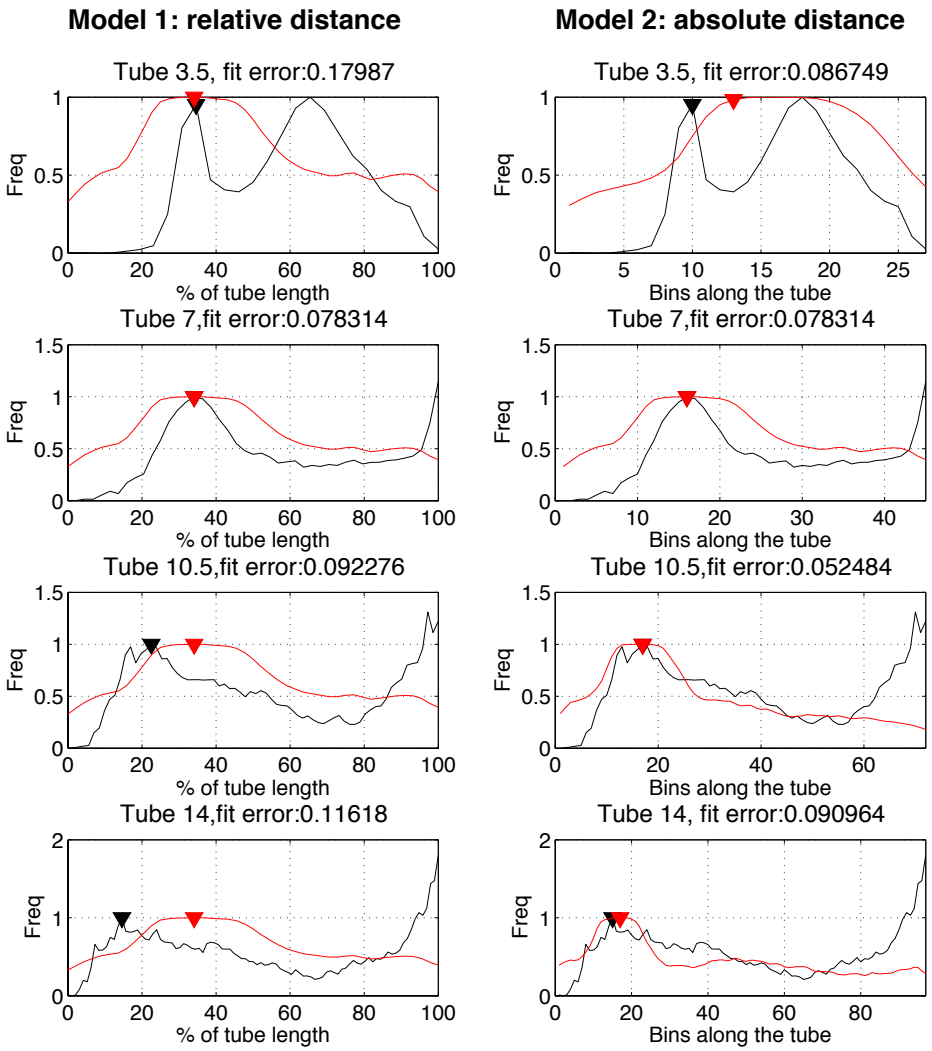

Figure 5.11 Testing two models of fly orientation using simulated data. Y-axis: averaged residency of control flies within the tubes for experimental data (black) and simulated data (red). X-axis: either a percentage of the tube length (model 1) or bins along the tube (model 2). Triangles represent the location of the residency peak.

data for all the flies in tubes of different lengths. In contrast, according to the model 1 lengthening the tube have resulted the progressive divergence from the actual positional preferences as demonstrated by the experimental data.

Accordingly, for model 2 the mean-squared errors (MSE) of fitting the simulated data to experimental data (control group of flies) is significantly lower than the MSE for model 1 (Figure 5.12, A). The linear regression line fitted to the distribution of MSE for model 2 showed the steady horizontal trend (black dashed line, Figure 5.12 A), meaning that on average the errors of model fit are on the same level when fitted to the experimental data from the tubes of different lengths. ANOVA test of distribution of MSE errors for tubes of different lengths have indicated that the mean and variance of MSEs for model 2 is significantly different from the MSEs for model 1.

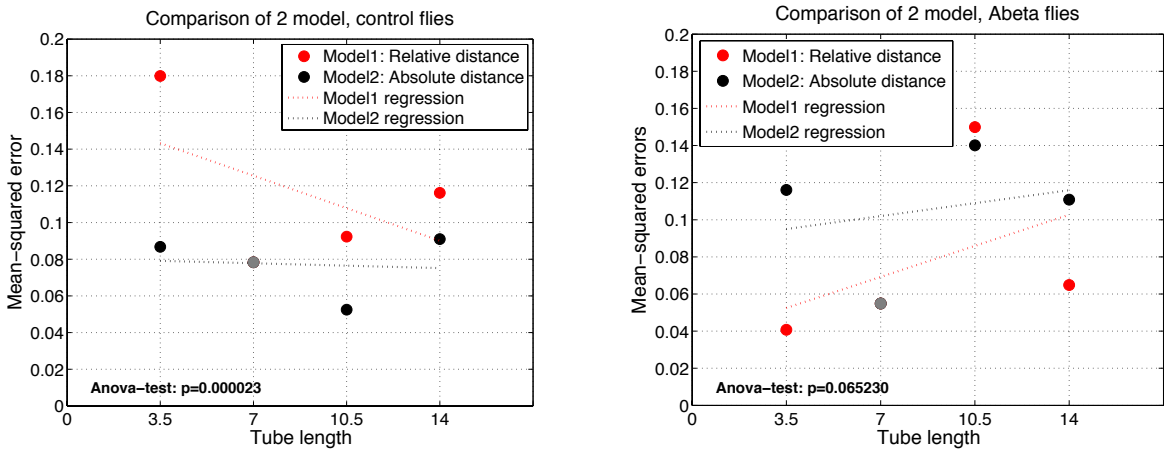

Figure 5.12 Mean-squared errors of fit distribution for two models. MSEs of fitting the simulated data to experimental were calculated and plotted on the y-axis for control (A) and $A\beta_{42}$-expressing flies (B). Dashed lines represent the linear regression lines for model 1(red) or model 2(black). The significance of the difference between two models is indicated by the p-value derived from ANOVA test.

As can be expected, fitting the simulated data to the average residency of $A\beta_{42}$-expressing flies in tubes of different lengths have demonstrated a worse match to model 2 of fly spatial orientation as compared to the control population of flies (Figure 5.12 B), with mean values of MSE for model 2 is 0.11 versus 0.08 for control flies. However, in $A\beta_{42}$-expressing flies a match to the model 1 was better than to model 2 and on a similar level as the match to model 2 in control flies, with the average values of MSE for model 1 is 0.08. In addition, according to the ANOVA test the distribution of MSEs for two models did not differ from each other.

5.4 DISCUSSION

As it has been shown previously, in the early stage of Alzheimer's disease neurodegeneration starts at the hippocampus and leads to the inability to set down new memories. In addition, the recent discovery of the "inner GPS", or positioning system, to be situated in the hippocampus in mammals has shed a light on why a spatial orientation is affected on the early stage in patients with AD (O'Keefe and Dostrovsky 1971, Fyhn, Molden et al. 2004). However, the mechanisms underpinning the spatial memory loss are not yet understood. Using the model organisms of the disease can help us investigate this complex problem and gain valuable insights.

Most motile organisms have the similar ability to navigate and orient themselves in the environment. For example, bacteria *E. coli* use the chemotactic approach to move towards the target and other organisms may use similar strategy. In particular, it has been demonstrated that *Drosophila* may use stereo chemosensation to locate the olfactory signal and use a "run-and-tumble" mechanism for locomotion. However, the sensorimotor pathways by which the fly integrates the input from different modalities, especially in the constant darkness, is unknown.

In this project we proposed to use the FLYGLOW system to investigate the flies' behaviour in the absence of visual cues to guide their positional preferences within their tubes. Two models of fly response to the available stimuli were developed with the underlying assumptions that fly can estimate its position by either olfactory gradient, which results in estimating the relative distance from the food, or self-motion cues, as if the fly counts the number of steps it has taken away from the food source. These two models were tested for the flies in tubes of different lengths for both control and $A\beta_{42}$-expressing flies. Moreover, two stochastic models of fly movement in constant darkness was developed and evaluated against the experimental data.

Consistently with the previous findings (Donelson, Kim et al. 2012, Faville, Kottler et al. 2015), it was found that control flies indeed have positional preferences within the tubes with the clear bias to stay near the air end of the tubes in the awake state and near the food source when they falling asleep. In contrast to the previous statement (Kim 2012) that the flies prefer to stay quarter of the tube closest to the

food suggesting the ability of the flies to judge the relative distance within the tubes, it was discovered, that irrespective of the tube length, the flies stay approximately 16 bins away from the food source (equivalent to 2 cm in real distance). In contrast, $a\beta_{42}$-expressing flies have demonstrated more dispersed and chaotic location preferences, especially during the sleep periods when they sometimes to prefer to fall asleep near the air end of the tubes.

The interesting observation was that flies in shortest tubes have a quite distinct pattern of positional preferences in both sleep or awake states. One explanation could be that being in the close proximity to the food source might determine their sleep-wake behaviour. In contrast to the commonly held belief that circadian rhythms determine the fly's sleep onset, and prior to falling asleep the fly prefers to be near the food, it may be that an alternative explanation is valid. Indeed, we saw clearly that a fly's preferred location for sleep was about 2 cm away from the food, which can be defined as the sleep region. It appears that when the flies were confined to the sleep region (shortest tubes 3.5 cm long have in fact have effective length around 2.5 cm) the time spent sleeping was more than double than in the longest tubes. This implies that the widely-accepted causation may be reversed, so that circadian rhythms govern locomotor behaviour and hence the particular patterns of residency within the tube; I propose that it is the behavioural localisation of flies within that sleep zone that induces sleep behaviour. Furthermore, it is well known, although not widely reported, in the field that the optimal tube length for observing clear circadian behaviour in the fly is 7 cm. Other tube lengths, particularly shorter, have resulted in deterioration in circadian sleep-wake behaviour. I propose that forced proximity to the food source in shorter tubes drives flies to sleep even in the absence of a circadian clock.

With regards to the models of behavioural control in constant darkness, it was revealed that the flies use self-motion cues to estimate their location within the tube, meaning that the fly has an absolute spatial awareness in constant darkness. Indeed, using the stochastic model of fly walking behaviour, where the persistence of movement varied according to where the fly is in the tube, I have demonstrated that in tubes of different lengths flies navigate by judging the absolute distance away from the food source rather finding an optimal position along an olfactory gradient. Notably, $a\beta_{42}$-expressing flies are characterized by the higher inter-individual

differences underpinning the averaged positional data. Indeed, the $A\beta_{42}$ fly behaviour approximates to random behaviour. The mean-squared errors for the simulated data for $A\beta_{42}$-expressing flies are similar for both control models. This suggests that the control function that allows control flies to maintain their position according to the absolute distance from the food is degraded in $A\beta_{42}$-expressing flies.

The more precise confirmation of these findings could be acquired in the future by repeating the similar experiment with the mutant flies, that lack odorant receptors, to formally exclude the possibility of influence of the olfactory stimuli on the fly's behaviour.

5.5 CONCLUSION

In this chapter, the spatial orientation of the flies was investigated by examining the fly spatial distribution within the custom made capillary tubes of different lengths using the FLYGLOW system. It was concluded that the flies use self-motion cues in order to estimate their position within the tube and it seems that in $A\beta_{42}$-expressing flies this underlying controlling mechanism is disrupted. Moreover, it was proposed that confinement of the flies in a defined sleep region result in the increased amount of sleep and disruption of normal sleep-wake cycles.

CONCLUSION

In this work the analysis of Alzheimer's disease was focused on the pathological role of the aggregation and accumulation of Aβ peptide. In the thesis I have aimed to investigate the different aspects of Aβ toxicity and so dissect the pathways of disease pathogenesis. The main outcomes of this work are as follows:

Using a worm model of Aβ toxicity to derive more information from human GWAS in Alzheimer's disease.

The role of Aβ toxicity in sporadic AD was evaluated by using the results of previously performed knockdown screen in *C. elegans* model system that finds the genes that modify the Aβ-induced paralysis phenotype. Comparing the human orthologues of genes identified by this worm modifier screen with the genes implicated by GWAS in sporadic AD I have demonstrated no identity between the two lists of genes, suggesting that mechanisms of Aβ toxicity, at least in worm model, are not equivalent to sporadic AD. However, further analysis of the direct physical interactors of those genes has revealed 7 worm genes that have a non-random overlap with GWAS. Four of those worm genes had larger than expected by chance overlap with interactors of GWAS genes, with three of them forming TRiC/CCT chaperone complex and fourth being a major substrate of the same chaperone complex. At the present time, the exact role of TRiC/CCT in the pathogenesis of sporadic AD is unclear, but the fact that RNAi knockdown of those genes in worm model rescues Aβ-induced paralysis phenotype might implicate that this chaperone complex in some way promotes Aβ toxicity. In conclusion, it has been shown that using the animal model of Aβ toxicity can provide the means to extract further information from human GWAS in AD and find the potential candidate genes.

Study of circadian disruptions in Alzheimer's disease using *Drosophila* as a model organism.

The involvement of Aβ toxicity in the circadian disruptions, associated with AD, has been evaluated by developing the new recording system to simultaneously observe the oscillations of peripheral clocks and behavioural rhythms in *Drosophila*. Careful examination of the results has revealed that in Aβ-expressing flies the normal

consolidated structure of the sleep is disrupted, leading to the fragmented sleep architecture. Taking advantage of simultaneous recording and comparing the rhythms robustness has demonstrated that oscillations of behavioural rhythms and peripheral clocks are independent of each other, suggesting that Aβ toxicity damages the communication pathways between molecular clock and circadian behaviour. In conclusion, the developed system has provided the opportunity to dissect the relationship between molecular and behavioural rhythms in *Drosophila* model of Aβ toxicity.

Study of positional preference in *Drosophila* and how this is degraded in a model of Alzheimer's disease.

The control mechanism of spatial orientation has been studied by examining the positional preferences of fruit flies within the tubes of different length. It has been established that in absence of visual stimuli, wild-type flies use self-moving cues to orientate themselves in the tubes and in the presence of Aβ toxicity the underlying control function of spatial navigation is disrupted. Moreover, it has been observed that the flies confined to the sleep region in the shortest tubes slept significantly more than flies in longer tubes, suggesting that the forced proximity to the food source induces the sleep behaviour.

Future Perspectives

The further evaluation of the genes forming the TRiC/CCT chaperone complex can shed the light on its involvement in the pathogenesis of sporadic AD and provide additional insights into the mechanisms of Aβ toxicity. The similar procedure can be applied to uncover additional information from the findings of GWAS in many complex neurodegenerative diseases.

The method developed in chapter 4 (FLYGLOW) can be applied to more broadly based behavioural, genetic and drug-based screens. For example, evaluating the circadian disruptions in mutant flies with the expression of sleep-regulation transgenes can provide the insight into the role of those particular genes in the sleep homeostasis. Another possibility would be to initiate the genetic screen to identify genes involved in the circadian disruption mechanism.

In order to validate the control mechanism of spatial orientation in flies, the experiment described in chapter 5 can be performed for the mutant flies lacking the odorant receptor in order to exclude the presence of olfactory stimuli.

APPENDIX. SUPPLEMENTARY TABLES

A.1 Suppressors of the Aβ-paralysis phenotype (Khabirova, Moloney et al. 2014)

The list consists of the human orthologues of worm genes that, when targeted by RNAi, suppress the paralysis phenotype in Aβ-expressing worms (n=78 genes).

C. elegans gene symbol	Human gene name	Human gene symbol	Accession number
F46E10.1	Long chain fatty acid acyl-CoA ligase	ACSF2	NM_001028767
T04C12.6	Actin and related proteins	ACTB	NM_073416
T25C8.2	Actin and related proteins	ACTG1	NM_067408
K07C5.1	Actin-related protein Arp2/3 complex, subunit Arp2	ACTR2	NM_073256
F55A12.7	AP-1 complex subunit mu-1	AP1M1	NM_059171
F29G9.3	Clathrin adaptor complex, small subunit	AP1S2	NM_072158
C13B9.3	Medium subunit of clathrin adaptor complex	ARCN1	NM_066062
C34E10.6	F0F1-type ATP synthase, beta subunit	ATP5B	NM_065710
R10E11.8	Vacuolar H+-ATPase V0 sector, subunits c/c'	ATP6V0C	NM_066764
Y49A3A.2	Vacuolar H+-ATPase V1 sector, subunit A	ATP6V1A	NM_074158
F20B6.2	Vacuolar H+-ATPase V1 sector, subunit B	ATP6V1B2	NM_076310
Y55F3AR.3	Chaperonin complex component, TCP-1 theta subunit (CCT8)	CCT8	NM_067634
F09G2.4	mRNA cleavage and polyadenylation factor II complex, subunit CFT2 (CPSF subunit)	CPSF2	NM_072421
M03F8.3	Cell cycle control protein (crooked neck)	CRNKL1	NM_001129507
B0464.1	Aspartyl-tRNA synthetase	DARS	NM_066688
C55B6.2	dsRNA-activated protein kinase inhibitor P58, contains TPR and DnaJ domains	DNAJC3	NM_076808
D2085.3	Translation initiation factor 2B, epsilon subunit (eIF-2Bepsilon/GCD6)	EIF2B5	NM_063440
F22B5.2	Translation initiation factor 3,	EIF3G	NM_001276737

	subunit g (eIF-3g)		
C40H1.4	Elongation of very long chain fatty acids protein 3	ELOVL3	NM_066655
H19N07.1	Eukaryotic Peptide Chain Release Factor	GSPT2	NM_001269363
T10C6.11	Histone H2B	HIST1H2BA	NM_074630
F26D10.3	Molecular chaperones HSP70/HSC70, HSP70 superfamily	HSPA8	NM_070667
C49F5.1	S-adenosylmethionine synthetase	MAT1A	NM_077601
C30A5.3	MOB-Like Protein Phocein	MOB4	NM_066397
Y57G11C.12	NADH:ubiquinone oxidoreductase, NDUFA6/B14 subunit	NDUFA6	NM_070389
K07C5.4	Ribosome biogenesis protein - Nop56p/Sik1p	NOP56	NM_073259
T22B11.5	2-oxoglutarate dehydrogenase, E1 subunit	OGDHL	NM_068216
D1054.15	Isoform 1 of Pleiotropic regulator 1	PLRG1	NM_001269330
F36A4.7	RNA polymerase II, large subunit	POLR2A	NM_068122
K02B12.3	Prolactin regulatory element binding	PREB	NM_059904
R07E4.6	CAMP-dependent protein kinase type I-Alpha, regulatory subunit	PRKAR1A	NM_076598
C50C3.6	U5 snRNP spliceosome subunit	PRPF8	NM_066384
CD4.6	20S proteasome, regulatory subunit alpha type PSMA1/PRE5	PSMA1	NM_072071
Y38A8.2	20S proteasome, regulatory subunit beta type PSMB3/PUP3	PSMB3	NM_062512
C52E4.4	26S proteasome regulatory complex, ATPase RPT1	PSMC2	NM_073604
Y49E10.1	26S proteasome regulatory complex, ATPase RPT6	PSMC5	NM_067208
C23G10.4	26S proteasome regulatory complex, subunit RPN2/PSMD1	PSMD1	NM_065946
C39F7.4	GTPase Rab1/YPT1, small G protein superfamily, and related GTP-binding proteins	RAB1A	NM_070996
F10B5.1	60s ribosomal protein L10	RPL10L	NM_063306
T22F3.4	60S ribosomal protein L11	RPL11	NM_071607
C27A2.2	60S ribosomal protein L22	RPL22	NM_062531
C09H10.2	60S ribosomal protein L44	RPL36AL	NM_063974

B0250.1	60s ribosomal protein L2/L8	RPL8	NM_075539
R13A5.8	60S ribosomal protein L9	RPL9	NM_066259
F54E7.2	40S ribosomal protein S12	RPS12	NM_065820
C16A3.9	40S ribosomal protein S13	RPS13	NM_065992
T01C3.6	40S ribosomal protein S16	RPS16	NM_074289
T05F1.3	40S ribosomal protein S19	RPS19	NM_060154
C23G10.3	40S ribosomal protein S3	RPS3	NM_065948
B0393.1	40S ribosomal protein SA (P40)/Laminin receptor 1	RPSA	NM_065577
F53E10.6	RRP15-like protein	RRP15	NM_071312
VZK822L.1	Fatty acid desaturase	SCD	NM_001268666
F43D9.3	Sec1 family domain-containing protein 1	SCFD1	NM_001129225
Y113G7A.3	Vesicle coat complex COPII, subunit SEC23	SEC23B	NM_075476
Y57E12AL.1	Tumor differentially expressed (TDE) protein	SERINC1	NM_171479
T08A11.2	Splicing factor 3b, subunit 1	SF3B1	NM_065452
Y116A8C.42	Small nuclear ribonucleoprotein Sm D3	SNRPD3	NM_070626
ZK652.1	Small Nuclear Ribonucleoprotein Polypeptide F	SNRPF	NM_066307
T27F2.1	SNW domain-containing protein 1	SNW1	NM_073549
F55A11.2	SNARE protein SED5/Syntaxin 5	STX5	NM_073567
T05C12.7	Chaperonin complex component, TCP-1 alpha subunit (CCT1)	TCP1	NM_063321
Y116A8C.35	U2 small nuclear RNA auxillary factor 1 isoform b	U2AF1	NM_070635
C47E12.5	Ubiquitin-like modifier activating enzyme 6	UBA6	NM_001268520
*C46G7.1	-	-	NM_068504
*F28C6.7a	-	-	NM_063422
*Y105E8C.e	-	-	N/A
*K04E7.2	-	-	NM_076686
*F58G6.7	-	-	NM_069313
*ZK858.1	-	-	NM_060045
*Y41D4B.19	-	-	NM_067703
*F42D1.3	-	-	NM_078054
*F55C5.4	-	-	NM_073679
*F17C8.5	-	-	NM_065572
*C41C4.7a	-	-	NM_063303
*W09B6.1	-	-	NM_001267098

*D2024.1	-	-	NM_068750
*W06H8.8	-	-	NM_001029030
*F25B5.4	-	-	NM_171139

A.2 Gene ontology enrichment for worm screen hits (Khabirova, Moloney et al. 2014)

The worm screen hits were analysed for enrichment in molecular function, cellular component and biological process ontology labels.

Molecular function
Proton transporting ATP synthase activity, rotational mechanism
Proton transporting ATPase activity, rotational mechanism
Cellular component
Cytosolic small ribosomal subunit
Eukaryotic translation initiation factor 3 complex
Chaperonin-containing T-complex
Biological process
Protein binding
Structural constituent of ribosome
Purine nucleotide binding

A.3 The 63 human genes in the AD GWAS white+grey zones (Khabirova, Moloney et al. 2014)

1	ABCA1
2	ABCA7
3	ADCY8
4	AEN
5	ANK1
6	APOC1
7	APOE
8	BIN1
9	CCDC134
10	CCDC76
11	CD2AP
12	CD33
13	CLU
14	CLVS1
15	CNTNAP2

16	CR1
17	CUBN
18	CYP7B1
19	DBT
20	DCHS2
21	DIAPH3
22	DIP2C
23	DISC1
24	DNAH11
25	EML1
26	ENSAP3
27	F13A1
28	FARP1
29	GAB2
30	GFRA2
31	GRIN2B
32	LRRC39
33	MGST3
34	MPP7
35	MS4A6A
36	MTHFD1L
37	NCAM1
38	NCR2
39	NDP
40	NEDD9
41	NHSL1
42	ODZ4
43	PAK2
44	PCDH11X
45	PITPNC1
46	POLN
47	PVRL2
48	RBFOX1
49	RORA
50	RYR2
51	SLC28A1
52	SLC4A1AP
53	SLC44A5
54	ST3GAL1
55	STK11
56	STK24
57	TOMM40

58	TXNDC6
59	UTS2D
60	VSNL1
61	WBSCR17
62	ZNF292
63	ZNF320

A.4 Overlap between the +1 interactome for the products of genes in the GWAS white+grey zone and those of genes in the +1 interactome of the worm RNAi screen (Khabirova, Moloney et al. 2014)

The human AD GWAS white+grey zone genes, and their +1 interactors, overlap with the human orthologues of worm-screen hits, and their +1 interactors. Worm screen genes that have human orthologues with a significantly non-random +1 interactome overlap with the GWAS list are shown in bold. Worm screen genes that have human orthologues that interact directly with GWAS genes are marked with an asterisk.

Gene	p-value	+1 interactome overlap [experimental vs MC]
MOB4	**0.003**	**larger ***
PSMA1	**0.014**	**smaller**
TCP1	**0.014**	**larger ***
NOP56	**0.014**	**smaller**
CCT8	**0.015**	**larger ***
ACTB	**0.018**	**larger**
RPL10L	**0.033**	**smaller**
RPL9	0.058	smaller
ATP6V1B2	0.064	smaller
HSPA8	0.065	larger
RPL22	0.087	smaller
PSMB3	0.089	smaller
RPL11	0.097	smaller
RPS13	0.12	smaller
PSMC2	0.131	smaller
PSMD1	0.137	smaller
PRPF8	0.141	smaller *
ACTR2	0.153	larger
ACTG1	0.159	larger *
PLRG1	0.168	smaller
SCFD1	0.189	smaller

AP1M1	0.196	smaller
RRP15	0.199	larger
U2AF1	0.199	smaller
RPS12	0.233	smaller
SNW1	0.253	larger
RPSA	0.272	smaller
ARCN1	0.289	smaller
SCD	0.289	larger
ATP5B	0.299	larger
RPS19	0.31	smaller
RPS16	0.313	smaller
DARS	0.314	smaller
RPL8	0.322	smaller
STX5	0.33	smaller
POLR2A	0.331	smaller
GSPT2	0.333	larger
EIF3G	0.334	larger
DNAJC3	0.343	smaller
PSMC5	0.346	smaller *
CPSF2	0.362	smaller
SEC23B	0.371	larger
RAB1A	0.382	smaller
SF3B1	0.421	smaller
PRKAR1A	0.435	smaller
RPS3	0.437	smaller
OGDHL	0.449	larger
SNRPD3	0.458	smaller
ACSF2	0.467	larger
NDUFA6	0.486	larger
UBA6	0.515	larger
CRNKL1	0.535	larger
MAT1A	0.539	smaller
RPL36AL	0.55	larger
ATP6V1A	0.557	larger
SERINC1	0.577	larger
EIF2B5	0.58	smaller
HIST1H2BA	0.58	smaller
AP1S2	0.655	smaller
ATP6V0C	0.759	larger

REFERENCES

Alzheimer's, A. (2014). "2014 Alzheimer's disease facts and figures." Alzheimers Dement **10**(2): e47-92.

Ballatore, C., V. M. Lee and J. Q. Trojanowski (2007). "Tau-mediated neurodegeneration in Alzheimer's disease and related disorders." Nat Rev Neurosci **8**(9): 663-672.

Balsalobre, A., S. A. Brown, L. Marcacci, et al. (2000). "Resetting of circadian time in peripheral tissues by glucocorticoid signaling." Science **289**(5488): 2344-2347.

Berg, H. C. and E. M. Purcell (1977). "Physics of chemoreception." Biophys J **20**(2): 193-219.

Bertram, L. and R. E. Tanzi (2005). "The genetic epidemiology of neurodegenerative disease." J Clin Invest **115**(6): 1449-1457.

Bertram, L. and R. E. Tanzi (2009). "Genome-wide association studies in Alzheimer's disease." Hum Mol Genet **18**(R2): R137-145.

Bettens, K., K. Sleegers and C. Van Broeckhoven (2010). "Current status on Alzheimer disease molecular genetics: from past, to present, to future." Hum Mol Genet **19**(R1): R4-R11.

Bindea, G., B. Mlecnik, H. Hackl, et al. (2009). "ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks." Bioinformatics **25**(8): 1091-1093.

Brandes, C., J. D. Plautz, R. Stanewsky, et al. (1996). "Novel features of *Drosophila* period Transcription revealed by real-time luciferase reporting." Neuron **16**(4): 687-692.

Brandt, R., M. Hundelt and N. Shahani (2005). "Tau alteration and neuronal degeneration in tauopathies: mechanisms and models." Biochim Biophys Acta **1739**(2-3): 331-354.

Brenner, S. (1974). "The genetics of *Caenorhabditis elegans*." Genetics **77**(1): 71-94.

Bu, G. (2009). "Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy." Nat Rev Neurosci **10**(5): 333-344.

Burns, A. and S. Iliffe (2009). "Alzheimer's disease." BMJ **338**: b158.

Buschert, V., A. L. Bokde and H. Hampel (2010). "Cognitive intervention in Alzheimer disease." Nat Rev Neurol **6**(9): 508-517.

Bushey, D., G. Tononi and C. Cirelli (2011). "Sleep and synaptic homeostasis: structural evidence in *Drosophila*." Science **332**(6037): 1576-1581.

Campbell, S. S. and I. Tobler (1984). "Animal sleep: a review of sleep duration across phylogeny." Neurosci Biobehav Rev **8**(3): 269-300.

Chen, J., B. J. Aronow and A. G. Jegga (2009). "Disease candidate gene identification and prioritization using protein interaction networks." BMC Bioinformatics **10**: 73.

Chen, K. F., B. Possidente, D. A. Lomas, et al. (2014). "The central molecular clock is robust in the face of behavioural arrhythmia in a *Drosophila* model of Alzheimer's disease." Dis Model Mech **7**(4): 445-458.

Chiti, F. and C. M. Dobson (2006). "Protein misfolding, functional amyloid, and human disease." Annu Rev Biochem **75**: 333-366.

Citron, M. (2010). "Alzheimer's disease: strategies for disease modification." Nat Rev Drug Discov **9**(5): 387-398.

Cleary, J. P., D. M. Walsh, J. J. Hofmeister, et al. (2005). "Natural oligomers of the amyloid-beta protein specifically disrupt cognitive function." Nat Neurosci **8**(1): 79-84.

Cong, Y., M. L. Baker, J. Jakana, et al. (2010). "4.0-A resolution cryo-EM structure of the mammalian chaperonin TRiC/CCT reveals its unique subunit arrangement." Proc Natl Acad Sci U S A **107**(11): 4967-4972.

Consortium, C. e. S. (1998). "Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology." Science **282**(5396): 2012-2018.

Coogan, A. N., B. Schutova, S. Husung, et al. (2013). "The circadian system in Alzheimer's disease: disturbances, mechanisms, and opportunities." Biol Psychiatry **74**(5): 333-339.

Crowther, D. C., K. J. Kinghorn, R. Page, et al. (2004). "Therapeutic targets from a *Drosophila* model of Alzheimer's disease." Curr Opin Pharmacol **4**(5): 513-516.

Crowther, D. C., R. Page, D. Chandraratna, et al. (2006). "A *Drosophila* model of Alzheimer's disease." Methods Enzymol **412**: 234-255.

Daigle, I. and C. Li (1993). "apl-1, a *Caenorhabditis elegans* gene encoding a protein related to the human beta-amyloid protein precursor." Proc Natl Acad Sci U S A **90**(24): 12045-12049.

De Strooper, B., R. Vassar and T. Golde (2010). "The secretases: enzymes with therapeutic potential in Alzheimer disease." Nat Rev Neurol **6**(2): 99-107.

Dechant, G. and Y. A. Barde (2002). "The neurotrophin receptor p75(NTR): novel functions and implications for diseases of the nervous system." Nat Neurosci **5**(11): 1131-1136.

Donelson, N. C., E. Z. Kim, J. B. Slawson, et al. (2012). "High-resolution positional tracking for long-term analysis of *Drosophila* sleep and locomotion using the "tracker" program." PLoS One **7**(5): e37250.

Dostal, V. and C. D. Link (2010). "Assaying beta-amyloid toxicity using a transgenic *C. elegans* model." J Vis Exp(44).

Douglas, P. M. and A. Dillin (2010). "Protein homeostasis and aging in neurodegeneration." J Cell Biol **190**(5): 719-729.

Dowse, H. B. and J. M. Ringo (1992). Do Ultradian Oscillators Underlie the Circadian Clock in *Drosophila*? Ultradian Rhythms in Life Processes. D. Lloyd and E. Rossi, Springer London**:** 105-117.

Drake, J., C. D. Link and D. A. Butterfield (2003). "Oxidative stress precedes fibrillar deposition of Alzheimer's disease amyloid beta-peptide (1-42) in a transgenic *Caenorhabditis elegans* model." Neurobiol Aging **24**(3): 415-420.

Faville, R., B. Kottler, G. J. Goodhill, et al. (2015). "How deeply does your mutant sleep? Probing arousal to better understand sleep defects in *Drosophila*." Sci Rep **5**: 8454.

Fire, A., S. Xu, M. K. Montgomery, et al. (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." Nature **391**(6669): 806-811.

Flicek, P., I. Ahmed, M. R. Amode, et al. (2013). "Ensembl 2013." Nucleic Acids Res **41**(Database issue): D48-55.

Forman, M. S., J. Q. Trojanowski and V. M. Lee (2004). "Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs." Nat Med **10**(10): 1055-1063.

Fraser, A. G., R. S. Kamath, P. Zipperlen, et al. (2000). "Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference." Nature **408**(6810): 325-330.

Frye, M. A. and M. H. Dickinson (2004). "Motor output reflects the linear superposition of visual and olfactory inputs in *Drosophila*." J Exp Biol **207**(Pt 1): 123-131.

Fuentealba, R. A., Q. Liu, J. Zhang, et al. (2010). "Low-density lipoprotein receptor-related protein 1 (LRP1) mediates neuronal Abeta42 uptake and lysosomal trafficking." PLoS One **5**(7): e11884.

Fyhn, M., S. Molden, M. P. Witter, et al. (2004). "Spatial representation in the entorhinal cortex." Science **305**(5688): 1258-1264.

Games, D., D. Adams, R. Alessandrini, et al. (1995). "Alzheimer-type neuropathology in transgenic mice overexpressing V717F beta-amyloid precursor protein." Nature **373**(6514): 523-527.

Gandhi, S. and N. W. Wood (2010). "Genome-wide association studies: the key to unlocking neurodegeneration?" Nat Neurosci **13**(7): 789-794.

Gatz, M., C. A. Reynolds, L. Fratiglioni, et al. (2006). "Role of genes and environments for explaining Alzheimer disease." Arch Gen Psychiatry **63**(2): 168-174.

Giannakopoulos, P., G. Gold, M. Duc, et al. (2000). "Neural substrates of spatial and temporal disorientation in Alzheimer's disease." Acta Neuropathol **100**(2): 189-195.

Giebultowicz, J. M., R. Stanewsky, J. C. Hall, et al. (2000). "Transplanted *Drosophila* excretory tubules maintain circadian clock cycling out of phase with the host." Curr Biol **10**(2): 107-110.

Gilestro, G. F. (2012). "Video tracking and analysis of sleep in *Drosophila* melanogaster." Nat Protoc **7**(5): 995-1007.

Glossop, N. R. and P. E. Hardin (2002). "Central and peripheral circadian oscillator mechanisms in flies and mammals." J Cell Sci **115**(Pt 17): 3369-3377.

Goedert, M., B. Ghetti and M. G. Spillantini (2000). "Tau gene mutations in frontotemporal dementia and parkinsonism linked to chromosome 17 (FTDP-17). Their relevance for understanding the neurogenerative process." Ann N Y Acad Sci **920**: 74-83.

Goh, K. I., M. E. Cusick, D. Valle, et al. (2007). "The human disease network." Proc Natl Acad Sci U S A **104**(21): 8685-8690.

Gomez-Marin, A., B. J. Duistermars, M. A. Frye, et al. (2010). "Mechanisms of odor-tracking: multiple sensors for enhanced perception and behavior." Front Cell Neurosci **4**: 6.

Gotz, J. and L. M. Ittner (2008). "Animal models of Alzheimer's disease and frontotemporal dementia." Nat Rev Neurosci **9**(7): 532-544.

Goudreault, M., L. M. D'Ambrosio, M. J. Kean, et al. (2009). "A PP2A phosphatase high density interaction network identifies a novel striatin-interacting phosphatase and kinase complex linked to the cerebral cavernous malformation 3 (CCM3) protein." Mol Cell Proteomics **8**(1): 157-171.

Guisbert, E., D. M. Czyz, K. Richter, et al. (2013). "Identification of a tissue-selective heat shock response regulatory network." PLoS Genet **9**(4): e1003466.

Hall, A. M. and E. D. Roberson (2012). "Mouse models of Alzheimer's disease." Brain Res Bull **88**(1): 3-12.

Hansen, B. E. (1995). "TIME SERIES ANALYSIS James D. Hamilton Princeton University Press, 1994." Econometric Theory **11**(03): 625-630.

Hardy, J. (2006). "A hundred years of Alzheimer's disease research." Neuron **52**(1): 3-13.

Hardy, J. and D. J. Selkoe (2002). "The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics." Science **297**(5580): 353-356.

Harold, D., R. Abraham, P. Hollingworth, et al. (2009). "Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease." Nat Genet **41**(10): 1088-1093.

Hartl, F. U. and M. Hayer-Hartl (2009). "Converging concepts of protein folding in vitro and in vivo." Nat Struct Mol Biol **16**(6): 574-581.

Hatfield, C. F., J. Herbert, E. J. van Someren, et al. (2004). "Disrupted daily activity/rest cycles in relation to daily cortisol rhythms of home-dwelling patients with early Alzheimer's dementia." Brain **127**(Pt 5): 1061-1074.

Hege, D. M., R. Stanewsky, J. C. Hall, et al. (1997). "Rhythmic expression of a PER-reporter in the Malpighian tubules of decapitated *Drosophila*: evidence for a brain-independent circadian clock." J Biol Rhythms **12**(4): 300-308.

Henderson, V. W., W. Mack and B. W. Williams (1989). "Spatial disorientation in Alzheimer's disease." Arch Neurol **46**(4): 391-394.

Hendricks, J. C., S. M. Finn, K. A. Panckeri, et al. (2000). "Rest in *Drosophila* is a sleep-like state." Neuron **25**(1): 129-138.

Hillier, L. W., A. Coulson, J. I. Murray, et al. (2005). "Genomics in *C. elegans*: so many genes, such a little worm." Genome Res **15**(12): 1651-1660.

Hindle, J. V. (2010). "Ageing, neurodegeneration and Parkinson's disease." Age Ageing **39**(2): 156-161.

Hollingworth, P., D. Harold, R. Sims, et al. (2011). "Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease." Nat Genet **43**(5): 429-435.

Huang, X. G., B. K. Yee, S. Nag, et al. (2003). "Behavioral and neurochemical characterization of transgenic mice carrying the human presenilin-1 gene with or without the leucine-to-proline mutation at codon 235." Exp Neurol **183**(2): 673-681.

Iliff, J. J., M. Wang, Y. Liao, et al. (2012). "A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid beta." Sci Transl Med **4**(147): 147ra111.

Ingram, E. M. and M. G. Spillantini (2002). "Tau gene mutations: dissecting the pathogenesis of FTDP-17." Trends Mol Med **8**(12): 555-562.

Kamath, R. S. and J. Ahringer (2003). "Genome-wide RNAi screening in *Caenorhabditis elegans*." Methods **30**(4): 313-321.

Khabirova, E., A. Moloney, S. J. Marciniak, et al. (2014). "The TRiC/CCT chaperone is implicated in Alzheimer's disease based on patient GWAS and an RNAi screen in Abeta-expressing *Caenorhabditis elegans*." PLoS One **9**(7): e102985.

Konopka, R. J. and S. Benzer (1971). "Clock mutants of *Drosophila melanogaster*." Proc Natl Acad Sci U S A **68**(9): 2112-2116.

Koonin, E. V. (2005). "Orthologs, paralogs, and evolutionary genomics." Annu Rev Genet **39**: 309-338.

Koudounas, S., E. W. Green and D. Clancy (2012). "Reliability and variability of sleep and activity as biomarkers of ageing in *Drosophila*." Biogerontology **13**(5): 489-499.

Kraemer, B. C., J. K. Burgess, J. H. Chen, et al. (2006). "Molecular pathways that influence human tau-induced pathology in *Caenorhabditis elegans*." Hum Mol Genet **15**(9): 1483-1496.

Kraemer, B. C., B. Zhang, J. B. Leverenz, et al. (2003). "Neurodegeneration and defective neurotransmission in a *Caenorhabditis elegans* model of tauopathy." Proc Natl Acad Sci U S A **100**(17): 9980-9985.

Kunst, M., M. E. Hughes, D. Raccuglia, et al. (2014). "Calcitonin gene-related peptide neurons mediate sleep-specific circadian output in *Drosophila*." Curr Biol **24**(22): 2652-2664.

Labbadia, J. and R. I. Morimoto (2014). "Proteostasis and longevity: when does aging really begin?" F1000Prime Rep **6**: 7.

LaFerla, F. M., K. N. Green and S. Oddo (2007). "Intracellular amyloid-beta in Alzheimer's disease." Nat Rev Neurosci **8**(7): 499-509.

Lambert, J. C., S. Heath, G. Even, et al. (2009). "Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease." Nat Genet **41**(10): 1094-1099.

Lambert, J. C., C. A. Ibrahim-Verbaas, D. Harold, et al. (2013). "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease." Nat Genet **45**(12): 1452-1458.

Lechner, M., S. Findeiss, L. Steiner, et al. (2011). "Proteinortho: detection of (co-)orthologs in large-scale analysis." BMC Bioinformatics **12**: 124.

Lee, V. M., T. K. Kenyon and J. Q. Trojanowski (2005). "Transgenic animal models of tauopathies." Biochim Biophys Acta **1739**(2-3): 251-259.

Levine, J. D., P. Funes, H. B. Dowse, et al. (2002). "Signal analysis of behavioral and molecular cycles." BMC Neurosci **3**: 1.

Lewis, J., D. W. Dickson, W. L. Lin, et al. (2001). "Enhanced neurofibrillary degeneration in transgenic mice expressing mutant tau and APP." Science **293**(5534): 1487-1491.

Lim, C. and R. Allada (2013). "Emerging roles for post-transcriptional regulation in circadian clocks." Nat Neurosci **16**(11): 1544-1550.

Lindner, A. B. and A. Demarez (2009). "Protein aggregation as a paradigm of aging." Biochim Biophys Acta **1790**(10): 980-996.

Link, C. D. (1995). "Expression of human beta-amyloid peptide in transgenic *Caenorhabditis elegans*." Proc Natl Acad Sci U S A **92**(20): 9368-9372.

Link, C. D., A. Taft, V. Kapulkin, et al. (2003). "Gene expression analysis in a transgenic *Caenorhabditis elegans* Alzheimer's disease model." Neurobiol Aging **24**(3): 397-413.

Long, D. M., M. R. Blake, S. Dutta, et al. (2014). "Relationships between the circadian system and Alzheimer's disease-like symptoms in *Drosophila*." PLoS One **9**(8): e106068.

Lorber, B., M. L. Howe, L. I. Benowitz, et al. (2009). "Mst3b, an Ste20-like kinase, regulates axon regeneration in mature CNS and PNS pathways." Nat Neurosci **12**(11): 1407-1414.

Lu, B. and H. Vogel (2009). "*Drosophila* models of neurodegenerative diseases." Annu Rev Pathol **4**: 315-342.

Luo, W., W. F. Chen, Z. Yue, et al. (2012). "Old flies have a robust central oscillator but weaker behavioral rhythms that can be improved by genetic and environmental manipulations." Aging Cell **11**(3): 428-438.

MacDonald, M. E., C. M. Ambrose, M. P. Duyao, et al. (1993). "A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes." Cell **72**(6): 971-983.

Mailman, M. D., M. Feolo, Y. Jin, et al. (2007). "The NCBI dbGaP database of genotypes and phenotypes." Nat Genet **39**(10): 1181-1186.

Mayeux, R. (2010). "Clinical practice. Early Alzheimer's disease." N Engl J Med **362**(23): 2194-2201.

Mendelsohn, A. R. and J. W. Larrick (2013). "Sleep facilitates clearance of metabolites from the brain: glymphatic function in aging and neurodegenerative diseases." Rejuvenation Res **16**(6): 518-523.

Metropolis, N. and S. Ulam (1949). "The Monte Carlo method." J Am Stat Assoc **44**(247): 335-341.

Mohawk, J. A., C. B. Green and J. S. Takahashi (2012). "Central and peripheral circadian clocks in mammals." Annu Rev Neurosci **35**: 445-462.

Monacelli, A. M., L. A. Cushman, V. Kavcic, et al. (2003). "Spatial disorientation in Alzheimer's disease: the remembrance of things passed." Neurology **61**(11): 1491-1497.

Mourao, M., L. Satin and S. Schnell (2014). "Optimal experimental design to estimate statistically significant periods of oscillations in time course data." PLoS One **9**(4): e93826.

Musiek, E. S. (2015). "Circadian clock disruption in neurodegenerative diseases: cause and effect?" Front Pharmacol **6**: 29.

Musiek, E. S., D. D. Xiong and D. M. Holtzman (2015). "Sleep, circadian rhythms, and the pathogenesis of Alzheimer disease." Exp Mol Med **47**: e148.

Nollen, E. A., S. M. Garcia, G. van Haaften, et al. (2004). "Genome-wide RNA interference screen identifies previously undescribed regulators of polyglutamine aggregation." Proc Natl Acad Sci U S A **101**(17): 6403-6408.

O'Brien, R. J. and P. C. Wong (2011). "Amyloid precursor protein processing and Alzheimer's disease." Annu Rev Neurosci **34**: 185-204.

O'Keefe, J. and J. Dostrovsky (1971). "The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat." Brain Res **34**(1): 171-175.

Oddo, S., A. Caccamo, M. Kitazawa, et al. (2003). "Amyloid deposition precedes tangle formation in a triple transgenic model of Alzheimer's disease." Neurobiol Aging **24**(8): 1063-1070.

Oren, M., A. M. Tarrant, S. Alon, et al. (2015). "Profiling molecular and behavioral circadian rhythms in the non-symbiotic sea anemone Nematostella vectensis." Sci Rep **5**: 11418.

Partch, C. L., C. B. Green and J. S. Takahashi (2014). "Molecular architecture of the mammalian circadian clock." Trends Cell Biol **24**(2): 90-99.

Partridge, L. and M. Mangel (1999). "Messages from mortality: the evolution of death rates in the old." Trends Ecol Evol **14**(11): 438-442.

Pearson, T. A. and T. A. Manolio (2008). "How to interpret a genome-wide association study." JAMA **299**(11): 1335-1344.

Peschel, N. and C. Helfrich-Forster (2011). "Setting the clock--by nature: circadian rhythm in the fruitfly *Drosophila melanogaster*." FEBS Lett **585**(10): 1435-1442.

Peter-Derex, L., P. Yammine, H. Bastuji, et al. (2015). "Sleep and Alzheimer's disease." Sleep Med Rev **19**: 29-38.

Plautz, J. D., M. Kaneko, J. C. Hall, et al. (1997). "Independent photoreceptive circadian clocks throughout *Drosophila*." Science **278**(5343): 1632-1635.

Prinz, P. N., E. R. Peskind, P. P. Vitaliano, et al. (1982). "Changes in the sleep and waking EEGs of nondemented and demented elderly subjects." J Am Geriatr Soc **30**(2): 86-93.

Prussing, K., A. Voigt and J. B. Schulz (2013). "*Drosophila melanogaster* as a model organism for Alzheimer's disease." Mol Neurodegener **8**: 35.

Pryszcz, L. P., J. Huerta-Cepas and T. Gabaldon (2011). "MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score." Nucleic Acids Res **39**(5): e32.

Querfurth, H. W. and F. M. LaFerla (2010). "Alzheimer's disease." N Engl J Med **362**(4): 329-344.

Rakshit, K., A. P. Thomas and A. V. Matveyenko (2014). "Does disruption of circadian rhythms contribute to beta-cell failure in type 2 diabetes?" Curr Diab Rep **14**(4): 474.

Reddy, A. B. and J. S. O'Neill (2010). "Healthy clocks, healthy body, healthy mind." Trends Cell Biol **20**(1): 36-44.

Reppert, S. M. and D. R. Weaver (2002). "Coordination of circadian timing in mammals." Nature **418**(6901): 935-941.

Reza, A. M. (2004). "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement." Journal of VLSI signal processing systems for signal, image and video technology **38**(1): 35-44.

Rice, J. (2006). Mathematical statistics and data analysis, Cengage Learning.

Richards, J. and M. L. Gumz (2012). "Advances in understanding the peripheral circadian clocks." FASEB J **26**(9): 3602-3613.

Roberson, E. D., K. Scearce-Levie, J. J. Palop, et al. (2007). "Reducing endogenous tau ameliorates amyloid beta-induced deficits in an Alzheimer's disease mouse model." Science **316**(5825): 750-754.

Rosen, D. R., L. Martin-Morris, L. Q. Luo, et al. (1989). "A *Drosophila* gene encoding a protein resembling the human beta-amyloid protein precursor." Proc Natl Acad Sci U S A **86**(7): 2478-2482.

Roy, S., B. Zhang, V. M. Lee, et al. (2005). "Axonal transport defects: a common theme in neurodegenerative diseases." Acta Neuropathol **109**(1): 5-13.

Saini, C., A. Liani, T. Curie, et al. (2013). "Real-time recording of circadian liver gene expression in freely moving mice reveals the phase-setting behavior of hepatocyte clocks." Genes Dev **27**(13): 1526-1536.

Seelig, J. D. and V. Jayaraman (2015). "Neural dynamics for landmark orientation and angular path integration." Nature **521**(7551): 186-191.

Shankar, G. M. and D. M. Walsh (2009). "Alzheimer's disease: synaptic dysfunction and Abeta." Mol Neurodegener **4**: 48.

Shankar, G. M. and D. M. Walsh (2009). "Alzheimer's disease: synaptic dysfunction and Aβ." Molecular Neurodegeneration **4**(1): 1-13.

Shannon, P., A. Markiel, O. Ozier, et al. (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks." Genome Res **13**(11): 2498-2504.

Shulman, J. M., P. L. De Jager and M. B. Feany (2011). "Parkinson's disease: genetics and pathogenesis." Annu Rev Pathol **6**: 193-222.

Sleegers, K., J. C. Lambert, L. Bertram, et al. (2010). "The pursuit of susceptibility genes for Alzheimer's disease: progress and prospects." Trends Genet **26**(2): 84-93.

Snyder, E. M., Y. Nong, C. G. Almeida, et al. (2005). "Regulation of NMDA receptor trafficking by amyloid-beta." Nat Neurosci **8**(8): 1051-1058.

Soto, C. (2003). "Unfolding the role of protein misfolding in neurodegenerative diseases." Nat Rev Neurosci **4**(1): 49-60.

Spira, A. P., A. A. Gamaldo, Y. An, et al. (2013). "Self-reported sleep and beta-amyloid deposition in community-dwelling older adults." JAMA Neurol **70**(12): 1537-1543.

Spires, T. L. and B. T. Hyman (2005). "Transgenic models of Alzheimer's disease: learning from animals." NeuroRx **2**(3): 423-437.

St Johnston, D. (2002). "The art and design of genetic screens: *Drosophila melanogaster*." Nat Rev Genet **3**(3): 176-188.

Stark, C., B. J. Breitkreutz, T. Reguly, et al. (2006). "BioGRID: a general repository for interaction datasets." Nucleic Acids Res **34**(Database issue): D535-539.

Steurer, J. (2014). "Taking benzodiazepine is associated with an increased risk of Alzheimer disease." Praxis (Bern 1994) **103**(25): 1530.

Stokkan, K. A., S. Yamazaki, H. Tei, et al. (2001). "Entrainment of the circadian clock in the liver by feeding." Science **291**(5503): 490-493.

Strange, K., M. Christensen and R. Morrison (2007). "Primary culture of *Caenorhabditis elegans* developing embryo cells for electrophysiological, cell biological and molecular studies." Nat Protoc **2**(4): 1003-1012.

Tabuchi, M., S. R. Lone, S. Liu, et al. (2015). "Sleep interacts with abeta to modulate intrinsic neuronal excitability." Curr Biol **25**(6): 702-712.

Tillement, J. P., L. Lecanu and V. Papadopoulos (2010). "Amyloidosis and neurodegenerative diseases: current treatments and new pharmacological options." Pharmacology **85**(1): 1-17.

Tomioka, K., O. Uryu, Y. Kamae, et al. (2012). "Peripheral circadian rhythms and their regulatory mechanism in insects and some other arthropods: a review." J Comp Physiol B **182**(6): 729-740.

Torroja, L., H. Chu, I. Kotovsky, et al. (1999). "Neuronal overexpression of APPL, the *Drosophila* homologue of the amyloid precursor protein (APP), disrupts axonal transport." Curr Biol **9**(9): 489-492.

van Ham, T. J., K. L. Thijssen, R. Breitling, et al. (2008). "*C. elegans* model identifies genetic modifiers of alpha-synuclein inclusion formation during aging." PLoS Genet **4**(3): e1000027.

Veleri, S., C. Brandes, C. Helfrich-Forster, et al. (2003). "A self-sustaining, light-entrainable circadian oscillator in the *Drosophila* brain." Curr Biol **13**(20): 1758-1767.

Volicer, L., D. G. Harper, B. C. Manning, et al. (2001). "Sundowning and circadian rhythms in Alzheimer's disease." Am J Psychiatry **158**(5): 704-711.

Walsh, D. M., I. Klyubin, J. V. Fadeeva, et al. (2002). "Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo." Nature **416**(6880): 535-539.

Weggen, S. and D. Beher (2012). "Molecular consequences of amyloid precursor protein and presenilin mutations causing autosomal-dominant Alzheimer's disease." Alzheimers Res Ther **4**(2): 9.

Welsh, D. K., S. H. Yoo, A. C. Liu, et al. (2004). "Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression." Curr Biol **14**(24): 2289-2295.

White, J. G., E. Southgate, J. N. Thomson, et al. (1986). "The structure of the nervous system of the nematode *Caenorhabditis elegans*." Philos Trans R Soc Lond B Biol Sci **314**(1165): 1-340.

Wu, M. N., K. Koh, Z. Yue, et al. (2008). "A genetic screen for sleep and circadian mutants reveals mechanisms underlying regulation of sleep in *Drosophila*." Sleep **31**(4): 465-472.

Wulff, K., S. Gatti, J. G. Wettstein, et al. (2010). "Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease." Nat Rev Neurosci **11**(8): 589-599.

Xie, L., H. Kang, Q. Xu, et al. (2013). "Sleep drives metabolite clearance from the adult brain." Science **342**(6156): 373-377.

Yoo, B. C., M. Fountoulakis, M. Dierssen, et al. (2001). "Expression patterns of chaperone proteins in cerebral cortex of the fetus with Down syndrome: dysregulation of T-complex protein 1." J Neural Transm Suppl(61): 321-334.

Yoo, S. H., S. Yamazaki, P. L. Lowrey, et al. (2004). "PERIOD2::LUCIFERASE real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues." Proc Natl Acad Sci U S A **101**(15): 5339-5346.

Zhang, B., S. Kirov and J. Snoddy (2005). "WebGestalt: an integrated system for exploring gene sets in various biological contexts." Nucleic Acids Res **33**(Web Server issue): W741-748.

Zielinski, T., A. M. Moore, E. Troup, et al. (2014). "Strengths and limitations of period estimation methods for circadian data." PLoS One **9**(5): e96462.