miR-137, schizophrenia and sleep regulation in

Drosophila melanogaster

Kalon P. Grimes

A thesis submitted in partial fulfilment of the requirements of Bournemouth University for

the degree of Doctor of Philosophy

January 2020

Copyright statement

miR-137, schizophrenia and sleep regulation in *Drosophila melanogaster*

- Kalon P. Grimes

Introduction – microRNAs are non-coding DNA sequences, which regulate gene expression by promoting the destruction of target mRNAs. The evolutionarily conserved *miR-137* is expressed in invertebrate and mammalian neuronal tissue and there is experimental evidence for *mir-137* regulating neurological development, synaptic plasticity, and cognitive function. Other genes likely regulated by *miR-137* include members of the dopamine signalling pathway. Furthermore, genome wide association studies (GWAS) have linked SNPs in *mir-137* with schizophrenia (SZ). Modulation of dopaminergic signalling and SZ are associated with abnormal sleep quality, suggesting that *miR-137* may mediate sleep-wake behaviour.

Method – This prediction was tested by assessing the diurnal and circadian sleep-wake behaviour of *Drosophila melanogaster* homozygous for a null *miR-137* allele (*miR-137^{KO}*) or in flies where *miR-137* expression had been silenced (*miR-137^{sponge}*). Further investigation involved reintroduction of *miR-137* expression in selected brain regions through use of the *Drosophila* UAS/Gal4 genetic construct. Sleep-wake behaviour was monitored and quantified using a well characterised Drosophila activity monitoring system (DAMS). Locomotor behaviour was assessed by video tracking flies and quantifying climbing ability in a negative geotaxis assay. For gene expression, qualitative PCR (qPCR) Taqman assays were used, and PCR genotyping was completed with custom designed primer sets and gel electrophoresis. A total transcriptomics assay was conducted and analysed using bioinformatical tools, then compared to GWAS datasets.

Results – *miR-137* null and knockdown genotypes had an extreme sleep phenotype characterised by increased total sleep amount. The phenotype was attributed to the homeostatic sleep control pathway through activation of the fan-shaped and mushroom bodies in the brain. Successful knockdown of *miR-137* expression in the brain by the *miR-137^{sponge}* was verified in qPCR, and PCR genotyping confirmed the replacement of *miR-137* locus with an inserted $w^{mW.hs}$ sequence. There was also a moderate locomotor defect, though this did not account for the severity of the sleep phenotype. Additionally, there was a developmental delay along with an increased mortality in the pupal stage. There was no evidence for circadian disruption or shortening of lifespan because of *mir-137* loss of function. Comparison of the transcriptome expression changes with sleep and psychiatric disorder related GWAS identified a select set of genes which provide putative mechanisms for *miR-137* function.

Conclusion – *miR-137* is an important conserved microRNA with alleles already significantly associated with a major psychiatric disorder in humans. Currently, this research demonstrates that *miR-137* is responsible for regulating pathways ultimately controlling sleep amount, and causing a slight locomotor phenotype, both of which are synonymous with some symptoms in the human SZ disorder. The project also acts to validate the high efficiency and ability of using *Drosophila melanogaster* as a model for future research into conserved microRNAs.

Table of Contents

List of Figures

List of Tables

List of abbreviations

Abbreviation	Word
3IY	3-iodo-tyrosine
5' 6nt	nucleotide
ADHD	Attention deficit hyperactive disorder
ASD	Autism spectrum disorder
BD	Bipolar Disorder
BDSC	Bloomington Drosophila stock centre
Bp	Base pairs
cDNA	Complementary DNA
CNS	Central nervous system
CO_2	Carbon dioxide
CT	Circadian time
DAMS	Drosophila activity monitoring system
DEG	Differentially expressed gene
DF	Deficiency
DNA	Deoxyribonucleic acid
DNTP	Deoxyribonucleotide triphosphate
EEG	Electroencephalogram
FB	Fan-shaped body
g	gram
GITC	Guanidine isothiocyanate
GWAS	Genome wide association study
KO	knockout
LiCl	Lithium chloride
LOF	Loss of function
M	Molar
MB	Mushroom body
MDD	Major depressive disorder
mg	milligram
min	Minute
miR-137	microRNA-137
miRNA	microRNA
mL	millilitre
mM	millimolar
mm	millimetre
mRNA	Messenger RNA
PCR	Polymerase chain reaction
poly-A	Poly-adenosine
PRS	Polygenic risk score

qPCR	Quantitative polymerase chain reaction
RISC	RNA induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Revolutions per minute
RT	Reverse transcription
s	Second
SNP	Single nucleotide polymorphism
SZ	Schizophrenia
UAS	Upstream activating sequence
UTR	Untranslated region
WHO	World health organisation
VDRC	Vienna Drosophila research centre
ZT	Zeitgeber Time
μL	Microliter
°C	Degrees centigrade

Acknowledgements

Author's declaration

Sequencing of amplified DNA was performed as part of a paid service by the Medical Research Council Protein phosphorylation and ubiquitylation unit (MRC PPU) in Dundee University, UK.

The FlyChip transcriptomics expression array and data analysis was performed as part of a paid service by Cambridge Systems Biology Centre in Cambridge University, UK.

Open source UKBiobank data was downloaded and used from the Neale Lab.

Pre-analysed GWAS results and bespoke MAGMA enrichment analysis was conducted by Dr Richard Anney, Cardiff University.

All other work in this these belongs to the author alone.

1 Introduction

The five main psychiatric disorders investigated by the Psychiatric Genomics Consortium were: schizophrenia (SZ); bipolar disorder (BD); major depression disorder (MDD); attention deficit hyperactive disorder (ADHD) and autism spectrum disorder (ASD) (Smoller *et al.* 2019). Lifetime prevalence rates for these disorders range from 0.7% for SZ to 12.5% for MDD (Doherty and Owen 2014), with SZ alone having an estimated total of 20 million sufferers worldwide (GBD 2017). Originally following the introduction of antipsychotic drugs in the 1950s, the biological hypotheses of SZ was dominated by the dopamine hypothesis (Wise and Stein 1973) before the introduction of newer antipsychotics suggested serotonin, GABA and glutamate involvement (McGuffin Owen and Gotesman 2002). However, since the late 1980s (Murray and Lewis 1987; Weinberger 1987), when the neurodevelopmental hypothesis of SZ was re-proposed, it has gained momentum with recent genetic studies suggesting a developmental risk factor model of psychosis (Murray *et al.* 2017). A little earlier to the neurodevelopmental hypothesis from the 1980s, it was suspected that genetic factors were involved because psychiatric illness was known to run in families. A European study by Gottesman and Sheilds (1991) studying families and twins showed this disorder is common in the population, with a lifetime risk of 1% and an increased risk for relatives of probands with SZ (McGuffin Owen and Gotesman 2002). As our knowledge over the last 50 years has grown, it is now clear that we are dealing with highly complex, polygenic variants harbouring a range of susceptibilities that when combined with multiple environmental, or non-genetic factors, leads to various psychiatric disorders (Lee *et al.*2013; Smoller *et al.* 2019; FitzGerald *et al.* 2020).

One recent SZ genome wide association study (GWAS) (Ripke *et al.* 2011) identified variants from the microRNA *miR-137* as statistically associated with this disorder. *miR-137* is a brain enriched microRNA with recent literature linking it to certain symptoms associated

with other neurological diseases and psychiatric disorders (Devanna *et al*. 2014; Valles *et al.* 2014; Duan *et al.* 2014). In addition, many of the proven and predicted targets of miR-137 also contain reported SZ risk alleles (Ripke *et al.* 2011; Duan *et al.* 2014). However, what GWAS studies fail to elucidate is the function of this or other genes within a biological system. Currently, the pathways through which *miR-137* is involved in SZ risk and pathogenesis is unclear and despite its obvious importance, few of its predicted targets have been proven or tested *in vivo*. Equally, the reported associations with ASD and ID (Thomas et al 2018), neurological disorders (Chen *et al.* 2017; Mahmoudi and Cairns 2017; Liang *et al.* 2016) and fundamental cellular functions (He *et al.* 2018; Chen *et al.* 2017; Tamim *et al.* 2014) make *miR-137* an ideal candidate for *in vivo* investigation.

There are obviously several model organisms available to researchers, e.g. *Mus musculus, Rattus norvegicus, Danio rerio, Saccharomyces cerevisiae*, and all come with benefits and caveats. Selecting the correct one requires careful consideration including ease or speed of culturing experiments and the percentage of homology with human genes. The aim is then to compare model organism results with human genetics studies through DNA sequence homolog, inferred similar genetically influenced behaviour or social interactions.

One way to understand the function of *miR-137* is through the use of model organisms. In this thesis, the model organism *Drosophila melanogaster* was used to investigate the function of *miR-137* by manipulating its gene expression *in vivo.* The fruit fly was chosen because it has an evolutionarily conserved ortholog of the human *MIR-137*, a quick lifecycle, extensive published research, relatively low cost, and many other benefits of a model organism (chapter 1.6). It is therefore an ideal organism to investigate the putative role *miR-137* might have within neurons without using costlier and ethically challenging resources such as human stem cells.

In order to explain the reason *miR-137* was chosen for investigation, this chapter will review recent literature. Obviously, one of the first questions one might ask is 'how do genes influence risk for psychiatric illness?'. There are various theories postulated, but two that show prominence and strengthened by current literature are the neurodevelopmental hypothesis (Weinberger *et al.* 1995) and the threshold model (Gottesman and Gould 2003). How is it possible that genetic variation increases or decreases risk associated with disease? The following sections will first detail *miR-137* and then relate that to the neurodevelopmental hypothesis of SZ. There are obvious overlaps with neurological disorders, phenotypically distinct yet both caused by dysregulation of the brain. Perhaps the threshold model of SZ, suggested to have two levels of risk (Lee *et al.* 2013) might share similar overlap for generalised risk of any neurological or psychiatric disorder, before additional risk alleles specify which disease phenotype emerges.

The power of recent GWAS is reviewed as is the current status of model organisms using gene knockout or sequestering methods. Finally, in preparation for the results presented in Chapters 4-8, the final section in this introductory chapter presents a brief review of how sleep regulation is controlled at a cellular level and how this specific phenotype shares dysfunction in both SZ and neurological disorders.

1.1   Transcription and regulation of DNA

Deoxyribonucleic acid (DNA) is the biological equivalent of computer code. Essentially, DNA is a long sequence contained in each individual cell nucleus that encodes the development and maintenance of an organism's life. Much like computer code, without the correct instruments to decipher and utilise the information, DNA is useless by itself. The DNA sequence is made of a combination of 4 different nucleotides; Adenosine, Guanine,

Cytosine, and Thymine. It is known to have a double helix structure made from two complimentary strands bound together in a spiral shape, which is very robust and protected from degradation or external influences. In humans, the total length of the DNA sequence is around 2 billion nucleotides in somatic cells and a copy of it is contained in every cell other than red blood cells which do not have a nucleus.

As genes, are translated into enzymes they then usually leave the protection of the nucleus to undergo further modification or directly perform a function. With DNA separated into coding and non-coding regions, the latter regarded as junk DNA, each have important roles in deciphering and modifying the coded instructions from DNA sequence.

1.1.1 Coding DNA

Coding DNA is the name given to sections of DNA that contain genes which undergo transcription into an mRNA molecule and then translation into a protein before it can perform its role. These sections account for approximately only 1% of the human genome, producing around 20,000 different genes (Encode consortium 2012).

In the protein coding regions of DNA, the nucleotide sequence is read in codons (a triplet of consecutive nucleotides), which encode 21 different amino acid residues. DNA also includes start and stop codons, which are read by RNA-polymerase enzymes in the nucleus. Once the RNA-polymerase binds a start codon, it begins synthesis of a complimentary RNA molecule (a process called transcription). Once the RNA molecule is complete it undergoes splicing to remove unnecessary regions. The resulting messenger RNA (mRNA) molecule then leaves the nucleus, travelling to the ribosomes where it is translated into amino acids and strung together into proteins. These proteins often undergo post-translational modifications which include binding other molecules like phosphate groups or folding the protein into its

19

final functional shape. Examples of proteins include; enzymes, structural proteins, or transmembrane channels and receptors.

1.1.2 *Non-Coding DNA*

The non-coding regions were originally thought of as "junk DNA" with little or no purpose, but research has now that many of these regions have a very important function (Alles *et al.* 2019). Like the coding regions they contain genes, but these perform their biological function directly after transcription as an RNA molecule rather than undergoing translation into a protein. There are several types of non-coding RNA; which include microRNA, small interfering RNA, and transfer RNA.

1.1.2.1 *microRNA*

microRNA (miRNA) are one type of the non-coding genes that act as post-transcriptional regulators of mRNA expression (Hobert 2007; Willemsen *et al.* 2011; Guella *et al.* 2013). As the volume of research into miRNAs increases, they are becoming progressively more significant due to their involvement in key gene signalling pathways, dysfunction of which often leads to serious disorders and problems for an individual (Enright *et al.* 2003; Hobert 2007; Xiao *et al.* 2007; Yin *et al.* 2014; Juzwik *et al.* 2018). They are part of the genetic regulatory system and operate through associating with an RNA-induced silencing complex (RISC), acting as a target guide to bind to the 3' un-translated region (UTR) of specific mRNA (Enright *et al.* 2003; Lee *et al.* 2004; Murchison and Hannon 2004; Hobert 2007; Xiao *et al.* 2007; Thomson *et al.* 2011; Willemsen *et al.* 2011). This binding leads to destruction or modification of the mRNA, ultimately resulting in less translation of the mRNA (Lee *et al.* 2004; Yin *et al.* 2014). While not a complete ablation of the target

20

mRNA expression, this translational oppression has been shown to be significant enough to cause physiological effects (Hobert 2007). miRNAs are one of the largest classes of gene regulatory molecules in animals with an estimated 2,300 distinct miRNA in the human genome (Hobert 2007; Alles *et al.* 2019). Approximately 10% of these are heavily conserved across multiple species, demonstrating the biological importance of these molecules (Enright *et al.* 2003). Changes in the sequence or expression of some miRNA have already been associated with SZ, BD, cancer, cardiac hypertrophy, multiple sclerosis, and even diabetes (Xiao *et al.* 2007; Ripke *et al.* 2011; Kandemir *et al.* 2014; Valles *et al.* 2014; Duan *et al.* 2014; Juzwik *et al.* 2018).

As the mechanism of mRNA targeting by miRNA is now relatively well understood it has enabled the development of algorithms which generate a list of putative gene targets from the DNA sequence of the miRNA and the mRNA (Enright *et al.* 2003; Agarwal *et al.* 2015; Agarwal *et al.* 2018). There are several prediction algorithms run by different institutes, but they usually operate around the same principles. The seed sequence and mRNA sequence are known, and therefore it is possible to predict binding affinity of each miRNA to the mRNA molecule (Thomson *et al.* 2011). The effect of the miRNA binding is dependent upon where on the mRNA the binding region is located and degree of homology between the sequence and seed sequence. As with all computational methods, the results achieved vary depending on factors such as which algorithm or database of mRNA sequences.

Alongside putative targets created with algorithm calculations are experimental based research projects which utilise the outputs to direct research into proving the miRNA regulation on some mRNA. These studies vary in approach; some look for direct interactions of a select pairing of miRNA and mRNA, others examine the transcriptome with different levels of miRNA expression (Hill *et al.* 2014), and finally some observe phenotypical effects of manipulated miRNA expression (Hobert 2007; Karres *et al.* 2007; Xiao *et al.* 2007). There

are also differences in the situations of the experiments, whether in a whole living organism or through *in vitro* cell cultures. The theory of holistically assessing the effect of miRNA on a specific target is that in the absence of the miRNA expression, the phenotype exhibited by the organism should mimic that of one which has an overexpression of the target mRNA (Hobert 2007). Furthermore, overexpression of a miRNA molecule should mimic a phenotype resulting from downregulation of the target mRNA as there is more miRNA related expression regulation. However, miRNA overexpression assays could produce skewed results as the miRNA does not act directly or individually on a target mRNA. Therefore, overexpression of the construct may just saturate the RISC complexes available or cause fluctuations among other active miRNA levels (Thomson *et al.* 2011). Furthermore, the overexpression increase could likely not be greater than 20-30% variance that many miRNA molecules are known to fluctuate in normal conditions (Hobert 2007).

1.1.2.2 Transcription and processing of a microRNA gene

Many miRNAs start off as a small sequence of a few hundred base pairs after initial transcription. This then undergoes post-transcriptional modification until it becomes a "mature" sequence that can perform its regulatory roles. miRNA genes are mainly transcribed by the enzyme RNA Polymerase II, which results in primary-miRNA (pri-miRNA) molecules (Lee *et al.* 2004). These long pri-miRNA strands are then cleaved by nuclear RNase III Drosha enzymes (Lee *et al.* 2003). The resulting hairpin structures of precursor-miRNA are exported out of the cell nucleus by exportin-5 transporters (Bohnsack *et al.* 2004; Lund *et al.* 2004). Once in the cytoplasm the pre-miRNA hairpins are further cleaved by RNase III Dicer molecules into 19-25nt double-stranded miRNA strands (Bernstein *et al.* 2001; Lee *et al.* 2004). The duplex-miRNA consists of a passenger strand and a mature strand. The passenger strand degrades leaving a mature-miRNA which is recruited as part of an active RISC

complex (Rana 2007; Thomson *et al.* 2011). The process of passenger strand degradation is largely unknown, but the selection of passenger or mature strand is decided through strand thermodynamic stability and base pairing strength (Khvorova *et al.* 2003).

1.1.2.3 RISC and downregulation

RISC complexes are a cluster of several proteins and a mature miRNA construct. The RISC complex utilises the complementary binding of the miRNA and mRNA to target which sequences to affect. The miRNA binds to the mRNA using a set of 6 base pairs called the "seed sequence", which is commonly nucleotides 2 to 8 of the 5' region of the mature miRNA molecule (Devanna *et al.* 2014). Argonaute (Ago) proteins are responsible for binding to the guide miRNA strand and orientating it to bind with target mRNA (Murchison and Hannon 2004; Rana 2007; Alles *et al.* 2019). Ago proteins also feature an RNase H-like fold at the PIWI domain, which can directly cleave target mRNA (Pratt and MacRae 2009). Also assimilated within the RISC complex are: Dicer proteins, transactivating response RNA binding protein, protein activator of the interferon-induced protein kinase protein, tudor staphylococcal nuclease domain containing protein DNA helicase MOV10, and RNA recognition motif containing protein TNRC6B (MacRae *et al.* 2008; Thomson *et al*. 2011). The result of successful binding between the active RISC and target mRNA causes mRNA degradation through de-adenylation of the poly-A tail or cleavage by proteins, or reduced translation efficiency at the ribosomes (Thomson *et al.* 2011). This RNA interference is employed by each miRNA to successfully regulate multiple genes and pathways involved with key cell processes such as apoptosis, cell differentiation, and organ development (Bartel 2004). It has also been reported that several miRNAs can bind to the same target gene cooperatively (Enright *et al.* 2003; Lewis *et al.* 2003; Devanna *et al.* 2014; Juzwik *et al.* 2018), increasing the complexity of mRNA regulation by miRNA.

1.2   *microRNA miR-137*

miR-137 is a miRNA molecule associated with a multitude of diseases and disorders in humans. It is predicted to regulate over 1300 targets including some key genetic components like gene transcription factors and ion channels, which perform functions vital for cell or organism survival (Agarwal *et al.* 2015). *miR-137* is the most abundant miRNA in the brain and localises at the synapse where it regulates local target mRNA (Willemsen *et al.* 2011; Yin *et al.* 2014). Overall; the expression location, predicted and proven targets, and statistical links to several debilitating disorders highlight the potential for *miR-137* as a miRNA molecule of interest with regards to psychiatric disorders.

1.2.1   *Gene expression*

miR-137 is widely expressed in the mammalian brain and is used as a biomarker in the blood to detect cancer in humans. It is the most abundant miRNA in the brain and localises at the synapse where it regulates local target mRNA (Willemsen *et al.* 2011; Yin *et al.* 2014). In humans, the *MIR-137* gene is encoded by a stretch of DNA 102bp long, and then cleaved into a mature miRNA molecule just 23bp in length. Orthologs of *miR-137* can be found in a variety of organisms including fruit flies, mice, bees, and chimps (figure 1.1).

Description	Max Score	Total Score	Query cover	E Value	Per. Ident	Accession
Zebrafish DNA sequence from clone CH1073-362I17 in linkage gro...	41	41	29%	0.47	92.59	FQ377605.3
Danio rerio microRNA 137-1 (mir137-1), microRNA	41	41	29%	0.47	92.59	NR_030086.1
PREDICTED: Gorilla gorilla gorilla uncharacterized LOC10114598...	39.2	39.2	28%	1.6	92.31	XR_002007996.2
PREDICTED: Pongo abelii uncharacterized LOC100936419...	39.2	39.2	28%	1.6	92.31	XR_150921.3
Danio rerio microRNA 137-2 (mir137-2), microRNA	39.2	39.2	28%	1.6	92.31	NR_045229.1
Homo sapiens MIR137 host gene (MIR137HG), long non-coding RNA	39.2	39.2	28%	1.6	92.31	NR_046105.1
Pan troglodytes microRNA mir-137 (MIR137), microRNA	39.2	39.2	28%	1.6	92.31	NR_035671.1
Rattus norvegicus microRNA 137 (Mir137), microRNA	39.2	39.2	28%	1.6	92.31	NR_031883.1
Zebrafish DNA sequence from clone CH73-25E21 in linkage group...	39.2	39.2	28%	1.6	92.31	FP085414.9
Mus musculus microRNA 137 (Mir137), microRNA	39.2	39.2	28%	1.6	92.31	NR_029551.1
Homo sapiens microRNA 137 (MIR137), microRNA	39.2	39.2	28%	1.6	92.31	NR_029679.1
Homo sapiens cDNA, FLJ18442	39.2	39.2	28%	1.6	92.31	AK311400.1
Homo sapiens cDNA, FLJ99659	39.2	39.2	28%	1.6	92.31	AK309618.1
Homo sapiens pre-microRNA 137 gene, partial sequence and tande...	39.2	39.2	28%	1.6	92.31	EF660891.1
Homo sapiens cDNA FLJ37288 fis, clone BRAMY2014205	39.2	39.2	28%	1.6	92.31	AK094607.1
Homo sapiens chromosome 1 clone RP4-672J20, complete sequence	39.2	39.2	28%	1.6	92.31	AC104453.2
Homo sapiens chromosome 1 clone RP11-287O16, complete sequence	39.2	39.2	28%	1.6	92.31	AC027691.13
Mus musculus BAC clone RP23-14P23 from 3, complete sequence	39.2	39.2	28%	1.6	92.31	AC122183.4
Zebrafish DNA sequence from clone CH211-144G12 in linkage grou...	39.2	39.2	28%	1.6	92.31	AL929343.7
Homo sapiens genomic sequence surrounding NotI site, clone...	39.2	39.2	28%	1.6	92.31	AJ335312.1
Homo sapiens genomic sequence surrounding NotI site, clone...	39.2	39.2	28%	1.6	92.31	AJ335083.1
Homo sapiens genomic sequence surrounding NotI site, clone...	39.2	39.2	28%	1.6	92.31	AJ322674.1
Homo sapiens BAC clone RP11-374F15 from 2, complete sequence	38.3	38.3	30%	5.7	89.29	AC008173.2
Mouse DNA sequence from clone RP23-211E14 on chromosome 4,...	37.4	37.4	32%	5.7	86.67	BX572085.6

Distribution of the top 24 Blast Hits on 24 subject sequences
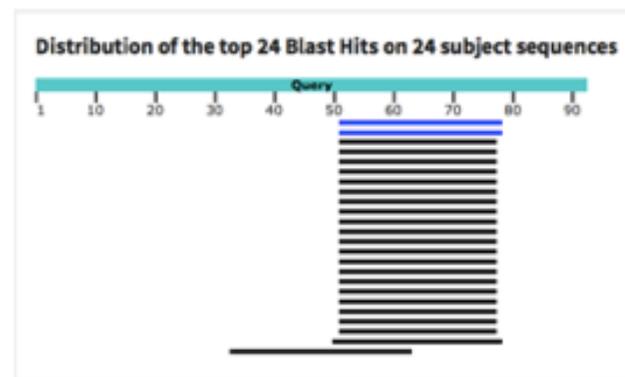
NR_048121 was BLASTed against other species to Identify homology between species of miR-137

Figure 1.1: BLAST results of *D. melanogaster miR-137* sequence.

The sequence for *D.melanogaster miR-137* was BLASTed against all other database held species to highlight the homology between species for *miR-137*. There is region between 50 and 78 base pairs highly conserved in all species bar one.

miR-137 is known to be active during foetal development (Howell and Law 2019) but shows an increase in expression beginning around the age of 25, and gradually increasing until old age. *miR-137* regulates adult neurogenesis, synaptic plasticity, and glutamatergic signalling, all processes thought to contribute to the pathobiology (Pacheco *et al.* 2019). This age also corresponds with onset of adult SZ (Foley *et al.* 2017).

1.2.2   Predicted and proven targets

There are many targets for each miRNA, and human *miR-137* is predicted to have a direct influence on over 1300 genes with evolutionarily conserved regions (Agarwal *et al.* 2015). Understanding the direct targets and subsequent gene networks is vital for understanding how miRNA influence gene expression and neurodevelopment.

1.2.3   Implications for disease

miR-137 has been reported in many diseases (OMIM 614304; Chen *et al.* 2020; Lee *et al.* 2019) *in vivo, in vitro* and *in silico*. miRNAs are involved in the regulation of multiple target mRNAs, therefore issues within a single miRNA gene can have the knock-on effect of expression changes of many downstream targets. Such changes are likely to have a phenotypical effect synonymous with disease symptoms. Differences in *miR-137* expression were reported in tumour growth and cancers (Chen *et al.* 2020; Lee *et al.* 2019)). Loss of *miR-137* expression due to hypermethylation was reported in patients with endometrial cancer and *miR-137* was reactivated with the introduction of epigenetic inhibitors (Zhang *et al.* 2018).

miR-137 is one of the most prevalent miRNA molecules expressed within the mammalian brain. As such it is not surprising that the gene has been implicated with the presentation of symptoms associated with neurological diseases and/or psychiatric disorders. Statistically, risk alleles of the human *miR-137* gene were also found to be significantly associated with SZ in GWAS studies (Ripke *et al.* 2011; Duan *et al.* 2014), and rs1702294 (C) and rs1625579 (T) SNPs have been linked to poorer scores on neuropsychological tests of memory performance (Cosgrove *et al.* 2017; Cummings *et al.* 2013). *miR-137* functions as a 'fine-tuning' agent for synaptic pruning and dendritic arborisation in neural cells and has been associated with altered cognitive function (Willemsen *et al.* 2011; Cummings *et al.* 2013; Green *et al.* 2013; Kuswanto et al. 2015). The miRNA also regulates processes that are potentially involved with cognitive dysfunction, in part by acting as a hippocampal gene network node, co-ordinating the expression of genes relating to nervous system function and development (Olde Loohuis *et al.* 2017)**.** SZ and BD sufferers have a 30% decrease in *miR-137* expression in prefrontal cortex (Guella *et al.* 2013), and significant enrichment in hippocampal synapses (Willemsen *et al.* 2011). Problems with the neuron maturation and physical development in the hippocampus and prefrontal cortex are associated with the development of SZ in humans (Heckers 2001), further associating the proper function of *miR-137* as necessary for prevention of psychiatric or neurological symptoms.

While *miR-137* itself is experimentally linked to neurological symptoms, some of its target mRNAs are also known to be involved with neurological function, development, and maintenance. These include; *RORα* (RAR-related orphan receptor alpha), *CACNA1C* (Voltage-gated calcium channel alpha 1c), *D2R* (Dopamine 2 receptor), and *TCF4* (Transcription factor 4). Many neurological symptoms such as hallucinations, cognitive defects, behaviour and emotion changes, and structural brain anomalies have already been associated with improper expression of these key genes (Cummings *et al.* 2013; Devanna and

27

Vernes 2014; Yin *et al*. 2014; Kong *et al.* 2015; Marschallinger *et al.* 2015; Pinggera *et al*. 2015; Wang *et al.* 2015; Gonzalez-Giraldo *et al.* 2016). *miR-137* has also been shown to be a major player in regulating synaptic plasticity (Valles *et al.* 2014), with many of the miRNA downstream targets being directly involved in the regulation and proper function of membrane potential or vesicle movement (Siegart *et al.* 2015). Known SZ risk genes *ZNF804A, CACNA1C, TCF4,* and *CSMD1* are proven direct targets of human *MIR-137* (Kwon *et al.* 2011; Kim *et al.* 2012; Guella *et al.* 2013), initially highlighted as significantly associated with SZ from GWAS studies (Ripke *et al.* 2011). As many neurological symptoms are prevalent in several neurological or psychiatric disorders, genes can be connected to a symptom associated with several disorders. Proven targets for *miR-137* also include many risk factors linked to ASD, such as *RORa*, *NRXN1*, *SHANK2*, *SCN2A* and *CACNA1:* many of which were previously mentioned as SZ or BD risk genes (Devanna *et al.* 2014).

Overall, there are many reports in the literature between neurological diseases, psychiatric disorders, and neurological symptoms and *miR-137* or its downstream target genes. While there are studies into the effect of *miR-137* improper expression, much of the data is statistical or inferred through research on the downstream targets. Assessing the holistic effect of complete loss of function would enable a better understanding of the genes *miR-137* regulates.

Expression of *miR-137* is necessary to regulate neurogenesis and maintain differentiation and proliferation of nerve cells. Neurogenesis is essential for memory, learning and mood regulation; disruption of which is seen in SZ (DeCarolis and Eisch, 2010). A paper by Wright *et al.* used TargetScan v6.2 to find 1141 putative targets for *miR-137* (2013). When compared with GeneCards and SZ Gene Resource, only 25 genes matched putative SZ genes.

Recent studies have investigated *miR-137* to elucidate function and target genes *in vitro* and *in vivo*. A study by Cheng *et al.* knocked out miR-137 in both mouse germline (gKO) and nervous system only (cKO) leading to postnatal lethality (2018). However, they found that partial loss of *miR-137* in heterozygous nervous system only (cKO) revealed dysregulated synaptic plasticity and behavioural phenotypes including repetitive behaviour, impaired learning and social behaviour. Further analysis revealed that PDe10a is elevated in heterozygous cKO mice and is reduced when a PDe10a inhibitor is administered. Interestingly, the behavioural phenotypes noted in the heterozygous cKO improve.

Conversely, over expression of *miR-137* appears to influence schizophrenia-type phenotypes in mice (Arakawa *et al.* 2019). This recent study tested pre-pulse inhibition, social and cognitive deficits analogous with SZ symptoms and found down-regulation of putative *miR-137* target genes. This makes sense, given microRNAs mainly down-regulate the expression of target genes by binding to mRNA, targeting transcripts for degradation or preventing translation (Pacheco *et al.* 2019). Although four SNPs were reported to be associated with *miR-137*: rs1625579; rs1198588; rs2660304; and rs1702294 (Ripke *et al.* 2011), VNTRs may also influence *miR-137* expression. Pacheco *et al.* investigated a VNTR with a repetitive sequence of 15 bases and high GC content (2019). Three repeats appear to be the shortest known length and the common allele. However, *in-vitro* studies show that mature *miR-137* expression is disproportionally affected by the length of the VNTR. They further applied a case-control study in adult and child onset SZ to find that the shorter VNTR (higher expression) comes with higher risk of SZ and the longer the VNTR (lower expression), the more protective the effect (Pacheco *et al.* 2019). Interestingly, there is a low level of linkage disequilibrium between the VNTR length and the *miR-137* SNP associated with SZ, with the VNTR not being haplotype dependent. In the presence of such SNPs, the SZ risk scores are relatively lowered when with VNTR over 3 repeats in length, an effect

likely due to the increase VNTR length shifting the miRNA splicing to exclude the mature *miR-137* coding sequence. This loss of the mature coding sequence ultimately acts to reduce *miR-137* expression, an effect that was naturally prevalent in foetal brain tissue which had higher frequencies of >3 VNTR repeats when compared to adult brain tissue (Pacheco *et al.* 2019). The SNP rs2660304 was found to contribute to risk independently, with recombination variants showing distinct changes to the SZ risk score (Pacheco *et al.* 2019). This information reinforces the idea of a complicated pattern of risk architecture in *miR-137* between SNP and VNTR haplotype in increasing susceptibility of SZ diagnosis.

miR-137 and other genes such as *AKT3, DRD2* and *AKT1* all point to signalling through the AKT/mTOR pathway (Howell and Law 2019). *miR-137* gain-of-function appears to be implicated in risk for SZ (Pacheco *et al.* 2019). Therefore, taken together, it suggests that *miR-137* plays critical roles in the running of cellular development at key stages in life, *miR-137's* biological role in neurogenesis and early neural development and further strengthens the neurodevelopmental hypothesis of SZ.

1.3   *Schizophrenia*

SZ has a prevalence of approximately 1% of the population, diagnosed using the International Classification of Diseases (ICD-11) produced by the World Health Organisation (WHO), by a psychiatrist assessing an individual for the presence of 'positive' and 'negative' symptoms. Positive symptoms of SZ are split into hallucinations, delusions, and disorganised thinking. The negative symptoms involve a loss of ability or enjoyment in life, including a lack of motivation, slow movement, sleep pattern changes, poor hygiene, reduced emotions, and low interest in socialising and activities (Jakubowski *et al.* 2012; Lee *et al.* 2013). There are various subtypes SZ, each with a distinct set of symptoms. For example, paranoid SZ is

one of the more common forms, with prominent hallucinations and delusions but largely unaffected speech and emotions. Catatonic SZ less common, mute and immobile. Hippocampus and prefrontal cortex abnormalities have been implicated in various studies (Yin *et al.* 2014; Aydin *et al.* 2019).

SZ is polygenic (Lee *et al.* 2013). Multiple risk alleles have been identified from GWAS. For example, *CACNA1C* encodes the alpha-1c subunit of the L-type voltage-gated calcium ion channel and is implicated in BD and SZ (Ferreira *et al.* 2008; Green *et al.* 2010, Ripke *et al.* 2011), enabling fundamental processes such as synaptic transmission and cell membrane depolarization (Zheng *et al.* 1995; Berger and Bartsch 2014). Unfortunately, as the field is very large and convoluted, research into the aetiology of SZ has been partially impacted by poorly defined phenotypes (Green et *al.* 2013).

The morbid risk in the general population of developing SZ is 1% (Gottesman 1991). Including SZ spectrum disorders such as schizoaffective disorder, schizotypal personality disorder and paranoid personality disorder, the morbid risk increases. The incidence of SZ is uniform across populations at 0.2 to 0.4 per 1000 and the lifetime prevalence (risk) is about 1% (McGuffin et al 1995). Most sufferers of SZ have onset in early adulthood with a similar incidence between both sexes, although women do tend to have a higher attainment of social functioning, most probably due to the later onset, which in turn confers a better outcome with less hospital admissions and a more benign course of illness (Gentile and Fusco 2019). Affected individuals may have long periods of illness and are usually unable to work, having difficulties sustaining interpersonal relationships. SZ has an enormous impact at a societal level through the economic cost of treatment and the social cost: the morbidity and the mortality (Costa and Steffen 2019).

The first comprehensive descriptions of SZ date from the 19th century by the German psychiatrist Emil Kraepelin. He noted syndromes of catatonia and hebephrenia and distinguished between manic depressive insanity and dementia praecox. He viewed SZ as a single process occurring in youth with dementia as the outcome. Swiss psychiatrist Eugene Bleuler used the term SZ in 1911. He felt SZ was more psychological than pathological with the core symptoms characterised by deficits in language, deficits in volition and attention and by incongruity of affect, ambivalence and ASD (Kaplan *et al.* 2008). Psychosis was a secondary effect.

Later saw Kurt Schneider publish a list of symptoms that led to a more reliable diagnosis of SZ known as first rank symptoms in 1959. Psychiatrists now use two diagnostic systems for SZ: ICD-11 and DSM-5. These diagnostic systems are highly complementary and have improved reliability of diagnosis of SZ.

The essential elements of SZ are a mixture of characteristic signs and symptoms that have been present for a significant proportion of time during a 1-month period with some signs of the disorder persisting for at least 6 months. They involve a range of cognitive and emotional dysfunctions in domains that include perception, inferential thinking, language and communication, behavioural monitoring, affect, fluency and productivity of thought and speech, volition, drive and attention. Individuals differ in symptomology, but diagnosis is conferred using a group of these characteristics.

SZ is characterised by the three overlapping forms of deficit termed positive symptoms, negative symptoms and cognitive impairment. Positive symptoms include distortions in thought content (delusions), perception (hallucinations), language and thought processes (disorganised speech) and self-monitoring of behaviour (grossly disorganised or catatonic behaviour). Negative symptoms include restrictions in the range and intensity of

emotional expression (affective flattening), in the fluency and productivity of thought and speech (alogia) and in the initiation of goal-directed behaviour (avolition). Cognitive deficits in performance of tasks involving attention, working memory and executive function are also a feature of SZ.

1.3.1   Psychotic symptoms

Psychotic symptoms involve the loss of contact with reality, including delusions, hallucinations or odd behaviour. Hallucinations might be auditory, visual, olfactory, gustatory or tactile. Commonly delusions in SZ include persecutory. The severity and presence of psychotic symptoms seems to be episodic over time (Lui *et al.* 2020).

1.3.2   Negative symptoms

Negative symptoms are deficit states in emotional and behavioural processes, diminished or absent influencing affective response (such as immobile facial expression, monotonous voice tone), a lack of emotion or pleasure, the ability to initiate and follow through plans and a reduced quality of speech. Negative symptoms are more pervasive, tend to be stable over time and are strongly associated with poor psychological functioning (Lui *et al.* 2020; Blanchard *et al.* 2020).

1.3.3   *Cognitive impairment*

Cognitive impairment in SZ includes problems in attention and concentration, psychomotor speed, learning and memory and executive functioning. A decline in cognitive performance is present in most people with SZ. This impairment remains stable over time (Mollon *et al.* 2020).

SZ is effectively a syndrome. At present, researchers have suggested various pathways with no firm biological marker to aid in diagnosis. It is accepted that stratifying diagnosis on the basis of polygenic risk scores might help in identifying genes and other aetiological factors quicker (Ohi *et al.* 2020). Although there is no clinical utility at the moment for aiding ICD-11 or DSM 5.

Results from family, twin and adoption studies have shown that the phenotype of SZ may include a broad spectrum of the other disorders such as schizophreniform disorder and schizotypal personality disorder. There is also uncertainty over the relationship of SZ and other affective psychoses and nonpsychotic affective disorders (Brink *et al.* 2019).

1.3.4   *Neurodevelopmental hypothesis of schizophrenia*

The neurodevelopmental hypothesis of SZ is now largely accepted. It suggests that an early brain lesion in the foetus during pregnancy or at birth, together with a combination of polygenic risk alleles gives rise to an increased risk of SZ. A combination of this risk might be environmental (Kelleher *et al.* 2014) or biological (Murray *et al.* 2017) but genetically is the result of subtle abnormal development in the central nervous system *in utero* or early infancy. These lesions seem to be latent until normal nerve cell maturation and neuronal pruning, switched on by various pathways, that results in prodromal and eventually clinical symptoms of psychosis leading not only to SZ (Jablensky *et al.*2017) but also depression,

suicidal ideas, and anxiety (Kelleher *et al.*2014). Therefore, if the neuronal scaffolding is compromised, you would expect to see deficits in psychophysiological and neurological functioning during childhood and early adolescence (Jablensky *et al.*2017); phenotypes originally observed and reported by Barbara Fish in the 1950s.

At the cellular level, the neurodevelopmental theory of SZ includes oxidative stress, excitotoxicity, proinflammatory cytokines during pregnancy and neuronal migration, suggesting a 'polygenic neurodevelopmental diathesis-stress model' (Jablensky *et al.* 2017). Genetic studies reporting common genetic variation each of very small effect size, with a few reported CNVs of larger effect size overlapping between SZ, ASD and learning disability, suggest a common neurodevelopmental genetic risk (Owen *et al.*58). This would be consistent with genes involved in basic neuronal scaffolding, migration and synaptic pruning. We now know that polygenic risk scores now account for 9% of variance in psychotic cases (Vasos *et al.*2017). So, the remaining variation might be environmental, such as risk of psychosis might be increased by adverse social, drug, immunological or hypoxic factors, most of which will occur during neural development (Murray *et al.*2017).

Therefore, *miR-137* fits the neurodevelopmental theory of SZ because 70% of miRNAs identified are expressed in the central nervous system (CNS), where specific subtypes are brain or brain-region specific (Rizos *et al.*2016). *miR-137* is crucial to all cells and expressed at different stages in the life cycle (Yin et al 2014) but reported to be greater in the amygdala and hippocampus at higher expression levels (Zamore *et al.*2005), and in adult brain regions with active neurogenesis, the subgranular cell layer in the hippocampal dentate gyrus (Nakahara and Carthew 2004; Yin *et al.*2014) with overexpression decreasing the number of astrocytes and inhibition reducing the number of neurons (Furuta *et al.*2010).

Specifically, *miR-137* embryonic expression is controlled by a two gene complex, in a negative feedback loop, consisting of the nuclear receptor transcription factor TLX and the histone demethylase LSD1 (lysine specific demethylase 1). Binding of this complex inhibits the transcription of *pre-miR-137* and binding of *miR-137* inhibits LSD1, maintaining expression at appropriate levels (Yin *et al.* 2014; Kiyohito *et al.* 2015). However, in adult neural stem cells, SOX2 and MeCP2 regulate *miR-137* expression (Yin *et al.* 2014).

So, we must be mindful of the many genetic interactions, reportedly associated with SZ, and the impact on brain development particularly in cognitive development, at various stages of life (Toulopoulou *et al.* 2015). Depending upon the stage of life cycle, genes associated with SZ might influencing upstream factors or downstream factors (Murray *et al.* 2017). Therefore, as stated by Murray *et al.* (2017), the neurodevelopmental hypothesis should now be regarded as the Developmental Risk Factor Model (Howes *et al.* 2017; Murray and Fearon, 1999):

"Thus, liability to psychosis is distributed in the same way as liability to hypertension or obesity. If an individual's blood pressure is persistently above a certain arbitrary level (90 mmHg in many countries), they are considered hypertensive; if the hypertension is not readily responsive to treatment, they may be further diagnosed as having severe or malignant hypertension. Similarly, if psychotic symptoms go beyond a certain threshold then a diagnosis of clinical psychotic disorder is appropriate, and if this persists and is associated with cognitive impairment of developmental origin, then a diagnosis of schizophrenia is made."

Murray *et al.* 2017

1.4   *Neurological diseases and psychiatric disorders*

"Neurological disease" is an umbrella term for disorders or diseases which affect the functionality of the brain. Stereotypically these result in behavioural oddities and ultimately a lack of ability to function normally within society without intervention.

A psychiatric disorder is an illness which affects the physiological, physical, or mental ability of an individual. It is known that both genetic and epigenetic risk factors contribute to the onset of these disorders (Lee *et al.* 2013), which makes understanding them more difficult. Neurological disorders have a multitude of risk alleles associated with their development from GWAS. These studies look at a collection of human genomes and comparatively assess which alleles are likely to be a factor of the disorder (Lee *et al.* 2013). From this computational analysis, many genes have been implicated and used in further research. However, it is important to note that in such a broad field there is still relatively little understood.

The main psychiatric disorders are largely accepted as SZ, ADHD, BD, and MDD (Smoller *et al.* 2013). Each disorder has a set of symptoms an individual must possess to be deemed a clinical sufferer. Unfortunately, the disorders often have overlapping symptoms, such as impaired motility and behavioural changes, which blur the boundaries between them. This overlap can make it difficult to give a correct diagnosis and causes some debate on the classification variants of each syndrome.

1.5   *Genome-wide association studies*

Candidate genes in psychiatry have been studied for decades (Horwitz *et al.* 2019). However, before 2007, many of the studies published had small sample sizes and low statistical power to detect variants of major effect. One method of increasing statistical power was to pool data sets together to run meta-analyses. However, even these sample sets did not provide reproducible results.

One method of identifying potential risk genes for these disorders is to compare a large collection of genomes of cases against the genomes of controls and analyse the allele frequency of variants between the two samples sets to determine any statistically significant differences. These GWAS are a powerful approach for the discovery of genes for many illnesses. This method of scaling-up candidate gene association studies to investigate genome-wide association with diseases was first published by a Japanese group (Ozaki *et al.* 2002) investigating susceptibility to Myocardial Infarction with Single Nucleotide Polymorphisms (SNPs). Using the resources available and understanding at that time of the statistical nuances of GWAS (Thomas *et al.* 2005) gave rise to the possibility of investigating other complex genetic conditions (Zeggini *et al.*2007; Parkes *et al.* 2007; Todd *et al.* 2007; Samani *et al.* 2007; Thomson *et al.* 2007; Sandhu *et al.* 2008; Barrett *et al.* 2008) including the major psychiatric disorders (WTCCC 2007; Ferreira *et al.* 2008; Ripke *et al.* 2011). Over the past 12 years, GWAS for psychiatric conditions (Horwitz *et al.* 2019) has served to elucidate the genetic architecture of these disorders by improving classifications and leading to targeted interventions, including the genes involved in the pathophysiology of SZ (Dennison *et al.* 2019). There are now many reproducible findings (Sullivan *et al.* 2019) but the field is now at a stage where there is a requirement for the understanding of the functional genomic architecture, loci that impact regulatory processes, influencing gene expression and the functional co-ordination of genes that control biological processes (Sullivan *et al.* 2019).

1.6   Model organisms

Model organisms are species which have been selected for research to obtain genetic information on humans which cannot be investigated directly. There are many reasons why researching a specific organism would be problematic; such as toxicity for viral or bacterial cultures, or ethical constraints for many animals including humans. There are several different model organisms that have been used to investigate gene function, genetic pathways, and molecular interactions.

For an organism to be useful in such experiments, they must have orthologous genes to the human gene targeted in the research. An ortholog is a gene which has a DNA sequence or similarity that indicates conserved functions despite being in different organisms, and by investigating this gene we can infer a similar behaviour will be seen in the human counterpart. Other benefits usually include shorter life cycle and reproduction time, relatively inexpensive culturing, and a wealth of knowledge already published on the genome to reinforce novel results. Ethical constraints aside, some model organisms are particularly useful for developmental research as they have a significantly shorter life cycle and faster development rate. Ultimately this would mean that research can progress much further and look at multiple generations quickly. There are several "well known" model organisms that are used in genetic research, many of which with specific advantages over the others covering much of the classification.

Mice (*Mus musculus*) and rats (*Rattus norvegicus*) are classic model vertebrates, used in many research fields including toxicology, behavioural analysis, development, and human disease modelling. They are also often used for neuroscience and investigating more evolutionarily advanced features such as the immune system which is not apparent in simpler model organisms.

Caenorhabditis elegans (C. elegans) is a nematode worm used to investigate genetic control over physiology and development since they normally have a fixed number of cells, 1031. This species was also the first multicellular organism to have its entire genome sequenced in 1998, revealing over 19000 genes and a 40% significant match with proteins in other organisms (The C. elegans sequencing consortium 1998).

Zebrafish (Danio rerio) is a species of fresh-water fish, which have a near-transparent body during early development. This provides a unique ability to visually examine the internal anatomy while the organism is developing.

1.6.1   *Drosophila melanogaster*

One of the first models used was *Drosophila melanogaster* (*D. melanogaster*), with published studies dating from 1906. Currently this is one of the most researched and well-documented models, with relevance to a huge variety of research fields (Adams and Sekelsky 2002; Duffy 2002; St Johnston 2002; Nichols *et al.* 2012). *D. melanogaster* feature many benefits useful for genetic research; a new generation of flies every 10-14 days at 25°C, a small culture space and simple food media, and a well-researched and documented genome (Adams *et al.* 2000; Linford *et al.* 2013). They also have no meiotic recombination in male flies, plenty of external features to assess genetic physiological importance in experiments, and only 4 chromosomes which can be easily visualized in the large polytene chromosomes of the larval salivary glands (St Johnston 2002). There are also many different tried and tested assays and behavioural tests that have been published for future research to be comparable to existing data (Nichols *et al.* 2012; Shaltiel-Karyo *et al.* 2012). Altogether, the wealth of benefits featured in *D. melanogaster* make it suitable for both large scale genetic screens and specific molecular-genetic interaction research.

Alongside the above traits, another useful feature used in *D. melanogaster* is the UAS/Gal4 mechanism taken from yeast (*Saccharomyces cerevisiae*). This is a bipartite system that can be controlled to drive expression of foreign DNA transcripts within a living fly (Brand and Perrimon 1993; Gao *et al.* 1999; Duffy 2002; St Johnston 2002). These transcripts can code for functional proteins, reporter proteins, missense RNA molecules, and more. Additionally, attaching the Gal4 DNA sequence to different genomic promoters can drive tissue specific expression allowing targeted research into localised genetic effects (Fischer *et al.*1988; St Johnston 2002). As the UAS/Gal4 system exists in 2 parts, the components are usually carried separately in two fly strains, whereupon crossing of an adult fly from each result in progeny with both parts of the system. Individually the UAS or Gal4 regions tend to have no effect on the organism, and therefore stable stocks are possible even when driving expression of a known lethal protein (Duffy 2002). Additionally, there are constructs which can increase the expression of the UAS construct or even utilise external temperatures to selectively switch on or off the expression. While the UAS/Gal4 system is also used in cell cultures and select other organisms (Xenopus laevis and zebrafish), the fruit fly is the organism which requires the least time and resources to successfully culture.

D. melanogaster can be cultured with relative ease on a food medium consisting of a mix of cornmeal, yeast, agar, and sugars in an incubator usually at a standard temperature of 25°C. The flies are housed in a culture vial consisting of a base of food media and a cotton wool plug, enabling the adults to be moved to new vials as required. In these conditions, the fruit fly cultures can progress from egg fertilisation to new adult fly within 10-12 days (figure 1.2). A significant amount of development happens within the first 24 hours, with the larva developing from a fertilised egg to hatching (Singh and Singh 1999; Linford *et al.* 2013). Over the following 5-7 days the larvae grow through 3 instar phases, living within the food media of the culture vial until they reach a size of up to approximately 5mm in length. At the

end of the 3rd instar stage, the larvae climb the sides of the culture vial and create a chrysalis which protects them while they undergo metamorphosis until around days 10-12 (Committee of developmental Toxicology 2000; St Johnston 2002). At this point, a new adult fly will eclose from the pupa and begin to reproduce in the following 8 hours (Committee of developmental Toxicology 2000). At the point where the fly metamorphoses into an adult, it undergoes a huge transformation in terms of general anatomy, neurological structure and physiology, and behaviour. The physiological changes to the central and peripheral nervous systems during pupation have been well studied, as this provides a new aspect to assess the effect of genes in neurological development through such a vital stage for the fly (Gao *et al.* 1999; Singh and Singh 1999; St Johnston 2002).
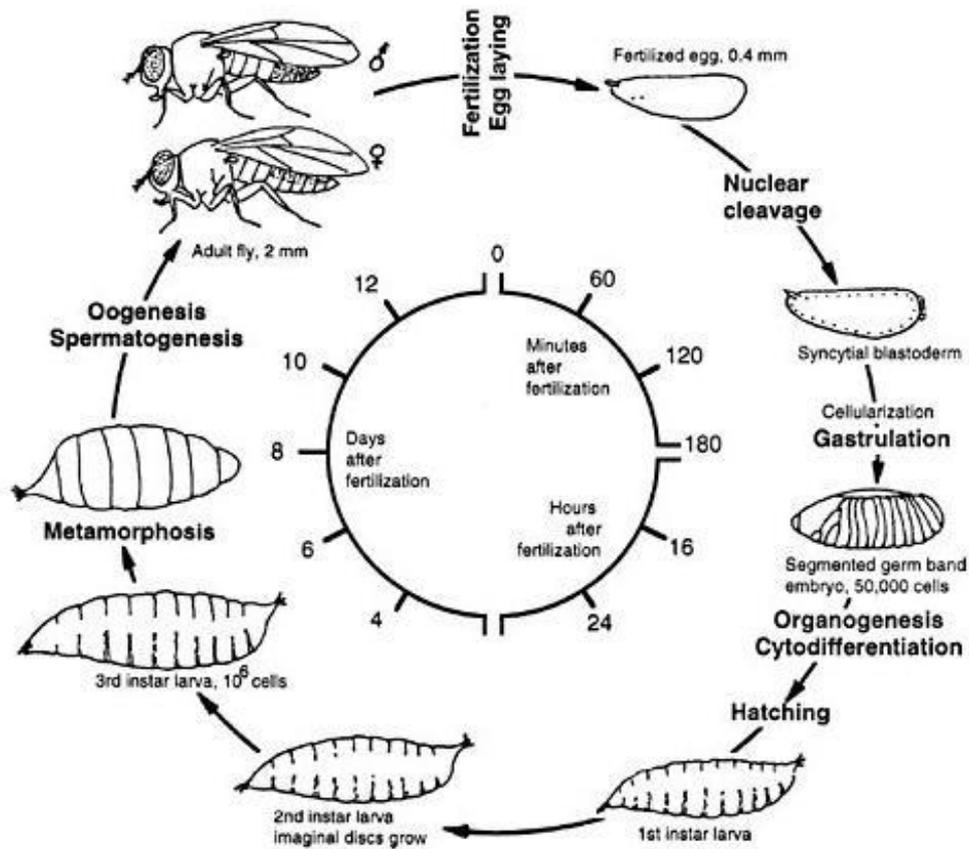
Figure 1.2: The life cycle of *Drosophila melanogaster*.

The life cycle of a fruit fly (Drosophila melanogaster) at 25°C (Committee of developmental Toxicology 2000). The fruit fly has a short life cycle of approximately two weeks. Following fertilization, the embryo develops and eventually hatches as a 1st instar larva after 24 hours. The larvae develop for the next 5 - 6 days through 2nd and 3rd instar stages, before pupating. The adult flies eclose after approximately 10 days since fertilization.

1.6.1.1 D. melanogaster genetics

Humans and *D. melanogaster* differ hugely in appearance, but in terms of genetics the differences are not so extreme. The fruit fly genome consists of three chromosomes and an approximate length total of 180mb and only around 13600 genes (Adams *et al.* 2000). This makes the *D. melanogaster* genome much smaller than humans', but they contain highly similar orthologous genes of around 77% of distinct human disease related genes (Reiter *et al.* 2001). In addition to this, when expressing the remaining human genes as a foreign insertion in the flies, they demonstrate similar phenotypical symptoms (St Johnston 2002).

Furthermore, *D. melanogaster* have hundreds of miRNA molecules within their genome (Ruby *et al.* 2007; Lu *et al.* 2008; Kozomara *et al.* 2014; Agarwal *et al.* 2018; Wessels *et al.* 2019). To date, 466 individual miRNAs have been identified in flies which are predicted to regulate ~30% of genes (Kozomara *et al.* 2014; Agarwal *et al.* 2018; Wessels *et al.* 2019). This is a significantly smaller number than the 2588 miRNA in humans, however the expression and RNA-downregulation mechanisms are similar (Ruby *et al.* 2007; Lu *et al.* 2008; Kim *et al.* 2009; Ha and Kim 2014; Agarwal *et al.* 2018; Wessels *et al.* 2019). Due to their usefulness as a genetic model, *Drosophila* have been the focus of many studies into the function of the conserved miRNA. Deletions of miRNA rarely cause lethal phenotypes or severe morphological changes in mammals as in *Drosophila*, where the effect is often more subtle and manifests when the organism is challenged (Wessels *et al.* 2019). Examples of key miRNA, *bantam* and *miR-14*, were identified in *drosophila* to regulate apoptosis due to null genotypes that were developmentally lethal (Brenneke *et al.* 2003; Xu *et al.* 2003). *miR-8* was shown to finetune the translation of *atrophin* and prevent neurodegeneration, a function that is believed to be conserved in mammals (Karres *et al.* 2007). Since, miRNA molecules have been implicated in almost all biological pathways in flies ranging from embryogenesis to regulation of the circadian rhythm (Kim *et al.* 2009; Carthew *et al* 2017; Xue and Zhang 2018).

The fruit fly *miR-137* gene holds an overall sequence identity of 68% with the human ortholog, while the most prevalent mature-3p miRNA sequence of "uauugcuugagaauacacguag" is 91% identical with only 2 differing residues. There are a limited number of studies into the effect of proper *miR-137* regulation in *D. melanogaster.* One of the earlier studies, conducted by Chen *et al.* (2014), performed a screening of miRNA deficient flies and found that loss of endogenous *miR-137* was accompanied by a significant decrease in climbing ability but no significant change in survival or lifespan. Later, in 2017,

Larval locomotion, specifically righting after being inverted, was not found to be impacted by *miR-137* loss (Picao-Osorio *et al.* 2017). Regarding fly health, Atilano *et al.* reported that loss of *miR-137* in flies resulted in increased susceptibility to infection of *C. albicans* and concluded that the miRNA played an important role in coordinating *Drosophila* immune response. *miR-137* is increased 3-fold in *Drosophila* epithelial tumours, however reversing the expression did not suggest a role in tumorigenesis (Shu *et al.* 2017). Comparison of the predicted *miR-137* targets show approximately 20% (278/1315) of human targets have an ortholog also regulated by miR-137 in the fly. Considering the miRNA is classed as highly evolutionarily conserved it might have been expected of a higher crossover between the two species, however the list of 283 genes include many from key pathways around neurological development, function, and maintenance. This includes calcium and potassium channels (CACNA1C/D/G/H/I, KCNAB3, KCNC1/2/3), transcription factors (TCF3/4), key genes previously mentioned in relation to psychiatric disorders (RORA, SCN2A), and more (full list of human genes and fly orthologs in supplementary material). This highly conserved gene identity and comparative regulation of key genes demonstrates *miR-137* evolutionary importance. Furthermore, this reinforces the suitability of using *D. melanogaster* as a model to investigate this largely unknown miRNAs function further.

The use of fruit flies in scientific research has pioneered many advances since their introduction in the early 1900s. Studies have found that there are multiple features of the fly which allow for comparable traits to human behaviour or genetic counterparts, across a wide range of fields from cardiovascular maintenance to transcriptional regulation of cell specification in development (Duffy 2002; St Johnston 2002; Nichols *et al.*2012; Linford *et al.* 2013). The fruit fly has proven itself particularly useful as an organism to investigate the genetics behind maintenance and development of the brain and central nervous system, with studies using the model to better understand fundamental biological processes within the

brain like synaptic plasticity, axon pathfinding, and dendritic morphology (Zhong *et al.* 1992; Gao *et al.* 1999; St Johnston 2002; Venken *et al.* 2011; Jakubowski *et al.* 2012). Fruit flies have also been used to evaluate the genetic components of memory and learning, with many established protocols for evaluating these behaviours individually (Venken *et al.* 2011; Nichols *et al.* 2012; Baggio *et al.* 2013). The fly brain is estimated to have one tenth of the neurons of a human brain, with a total of approximately 90000 neurons (Venken *et al.* 2011). Other differences between the flies and the humans CNS also include a significantly reduced count of synapses per neuron in the fly, and the primary transmitter in flies is glutamate unlike vertebrates (Venken *et al.* 2011). However, the fly brain exhibits a similar complexity regarding multiple neuron types and brain sub regions working in sync (Bullock 1978), which gives reason to suggest that research into *D. melanogaster* will yield results with relevant insight to the mechanisms underlying vertebrate neurology.

Often, as genes are linked to human illnesses, the *D. melanogaster* ortholog is investigated for a phenotypical effect. Altered expression of the human Alzheimer's disease risk gene *Glycogen synthase kinase-3* (*GSK-3*) ortholog in flies induces progressive locomotor impairment in larval stages, a hallmark symptom of many neurodegenerative diseases (Jakubowski *et al.* 2012). Expressing human genes not normally in the fly genome is another approach available for neurological research. The fruit fly has a documented "Parkinson's disease" model that expresses a foreign DNA insertion containing the human α-synuclein gene. This model develops symptoms synonymous with Parkinson's disease such as; loss of dopaminergic neurons, α-synuclein aggregates, and age-dependent locomotor defects (Shaltiel-Karyo *et al.* 2012).

1.7   Sleep

Sleep is a process which is defined as a consolidated period of rest and inactivity coupled with reduced response to external stimuli (Cirelli *et al.* 2005; Cirelli 2009; Bushey *et al.* 2010; Crocker and Sehgal 2010). Sleep or sleep-like behaviour is a recurring state in many organisms and is a unique behavioural trait that contradicts the basis of survival, temporarily reducing awareness and putting muscles in a state of inactivity. While in such a state, the organism is vulnerable to risks such as predation and therefore to be prevalent in such a diverse set of species sleep must serve an important purpose

1.7.1   Sleep Regulation

It is known that sleep is governed by genetics, though environmental factors can have an impact on the quality of a sleep episode (Cirelli 2009; Sehgal and Mignot 2011). Discovery of hereditable sleep traits led to investigating suitable model organisms for sleep research, from which some of the molecular mechanisms behind sleep have been uncovered (Cirelli 2009, Sehgal and Mignot 2011). Sleep is regulated by two independent mechanisms, circadian and homeostasis. Circadian rhythm dictates the cycle of sleep within a 24-hour period, restricting the behaviour to times that are ecologically suitable (Cirelli 2009; Crocker and Sehgal 2010). The homeostatic mechanism is reflected in the perceived increase of pressure to sleep throughout the waking period, which is reset when having slept (Crocker and Sehgal 2010; Donlea *et al.* 2013). Both mechanisms have a set of genes already associated with them, some of which overlap. Circadian genes interact and determine circadian rhythmicity, mutations of which result in phenotypes such as abnormal sleep duration but no difference in sleep pattern following deprivation (Cirelli 2009).

Some genes' level of expression changes in the brain between sleep and waking states, lending further evidence to a molecular function of sleep (Cirelli 2009). This molecular function is likely to be conserved across many organisms, as has been proven with other genetic pathways. Sleep-like behaviour has been documented in widely used model organisms such as mice, zebrafish, *D. melanogaster*, and even worms (Crocker and Sehgal 2010; Bidaki *et al.* 2011; Sehgal and Mignot 2011). In mammalian and avian species, sleep can be recognised through criteria detected on an electroencephalogram (EEG) which cannot easily be done for many other species (Crocker and Sehgal 2010). Many sleep disorders, such as insomnia and narcolepsy, have had some underlying genetic mechanisms exposed through genome analysis and model organism research (Bidaki *et al.* 2011; Sehgal and Mignot 2011). Some disorders are also found to be caused by singular gene mutations such as chronic primary insomnia and sleep-phase syndrome. Many neurological disorders also have sleep variations reported as a symptom (Lane *et al.* 2017). While individual genes have been statistically or experimentally linked to the onset of sleep disorders, not all have had the molecular mechanisms behind their involvement exposed.

The underlying process of sleep being governed by genetics but heavily influenced by external factors implies that the purpose of sleep is simply to recover from a bout of activity and prolonged consciousness. However, the exact purpose is unclear.

Genetically and chemically, sleep and the abundance of certain chemicals or gene expression can be correlated. Furthermore, there are sleep promoting drugs which act as sedatives and other substances which act as stimulants that temporarily help reduce sleep pressure.

There are several medical disorders which have sleep dysfunction as either a cause or symptom. Many psychiatric disorders such as SZ and BD exhibit differences in sleep

behaviour than would be expected of "normal" sleep. Alongside differences in sleep/wake cycle, symptoms such as catatonia are comparable to being sleepy and fatigued. It is also possible that dysfunctional sleep can lead to the deterioration of proper neurological functions, potentially describing a cause of the neurological disorders rather than symptom.

Sleep dysregulation can cause a multitude of problems for the function of the body. Simply delaying sleep results in immediate negative impacts on performance and health (Bidaki *et al.* 2011), followed with a rebound effect of increased sleep duration in the next sleep period. Sleep issues have been linked to poor memory retention (Tomita *et al.* 2015), and abnormal total sleep amount has been linked to a reduced lifespan (Bushey *et al.* 2010). Disordered sleep may also play an important role in the aetiology and maintenance of physical and mental health (Lane et al. 2017). Links between fractured or curtailed sleep behaviour and type 2 diabetes are demonstrated by a rise in blood-glucose levels in individuals chronically deprived of sleep (Spiegel *et al.* 2005; Knutson and Cauter 2008). Sleep curtailment has also been linked to dysregulation of appetite, higher body mass, and a decrease in energy expenditure (Spiegel *et al.* 2005; Knutson and Caulter 2008). Chronic sleep deprivation has even been statistically correlated with increased suicide ideation in teenagers (Whitmore and Smith 2018).

1.7.2 Sleep in Drosophila melanogaster

While physiologically very different to humans, the fruit fly has a sleep state which demonstrates some clear parallels to human sleep. In flies, the documented sleep state is classified as a period of quiescence for 5 minutes or longer (Cirelli *et al.* 2005; Cirelli 2009; Bushey *et al.* 2010). Numerous studies have identified similarities between the human and *D. melanogaster* sleep states, including both behavioural features like increased arousal

thresholds and molecular mechanisms like Dopamine signalling (Shaw *et al.* 2000; Bushey *et al.* 2010; Tomita *et al.* 2015). If deprived of sleep through continuous stimuli, *D. melanogaster* exhibit decreased vigilance and upon being allowed to rest, display a sleep-rebound effect of a longer and less fragmented sleep following sleep deprivation (Shaw *et al.* 2000; Bushey *et al.* 2010). Another similarity with humans is that fruit flies are also predominantly diurnal, with most of the rest periods being consolidated in the night/lights off period (Bushey *et al.* 2010; Liu *et al.* 2015).

Circadian rhythms and the regulating clock neurons which dictate the daily oscillations of most physiological and behavioural processes have been widely researched in the *D. melanogaster* model. The circadian rhythm or biological clock enables an organism to anticipate changes in environmental factors such as lights on or sunrise (Renn *et al.* 1999). The standard activity graph for flies in controlled conditions over a 24-hour period includes clear pre-emptive behaviour of increased activity before the programmed lights on or lights off timing (Cirelli *et al.* 2005; Cirelli 2009; Bushey *et al.* 2010). In mammals, the pacemaker for daily activity cycles is in the hypothalamus while the fly neurons regulating this cycle are associated with the brain's visual centres (Renn *et al.* 1999). The fly brain contains approximately 150 dedicated clock neurons, divided into distinct subsets by function and anatomical location (Cavanaugh *et al.* 2014). Of these, the small lateroventral neurons appear to be one of the most dominant regulators of the sleep/wake cycle as loss of these neurons result in arrhythmic behaviour patterns in constant dark conditions, yet ablation of all but these neurons can allow a robust activity pattern even in total darkness (Renn *et al.* 1999; Cavanaugh *et al.* 2014). The neuropeptide *pigment-dispersing factor* (*pdf*) is expressed within most of the lateroventral neurons, and is implicated as the principal circadian transmitter, with flies deficient in this gene expression showing entrainment to the light/dark cycle but losing regular patterns in constant darkness (Renn *et al.* 1999). The requirement of *rho* for

sleep is localized to the pars intercerebralis part of the fly brain that is developmentally and functionally analogous to the hypothalamus in vertebrates (foltenyi k 2007; Cavanaugh *et al.* 2014). This region is also proven to be responsible for the *D. melanogaster* ortholog of human *corticotrophin releasing factor* regulation of proper sleep/wake cycles, dysfunction of which results in arrhythmicity (Cavanaugh *et al.* 2014).

The alternative, and less understood, component of sleep regulation is the sleep homeostat. This is responsible for creating an increased need to sleep in response to stimuli or prolonged periods of wakefulness, and the cause for the sleep rebound effect seen after sleep-deprivation (Donlea *et al.* 2014). This function in *D. melanogaster* has been linked to fan-body region of the brain, which is similar in function to the mammalian hypothalamic ventrolateral preoptic nuclei (Donlea *et al.* 2014). The fan-body is modulated by dopaminergic neurons, and artificial stimulation induces sleep on command (Lui *et al.* 2012).

Sleep in both fruit flies and humans has been linked to longevity, with numerous genetic mutations in *D. melanogaster* affecting sleep amount often having a negative effect on average lifespan (Bushey *et al.* 2010).

1.8   *Hypotheses, Aims and Objectives*

The studies contained within this thesis investigated the functional role of *miR-137* in the *Drosophila melanogaster* nervous system and used bioinformatic analyses of publicly available human genomic data to suggest the influence this micro-RNA might play in contributing to an increased risk of developing SZ. The *miR-137* gene was selected for this study based upon an earlier GWAS by Ripke et al. (2013) that identified micro-RNA as statistically associated (p= 1.6×10^{-11}) in a sample size of (n SZ= 9,394, n controls = 12,642). Given that GWAS have no a priori hypothesis, any gene found to be statistically associated

must be validated through a biological experiment. *miR-137* was selected as the candidate gene of choice for this investigation because it was 1) statistically associated in a GWAS, 2) biologically plausible and 3) central to cellular regulation. Although there is only a limited number of published psychiatric genetic studies involving *D. melanogaster*, the fruit fly was chosen as the model organism because *miR-137* has an evolutionary conserved ortholog with 92% mature sequence homology, a rapid-life cycle, cheap-to-run experiments and acts as a solid 'first-pass' compared to stem cell studies (e.g. Hill et al. 2014).

Initially, a knockout strain of miR-137 was sourced and tested to measure phenotypes comparable with schizophrenia. Subsequently, a bioinformatics approach was used to determine putative molecular pathways.

Aim:

The aim of this work was to better understand the role of *miR-137* in regulating vital processes for development using *Drosophila melanogaster* as a model organism and relate this to phenotypic and putative molecular processes involved in schizophrenia pathogenesis.

Objective A:

Establish the likely importance of a loss of function of *miR-137* on the development and health in *Drosophila melanogaster* (See Chapter 4).

Hypothesis 1:

Due to the published importance of *miR-137* in development and neurological function, loss of genomic expression will have a negative impact on average lifespan in *Drosophila melanogaster.*

H_0: Loss of genomic *miR-137* expression does not negatively affect lifespan.

H_1: Loss of genomic *miR-137* expression negatively affects lifespan.

Hypothesis 2:

Due to the published importance of *miR-137* in development and neurological function, loss of genomic expression will slow development rate in *Drosophila melanogaster.*

H_0: Loss of genomic *miR-137* expression does not slow development rate

H_1: Loss of genomic *miR-137* expression slows development rate

Hypothesis 3:

Due to the published importance of *miR-137* in development and neurological function, loss of genomic expression will increase development lethality in *Drosophila melanogaster.*

H_0: Loss of genomic *miR-137* expression does not increase development lethality

H_1: Loss of genomic *miR-137* expression increases development lethality

Objective B:

Establish the likely importance of a loss of function of *miR-137* on the behaviour and physical ability of *Drosophila melanogaster* (See Chapter 5).

Hypothesis 1:

Due to published neurological importance of *miR-137* and the data from the locomotor screen by Chen *et al.* (2014), loss of genomic expression will decrease overall activity in *Drosophila melanogaster*.

H_0: Loss of *miR-137* does not decrease overall activity.

H_1: Loss of *miR-137* decreases overall activity.

Hypothesis 2:

Due to published neurological importance of *miR-137* and the data from the locomotor screen by Chen *et al.* (2014), loss of genomic expression will decrease movement velocity in *Drosophila melanogaster*.

H_0: Loss of *miR-137* does not decrease movement velocity.

H_1: Loss of *miR-137* decreases movement velocity.

Hypothesis 3:

Due to published data by Chen *et al.* (2014), loss of genomic *miR-137* expression will decrease locomotor ability in *Drosophila melanogaster*.

H_0: Loss of *miR-137* does not decrease locomotor ability.

H_1: Loss of *miR-137* decreases locomotor ability.

Hypothesis 4:

Due to the published neurological importance of *miR-137*, loss of genomic expression will increase total sleep amount in *Drosophila melanogaster*.

H_0: Loss of *miR-137* does not increase sleep amount.

H_1: Loss of *miR-137* increases sleep amount.

Objective C:

Establish the ability to phenocopy the increased sleep phenotype of the *miR-137^{KO}* *Drosophila melanogaster* strain (See Chapter 6).

Hypothesis 1:

Due to the sleep phenotype exhibited in the *miR-137^{KO}* strain, knockout of *miR-137* expression through the heterozygous crossing of another *miR-137* null strain and the *miR-137^{KO}* strain increases sleep amount in *Drosophila melanogaster*.

H_0: Heterozygous knockout of *miR-137* does not increase sleep amount.

H_1: Heterozygous knockout of *miR-137* increases sleep amount.

Hypothesis 2:

Due to the published importance and expression location of *miR-137,* plus the phenotype exhibited in the *miR-137^{KO}* strain, targeted genetic knockdown of *miR-137* expression in the central nervous system increases sleep amount in *Drosophila melanogaster*.

H_0: Knockdown of *miR-137* in the central nervous system does not increase sleep amount.

H_1: Knockdown of *miR-137* in the central nervous system increases sleep amount.

Hypothesis 3:

Due to the published importance and expression location of *miR-137,* plus the phenotype exhibited in the *miR-137^{KO}* strain, targeted genetic knockdown of *miR-137* expression in the fan-shaped body (brain subregion) increases sleep amount in *Drosophila melanogaster*.

H_0: Knockdown of *miR-137* in the fan-shaped body does not increase sleep amount.

H_1: Knockdown of *miR-137* in the fan-shaped body increases sleep amount.

Hypothesis 4:

Due to the published importance and expression location of *miR-137,* plus the phenotype exhibited in the *miR-137^{KO}* strain, targeted genetic knockdown of *miR-137* expression in the mushroom bodies (brain subregion) increases sleep amount in *Drosophila melanogaster*.

H_0: Knockdown of *miR-137* in the mushroom bodies does not increase sleep amount.

H_1: Knockdown of *miR-137* in the mushroom bodies increases sleep amount.

<u>Objective D:</u>

Establish the ability to recover the increased sleep phenotype of the *miR-137^{KO}*

Drosophila melanogaster strain (See Chapter 7).

<u>Hypothesis 1:</u>

Due to the published importance and expression location of *miR-137,* targeted

reintroduction of *miR-137* expression in the central nervous system reduces sleep amount in

miR-137^{KO} Drosophila melanogaster.

H$_0$: Reintroduction of *miR-137* in the central nervous system does not reduce sleep

amount.

H$_1$: Reintroduction of *miR-137* in the central nervous system reduces sleep amount

<u>Hypothesis 2:</u>

Due to the published importance and expression location of *miR-137,* targeted

reintroduction of *miR-137* expression in the fan-shaped body (brain subregion) reduces sleep

amount in *miR-137^{KO} Drosophila melanogaster.*

H$_0$: Reintroduction of *miR-137* in the fan-shaped body does not reduce sleep amount.

H$_1$: Reintroduction of *miR-137* in the fan-shaped body reduces sleep amount.

<u>Hypothesis 3:</u>

Due to the published sleep-disrupting effect of caffeine, chronic treatment reduces sleep

amount in *miR-137^{KO} Drosophila melanogaster.*

H$_0$: Treatment with caffeine does not reduce sleep amount.

H$_1$: Treatment with caffeine reduces sleep amount.

Hypothesis 4:

Due to the likelihood of dopamine sensitivity from loss of *miR-137* regulation and the use of 3-iodo-tyrosine to inhibit dopamine, treatment reduces sleep amount in *miR-137^{KO}* *Drosophila melanogaster*.

H_0: Treatment with 3-iodo-tyrosine does not reduce sleep amount.

H_1: Treatment with 3-iodo-tyrosine reduces sleep amount.

Hypothesis 5:

Due to the extensive use of lithium chloride to treat schizophrenia and the *miR-137* and sleep link to neurological disorders, treatment reduces sleep amount in *miR-137^{KO}* *Drosophila melanogaster*.

H_0: Treatment with lithium chloride does not reduce sleep amount.

H_1: Treatment with lithium chloride reduces sleep amount.

Objective E:

Establish the *miR-137^{KO}* exome change similarities between *Drosophila melanogaster* expression data and *Homo sapiens* expression data (See Chapter 8).

Hypothesis 1:

Due to the published evolutionarily conserved status of the *miR-137* gene sequence, there will be a high proportion of differentially expressed genes found in *Drosophila melanogaster* and in *Homo sapiens*.

H_0: There will not be a high proportion of matching differentially expressed genes.

H_1: There will be a high proportion of matching differentially expressed genes.

Objective F:

Establish the similarities between differentially expressed genes in *Drosophila melanogaster miR-137^{KO}* exome and human genome wide association study results (See Chapter 8).

Hypothesis 1:

Due to the published evolutionarily conserved status of the *miR-137* gene sequence and the phenotype exhibited by the *miR-137^{KO}* strain, there will be a high proportion of differentially expressed genes found in *Drosophila melanogaster* and in human sleep related genome wide association studies from the UKBiobank.

H_0: There will not be a high proportion of matching differentially expressed genes.

H_1: There will be a high proportion of matching differentially expressed genes.

Hypothesis 2:

Due to the published evolutionarily conserved status of the *miR-137* gene sequence and the original genome wide association study link to schizophrenia (Ripke *et al.* 2013), there will be a high proportion of differentially expressed genes found in *Drosophila melanogaster* and in the human schizophrenia-related genome wide association study by Ripke *et al.* (2018)

H_0: There will not be a high proportion of matching differentially expressed genes.

H_1: There will be a high proportion of matching differentially expressed genes.

Hypothesis 3:

Due to the published evolutionarily conserved status of the *miR-137* gene sequence and the phenotype exhibited by the *miR-137^{KO}* strain, there will be a high proportion of differentially expressed genes found in *Drosophila melanogaster* and in the human major depressive disorder genome wide association study by Pardinas *et al.* (2016).

H_0: There will not be a high proportion of matching differentially expressed genes.

H_1: There will be a high proportion of matching differentially expressed genes

2   Methods

2.1   Fly Stocks

Drosophila melanogaster strains were ordered online from the Bloomington Drosophila stock centre (BDSC) or the Vienna Drosophila resource centre (VDRC). Stocks at these institutes are donated by researchers worldwide and maintained/distributed to scientists for research or teaching purposes. Some stocks were already held on site from previous research by Dr Paul Hartley (Bournemouth University).

2.2   Fly rearing and cultivation

Drosophila strains were typically raised on a yeast cornmeal food blend made as in section 2.1.1. Stocks that were not in current use were kept at 18°C; while working stocks, crosses and experiments were kept at 25°C.

Routine stock maintenance was performed by flipping flies onto fresh food when required (typically once every 2 weeks at 25°C, and once every 6-8 weeks at 18°C) and visually observing general health. In the event of a failing stock, fresh food was supplied along with additional live yeast which provided additional nutrients to help boost health and stock proliferation. If this failed and the stock was down to minimal numbers, they were moved to a food which contained a high concentration of molasses and additional live yeast within the 25°C incubator.

2.2.1   Fly food

The food was made from a cornmeal, yeast, agar mix in batches made as described below. Each batch supplied approximately 200 vials, some of which was preserved in sealed glass bottles and reheated/poured when required. The vials were sealed with a cotton wool plug and stored in a plastic bag at room temperature until used for fly cultivation.

1. Mix 100g sucrose, 12g agar, and 700mL of distilled water. Heat until dissolved.

2. In a separate container, mix 35g active yeast, 102.2g cornmeal, and 14g soy flour. Add distilled water up to a total volume of 700mL.

3. Combine the two mixtures and heat in a microwave for 5x2 minute periods, with stirring in between.

4. Allow to cool slightly before adding 5mL propionic acid and 3.8g of Nipagin dissolved in 40mL ethanol.

5. Dispense into vials of approximately 8mL or store in airtight containers until needed.

2.2.2   Crosses

Genetic crosses were carried out to create specific genotypes for experimentation, such as the UAS/GAL4 system and deficiency crosses. These utilised phenotypic markers to properly identify the correct offspring. In this study, most of the markers were curly wing (CyO), stubble (Sb), tubby (Tb), white/red eyes, and scutoid (Scu) (figure 2.1). Crosses were set up with a mix of males and females to provide the correct genotype in the f1 offspring and were moved to a new vial within 4/5 days to prevent parent and progeny adults mixing.
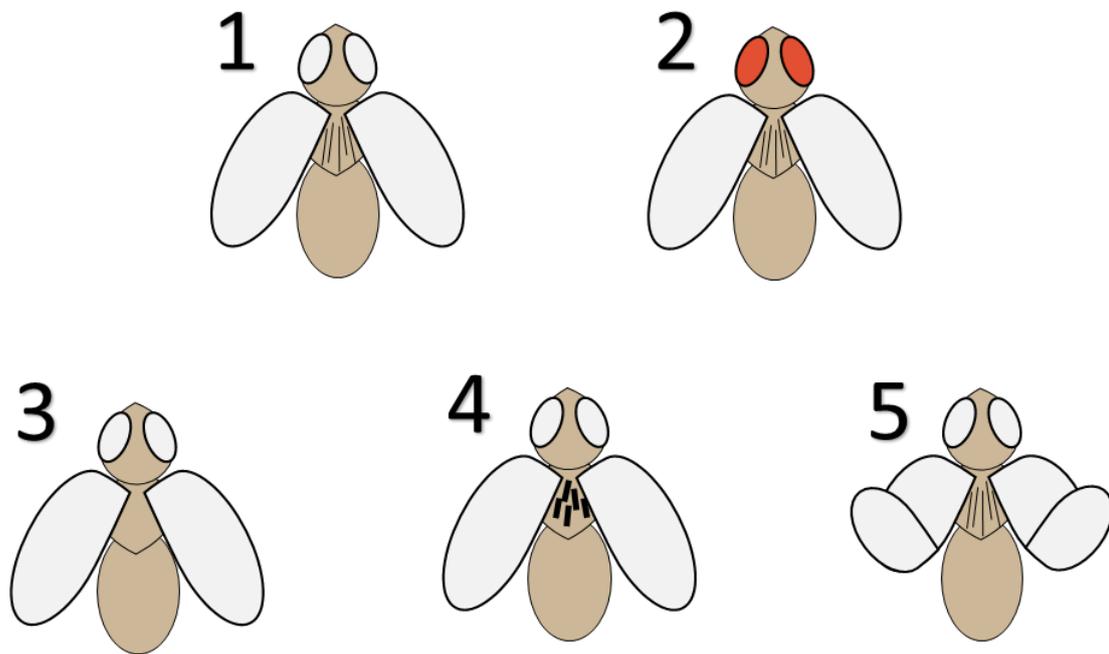
Figure 2.1: Frequently used *Drosophila melanogaster* genetic markers.

1) w^{1118} control strain with double recessive w⁻ mutation resulting in a white eye phenotype. 2) True "wild type" red eye phenotype, often used in w^{1118} background to show the presence of a mutation. 3) Scutoid: Hairs on the body are not present. 4) Stubble: Hairs on the body are shorter and thicker. 5) Curly: The tips of the wings curl upwards.

The selection of female virgins must be done within a few hours of eclosion as the markers become more difficult to identify after a longer period. Virgins are selected with noticeably pale and soft bodies, furled wings, and the meconium present. This is the most important marker for virginity and is a black spot seen in the abdomen. The differentiation of male and female flies was done by observing the obvious genitalia differences at the end of the body.

2.3   DNA extraction

To begin DNA analysis, the DNA must be removed from the organism. DNA is a fairly robust structure, and there are many available protocols for successful extraction. For this *Drosophila* work, a phenol-chloroform protocol was used to extract the DNA and separate it from proteins and other cell debris. The method outlined in http://www.protocol-online.org/prot/Protocols/Isolation-of-micro-RNA--miRNA--3342.html was used. While there are several commercially available kits or reagents to easily extract DNA, this method was chosen as it was both able to separate small RNA molecules along with the DNA and cost effective. As the project orientates around microRNA (~22bp) the ability to effectively collect small RNA molecules was imperative. The protocol utilises the solubility of DNA in a polar solvent, and the protein denaturing action of phenol to separate and protect the nucleic acid. Upon introducing phenol-chloroform to a lysed (broken up) biological sample, the nucleic acid dissolves in the water component while the proteins are denatured and soluble within the phenol environment. Following a spin in the centrifuge, the two layers clearly separate with the nucleic acid in aqueous solution that can be carefully pipetted out. The nucleic acid solution is then added to pure ethanol or isopropanol which changes the solutions polarity and forces the DNA and RNA out of solution, this process can have better results if performed on ice or including a brief incubation period in -20°C.

Protocol:

1. Collect sample

2. Homogenize sample in 100μl denaturing solution (4M Guanidine Isothiocyanate (GITC), 0.02M Sodium citrate, 0.5% sarcosyl) in a 200μl ependorf.

3. Add 10μl D of 2 M sodium acetate.

4. Add 100μl of phenol:chloroform:isoamyl alcohol (25:24:1). Cap and vortex thoroughly.

5. Spin the sample in a microcentrifuge at 12500 rpm for 5 minutes

6. Transfer top aqueous layer to a tube containing 100μl of isopropanol. Mix 1μl of glycogen with the nucleic acid solution before adding the isopropanol. Leave at -20°C for 30 minutes.

7. Spin the sample in a microcentrifuge at 12500 rpm for 5 minutes

8. Remove supernatant completely

9. Wash pellet in 2x 200μl of 75% ethanol–25% DEPC-treated water

10. Remove supernatant completely

11. Dry for 5 minutes under a fume hood

12. Resuspend pellet in DEPC-treated water (volume dependent upon concentration required)

Initially samples of whole flies were used and homogenised by crushing in denaturing solution within the Eppendorf. For PCR genotyping; samples of *miR-137^{KO}, w^{1118}*, and heterozygous (*w^{1118}/miR-137^{KO}*) were used to provide a complete result. In this setup it was predicted that primer sets for both *miR-137* and *w$^{+mW.hs}$* would bind and produce a result in the heterozygous knockout flies as they carried both DNA regions, while primer specificity could be tested by presence-absence of relevant primer sets in the homozygous stocks.

2.4   PCR Genotyping

To ascertain the presence or absence of a specific gene, the region must be amplified and run through a gel to visualize the DNA. To do so, the DNA extracted needed to be quantified and then undergo polymerase Chain reaction in a thermocycler to exponentially amplify a desired region to the concentrations which can be visually shown on an electrophoresis gel.

2.4.1   DNA quantification

Following DNA extraction and purification, the quantity of nucleic acid was measured to ascertain the procedure was successful. Using a Thermofisher Nanodrop 2000 machine, the nucleic acid concentration and purity can be measured using 1µl of PCR product per measurement. The machine utilises spectrophotometer readings to calculate the absorbance of certain molecules, which is then used to ascertain concentration and a 260/280nm wavelength ratio useful for judging purity. The results from a Nanodrop reading is a wavelength/absorbance graph, alongside a results table for DNA concentration (ng/µl), 260nm, and 280nm wavelength readings. A 260/280 ratio of ~1.8 is deemed as "pure" for DNA, while a measurement of ~2.0 is "pure" RNA. Ratios of significant difference usually indicates that contaminants such as phenol or proteins are present, and additional steps taken to purify if contaminants will affect downstream procedures.

2.4.2   Primer Design

To provide evidence of the genetic substitution, PCR genotyping through selected DNA region multiplication and gel electrophoresis was used for presence and absence analysis. Several sets of primers were designed to bind to the extended gene region of the endogenous *miR-137*, and some were made to bind to parts of the expected insertion (figure 2.2).
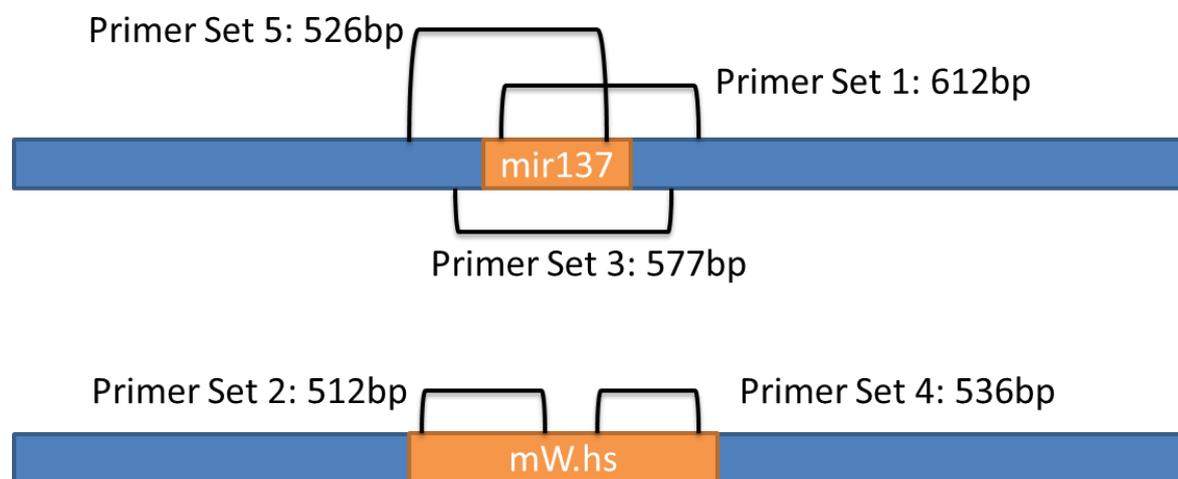
Figure 2.2: Regions of predicted DNA amplification with designed primer sets.

Primers were designed for use in PCR to bind to regions of genomic *miR-137* and the $w^{+mW.hs}$ insertion.

Primers are small sequences of 18-22 nucleotides which are specifically designed to be complementary to parts of the forward and reverse strands of a DNA sequence. They work as part of the polymerase chain reaction (PCR) protocol for amplifying regions of DNA. During the PCR process, the primers bind to the specific target regions and are then used to kick start an extension process which creates a complementary strand of the target region between the primers.

Correct primer design also has other important factors, such as melting temperature and nucleotide frequency. The sequence for the extended gene region of *D. melanogaster miR-137* was taken from FlyBASE (Gene ID: FBgn0262446), while the sequence for the synthetic $w^{+mW.hs}$ insertion was obtained from Klemenz *et al*. (1987). These sequences were then used to design the complementary primer sets. Initially 5 primer sets were made, 2 for the *miR-137* region, 2 for the $w^{+mW.hs}$ insert, and one that may have spanned the whole insertion (figure 2.2). Primer pairs were designed to span regions of approximately 5-600 base pairs (bp) so the amplificated DNA sequences were not too long which also kept the PCR cycles shorter (table 2.1). The primers were selected to be 18-22 nucleotides in length while also having a suitable melting and annealing temperature. They were also designed to include more G or C nucleotides towards the help with annealing stability. Primers also had to be designed to run 5' to 3' orientation, so the reverse pair had to be the reverse complement of the reported sequence.

The T_m or melting temperature was predicted using an online programme available at: http://www6.appliedbiosystems.com/support/techtools/calc/. While the reverse complement was computed via http://reverse-complement.com/.

Table 2.1: Primer pair sequences and expected amplicon length

Primer	Forward sequence	Reverse sequence	Amplicon length
1	GCCACGTGTATGCTCGTAG	TGGTTCAAACTACAACTGACAG	612bp
2	TAACATAAGGTGGTCCCG	TTCATTAAACAATGAACAGGAC	512bp
3	TTATAGTGCAATTAATTCTGCG	TTCCAGCGTAGAGCAAATG	577bp
4	AGACGACCCTGCTGAATG	CATCAGAAGGATCTTGTCAAAG	536bp
5	TGTACAGTTATGGCAGCAGG	ACAAATCTCGGTGAACTACG	526bp

2.4.3   PCR machine

Once DNA has been extracted and the relevant primers have been designed, they are used within a PCR machine. The machine is a thermocycler, which can be used to change the temperature of a solution throughout a programmed cycle. The programme is set up to mimic optimum conditions for the relevant reagents to work in cycles that ultimately result in an exponential amplification of the desired region (amplicon). Along with the DNA sample (template) and designed primers mentioned previously; Taq polymerase, buffers, free nucleotides (DNTPs), and magnesium chloride is added to a 200μl PCR tube (table 2.2).

Table 2.2: Reagent and volume mix for PCR reaction mixture.

Reagent	Volume per 10μL reaction (μL)
DNTPs (10mM)	0.2
Taq Enzyme	0.1
Taq Buffer (10x)	1
$MgCl_2$ (25mM)	0.6
Reverse Primer	0.2
Forward Primer	0.2
DNA Template	1.5
H_2O	6.1
Total	10

Taq polymerase is a thermostable DNA polymerase enzyme extracted from the bacteria *Thermus aquaticus*, that operates at relatively high temperatures of 72°C. The enzyme binds to the primers following annealing and uses the free nucleotides to extend the complementary sequence, ultimately creating a complementary strand of the template. Following the extension period, the mixture heats up and the strands come apart allowing the next cycle of DNA amplification (table 2.3). The cycle was repeated 35x, which if successful can produce over a billion copies of the amplicon.

Step	Temperature (°C)	Time	Purpose
Start	94	2 min	TAQ Polymerase activation
Cycle i	94	15 sec	Separation of DNA strands
Cycle ii	56	15 sec	Annealing of primers and enzyme
Cycle iii	72	30 sec	Extension of primers by enzyme
Finish	72	5 min	Final extension period

2.4.4 Gel electrophoresis

For electrophoresis, gels were made and tested at 1%, 1.5%, and 2% for best results. As the DNA fragments are quite short (~600bp) the 2% gel provided the best separation when run at 80 Volts for 60 minutes. For a 2% gel, 0.8gms of agarose was added to 40mL of 1x TAE buffer in a 100mL conical flask, and then heated in a microwave for approximately 2x 30 seconds. SybrSafe was added at a concentration of 1uL per 10mL of gel, once the mix had cooled to approximately 50°C. The gel was poured into an electrophoresis tray and a desired well-comb was put in, before letting it cool completely. While cooling, the PCR product has 20% volume of 5x loading dye added. Once hardened, the gel is submersed in 1x TAE buffer and 5uL of sample/dye mix was added to each well. In the first well a 100bp ladder was used (100bp to 1kbp in 100bp increments). Following an hour electrophoresis, the gel was removed and put on a UV light viewing plate and photographed. In later experiments the gel was imaged using a Bio-Rad Chemidoc MP Imager.

2.5   Gene Expression

Gene expression amount can be assessed through the concentration of RNA found within a sample. Following RNA extraction, the RNA must be protected by converting into a cDNA via Reverse transcription (RT). Once the RNA is stabilised, its comparative expression amount can be measured utilising the specific Taqman microRNA expression assays by Applied Biosystems in a qPCR machine.

2.5.1   Reverse Transcription

For gene expression analysis, fluorescent quantitative polymerase-chain-reaction (qPCR) techniques are used to assess the level of target messenger RNA (mRNA) present in the template sample. Before the expression can be measured, extraction of total RNA and conversion into double stranded complimentary-DNA (cDNA) through reverse transcription (RT). These cDNA molecules are more resilient and degrade slower which makes them better for storing while all samples are accumulated for a qPCR run. RT requires a suitable reverse-transcriptase, a purified RNA sample, and a three-step programme in a thermocycler (tables 2.4 and 2.5). The initial incubation at 16°C enables the binding and annealing of the RNA, enzyme, and primers. The 42°C incubation period is the working temperature for reverse transcription enzyme to extend the primer tail and make cDNA, following this phase a short time at a high temperature to deactivate the enzyme. Following RT, the samples are stored at -20°C until ready to be used. To protect sample integrity, these were usually stored for a maximum of 2 weeks.

For miRNA molecules, the small size of around 20bp complicates the procedures and prevents conventional primers from converting and/or amplifying the target. For standard mRNA reverse transcription, a primer that binds to the poly-A tail of mRNA is used to convert all available mRNA into a cDNA library. However, miRNA do not have poly-A tails, therefore Stem-loop RT primers must be used to circumvent these problems and enable successful qPCR. The stem-loop RT primers contain a complimentary region to the target mature miRNA at the end region, attached to a synthetic stem-loop construct. The 5' 6nt of the primer construct binds to the 3' 6nt of the mature miRNA and the RT enzyme during the incubation period. Then in the extension phase the RT enzyme extends the primer tail with nucleotides complimentary to the mature miRNA sequence, resulting in a double stranded cDNA molecule that can be used as normal in qPCR (figure 2.3).
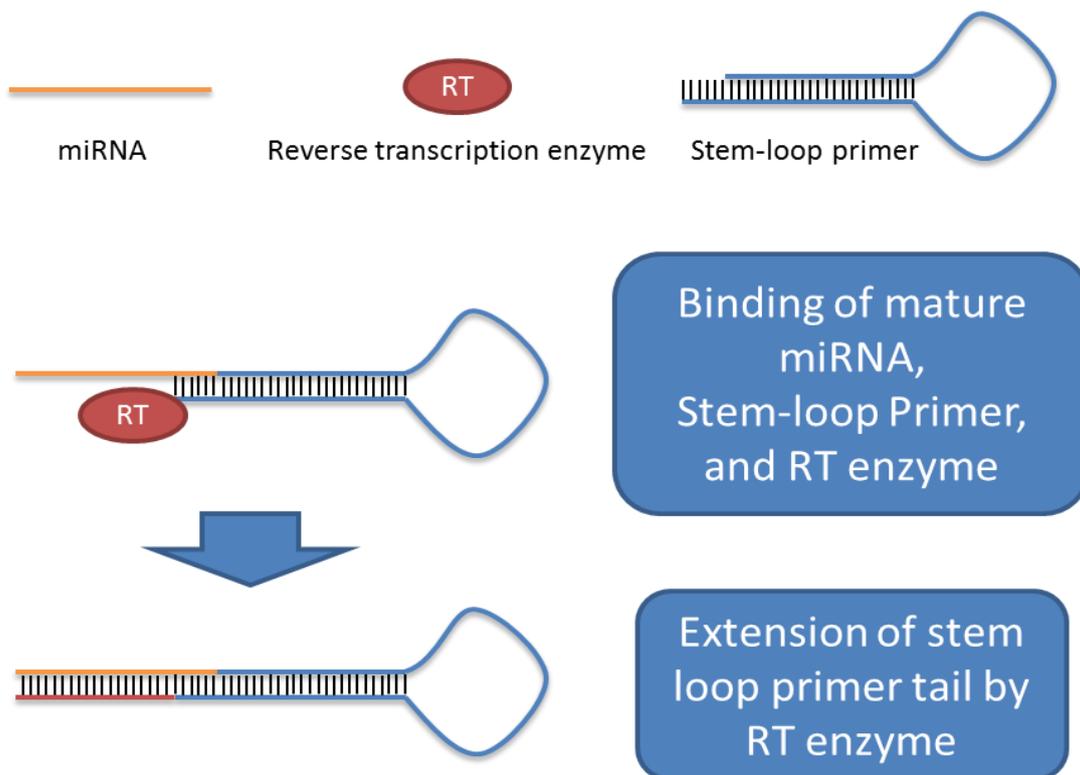
Figure 2.3: MiRNA Stem loop primer reverse transcription.

The mature miRNA molecule is 18-25 nucleotides in length, and therefore a standard primer cannot be used. To combat this, the stem loop primer has a 'sticky end' which binds to the last few nucleotides and then a reverse transcription enzyme completes the strand resulting in a hairpin loop integrating the miRNA molecule.

For this experiment, RT was done by splitting the same biological sample into two. One set used the hairpin-loop primers for converting *miR-137*, and another was used to make a total cDNA library with Poly-A primers for *rps17* analysis.

Table 2.4: Reverse Transcription thermocycler programme.

Step	Temperature (°C)	Time (Min)
1	16	30
2	42	30
3	85	5

Table 2.5: Reagent mixture used for reverse transcription

Reagent	for *miR-137*	for *RPS17*
DNTPs (10mM)	0.15	0.15
Taq Enzyme	1	1
Taq Buffer (10x)	1.5	1.5
Primer	3	1.5
Template	max within 5µL	
H2O	To 15µL total	

2.5.2   Comparative QPCR

Following RT, samples are loaded into a qPCR machine (Applied Biosystems Step-One) with specific primer sets and fluorescent probes and quenchers (table 2.6). The qPCR machine amplifies the desired cDNA regions exponentially for 30-40 cycles, and as the sequences are multiplied more fluorescence is given off by the binding and release of fluorescent markers (table 2.7, figure 2.4). This increase in fluorescence is detected by the machine for each sample well and calculated against a standard curve for a comparative value of target mRNA expression. The standard curve is made using a known dilution factor and a

mixture of all samples or an mRNA with a known starting concentration. Within each qPCR run, samples are run with duplicate or triplicates for test mRNA expression and a control mRNA expression. The control mRNA is usually a housekeeping gene which is unlikely to have changed in expression across the different samples and is used to normalize test mRNA measurements in case of accidental differences in starting template cDNA concentrations.

Table 2.6: Qualitative polymerase chain reaction reagent mix

Reagent	Volume (μL)
RT Product	2
Master Mix (2x)	10
Taqman Assay (20x)	1
H2O	7
Total	**20**

To analyse the expression of *miR-137* within *D. melanogaster* heads, ribosomal protein *rps17* was used as a control gene. Included on each qPCR plate were blanks for each primer set to check for of contamination and background fluorescence. Tests were run in duplicate, with 5 different genotypes in 48-well plates set up as shown in figure 2.5. The standard curves were made using an equal mixture of all genotypes in the experiment, and then diluted with a factor of 1:2 for 6 points. Genotype samples were all diluted 1:5 prior to being used in the qPCR mix, ensuring they were within the middle region of the standard curve and providing sufficient dilution to minimise contaminants.

Table 2.7: Quantitative PCR thermocycler programme steps

Step	Temperature (°C)	Time
1	95	10 min
2	95	15 seconds
3	60	60 seconds
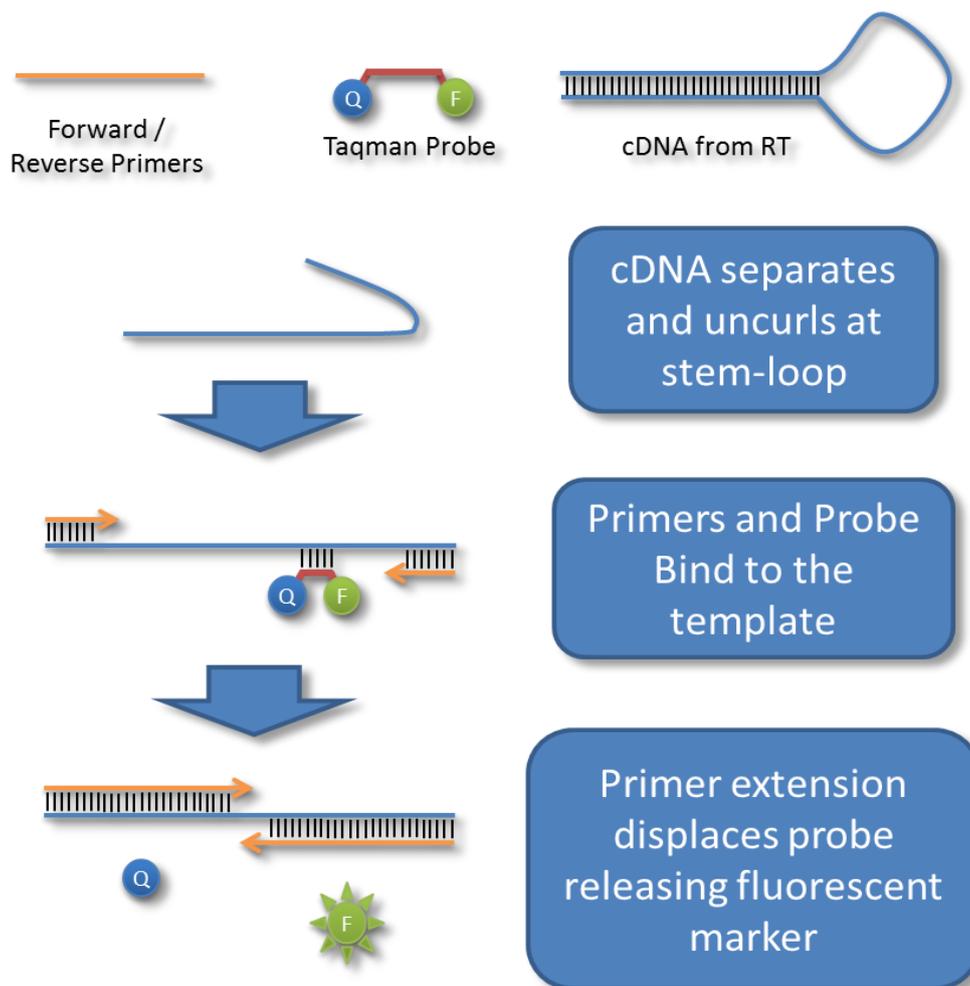
Figure 2.4: MiRNA quantitative PCR process.

The cDNA created from using a hairpin-loop primer with mature miRNA molecule uncurls as part of the melting part of the PCR cycle. In the annealing phase, a probe and a primer will bind to the single strand of cDNA. During the extension phase, RT will extend the primer tails and displace the probe, releasing a fluorescent marker that is measured by the PCR machine.

Figure 2.5: qPCR 48-well plate map.

Comparative qPCR of five genotype samples (blue "U") were run in duplicate and included assays for miR-137 and for rps17, a housekeeping gene. Two control blanks for each gene assay were included (grey "N") to check for background fluorescence. A standard curve for each gene was made using a mix of all genotype samples in the experiment, diluted for 6 points with a consecutive dilution ratio of 1:2 (orange "S").

2.6   *General Health of miR-137^{KO} Drosophila*

To begin investigation of the function of *miR-137* a set of general health assays were conducted to identify any obvious phenotype effect as a result of *miR-137* LOF.

2.6.1   *Lifespan*

The lifespan assay recorded the number of male and female flies at the moment of eclosion and note when the flies died. To start the experiment, flies that had eclosed. The flies were moved to new vials every 4-6 days to prevent new offspring mixing into the adults. The results were recorded and analysed using Microsoft Excel.

2.6.2   *Development lethality and timing*

To easily count individuals, flies were set up in culture pots on a fruit juice agar plate (see section 2.6.2.1). Culture pots were inverted stock bottles, with enough small holes punched into it to enable air flow. 14 females and 10 males were added to each culture pot, and placed on a grape agar plate at 25°C. In addition to the agar plate, the flies were fed a yeast paste made from a roughly 7:9 ratio of yeast to distilled water. Initially the cultures were left to feed and acclimatise for 3 days, and then fresh plates and yeast paste were used. To collect eggs, the plates were left for 2 hours before being changed again. Once eggs had been laid on the agar, they were counted and transferred to another agar plate. Then they were viewed each day to count the number of individuals which progressed to the next stage and when. The next stage was often transferred to a new plate and supplied fresh yeast paste to enable easier counting.

Instars were differentiated by key features. First instars were present approximately 24 hours after the eggs were laid and can be separated from second instars by analysing the mouthparts. In second instars the mouthparts have developed to look much more like hooks, whereas in the first instars the mouthparts appear to be 3 dark spots. Second instars are usually at 48 hours, and third instars are at approximately 72 hours. Third instars can be deduced by the fanning structure at the end of the trachea and the presence of dark orange rings around the spiracles. Once the larvae reached 3rd instar they were transferred to normal Bloomington food vials and watched until pupation occurred. The number of pupae and then the final number of adult flies were recorded.

2.6.2.1   *Agar plates*

To easily view and count individuals, flies were allowed to lay eggs on grape juice agar plates. The plates were made using a protocol mentioned in Featherstone et al. (2009). 70mL of distilled water was added to 3gms of agar in a 250ml conical flask and microwaved until boiling. In a 100mL conical flask, 30mL of grape juice was added to 3gms of sucrose and microwaved until boiling. The juice-sugar mix was added into the agar solution, and then 0.05gms of methyl-4-hydroxybenzoate was added and the solution was swirled until well mixed. Approximately 5mL of the mix was poured into 35mmx10mm petri-dishes and allowed to cool and set. Once set, the plates were stored at 4°C upside down until used.

2.7   *Drosophila Activity Monitoring System*

The *Drosophila* activity monitoring system (DAMS) by Trikinetics is a machine which is used to indirectly measure activity of individual flies housed in clear tubes. The machine communicates with a computer which is used to continuously record activity counts across periods of multiple days.

2.7.1   *Set-up and use*

Drosophila strains were put through DAMS analysis to assess circadian rhythm, sleep-wake activity, and motor ability. The system functions by sending an infra-red beam through a plastic tube, and houses 32 individual plastic tubes per tray (figures 2.5 and 2.6). The system records the number of beam crosses performed by the fly in a set time frame, along with the time and date.
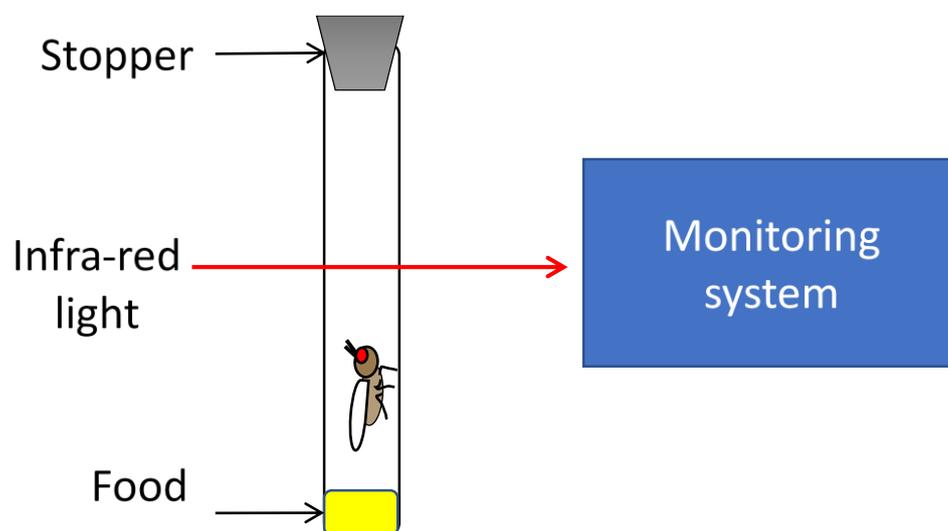
Figure 2.6: Drosophila activity monitoring system (DAMS) function.

The flies are individually housed in clear plastic tubes that have an infra-red light passing across the middle. The number of times the beam is broken is reported to a computer which records the data along with time and date stamp. This number can then be used to calculate activity or lack of activity of the individual flies across several days.

The DAMS tubes are 65mm in length, and with a 5mm outer diameter and 3mm inner diameter. They are set up to contain approximately 10mm of Bloomington Diet food at one end, sealed with parafilm, and a cotton wool stopper at the other. The distance of which is remained open for the fly to move around is controlled and kept even across all tubes in the experiment. Once the flies are introduced to the vials, they are loaded into the DAMS tray in alternating genotypes to ensure fair controls against environmental factors such as shading or heat gradient (figure 2.6).

Figure 2.7: A complete DAMS array set-up.

The DAMS tray is set up with 32 individual tubes housing 32 individual flies. Each tube has a cotton plug at one end and a small volume of yeast-cornmeal based food sealed with parafilm at the other.

For sleep-wake behaviour and activity analysis, the DAMS tray was set up and put in 25°C. The first readings including at least one full day (7am to 7pm) was omitted from subsequent analysis as it was the acclimatisation period, a during which the environment change may affect behaviour. For circadian rhythm analysis, the DAMS monitor was set up in a box which could be shut to block out light. The flies were left to acclimatise for a full day, and then at the lights out of the incubator cycle the box was shut and left. The flies inside the box were subjected to no lights for 7 days.

2.7.2   Data analysis

The DAMS system produces a lot of raw data in the form of beam counts per 5-minute period per fly for an average run of 5-7 days. Depending upon how the data is converted it can be used to look at activity or sleep.

Raw data from the DAMS was copied into Microsoft Excel sheets that were pre-programmed to manipulate the data to provide activity or sleep-like behaviour results. Programming the cells to calculate each step enabled a "copy and paste" approach to analysing new DAMS data each time an experiment was run.

The activity excel sheet calculated the average activity counts for half hour periods that was then plotted as an actogram, showing the activity pattern across a 24-hour period.

The sleep excel sheet was programmed to turn the periods of inactivity into ascending counts of 1, and as a result was able to compute parameters such as: sleep bout length, total sleep amount, number of sleep bouts, and max sleep bout length. The data was also accumulated into an average sleep amount per 30 minutes to provide a sleep profiles across a 24-hour period.

2.8   Video Tracking

Due to some limitations of the DAMS system, additional verification of fly movement was undertaken to better understand behaviour. Direct video recording of *D. melanogaster* was taken and analysed to compare to the DAMS findings.

2.8.1   Video Set Up

Flies in DAMS vials were taken from the DAMS and placed on a transparent plastic tray atop a light box. The tray is necessary to prevent the flies overheating from prolonged direct contact with the light box. A video camera was set up on a quad-pod above the vials pointing down (figure 2.7). A recording was taken for approximately 50 minutes.
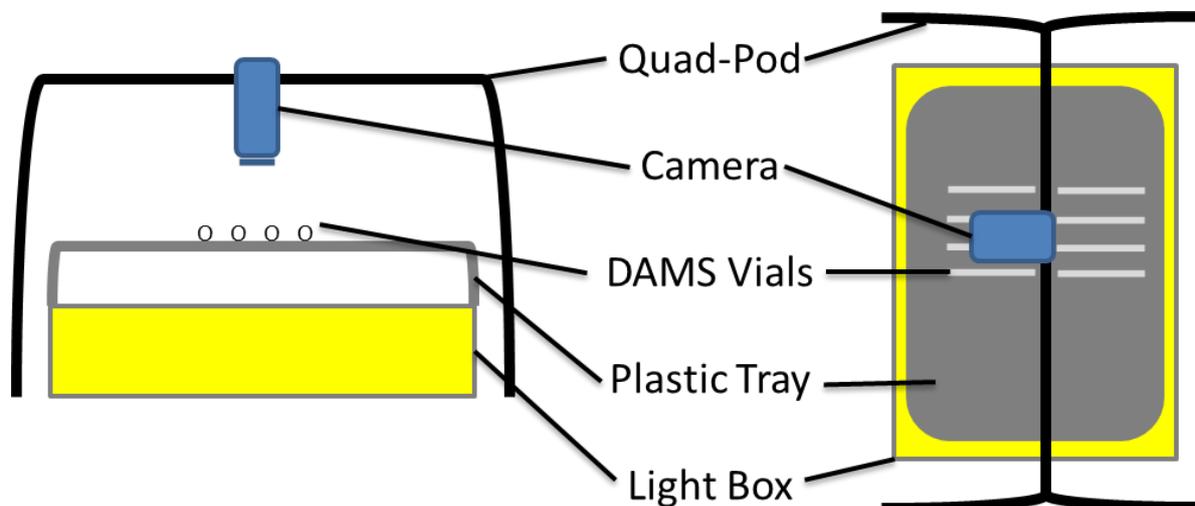
Figure 2.8: Video recording set up for tracking.

Video and subsequent computer analysis was used to directly monitor the fly's activity over an approximate 50-minute period. The backdrop of the light box gave a clear contrast which enabled the video camera to pick up the individual flies. The flies were taken directly from a DAMS set up and kept in the same vials.

2.8.2   *Video analysis*

Following recording, the video was converted into 1 frame per second and then imported into ImageJ, where the image stack was cropped and converted into simplistic black and white images (figure 2.8). Then measure mass was used to get a y-plot for the median fly location (simplified to black dot) across all slices of the image stack.

2.8.3   *Data analysis*

The y-plot data was then put into Microsoft Excel and normalised into 0-1 representative with 1 being cotton plug end and 0 being the food end. This was then converted into approximate mm measurements and each measurement was subtracted from the previous one to calculate the distance moved (mm), and subsequently velocity (mm/s).

Figure 2.9: Simplified image frame from video recording.

Flies were recorded in DAMS vials, and then the video was edited into 1 frame per second stacks and converted into a simplistic black/white image for tracking in ImageJ software. Due to the contrast editing, the sides of the tubes are not seen, however it is possible to pick out the food (large black dots at bottom of picture), the cotton wool (large black dots at top of picture), and the fly (small black dot) of each vial.

2.9   *Negative Geotaxis*

Flies of 1-2 days since eclosion were separated under CO_2, and cohorts of 10 male adults were stored in vials containing food until 5 to 7 days old. They were then transferred to an empty vial without using CO_2. A second tube was inverted and secured to the top with tape. The double vials were stood up against a light box and horizontal ascending lines with 20mm intervals, marking sections numbered 1 to 10. A camera was set up at a controlled distance and levelled to approximately midway up the tubes. Once set up, a video recording was started, and the vials were inverted and tapped sharply three times to ensure all the flies were on the base. The video was taken for 20 seconds to capture all fly movement during the assay and after each round of recording, the flies were left untouched for 1 minute before repeating. Two vials of 10 males were used for each genotype (*miR-137^{KO}* and *w^{1118}*) and each was repeated 5 times. Once the videos were all collected, it was reviewed and stopped at a climbing time of 5 seconds and the numbered section achieved by each individual fly was recorded in excel.

2.10   *Statistics*

Due to the simple nature of the experiments, t-tests were conducted to calculate the significance between two variables at a time to reinforce the results and discussions. The experiments were set up to directly associate a change in behaviour differing from the "norm", and as such compared one test genotype against the *w^{1118}* control at a time. To draw further comparisons, some tests (such as the recovery or phenocopying chapters 6 and 7) also included a significance test to assess the success of each experiment. To use a t-test, it assumes equal variances and a normal distribution of the data sets. For larger datasets the

distribution was assumed as normal in line with the Central Limit theorem, but for smaller datasets (n < 30) the distribution was checked as follows.

2.10.1   Testing for equal variance and normal distribution

Prior to any significance testing, the samples were put through f-test calculations using the prebuilt function in Microsoft Excel. The F-tests enable a measure of equal or unequal variance in the samples which can then be used to select the correct significance tests.

An F-test of variances assumes a normal distribution of the datasets. To test the distribution, you can use the Shapiro-Wilks formula which calculates the distribution and if found significant (p-value <0.05) then the null hypothesis of a normal distribution is rejected. There are several online calculators that will calculate the p-value from your dataset, or it can be done manually using Shapiro-Wilks coefficient tables. The Central Limit Theorem generally applies to samples numbering over 30 and dictates that it is acceptable to assume normal distribution for such sample sizes and larger. In this study, sample sizes of under 30 were subjected to the Shapiro-Wilks test for normality before significance testing.

2.10.2   Testing for significant differences

To test for significant differences between two datasets, a simple t-test was conducted utilising the prebuilt function in Microsoft Excel. The test chosen was either for equal or unequal variances dependent upon the result of the F-test. If the resulting p-value was smaller than 0.05, the null hypothesis of no significant differences between the samples was rejected and the two samples were deemed statistically significant until further correcting for false discovery rates.

2.10.3   Post-Hoc testing for false results

Due to repeatedly conducting t-tests, the likelihood of false positives (Type I errors) is amplified. Finding a result that is significant to a 5% level will happen by chance 5 times out of every 100 tests and to reduce the likelihood of producing errors by chance, p-values can be corrected using one of several measures. One of the simplest, but most conservative, of which is the single-step Bonferroni correction method (Bonferroni 1936), which divides the significant p-value by the number of t-tests conducted within a "family" of experiments to give a new significance threshold. This correction for family-wise error rate (FWER) reduces false positives but, in the event of a high number of tests, can increase false negatives (Type II error) therefore reducing statistical power. Studies of several thousand significance tests such as genomic microarrays, often use a method for controlling the False Discovery rate (FDR) known as the Benjamini-Hochberg test (Benjamini and Hochberg 1995). This allows for a degree of error and produces candidate results that can be further validated in follow-up studies.

In this thesis, a "family" of experiments were grouped in accordance to dataset regardless of how reported or grouped in graphs. As this study is not a large-scale screen and even the project-wide t-test count is under 1000, the Bonferroni method was used to calculate excessively conservative thresholds that were used as discussion points to evaluate if the result is likely truly significant. Only one test was conducted for lifespan, activity, and negative geotaxis so these needed no correcting. For the other experiments, the number of tests and most conservative calculated p-values used for discussion are as follows:

- Developmental rate:

  - Tests conducted = 3

  - Corrected 5% significance p-value = 0.0167 (3sf)

- Video tracking velocity:

  - Tests conducted = 3

  - Corrected 5% significance p-value = 0.0167 (3sf)

- DAMS sleep behaviour:

  - Tests conducted = 63

  - Corrected 5% significance p-value = 0.000794 (3sf)

- Pharmacological treatments (individually):

  - Tests conducted = 6

  - Corrected 5% significance p-value = 0.00833 (3sf)

3   Verification of genomic miR-137 knockout

Fly stock #58893 was ordered from Bloomington Drosophila Stock Centre which was reported to have a homologous arm substitution that replaced the endogenous *miR-137* coding region with an insertion which contained a $w^{+mW.hs}$ sequence. The stock was on a white-eyed (w^{1118}) background and the result of the exogenous sequence was a red eyed phenotype. To be sure the stock was supplied as reported, experiments were needed to check for the presence of the insertion and lack of the original *miR-137* gene.

3.1   Gel electrophoresis

Assuming successful amplification, the amplicons make up a significant portion of the PCR product. A loading dye can be added, and then the solution is put into a well along the top edge of an electrophoresis gel. The gel is made of a mixture of tris base, acetic acid, and ethylenediaminetetraacetic acid (TAE) buffer, agarose, and a DNA binding marker like Sybersafe. Once fully loaded, the gel is submersed in more TAE buffer and has an electrical current passed through it from a negative to positive electrode. As DNA is negatively charged, it will begin to travel through the gel towards the positive electrode. The gel impedes the DNA movement, and this separates the DNA in relation to size with the shortest sequences travelling faster than longer amplicons. The Sybersafe binds to the DNA as it travels and labels it with a fluorescent marker that can be seen under UV light. The result of a gel electrophoresis is a series of fluorescent bands that are separated in terms of size, allowing the identification of (or lack of) each amplicon.

The function of the loading dye is trifold; it weighs down the sample to prevent it floating into the buffer, it also colours the sample for ease of locating, and the two visible dye components travel both faster and slower than the DNA sample to easily ascertain the

location the bands. Alongside the PCR product, the first well on the gel is usually a DNA

ladder, which are pre-made solutions of a mixture of known DNA sequence lengths and can

be used as a reference for the length of an amplicon.

Table 3.1: Electrophoresis gel layout and expected DNA bands.

Well number (left to right)	Primer	Genotype	Expected band
1	Ladder	Ladder	Ladder
2	1	w^{1118}	Y
3	1	miR-137^{KO}	N
4	1	Heterozygous knockout	Y
5	2	w^{1118}	N
6	2	miR-137^{KO}	Y
7	2	Heterozygous knockout	Y
8	3	w^{1118}	Y
9	3	miR-137^{KO}	Y
10	Ladder	Ladder	Ladder
11	3	Heterozygous knockout	Y
12	4	w^{1118}	N
13	4	miR-137^{KO}	Y
14	4	Heterozygous knockout	Y
15	5	w^{1118}	Y
16	5	miR-137^{KO}	N
17	5	Heterozygous knockout	Y

The set-up of the first gels were the 5 primer sets used on each of the genotype samples

(knockout, control, and heterozygous knockout) as laid out in table 3.1. Five primer sets were

created to be able to select the most successful for amplification and subsequent sequencing,

with methods described in chapter 2.4.2.

Initially there were issues with the results demonstrated on the gels, with some of the

expected bands missing (figure 3.1). PCR factors such as annealing temperature, extension

time, and number of cycles can be modified to produce clearer results in the gel. Also, any

contamination during or carried from the extraction or purification process can also impair the amplification. Therefore, an absence of bands within the first few attempts could not simply be put down to ineffective primers. Several repeats of the DNA extraction process and subsequent electrophoresis gels were conducted, with emphasis on the reagents used and dilution of the DNA template to prevent carry over of any ethanol.
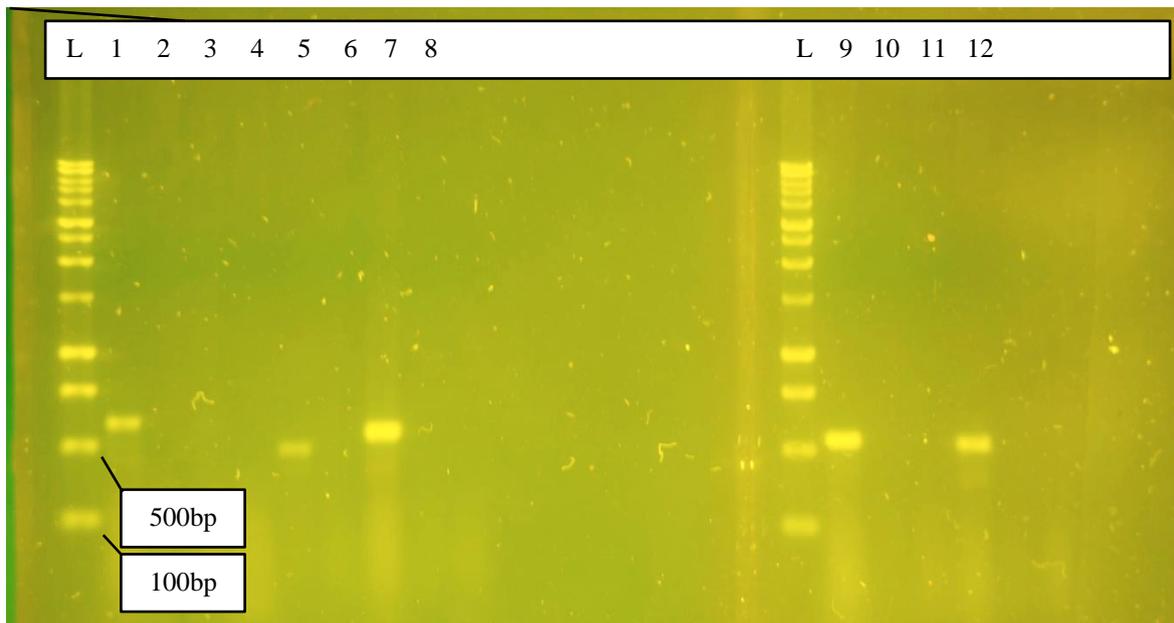
Figure 3.1: Electrophoresis gel result of PCR genotyping.

Clearly seen on the left-hand side of each is the molecular ladder and its increments used for sizing of the amplified bands. The bands are all in the lower region of the ladder (100 and 500bp were the smallest bands shown as the lowest marks in the "L" row).

After several failed attempts, new primers and reagents were used to eliminate contamination. The extracted DNA template was also diluted further before introduction to the PCR reagents to minimise the carry-over of toxic products from DNA extraction, and Actin primers were ordered from DSRC to amplify ~200bp regions of Actin-5c and Actin-57b. Genes from the Actin family are highly expressed across all cell types, and the primers acted as a positive control for presence of genomic DNA.

Following the troubleshooting, the results demonstrated that primer sets 1 and 2 were most effective at amplifying the correct target gene regions, *miR-137* and $w^{+mW.hs}$ respectively. Primer 4 designed to bind to $w^{+mW.hs}$ was also picked up in the w^{1118} control flies. The w^{1118} flies contain an allele that is a partial deletion of what would be a wild-type *white* gene. True wild-type flies have a red eye phenotype, which is not present in the w^{1118} control flies who have the recessive white eye phenotype. The $w^{+mW.hs}$ insertion is a mini version of the white gene which recovers the red eye phenotype in a w^{-} background.
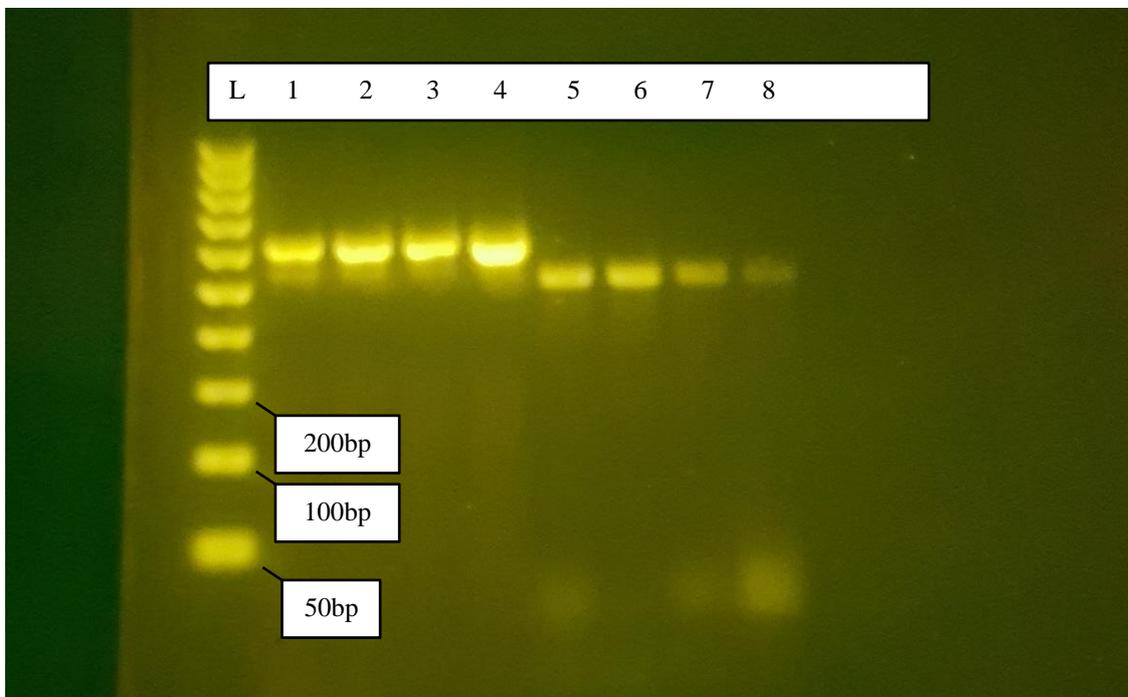
Figure 3.2: Electrophoresis gel result of PCR genotyping sent for sequencing.

Clearly seen on the left-hand side is the molecular ladder and its increments used for sizing of the amplified bands (row "L", with bands of 50, 100, 200 etc.). The next 4 bands (rows 1-4) are the amplified region from primer 1, a section of the genomic *miR-137* sequence. The last 4 (rows 5-8) are the amplified regions from primer 2, part of the $w^{+mW.hs}$ sequence.

3.2   *Amplicon Sequencing*

To verify the amplified regions of these primers, several individual PCR runs were collected for both selected primers (1 and 2) and checked for successful amplification (figure 3.2), then sent to Beckman Coulter Genomics for sequencing. The amplicons were sequenced, and results were returned as a nucleotide sequence list and nucleotide read graph for each (figures 3.3 and 3.4).
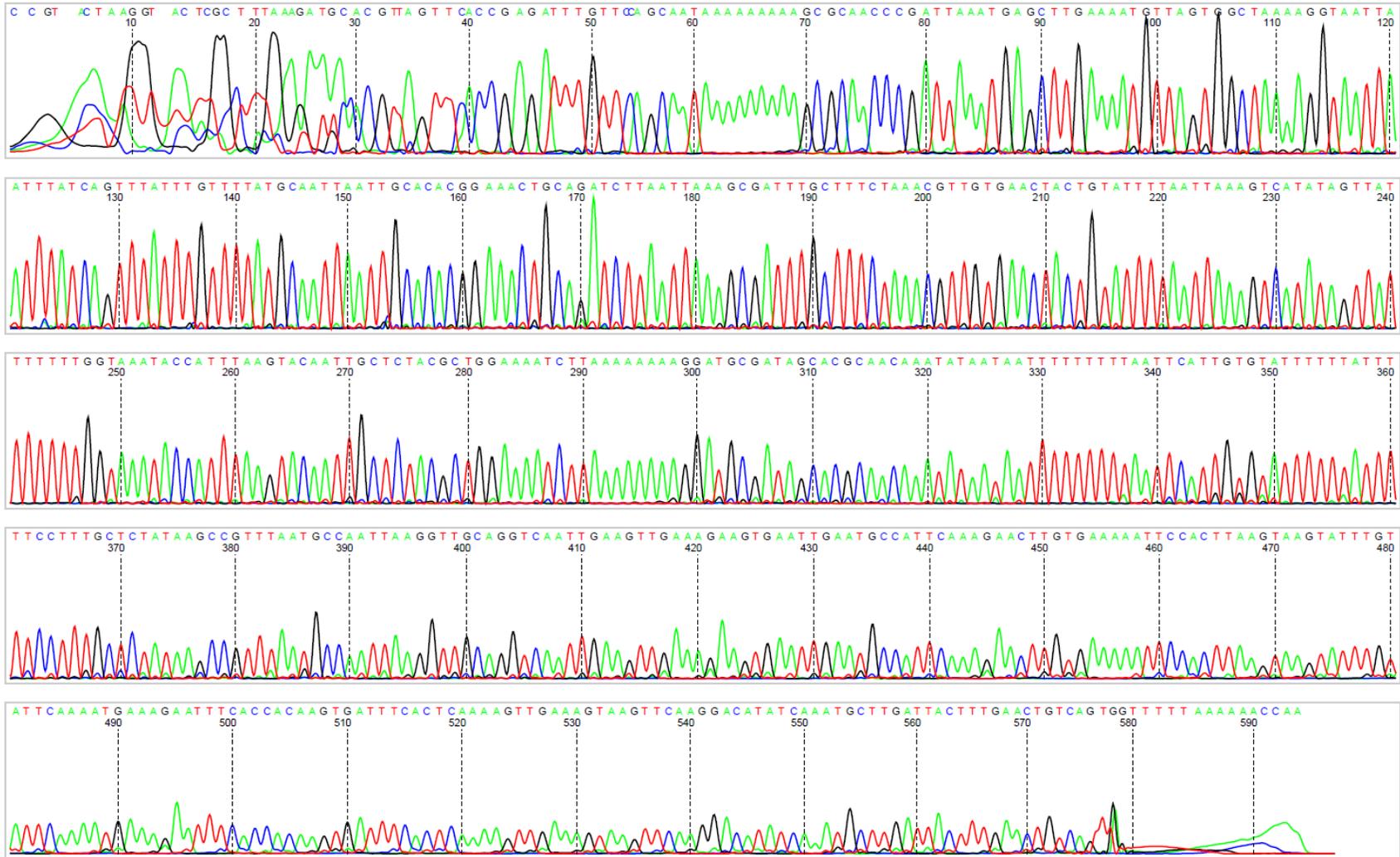
Figure 3.3: Sequencing result of primer set #1 amplicon (*miR-137*).

Primers were designed to bind to a selected region of *miR-137*, and then put through PCR amplification before being sent for sequencing.

Figure 3.4: Sequencing result of primer set #2 amplicon ($w^{+mW.hs}$).

Primers were designed to bind to a selected region of $w^{+mW.hs}$, and then put through PCR amplification before being sent for sequencing.

3.3   *Amplicon comparison to expected sequences*

The sequencing results were then compared to the predicted amplified sequence to check for high homology. Using the sequences and online programmes such as LALIGN (available from: https://embnet.vital-it.ch/software/LALIGN_form.html), the homology of the sequences can be assessed. The comparison shows a very high homology between the expected sequences and the amplified sequences (figures 3.5 and 3.6). Unfortunately, the primers were not designed to assess the mature *miR-137* homology specifically and due to its proximity to the primer binding site (41 – 63 base pairs in figure 3.5, identified with the blue bar) it's not possible to determine its identity. Similarly, the coding sequence of *miR-137* (<0 – 88 base pairs in figure 3.5, identified with the orange bar) cannot be fully seen, though there appears to be some discrepancies towards the end of the coding strand. While it is not clear what effects these differences could have, they are not part of the mature sequence and therefore not the area involved with RNA-regulation.
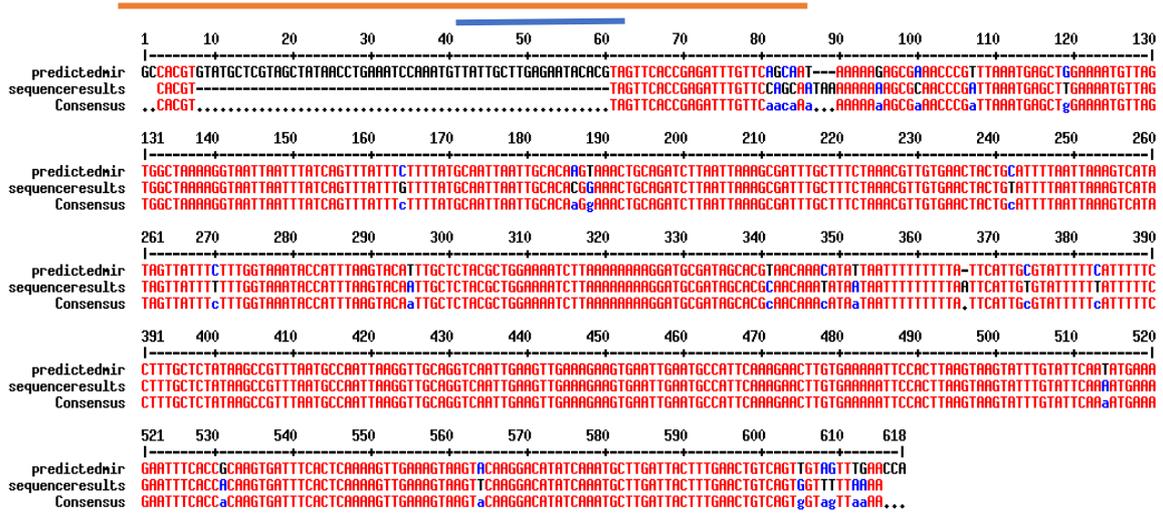
Figure 3.5: Alignment of *miR-137* DNA sequences.

Using Lalign, the amplified *miR-137* region by primer set 1 and the sequence from Flybase was compared for sequence homology. The red letters highlight identical nucleotides and therefore demonstrate the very high similarities between the two. The mature *miR-137* sequence responsible for the RNA-downregulation function is shown from base pairs 41-63 (blue bar), and the coding sequence for *miR-137* is shown from base pairs <0 -88 (orange bar).
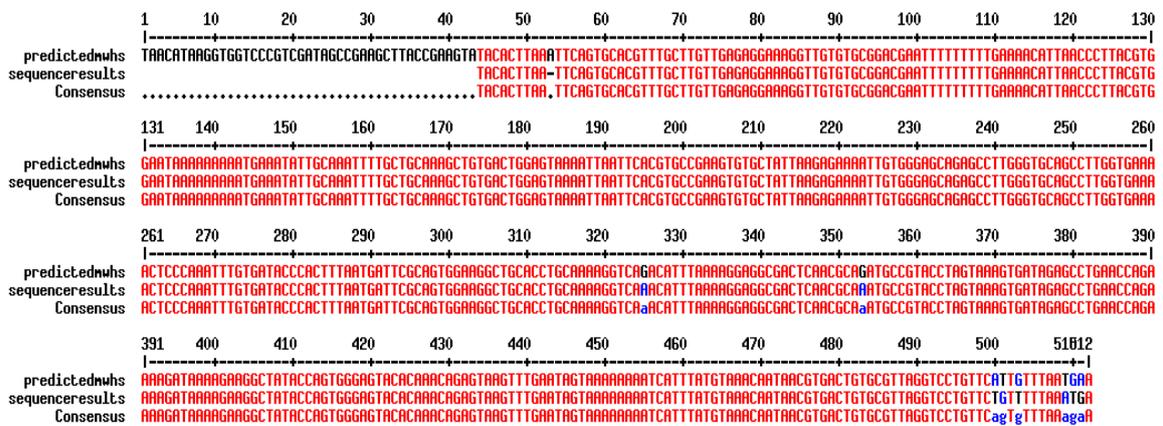
Figure 3.6: Alignment of *w^{mW.hs}* DNA sequences.

Using Lalign, the amplified *w^{mW.hs}* region by primer set 2 and the sequence from Flybase was compared for sequence homology. The red letters highlight identical nucleotides and therefore demonstrate the very high similarities between the two.

96

3.4   Conclusion

Overall, the primers are proven to successfully bind and promote the amplification of the region they were designed around. Furthermore, the amplicons are clearly the expected regions as proved by the high sequence homology between reported literature sequence and the sequence of the amplicon. Through presence-absence analysis, the PCR genotyping clearly shows lack of genomic *miR-137* DNA within the *miR-137^{KO}* stock. Not only this but, in the same knockout stock, there is the addition of a $w^{+mW.hs}$ sequence as reported by Bloomington.

4 *miR-137^{KO} effect on D. melanogaster health*

With all genetic manipulation of DNA *in vivo*, the resulting genotype can be detrimental to overall health or living capabilities of the organism. With the loss of a microRNA, and subsequent introduction of a larger construct, the knockout stock was measured for any major defects within parameters such as longevity, development, and other general behaviours. Observation of the viable stock did not give obvious reason to evaluate behaviours such as feeding, overall size, or ability to reproduce. The flies were capable of propagating, reproducing and living as would be expected of a non-mutant stock. However, lifespan and development lethality/rate are harder to determine from simple observation and were therefore investigated to see if loss of *miR-137* has implications for these factors.

4.1 *Lifespan*

A lifespan assay was conducted as *Drosophila* lifespan is reported to have been affected by many genetic manipulations previously. The test was conducted as described in Chapter 2.6.

4.1.1 *Results*

The lifespan assay shows that on average, both the *w^{1118}* and *miR-137^{KO}* strain had a similar mean lifespan of circa 51 and 49 days respectively (figure 4.1). The average lifespan is not statistically different when using a t-test (p-value > 0.05).
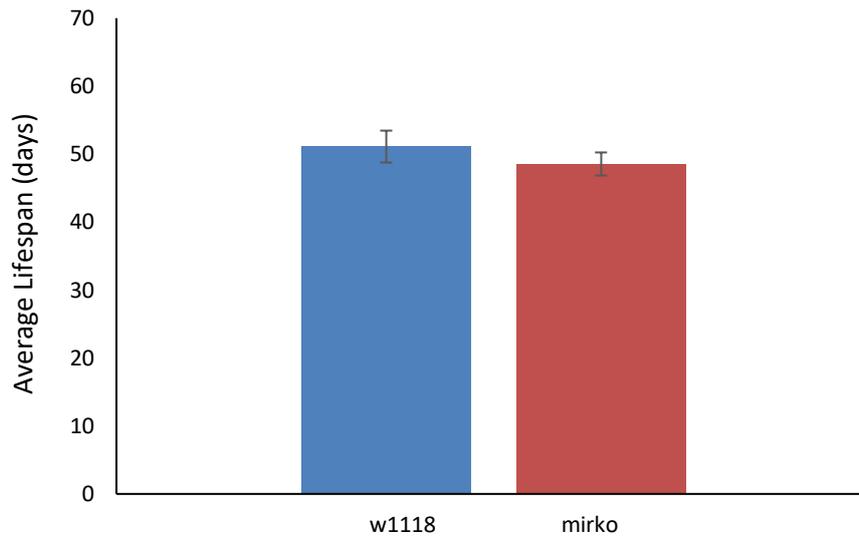
Figure 4.1: Mean lifespan of *w^1118* and *miR-137^KO*.

The *w^1118* flies did not show a significant difference to the *miR-137^KO* (51.1 days, n = 68 and 48.5 days, n = 101 respectively). Error bars are standard deviation. (t-test p-value > 0.05).

There is a clear drop in survival rate of the *miR-137^KO* strain once they reach about 42 days old (figure 4.2). This is shown with a dramatic increase in percentage death from ages 31-45 days old and then 45-59 days old, with an additional ~25% death at each stage in the knockout strain compared to controls (figure 4.3).
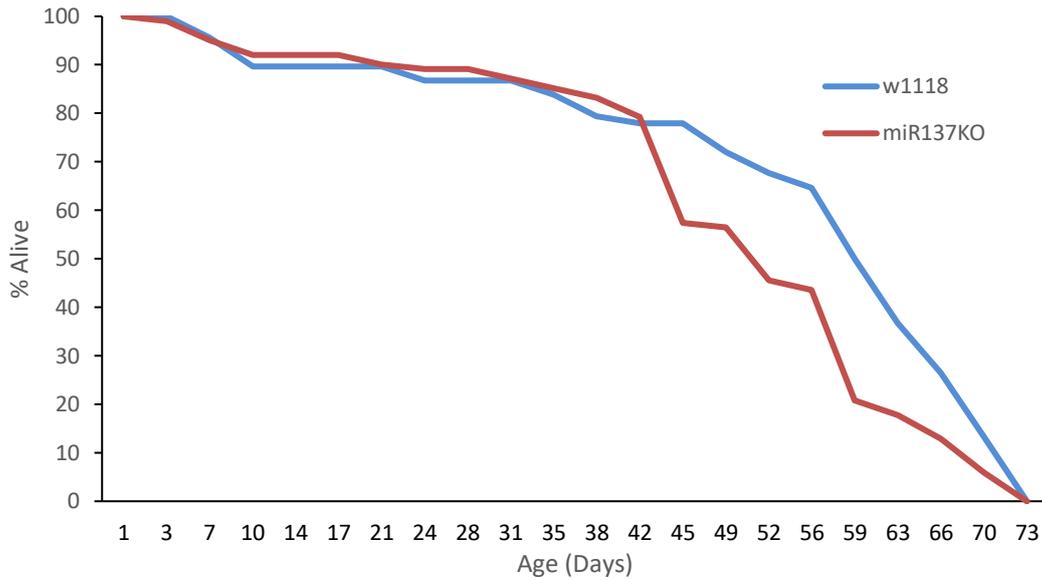
Figure 4.2: Survival rate of w^{1118} (n=68) and *miR-137^{KO}* (n=101) adults across age.

There was an apparent drop in survival rate of the *miR-137^{KO}* (n=101) at ~day 42 compared to w^{1118} (n=68).
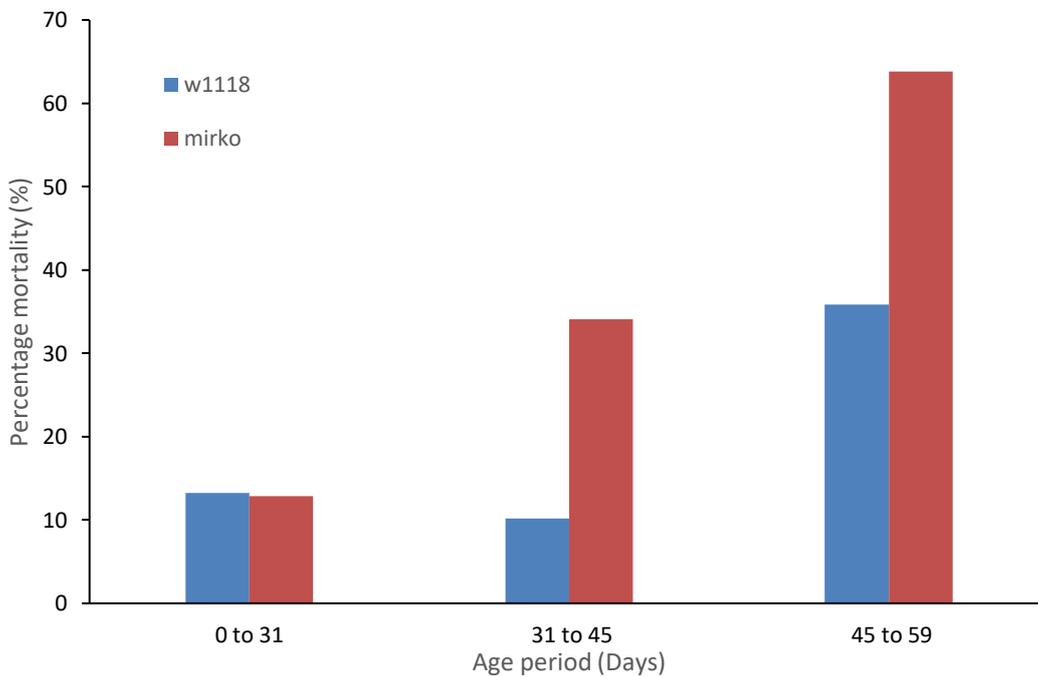
Figure 4.3: Mortality rate of w^{1118} and miR-137^{KO} by age range.

Percentage mortality for age ranges 0-31 days, 31-45 days, and 45-60 days for w^{1118} (n=68) and miR-137^{KO} (n=101). In the early stage both genotypes have a similar death percentage, whereas in the older stages the *miR-137^{KO}* strain has an increased death rate of 24% and 28% in 31-45 and 45-60 respectively.

4.2   Developmental rate and lethality

The rate at which an organism develops is governed by a multitude of factors from both environment and genetic roots. In controlled environments *D. melanogaster* generally keep to a generation cycle of approximately 10 days at 25°C, with new adults emerging from eggs laid by the previous generation. To assess if *miR*-137 LOF affects developmental rate or lethality, assays were conducted as described in Chapter 2.6.

4.2.1   Results

The rate of development shows a significant difference between the control strains w^{1118} and *miR-137^{KO}* subjects (figure 4.4). The knockout strain reaches 3rd instar stage significantly earlier than controls (t-test p-value < 0.05) yet takes longer in 3rd instar and pupal stages to reach adulthood later than controls (t-test p-value < 0.01).
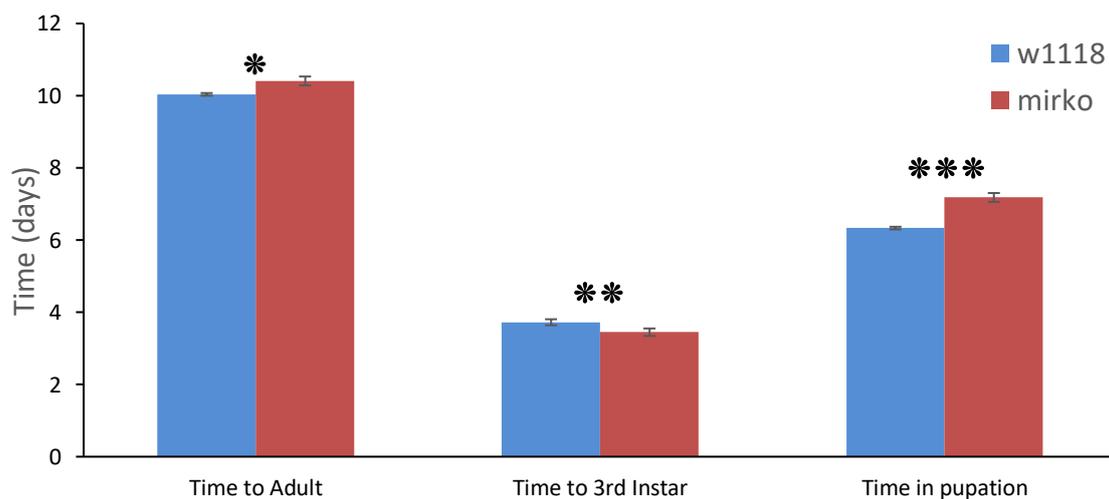
Figure 4.4: Average development time of w^{1118} and *miR-137^{KO}*.

 The w^{1118} (n = 48) flies reach third instar significantly slower than the *miR-137^{KO}* (n=48) strain. However, reach adulthood quicker and spend less time in pupation. Non-corrected t-test results: * p-value < 0.01. ** p-value < 0.05. *** p-value < 0.01. Following post-hoc Bonferroni significance corrections, the time to 3rd instar no longer satisfied the significant threshold.

The developmental survival assay shows increased lethality in the *miR-137^{KO}* strain, with 45.8% surviving from 1st instar to adulthood compared to 56.2% in *w^{1118}* controls (figure 4.5). The survival at each stage shows a change in survival percentage between the 3rd instar and adult, with 4% death in the controls and 14% death in the knockout strain (figure 4.6).
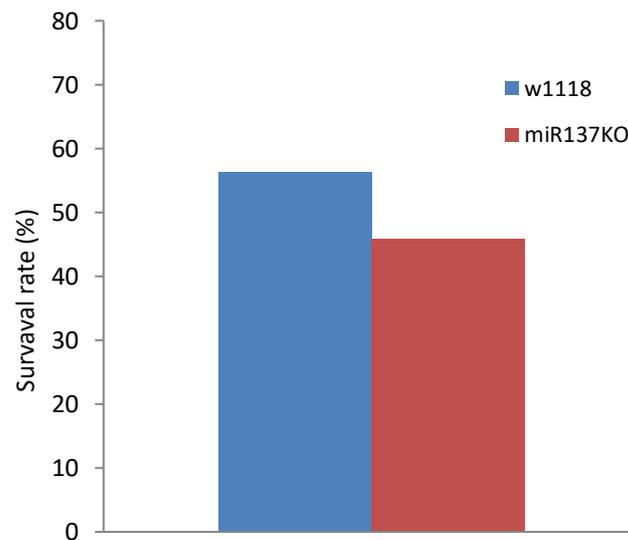
Figure 4.5: Survival rate from hatching to adulthood.

The *miR-137^{KO}* (45.8%) has a lower survival rate compared to *w^{1118}* controls (56.2%).
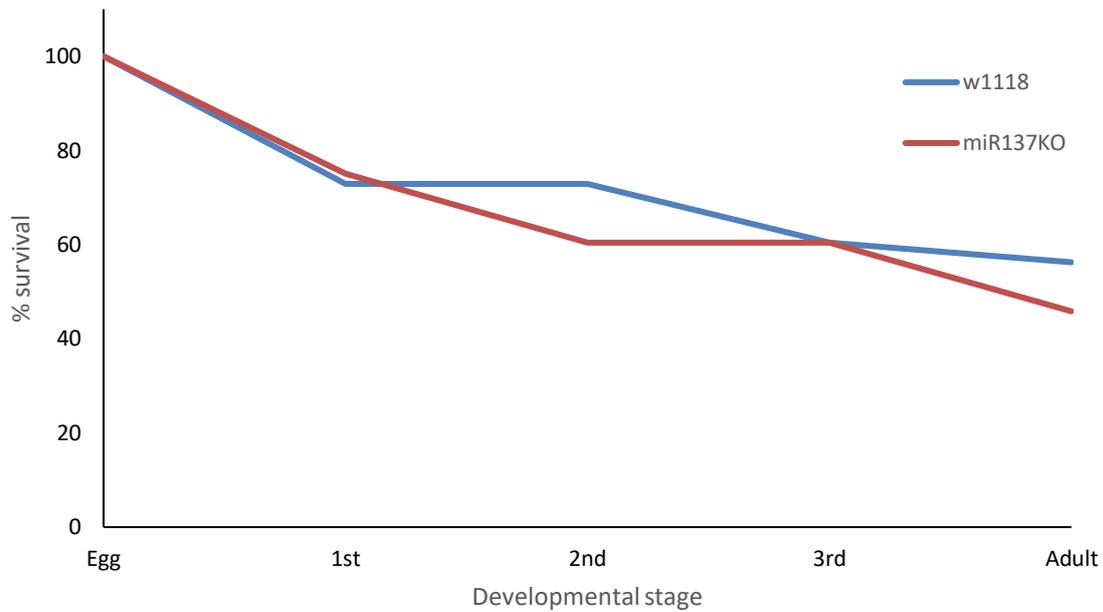
Figure 4.6: Survival rate across developmental stages.

The *miR-137^{KO}* strain (n=48) shows less overall survival to adulthood compared with *w^{1118}* controls (n= 48). The knockout flies also show an increase in mortality between 1st and 2nd instars.

4.3   Discussion

From the lifespan and developmental assays, there is little evidence of this stock having a severe longevity problem with genomic loss of the *miR-137* gene. The mean average ages are not significantly different (p-value > 0.05), and correlate with documented lifespan from various studies. When cultivated on a sugar yeast diet, Linford *et al.* (2013) found that wild types had a mean age of 50 days, and a Sun *et al.* (2002) reported 46-52 days being within acceptable parameters of control *D. melanogaster* lifespan. While investigating differences in diet effect on lifespan, Bass *et al.* (2007) reported that mixed gender vials of canton-S (another widely used control strain) produced an average individual lifespan of around 49 days when on standard laboratory food mix including maize, yeast, and sugar. In a screen for miRNA mutants, no significant change in lifespan was found for males or females with *miR-137* knockout (Chen *et al.* 2014).

103

As the flies age, there is a noticeable difference in the survival graph following ~41days old (figure 4.2). The *miR-137^{KO}* strain survival rate drops much faster at this stage than the controls, though this does not appear to influence the mean age (figure 4.1, which has a t-test p-value > 0.05). The rate of survival decrease looks to be more severe in the controls at a later stage, which could then be masking the average life span across the population. This drop ~41 days could infer a maintenance role for *miR-137* that helps delay or manage the onset of age-related mechanisms.

Developmentally, there appears to be a small change in the percentage of individuals which survive from imago hatching to post-eclosion adult (figure 4.5). The survival rate graph shows the expected overall trend for both, with less surviving individuals at each developmental stage. There is an overall lower yield of adults in the *miR-137^{KO}* strain, with 46% survival compared to 56% in the *w^{1118}* controls where the biggest decrease difference was during the pupal stage between 3rd instar and adulthood. The knockout strain had a 14% decrease in survival compared to 4% in the controls, potentially implying a role for *miR-137* through the metamorphosis period. During this stage it is known that the organism undergoes huge changes in terms of biological anatomy and neural development (Technau and Heisenberg 1982: Singh and Singh 1999: Pauls *et al.* 2010), so genes responsible for CNS maintenance or proliferation would be important. This developmental lethality during pupation is contrasting to results by Chen *et al.* (2014) which did not find any significant difference in fly survival to adulthood.

The rate at which *miR-137^{KO}* flies reach development stage is has a slight but significant difference when compared *w^{1118}*. This trend was reversed when looking at time to reach adulthood, where the controls were faster. This switch is because the knockout strain spends an increased amount of time in the 3rd instar and pupal stages, further inferring *miR-137* is required for an aspect of metamorphosis. When allowing for Bonferroni correction of

the t-test p-values, the significance cut-off value is changed to 0.0167. This new threshold means that the difference in the time taken for the flies to develop from egg to 3rd instar larva is no longer significant (non-corrected p-value of 0.0468). This however does not change the significance of time spent in pupation or overall time to adult fly.

The knockout strain has had the *miR-137* locus missing for a long period of time and with each generation it is more likely to have background mutations, which allow the stock to cope with the miRNA LOF. These modifiers could have a masking effect causing the flies to display a subdued phenotype. As longevity is widely research in *D. melanogaster,* there are many genes which are linked to a change in lifespan. For example, when overexpressing antioxidative enzymes, explicitly mitochondrial superoxide dismutase mnSOD, there is an increase in mean lifespan likely because of the increased enzyme expression reducing the accumulated effects of oxidative stress in an aging individual (Sun *et al.*2002). TOR signalling pathways are also widely believed to result in an increase in lifespan if reduced in adults, an effect linked to insulin signalling and dietary restriction (Stefana *et al.* 2017).

4.4   Conclusion

There is a noticeable change within some of the characteristics contributed to measuring drosophila health in the *miR-137* deficient flies. Most apparent is a significant change within developmental survival and overall longevity particularly in the old aged flies. The severity of the phenotype resultant of losing a ~100bp non-coding gene demonstrates the biological importance of *miR-137*, further backed up with a high evolutionary conservation of the gene across a multitude of species. The lack of significant difference in average lifespan means that objective A hypothesis 1 must reject the alternate hypothesis and accept H_0 (Chapter 1.8).

Developmentally, the *miR-137^{KO}* flies exhibit a developmental delay and an increased mortality rate within the 3rd instar and pupal stages when compared to controls, which means that null hypothesis from objective A hypotheses 2 and 3 can be rejected in favour of the alternate hypotheses (Chapter 1.8). The pupal stage is a natural phenomenon which puts excessive stress on the organism as it undergoes metamorphosis. It is likely vital for the correct expression of genes responsible for maintenance and proliferation of the CNS at this stage, and the loss of *miR-137* could be affecting successful CNS development.

5 *miR-137^{KO} effect on D. melanogaster behaviour*

Drosophila have been widely used as a model organism to investigate genes for human relevance in diseases. The most obvious effects of gene expression manipulation are phenotypical changes, which can be anatomical, behavioural or both. To investigate any effect loss of genomic *miR-137* may have, the stock was put through some behavioural tests.

5.1 *Drosophila activity monitoring system*

The *Drosophila* activity monitoring system (DAMS) is a well-used system that indirectly measures activity of individual flies across periods of time. The DAMS is an array which holds 32 flies in individual 75mm long plastic tubing and passes a perpendicular infrared beam through the centre. To measure activity, the number of beam crosses is collated across periods of time to provide average activity counts. The data was then manipulated in Microsoft excel as described in Chapter 2.7.

5.1.1 *Results*

The comparison of *w^{1118}* controls and *miR-137^{KO}* flies show a drastic decrease in overall activity in the knockouts (figure 5.1). The circadian rhythm is apparent in both strains, with signature peaks at ZT 0 and ZT 12 plus a pre-emptive gradual increase in activity throughout the night phase and inverted parabola during the day. Wild types reach 9 and 7 counts per 5 minutes for the morning and evening peaks respectively, whereas the *miR-137* knockout strain reaches only 3.3 and 4.2 counts per 5 minutes.
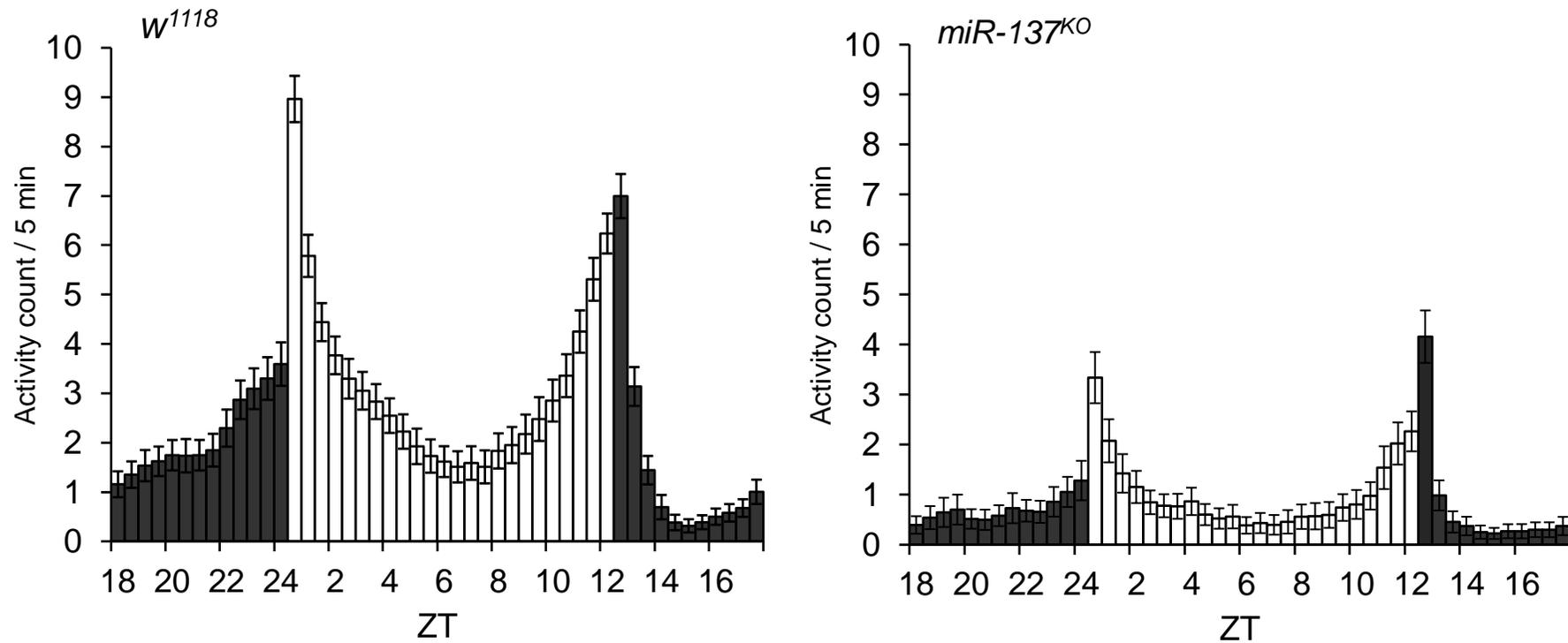
Figure 5.1: Actograms of *miR-137^{KO}* flies and controls across a 24-hour period.

The bars represent average beam crosses in 5 minutes for each 30-minute slot, averaged from 5 days of monitoring in the Drosophila activity monitoring system (DAMS). ZT is Zeitgeber time, which is normalised to 0 for lights on and 12 for lights off. The light and dark bars represent day and night as appropriate. The activity of the *w^{1118}* flies (n=398) represent a normal circadian pattern of peaks at ZT 0 and ZT12, and a gradual increase in activity in the night period (ZT 16-24). The *miR-137^{KO}* (n=114) strain shows a similar overall pattern but is severely stunted in terms of activity count across the whole 24-hour period. Error bars are standard error.

Total average activity per day shows a significant difference between control and test subjects (figure 5.2). Average beam crosses for w^{1118} was 727.5 while $miR\text{-}137^{KO}$ flies had a significantly lower average of 251.1 beam crosses (t-test p-value < 0.01).
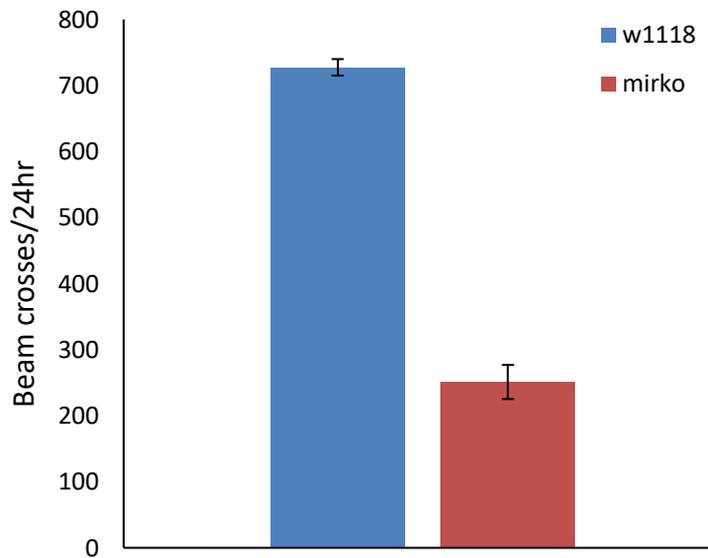
Figure 5.2: Average number of beam crosses per 24-hour period.

The bars represent total number of beam crosses per 24 hours activity count for w^{1118} (w1118, n=398) and $miR\text{-}137^{KO}$ (mirko, n=114) strains. The data is obtained from 5 days monitoring in the Drosophila activity monitoring system (DAMS) by Trikinetics. The w^{1118} strain had a mean average of 727.5 while $miR\text{-}137^{KO}$ was 251.1, a significantly different value (T-test p-value < 0.01). Error bars are standard error.

5.2   *Video Tracking*

Due to the DAMS inability to determine lack of movement and lack of ability to move, the flies were subjected to periods of direct video recording and subsequent computer tracking in ImageJ as described in Chapter 2.8. By direct analysis of the fly movement a pattern of travel can be determined to verify the use of the beam-pass method of the DAMS. It can also be used to calculate velocity for each fly to assess locomotor ability, by calculating distance moved per second across the recording. This can then be used to create average velocity for each fly and the genotype.

5.2.1   Results

The fly plots for each fly show that their movements frequently cross the mid-point and therefore validates the findings and future use of the DAMS array (figure 5.3). Furthermore, the knockout flies appear to have "flat" periods which demonstrate a pause or period of inactivity that may be indicative of a sleep-like episode.
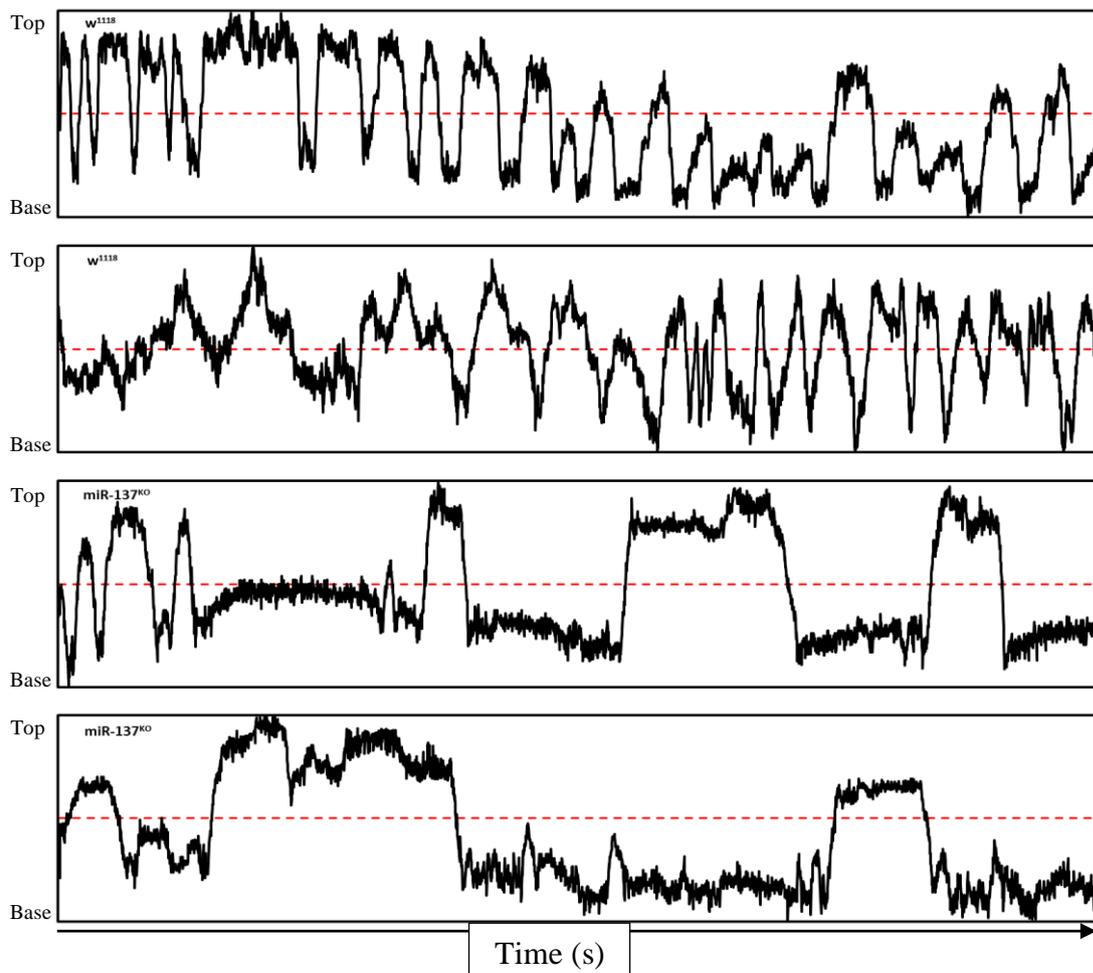
Figure 5.3: Fly tracking plots of *w*^{1118} and *miR-137*^{KO}.

The plots represent location for every second across a continuous period of 51 minutes (x-axis). The flies were placed in single tubes identical to the Drosophila activity monitoring system (DAMS) by Trikinetics. The vials had cornmeal, yeast, agar food at one end (base of plot) and a cotton wool plug at the other (top of plot). The dotted red line is representative of where the infra-red beam would pass through the vial and can be seen that almost all movement resulted in a crossing of the beam, validating the DAMS use for activity monitoring. As can be seen, the w^{1118} strains have a rhythmic and continuous up and down movement, while the miR-137^{KO} strain also traverses the entire tube but has "resting" periods of little or no activity. Two plots for each genotype represent two different flies recorded for the same period of time and are included as two to indicate similarities and that they are indicative of the genotype as a whole.

Analysis of the imagestacks showed a significant velocity change between w^{1118} and $miR\text{-}137^{KO}$ strains in both the evening experiment and when both morning and evening results were combined (t-test p-values < 0.01) (figure 5.4). The morning results have a greater difference in the average velocities than the evening experiment, although there is quite a variance in the 7am w^{1118} velocities potentially explaining the lack of significance found (t-test p-value > 0.05). The morning period was seen to have different velocities, with the w^{1118} controls averaging 1.94mm/s and the $miR\text{-}137$ knockout strain only achieving 1.64mm/s. The evening experiment showed the w^{1118} controls and $miR\text{-}137^{KO}$ achieving average velocities of 1.90 and 1.72 mm/s respectively.
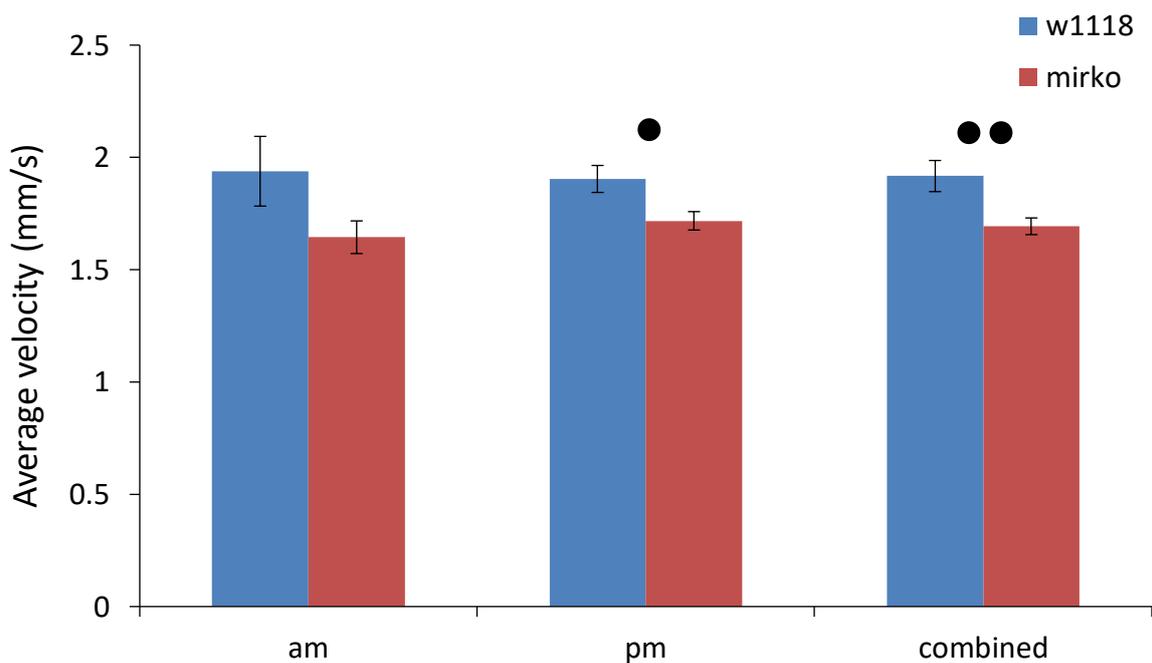
Figure 5.4: Mean velocity of w^{1118} and $miR\text{-}137^{KO}$.

The am period shows the controls and knockout strain averaging speeds of 1.94 and 1.64mm/s respectively, with no significance at 5% value (p-value > 0.05). When subjected to a t-test and subsequent Bonferroni corrections as described in Chapter 2.10, the pm period shows a significant difference with the controls and knockout strains averaging 1.90 and 1.72mm/s respectively (●: p-value < 0.01). Collectively the results are significant (●●: p-value < 0.01) with the w^{1118} control strain achieving a higher velocity of 1.92mm/s compared to the knockout strain at 1.69.

5.3   *Negative geotaxis*

Negative geotaxis is a test used to startle and assess locomotor response of *D. melanogaster*. The test is carried out as outlined in Chapter 2.9 and utilises an innate behaviour to run upwards (away from gravity) when prompted. While it is a simple assay, the results can be used to infer climbing or general locomotor ability.

5.3.1   *Results*

The *miR-137^{KO}* line did not perform as well as the *w^{1118}* stock in the climbing assay. There was a clear difference on the percentage of flies reaching each height region, as can be seen in the exceedance graph (figure 5.5). In the knockout strain there is a steep decline at the start of the plot, indicating a high percentage (27.7%) remained within the starting section compared to 8% in the controls.
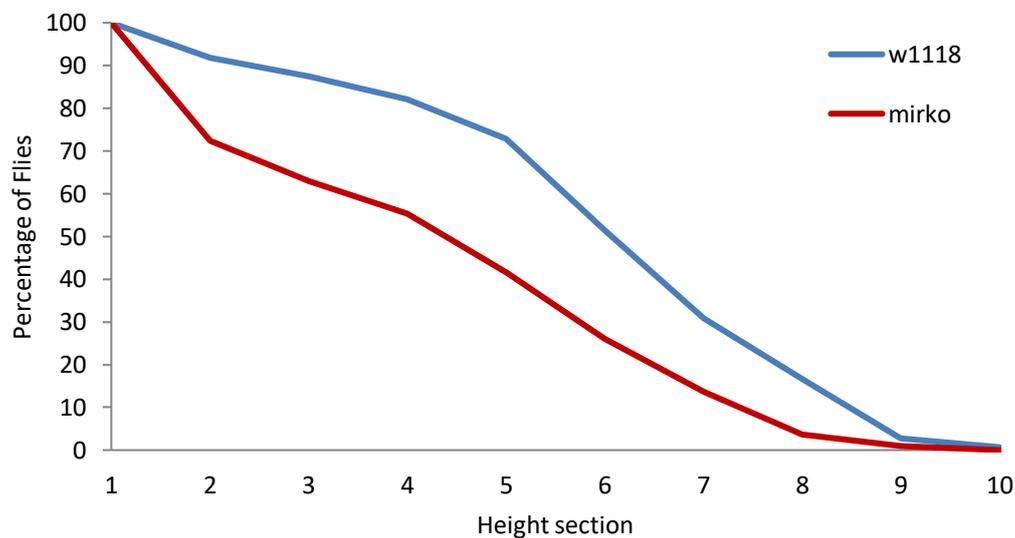
Figure 5.5: Percentage exceedance of height achieved in a negative geotaxis assay.

The miR-137^{KO} (n=300) performs poorly compared to the w^{1118} (n= 295), with a noticeable drop at the start (27.7% compared to only 8% in controls). Just 1% of the knockout flies reached the top two sections of the test apparatus, compared to 3.4% in controls.

The average climbing ability of the *miR-137^{KO}* strain was significantly lower than that of the *w^{1118}* controls (figure 5.6) (t-test p-value < 0.01). The controls averaged 5.38, while knockouts reached 3.77.
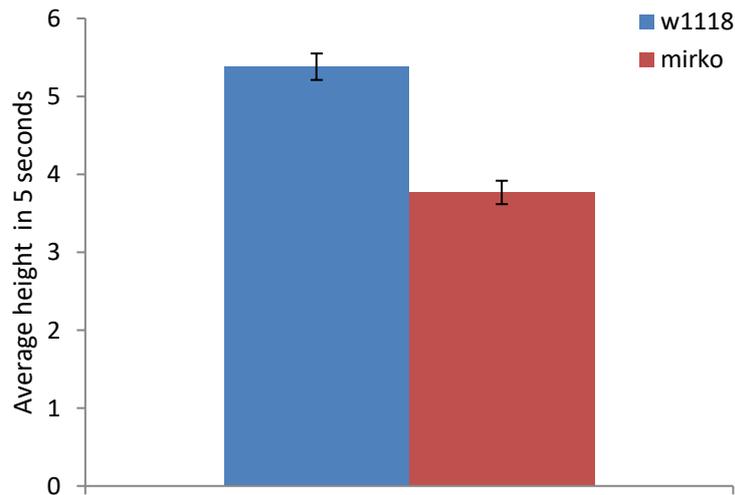
Figure 5.6: Mean height achieved in a negative geotaxis assay

Average height achieved of w^{1118} (n=59, repeated x5) and miR-137^{KO} (n=60, repeated x5) in a negative geotaxis assay. The assay was separated into 10 different ascending sections that were approximately 2cm wide. The control flies averaged 5.38 while the *miR-137* knockout flies reached 3.77, which is a significant difference (t-test p-value < 0.01).

5.4   Sleep-like behaviour

D. melanogaster is well known for having a sleep-like state comparable to human sleep. Like in humans, the fly sleep state is associated with an increased arousal threshold and less environment awareness (Cirelli *et al.* 2005). Another comparison to human sleep is the two separate regulatory systems, circadian rhythm and homeostatic regulation. Circadian pathways are relatively well understood and enables an organism to anticipate changes in environmental factors such as lights on or sunrise (Renn *et al.* 1999). The homeostatic system is less well known and is an area of current interest in scientific communities, responsible for

113

creating an increased need to sleep in response to stimuli or prolonged periods of wakefulness, and the cause for the sleep rebound effect seen after sleep-deprivation (Donlea *et al.* 2014). In literature, an inactive period of 5 minutes or more is associated with the above criteria and is therefore counted as a period of sleep-like behaviour (Cirelli et al. 2005; Cirelli 2009; Bushey et al. 2010). The sleep behaviour was calculated using the data from the DAMS system, with inactive 5 minutes periods counting as 1 sleep count. The DAMS was set up as explained previously in Chapter 2.7.

5.4.1   *Results*

The *miR-137^{KO}* flies have a sleep profile that shows high amounts of sleep throughout most of the 24-hour period when compared to controls (figure 5.7). At lights on and lights off (07:00 and 19:00 respectively) both the *w^{1118}* and *miR-137^{KO}* strains demonstrate a standard circadian sleep profile with lower sleep at the light change times, an inverted parabola of sleep during the light phase (07:00 – 19:00), and high sleep and gradual awakening throughout the night phase (19:00-7:00).
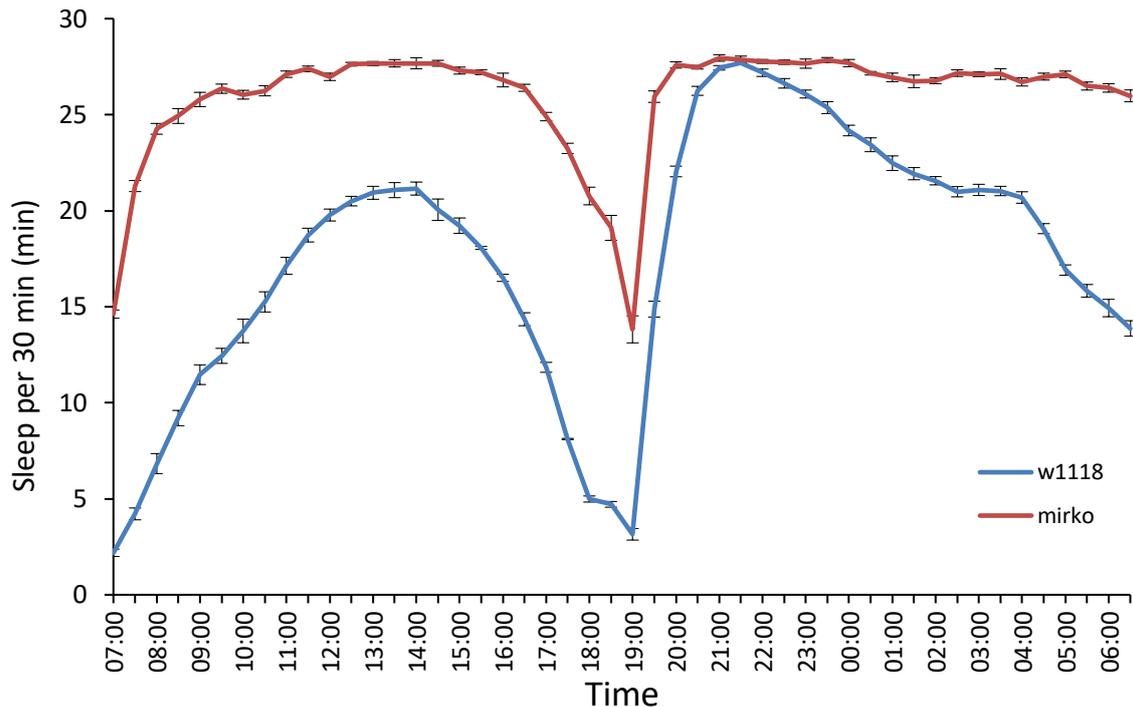
Figure 5.7: 24-Hour sleep profile in *w^{1118}* and *miR-137^{KO}*.

w^{1118} (n=410) and miR-137^{KO} (n=102). Sleep amount is shown for 30-minute bins across a 24-hour period, calculated using the Drosophila activity monitoring system (DAMS) by Trikinetics and Microsoft excel. The wild-type control flies can be seen to have a "standard" profile, with very low sleep amounts recorded at lights on and lights off (07:00 and 19:00 respectively), and an inverted parabola during daylight with high sleep gradually waking up towards the end of the night period. The knockout flies can be seen to have much higher sleep amounts throughout, with a much more stunted drop at the lights on and lights off times. Error bars are standard error.

Throughout a 24-hour period, *miR-137^{KO}* flies were shown to sleep for an average of 1240.4 minutes, compared to *w^{1118}* controls sleeping 816.7 minutes (figure 5.8). This was mostly a result of a marked increase in daytime sleep in knockouts, with controls sleeping a total of 316.4 minutes and knockouts 602.9 minutes (t-test p-value < 0.01). The total sleep during the night period was less severe but still significant (t-test p-value < 0.01), with controls sleeping 500.4 minutes and knockouts sleeping 637.5 minutes.
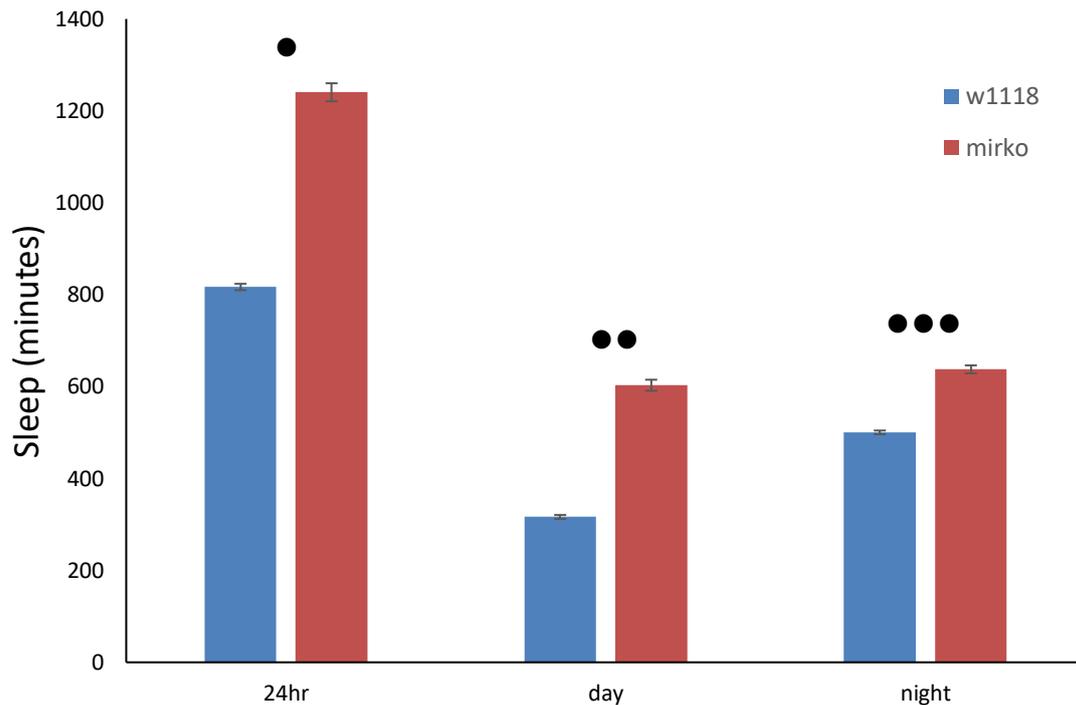
Figure 5.8: Total sleep amount in *w¹¹¹⁸* and *miR-137ᴷᴼ*.

w¹¹¹⁸ (n=322) and *miR-137ᴷᴼ* flies (n=77) throughout a 24-hour period and separated day and night 12-hour periods. The data is obtained by using the Drosophila activity monitoring system (DAMS) by Trikinetics and analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded on the DAMS, and measurements are taken across 5 days and averaged. When subjected to a t-test and subsequent Bonferroni corrections as described in Chapter 2.10, the control strain sleeps significantly less in total than the knockout strain (●: t-test p-value < 0.01), which is mostly due to a very high sleep amount during the day period for the miR-137ᴷᴼ flies. The differences in total sleep is significantly different in both the day and night periods (●● and ●●●: t-test p-values < 0.01). Error bars are standard error.

Sleep bout length is the measure of time between first recorded sleep behaviour and the next recorded activity in the DAMS system. This can vary from 5 minutes upwards. The control flies were found to have an average sleep bout length of 29.7 minutes for a 24-hour period, 19.4 minutes during the day, and 47.5 minutes during the night (figure 5.9). Comparatively the knockout strain was seen to have much longer sleep bout lengths of 95.6 minutes over a 24-hour period, 95.6 minutes in the day, and 113.9 minutes at night. Both the total 24-hour period and individual day/night are significantly different to controls when subjected to a t-test.
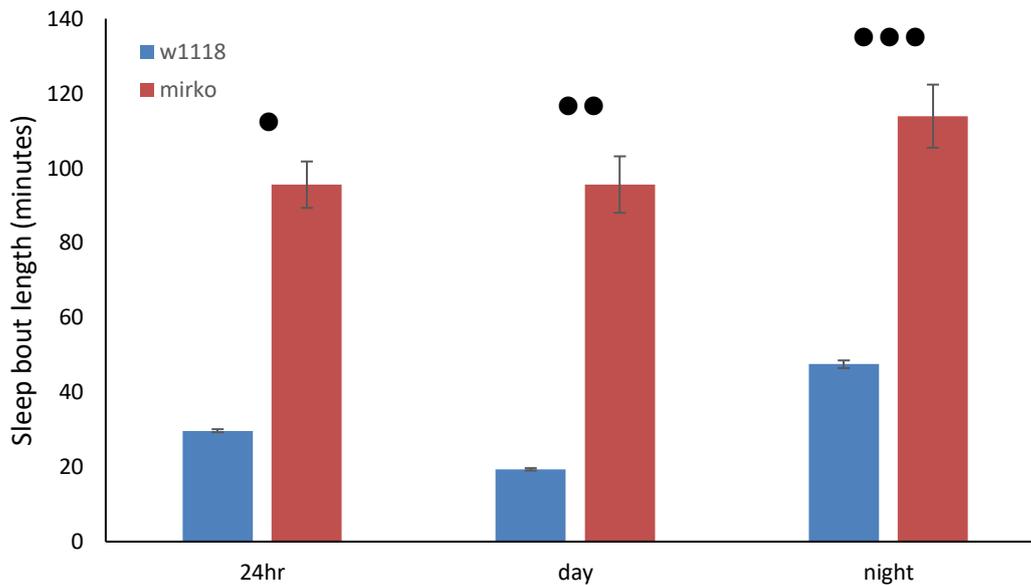
Figure 5.9: Mean sleep bout length in *w^1118* and *miR-137^KO*.

w^1118 (n=324) and *miR-137^KO* flies (n=77) throughout a 24-hour period and separated day and night 12-hour periods. The data is obtained by using the Drosophila activity monitoring system (DAMS) by Trikinetics and analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded on the DAMS, with recordings taken across 5 consecutive days and averaged. One sleep bout is the amount of time between first sleep period and when activity was recorded again. When subjected to a t-test and subsequent Bonferroni corrections as described in Chapter 2.10, the control strain sleep for significantly shorter periods when compared to the knockout strain (●: t-test p-value < 0.01). The differences in sleep bout length is significantly different in both the day and night periods (●● and ●●●: t-test p-values < 0.01). Error bars are standard error.

Alongside sleep bout length, the sleep bout number shows significant differences between control and knockout strains (figure 5.10). The knockouts have much lower number of sleep bouts across both periods of the day, particularly the day period with approximately 50% the number recorded. The knockouts demonstrate 18.3, 8.8, and 9.5 average sleep bouts for 24-hour, day, and night periods respectively. In comparison to 29.8, 18.3, and 12.8 in controls which displays significantly higher values.
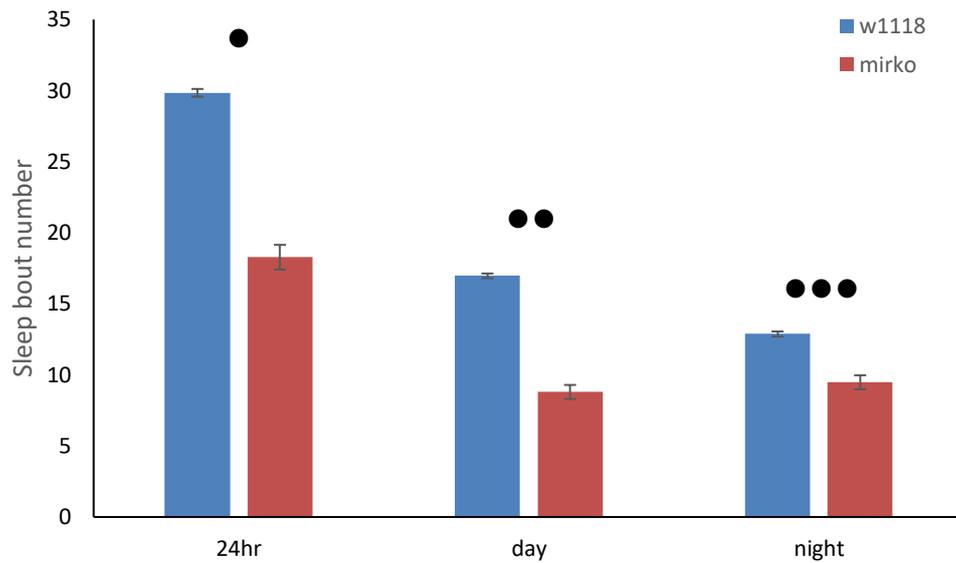
Figure 5.10: Mean sleep bout number in *w^{1118}* and *miR-137^{KO}*.

w^{1118} (n=324) and miR-137^{KO} flies (n=77) throughout a 24-hour period and separated day and night 12-hour periods. The data is obtained by using the Drosophila activity monitoring system (DAMS) by Trikinetics and analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings taken across 5 consecutive days and averaged. One sleep bout is the amount of time between first sleep period and when activity was recorded again. When subjected to a t-test and subsequent Bonferroni corrections as described in Chapter 2.10, the control strain sleep for significantly higher number of sleep bouts when compared to the knockout strain (*t-test p-value < 0.01). The differences in sleep bout number is significantly different in both the day and night periods (** and ***: t-test p-values < 0.01). Error bars are standard error.

To assess the internal circadian clock of the knockout strain, the flies were kept in 24-hour darkness for an extended period of time. The graphs show that there is a general confusion and diminished movement exhibited by both the control and knockout strains, but unfortunately no obvious trend seen due to the already minimal activity in the miR-137^{KO} flies (figure 5.11).
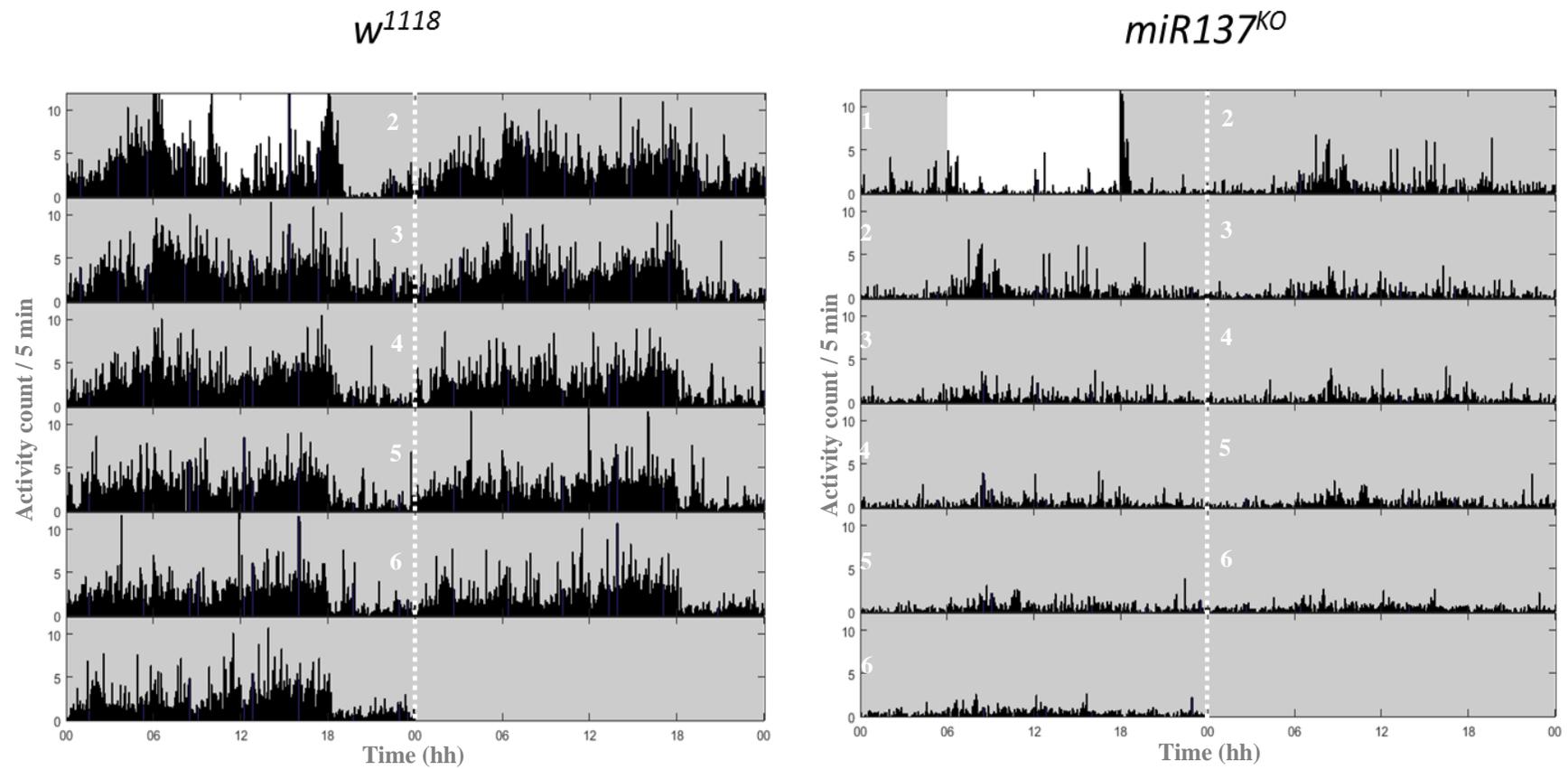
Figure 5.11: Circadian rhythm under constant dark (DD) conditions.

The lines represent average beam crosses in 5 minutes for each 30-minute slot measured in the Drosophila activity monitoring system (DAMS). A total of 16 males from each genotype were entrained to a normal LD rhythm with lights on and lights off at 6am and 6pm respectively. The experiment measured 1 day in normal 12:12 hour light/dark (LD) and following 5 days of constant darkness (DD). The days have been duplicated and put next to previous day to better visualise data, separated by the dotted line and are numbered accordingly. Shading denotes light and dark conditions.

5.5   Discussion

5.5.1   Activity

The hypoactive phenotype is a severe behavioural trait that could provide an important aspect of *miR-137* function. The 24-hour actograms clearly display minimal activity counts and the overall 24-hour activity change is circa 66% less in the $miR\text{-}137^{KO}$ strain than controls. This behavioural change signifies the importance of *miR-137* regulation, as the impact is obvious. Despite the overall activity defect, the circadian rhythm is not impacted and shows the usual peaks appearing at ZT 0 and ZT 12. There is also the gradual increase over night, inverted parabola during the day, and pre-emptive activity before the peaks (Renn *et al.* 1999; Cirelli et al. 2005; Cirelli 2009; Bushey et al. 2010). The biggest difference in activity is during the morning peak, which in controls was the highest point and therefore most active period of the day. In the *miR-137* knockout strain, the morning peak was stunted more so than the evening peak, implicating *miR-137* in pathways responsible for arousal from sleep.

The DAMS system is a quick and efficient set up that requires minimal researcher input and returns a wealth of data. This data can be used in different ways to measure multiple behaviours including sleep, circadian rhythm or sleep/wake pattern, and overall activity levels. It is also possible to look at specific periods of time, offering the ability to narrow down the results. The DAMS is, however, an indirect measure of activity which allows for some interpretation. A limitation of the DAMS system is the inability to separate sleep-like behaviour and locomotor behaviour, as defects within either or both will present similar data. To further define the contribution of sleep and locomotor behaviour to the phenotype, additional research is needed.

5.5.2   *Location and Velocity*

D. melanogaster have a well-documented circadian pattern with boosted activity at periods around light phase changes, resulting in actograms like the one produced by the DAMS. These periods offer the biggest comparative difference between the behaviour of w^{1118} and *miR-137^{KO}* flies and were therefore the times chosen to record to investigate the phenotype further.

Directly tracking fly location against time, as shown in the (figure 5.3), demonstrates an obvious difference between the two strains. The data from the DAMS can report a false hypoactive phenotype that is due to the DAMS array failing to detect the activity of flies not crossing the centre point of the tube. Less overall activity can also be indicative of reduced locomotor ability, with more time taken to traverse the length of the tube causing a decrease in the count of beam crosses. As can be seen from the video tracking graphs, the control w^{1118} flies display a rhythmic behaviour of a continuous up and down pattern. In comparison, the *miR-137^{KO}* flies have stationary periods and sporadic activity. However, despite the obvious difference in overall behaviour, neither strain demonstrates significant activity which do not cross the centre of the vial. This therefore validates the hypoactive phenotype and the use of the DAMS.

Relating the DAMS and the video tracking is not a fair comparison because of the obvious differences between the video recording and DAMS setup. The DAMS were set up within a climate-controlled incubator, with a constant temperature of 25°C and a fixed light source away from the vials. With the video recording, the external temperature was not controlled, and the light source was brighter and directly below the flies. There was also more background noise in the recordings as there was no protection like being within a closed

incubator. These are all factors which could have produced a change in the activity or behaviour of the flies between the two experiments.

The resolution of the video and subsequent editing to a black and white image-stack, then analysis in ImageJ software leads to some interference on pixel colour and location. This is the reason for the clearly never truly stationary fly, which is an artefact of the subtle light changes and fly extremities giving a different black pixel count for each image slice. This is a parameter which would be better to control for, but as the fly plots work to adequately locate and follow the movement of the fly and is similar in all experiments it is suitable for this experiment. A cleaner graph could be achieved with better equipment or software; however this is outside the scope of this project and the results shown are sufficient for evaluating the severe phenotype change in the *miR-137^{KO}* flies.

5.5.3   *Climbing*

The negative geotaxis test is a simple but effective assay to assess a variety of behavioural factors through climbing ability. However, like the DAMS, the negative geotaxis assay is not able to differentiate between more complex behaviours. The *miR-137^{KO}* are less capable of reaching a high section during the 5 second constraint of the test. The difference shown is significant when compared in a t-test, and the most obvious cause for achieving a lesser height is inability to climb or generally move as well as the controls, indicating a locomotor defect. There are other factors such as disordered climbing, lack of proper gravity sensing, or inability to cope with the initial stimulation of short but abrupt tapping/shaking. The decrease in climbing ability found in the *miR-137^{KO}* stock was also reported by Chen *et al.* 2014 and taken with the results from video tracking there is a locomotor defect as a result of *miR-137* LOF.

5.5.4   *Sleep-like behaviour*

Sleep in *D. melanogaster* is a widely accepted and understood concept that has been the subject of extensive research due to its similarities with human sleep and the organisms ease of use as a model. In this project, there is a severe increase in the sleep-like state of the *miR-137^{KO}* flies throughout a 24-hour period, particularly noticeable within the day (figure 5.7). There are two mechanisms behind sleep, the homeostatic pathways and the circadian pathways. The circadian rhythm is the biological clock which dictates the day/night activity or sleep cycle within a 24-hour period. The homeostat is the pathway which drives the accumulating need to sleep and is responsible for increased tiredness following an interrupted sleep period. *D. melanogaster* have both these pathways just as in humans, with a rebound effect of increased sleep following prolonged wakefulness and pre-emptive behaviour patterns around the light changes.

For flies, the circadian rhythm is well documented that in a 12:12 hour light-dark cycle, a "wild type" activity graph shows peaks at lights on and lights off, with a gradual build up before each. Dark period activity tends to decrease sharply after the evening peak and build gradually, whereas during the light period a uniform parabola shape dipping at ZT6 is most common (Renn *et al.* 1999; Cirelli et al. 2005; Cirelli 2009; Bushey et al. 2010). Any changes involving the circadian rhythm genes often result in a longer or shorter biological clock, altering the shape of the activity graph. While the *miR-137^{KO}* strain has decreased activity, they show the two characteristic peaks of activity when kept to a 12-hour light-dark cycle. To investigate possible circadian changes, *miR-137^{KO}* and *w^{1118}* flies were put into 24-hour darkness (DD) following entrainment to the standard 12:12 hour light-dark cycle. The low activity levels in the *miR-137^{KO}* strain means it is difficult to perceive the rhythm, but under DD conditions there does not appear to be any obvious circadian problem. This visual assessment points to the homeostatic sleep regulation pathways as the reason behind the

increased sleep amount and not issues within the circadian rhythm. However, further investigation is needed to validate such a conclusion.

Total sleep in this experiment set is an inverse measure of activity, which can be influenced by several factors that are not necessarily sleep-like behaviour. However, the total sleep data shown here has an extreme difference across both total 24-hour and daytime amount, with a lesser but still significant difference within the night period. The lesser difference at night-time is explained with a natural increase in sleep-like behaviour for controls during the night phase, as the organisms are diurnal. This increase brings the control total sleep amount closer to the threshold of 720 minutes (number of minutes in 12-hour period) and therefore dilutes the differences. However, the substantial difference within the daytime is a significant feature of *miR-137^{KO}* behaviour and with such a severe phenotype it is unlikely to be caused by other influences. The scale of the phenotype displayed in the knockout strain is severe even in comparison to screens performed in previous studies (Cirelli et al 2005).

Sleep bout length is the measure of amount of time spent asleep in each sleep period. This is a parameter less easily influenced by behavioural differences as it is more specific than simply a measurement of inactivity. The sleep bout length values show a similar trend to that of the total sleep amount, where all values are significantly higher in the *miR-137^{KO}* strain compared to the *w^{1118}* controls. The difference however is severe in both the 24-hour period and the day/night separated data, indicating that the sleep exhibited by the knockout strain is because of drastically increased sleep bout lengths. From this data alone, it is unclear whether this is due to an increased pressure to stay asleep or an issue within the arousal pathways consequently having issues waking up from a sleep period.

Average sleep bout number is not such a useful parameter by itself but can be used to deepen the understanding gained from total sleep and sleep bout length. Alongside these values the increased total sleep shown in the *miR-137^{KO}* is consolidated and much longer bouts of sleep accumulating to the high total sleep value. Furthermore, in controls the sleep is much more broken up (lots of short bouts), while the knockout strain displays a very similar pattern to the night period with a few long sleep bouts. Long periods of consolidated sleep reinforce the lack of waking ability or increase pressure to sleep hypothesized from the other graphs.

5.6   Conclusion

The loss of *miR-137* appears to have both a locomotor defect and a severe hyper-sleep effect on general fly behaviour. The most obvious result is a severe activity deficit within the *miR-137^{KO}* strain that, when compared to control flies within the same environment, gives a very high significance value and visually clear actogram. This outcome enables the rejection of the null hypothesis in objective B hypothesis 1 (Chapter 1.8). Initially the phenotype needed verification and dissection into different classes of behaviour, and through carefully selected experiments with the equipment available to this project the behaviour was determined to be both a locomotor defect and sleep-like behaviour issue.

The verification of the DAMS system via the video recording of peak activity times, performed a two-part function where the tracking of the fly within the vial verified the DAMS ability to read activity while also providing an ability to effectively measure average velocity for both test and control genotypes. With a further experiment of a simple but effective negative geotaxis test, the locomotor issue was well defined as a decrease in velocity shown by physical recording and lessened climbing ability within a short timeframe.

These results enable the rejection of the null hypotheses in objective B hypotheses 2 and 3 (Chapter 1.8).

The further investigation of sleep-like behaviour by inversely analysing the activity for sleep bouts, stated previously in literature as 5 minutes of inactivity. The result of which shows a drastic increase within overall sleep amount as well as average sleep bout length, both differences with high statistical significance. This outcome enables the rejection of the null hypothesis in objective B hypothesis 4 (Chapter 1.8). The video recording also displays that within the knockout flies there were several periods of quiescence where the flies appeared to stay still for short periods of time, despite the recordings being taken within the "most active" times of the day. This pause is likely a bout of sleep-like behaviour, further emphasising the effect of *miR-137* on sleep behaviour.

miR-137 is known to be enriched in the brain of an organism, potentially localising at synapse junctions. Many of the targets (both predicted and proven) are major players in brain neuron proliferation and maintenance, or vesicle release/signalling pathways. As such, lack of finetuning by the microRNA molecule appears to lead to issues that ultimately manifest as behavioural phenotypes that could potentially be explained with deficiencies within the neuron maintenance and synapse transmission regulation. Further investigation into the pathways that are governed by *miR-137* would be useful to better decipher the reasons behind the phenotype exhibited by the $miR\text{-}137^{KO}$ strain.

It is worth noting that, mainly due to limitations with the DAMS, these experiments were conducted solely on male *Drosophila* and further work to confirm the phenotype in females still needs to be completed.

6 *Phenocopying genomic loss of miR-137*

Further investigation of the effect of *miR-137* comes from trying to create the same phenotype in a different way. Phenocopying in this way can be a useful tool to further diagnose a behaviour and its cause. This study utilised both deficiency stocks from Bloomington, and the well published UAS/Gal4 system to attempt to reproduce the hyper-sleep phenotype to further understand the proper function of the *miR-137* gene.

6.1 *miR-137 knockout via deficiency strain*

The original *miR-137^{KO}* stock featured a substitution of the *miR-137* gene for a $w^{mW.hs}$ insertion. The result of which meant there was no genomic expression of *miR-137* at all, and the hyper-sleep and hypoactive phenotypes were displayed (chapter 5). As the stock was used straight from Bloomington, there was a chance that the results were false positives and a result of background mutations which exist within the culture. Normal protocols are to homogenise the DNA with the controls by backcrossing so that the DNA sequence except the region of interest is highly similar. In this project, many attempts at back crossing proved unsuccessful as the red eye marker did not behave normally and proved problematic and time consuming. In most other cases, copies of the red eye phenotype exhibit a darker colour in homozygous flies that can be easily identified and separated from the lighter eye colour of heterozygous individuals. During the repeated attempts to backcross, either no homozygous flies emerged due to homozygous *miR-137* fatality or the eye phenotype was indistinguishable between the two.

Therefore, to combat the likelihood of false results, the use of a deficiency cross by Bloomington was used to provide heterozygous and homozygous knockouts in cooperation with the original knockout line. The deficiency genotype was a partial deletion of several

genes which included the *miR-137* locus and surrounding areas. This would serve as a comparison to the original knockout line behaviour, while also being able to provide flies which have no *miR-137* expression due to heterozygous combinations of both strains. As such this provides a method through which to separate the likelihood of background mutations causing the sleep behaviour and further investigate homozygous and heterozygous knockout of the *miR-137* gene.

The DF(2R)ED2457 *D. melanogaster* strain was ordered directly from Bloomington (stock number: 8915). The stock exhibits a partial deletion on the 2R chromosome arm, that removes the coding regions of 25 genes including *miR-137* (table 6.1). While not a perfect substitution for backcrossing due to other gene implications, presence or absence of the hyper-sleep phenotype would indicate that loss of *miR-137* is the cause of such behaviour change or if the mutations required to enable surviving *miR-137* LOF are the cause.

Table 6.1: *D. melanogaster* **deficiency stock DF(2r)ED2457 affected genes.**

DF(2R)ED2457 affected genes	
Gene symbol	**Gene name**
Asph	Aspartyl β-hydroxylase
ATPCL	ATP citrate lyase
casp	caspar
CG30095	-
CG8370	-
CG8386	Ubiquitin-fold modifier conjugating enzyme 1
CG8388	-
CG8389	-
CG8397	-
CG8399	-
CG8401	-
CG8405	-
CG8414	-
CR43415	long-noncoding RNA
Dg	Dystroglycan
Jhe	Juvenile hormone esterase
Jhedup	Juvenile hormone esterase duplication
miR-137	microRNA-137
mRpL34	mitochondrial ribosomal protein L34
Prosbeta1	Proteasome β1 subunit
Rho1	-
Ric	Ras-related protein interacting with calmodulin
Rrp42	-
SP2353	-
spin	spinster

6.1.1   Method

Stocks that exhibited different combinations of *miR-137* knockouts were put into the DAMS system as described in Chapter 2.7. The genotypes used were controls, original knockout stock, heterozygous knockout stock and deficiency cross, and heterozygous deficiency control cross (table 6.2).

Table 6.2: Genotypes used in deficiency strain knockout experiment.

Strain	Full genotype	Expected *miR-137* expression
wild type	w^{1118}; +; +	normal ~100%
mirko	w^{1118}; $miR\text{-}137^{KO}$; +	none ~0%
hets	w^{1118}; $miR\text{-}137^{KO}$ /+; +	~50%
df/ko	w^{1118}; $miR\text{-}137^{KO}$ /DF(2R)ED2457; +	none ~0%
df/+	w^{1118}; DF(2R)ED2457/+; +	~50%

6.1.2   Results

Comparison of the two full knockout strains, $miR\text{-}137^{KO}$ and $DF(2R)ED2457/miR\text{-}137^{KO}$, show no significant difference across either the day or night period (table 6.3, figure 6.1). Both strains exhibit a severe hyper-sleep phenotype (over 1200 minutes per day) that is most obvious in the day period compared to w^{1118} controls. The heterozygous knockouts, $miR\text{-}137^{KO}/w^{1118}$ and $DF(2R)ED2457/w^{1118}$, showed a similarly increased sleep amount total at approximately half the increase shown by full knockouts over the 24-hour period. However, they separate in behaviour with the DF/+ strain having a significantly higher sleep amount during the day (t-test p-value < 0.01) and significantly lower at night (t-test p-value < 0.01). Full statistics can be found in supplementary information.

Table 6.3: Total sleep amount from deficiency strain knockout experiment.

	Total sleep amount (min [2dp])		
Strain	**24-hour**	**Day**	**Night**
w^{1118}	816.72	316.36	500.36
hets	966.25	413.75	552.50
mirko	1240.40	602.91	637.49
df/+	996.33	482.61	513.72
df/ko	1214.74	595.32	619.42
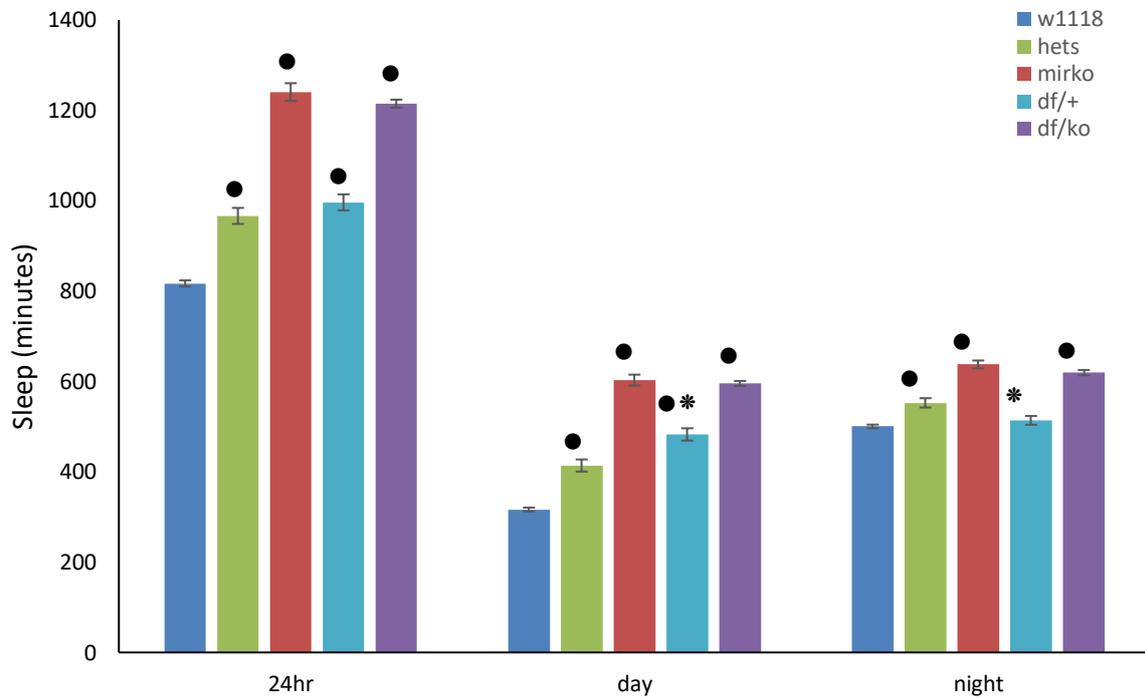
Figure 6.1: Total sleep amount in deficiency knockout experiment.

Average total sleep amount in minutes for w^{1118} (controls, n=324), miR-137^{KO} flies (mirko, n=77), miR-$137^{KO}/w^{1118}$ (hets, n=32), $DF(2R)ED2457/w^{1118}$ (df/+, n=18), and $DF(2R)ED2457/miR$-137^{KO} (df/ko, n=31) throughout a 24-hour period and separated day and night 12-hour periods. The data is obtained by using the Drosophila activity monitoring system (DAMS) by Trikinetics and analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings taken across 5 consecutive days and averaged. Genotypes with a sample size of less than 30 were subjected to the Shapiro-Wilk test for normal distribution, then all genotypes were compared to the w^{1118} strain individually using equal or unequal variance t-tests, with significant differences (p-value <0.05) signified by ●. The heterozygous knockouts (hets and df/+) were compared with a significant difference found in the day and night period separately, signified by ∗, though these differences were not significant following Bonferroni corrections as described in Chapter 2.10. Full knockouts were also compared but were not significantly different during any period. Error bars are standard error.

6.1.3 Discussion

The sleep phenotype shown in the *miR-137^{KO}* strain is of such a high severity that it was unlikely to have been caused by any random background modifier, but without backcrossing it still required verification. The results from original knockout stock and deficiency strain cross showed that with full loss of genomic *miR-137* there is a heightened sleep amount. Comparatively, the heterozygous knockouts show an increase that is half as severe as the full knockouts. When combining these results, it is not feasible for a background mutation to cause the phenotype as it would not be in both stocks. If the effect was solely due to a background mutation then the heterozygous knockouts and the full knockout/deficiency would be equivalent, each with one copy of the apparent background modifier. Instead the full knockout deficiency cross is comparable to the original homozygous *miR-137^{KO}* results, backing up the loss of *miR-137* as the cause. Following Bonferroni corrections as described in Chapter 2.10, the corrected p-value becomes 0.000794. All comparisons previously reported as significant meet this new threshold except the comparisons of the two heterozygous genotypes in the separated day and night. This loss of significance reinforces the similarities between the genotype behaviours.

The deficiency cross is not a perfect cover for lack of backcrossing, as there are multiple genes involved with the deletion. As such, not only is there a lack of *miR-137* expression in these stocks but there are also other genes missing. Some of the genes deleted with the deficiency strain have had links to behaviours such as hyperactivity and implicated within neural development.

6.2 RNA-sponge downregulation of miR-137

Drosophila melanogaster are a useful model organism for many reasons. One of which is the UAS/Gal4 system which has been widely used to manipulate genetic expression *in vivo* for research. The system is explained fully in Chapter 1, but essentially it is a bipartite system which composes of a promotor and a coding sequence of choice that are inserted separately into the genome of *D. melanogaster* at embryotic stages. These flies develop and can be cultured without much difficulty as the individual components of the system rarely have any effect. Upon crossing of the two parent flies, the offspring will have both parts of the system and will therefore have the inserted UAS sequence expressed as part of their genomic expression. This system can be used to express a copy of a genomic gene sequence or express a complementary section of a gene for RNA-interference downregulation. Marker genes such as fluorescence proteins or phenotype changes can also be included to validate the success of the UAS/Gal4 system.

For miRNA research, copies of the mature sequence can be inserted as the UAS sequence but, due to the small size and function of the miRNA, an RNAi approach is not possible. Therefore, to downregulate miRNA expression the UAS insertion encodes a sponge construct that binds and captures miRNA molecules, preventing them from functioning as normal. The sponge constructs are specific to individual miRNA molecules and can be expressed alongside other markers such as fluorescent proteins.

6.2.1   *Elav-Gal4*

Elav-Gal4 is a Gal4 promotor sequence attached to the genomic *embryonic lethal abnormal vision* gene (*elav*) gene, which is heavily expressed in the larval and adult CNS. If this construct is used alongside the UAS-miR-137-sponge insertion, it will drive the expression of the sponge construct alongside *elav* and downregulate *miR-137* expression in these areas. Genotypes used in the experiment are shown in table 6.4.

To help verify the success of the *elav-Gal4/UAS-miR-137-sponge* system in downregulation of *miR-137*, samples of dissected *D. melanogaster* brains were subjected to DNA extraction and subsequent comparative qPCR to measure *miR-137* expression. The procedure for extraction, quantification, reverse-transcription, and qPCR was carried out as described in Chapter 2.3-4.

Table 6.4: Genotypes used in *elav* driven knockdown experiment.

Strain	Full genotype	Description
w1118	w^{1118}*; +; +*	'Wild type' control flies
mirko	w^{1118}*; miR-137^{KO}; +*	*miR-137* knockout flies
Elav/sponge	*w*; +; elav-Gal4 / UAS-miR-137-sponge*	*miR-137* downregulation within the CNS
Elav/+	*w*; +; elav-Gal4 / +*	Gal4 heterozygous control
Sponge/+	*w*; +; UAS-miR-137-sponge / +*	UAS heterozygous control

6.2.2   Brain regions

Utilising different genes to attach the Gal4 construct enables a better targeted expression of the sponge construct and therefore more precise area of *miR-137* knockdown. *D. melanogaster* have a brain region known as the mushroom body that has been compared to the mammalian cerebral cortex and been linked to sleep regulation (Sitaraman *et al.* 2015). Another region of the *D. melanogaster* brain is the fan-shaped body, which is close in proximity to the mushroom body and similarly has been linked to sleep regulation in flies (Liu *et al.* 2012). Because of *miR-137* LOF on sleep behaviour and known high brain expression, two different Gal4 promotors were used to evaluate precise *miR-137* knockdown within these regions. Genotypes used in the experiment are shown in table 6.5.

The stocks used were GMR23E10-Gal4 (#49032) and R13F02-Gal4 (#69908) from Bloomington donated by Janelia farms. GMR23E10 causes expression near AstA-R1 which is highly localised to the fan-shaped body and nearby dopaminergic neurons, while R13F02-Gal4 expresses strongly within the mushroom bodies (Jenett *et al.* 2012).

Table 6.5: Genotypes used in brain region knockdown experiment.

Strain	Full genotype	Description
w1118	w^{1118}; +; +	'Wild type' control flies
mirko	w^{1118}; miR-137^{KO}; +	*miR-137* knockout flies
MB/sponge	w*; +; R13F02-Gal4 / UAS-miR-137-sponge	*miR-137* downregulation within the mushroom body
FB/sponge	w*; +; GMR23E10-Gal4 / UAS-miR-137-sponge	*miR-137* downregulation within the fan-shaped body
MB/+	w*; +; R13F02-Gal4 / +	Gal4 heterozygous control
FB/+	w*; +; GMR23E10-Gal4 / +	Gal4 heterozygous control
Sponge/+	w*; +; UAS-miR-137-sponge / +	UAS heterozygous control

6.2.3   Results

Comparative qPCR of brain samples was conducted as described in Ch2.5, with the samples run in duplicate on a 48-well plate to measure comparative expression values of *miR-137* and housekeeping gene *rps17*. In total there were 3 separate qPCR attempts with the same methodology, and the results repeatedly demonstrated that no expression levels of mature *miR-137-3p* were detected in the *miR-137^{KO}* sample. The standard curve and calculations for PCR efficiency, R^2 value, and slope gradient were automatically calculated by the Applied Biosystems Step-one software V2.3 (table 6.6). The calculated values for these parameters were not within applied biosystems recommended ranges and therefore the validity of the comparative gene expressions is questionable (table 6.6). The step-one software does allow for manual inclusion or exclusion of data points which can then be recalculated and offer better values for efficiency, R^2 and slope; however, as the samples were run in duplicates only this was not deemed suitable to provide clear results. Unaltered (fully included) gene expression graphs and standard curves are included in supplementary information (Chapter 12).

Table 6.6: Comparative qPCR efficiency, R^2, and slope results

qPCR Experiment Number	Standard Curve Slope	R^2 Value	PCR efficiency
Recommended values	-3.3	> 0.99	100%
1: miR-137	-3.59	0.92	91.66%
1: rps17	-4.25	0.97	71.91%
2: miR-137	-5.02	0.85	58.15%
2: rps17	-4.33	0.97	70.17%
3: miR-137	-3.70	0.90	86.99%
3: rps17	-3.59	0.98	92.42%

Flies with the sponge construct expressed under *elav* control demonstrate a hyper-sleep phenotype that is like that of the *miR-137^{KO}* flies (table 6.7; figure 6.2), but still significantly lower over the 24-hour period and during the day (p-values of 0.00017 and 3.57517E-07 respectively). The control strains *elav/+* and sponge/+ were largely comparable to the *w^{1118}* flies, with a small decrease in 24-hour *elav/+* sleep (p-value = 0.046). Full t-test results can be found in the supplementary section.

Table 6.7: Total sleep amount from *elav* driven knockdown experiment.

	Total sleep amount (minutes [2dp])		
Strain	**24-hour**	**Strain**	**24-hour**
w1118	816.72	316.36	500.36
mirko	1240.40	602.1	637.49
elav/sp	1151.53	516.80	634.73
elav/+	762.07	356.53	405.53
sponge/+	811.56	344.19	467.38
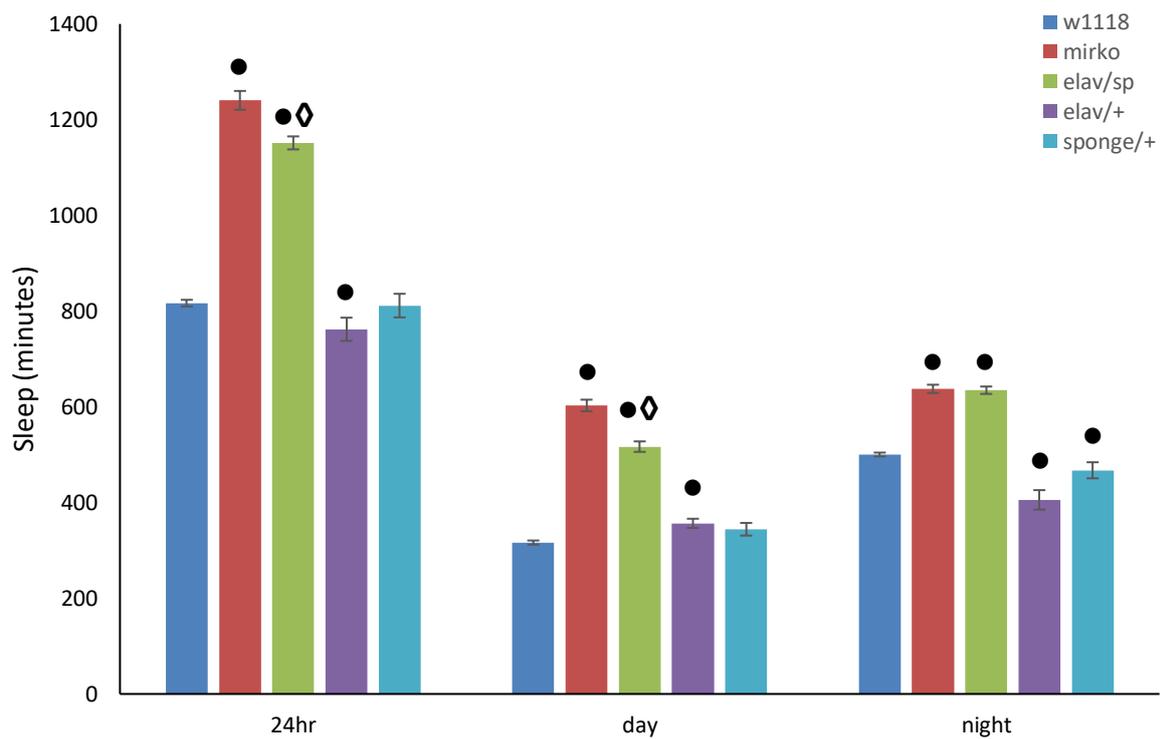
Figure 6.2: Total sleep amount in *elav* downregulation experiment.

The graph shows both the total 24-hour period, and the separation of the two 12-hour day/night phases. w^{1118} (n=324), $miR\text{-}137^{KO}$ (n=77), *elav*/sponge (n= 45), sponge/+ (n = 16), and *elav/+* (n= 15) flies were put into a Drosophila activity monitoring system (DAMS) by Trikinetics and the data was analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings taken across 5 consecutive days and averaged. Genotypes with sample sizes of less than 30 were subjected to the Shapiro-Wilk test for normality, and then all genotypes were compared to the w^{1118} strain individually using equal or unequal variance t-tests, with significant differences (p-value <0.05) signified by ●. The elav/sponge strain was compared to the $miR\text{-}137^{KO}$ strain in the same manner and significant differences are signified by ◊. Following Bonferroni corrections as described in Chapter 2.10, the difference shown across the 24-hour period between the *elav/+* control and the w^{1118} strain does not meet the new threshold. Error bars are standard error.

Expression of the UAS-sponge construct within the specific subregions of the brain produced similar effects on the total sleep, with an overall increase in sleep amount per 24 hours (figure 6.3; table 6.8). The amount of sleep is around halfway between the w^{1118} flies and the $miR\text{-}137^{KO}$ flies across all periods. The heterozygous control strains are comparable to the w^{1118} control flies, with any significant differences being a reduction in total sleep amount. Full t-test results can be found in the supplementary information.
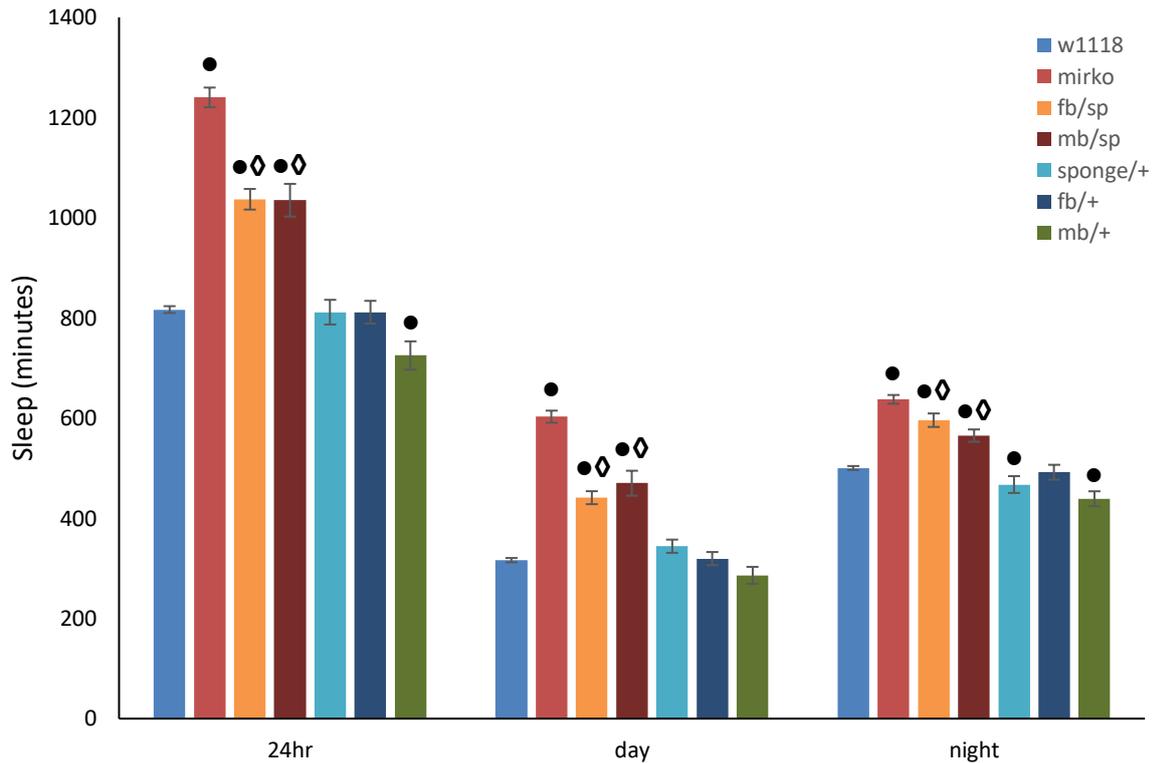
Figure 6.3: Total sleep amount in brain region downregulation experiments.

The graph shows both the total 24-hour period, and the separation of the two 12-hour day/night phases. w^{1118} (n=324), miR-137^{KO} (n=77), mb/sponge (n= 15), fb/sponge (n= 16), sponge/+ (n = 16), mb/+ (n = 16), and fb/+ (n= 16) flies were put into a Drosophila activity monitoring system (DAMS) by Trikinetics and the data was analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings taken across 5 consecutive days and averaged. Genotypes with sample sizes of less than 30 were subjected to the Shapiro-Wilk test for normal distribution, and then all genotypes were compared to the w^{1118} strain individually using equal or unequal variance t-tests, with significant differences (p-value <0.05) signified by ●. The fb/sponge and mb/sponge strains were compared to the miR-137^{KO} strain in the same manner and significant differences are signified by ◊. Error bars are standard error.

Table 6.8: Total sleep amount from brain region knockdown experiments.

Strain	Total sleep amount (minutes [2dp])		
	24-hour	Day	Night
w1118	816.72	316.36	500.36
mirko	1240.40	602.91	637.49
sponge/+	811.56	344.19	467.38
fb/sp	1037.00	441.13	595.88
fb/+	811.44	319.50	491.94
mb/sp	1035.07	470.00	565.07
mb/+	725.06	286.06	439.00

6.2.4   *Discussion*

The comparative qPCR results repeatedly demonstrated a complete loss of *miR-137* expression within the knockout strain, reinforcing the work done in chapter 3. However, the other data collected did not meet the recommended requirements (efficiency, R^2, and standard curve slope) and should not be considered. Additionally, results had large standard deviations and expected outcomes were not found, such as the heterozygous KO which should have shown approximately half the expression of *miR-137* as the controls further reducing confidence in the results. The method for the expression analysis was conducted as explained in chapter 5.2 and was designed to enable an investigative approach for several genotypes. The number of genotypes put onto one 48-well plate meant that duplicate samples were used instead of triplicates so they could all fit into one run. Further areas for complications include the intricate process for microRNA extraction using phenol-chloroform and separate procedures for amplification of the housekeeping gene (poly-adenosine primers for total cDNA) and the specialised microRNA amplification kits. There is also the potential for genomic differences across individual flies, with sponge or overexpression constructs having differing levels of effectiveness. Lack of robust results ultimately meant that the Oregon control strain and *miR-137* overexpression strain were not investigated further as it was impossible to evaluate the expression levels. The overexpression strain was put through a pilot test for altered behaviour patterns and did not appear to display any phenotype, so there was no reason to believe the constructs caused an expression change. This effect was not mimicked in the sponge regulated knockdown as there was a phenotype difference when compared to the *w1118* wild-type and the heterozygous construct controls. An obvious improvement to the qPCR method would be to drop the additional phenotypes and focus on using triplicates to increase validity before further trouble shooting. There was also no randomisation of sample location within the plate, which could have allowed machine errors

to occur across all runs without knowing. Utilising a different method of microRNA extraction, such as commercially available silica-based bind-wash-elute columns, may also increase the effectivity of extraction and reduce contaminants being taken through to the reverse transcription and amplification process. Should the comparative expression approach be discarded, it is also possible to purchase known quantity solutions that can be used as the compared sample to produce actual values of the expression in each sample which may have produced more finite results.

The *miR-137^{KO}* strain was donated to BDSC in 2014 by Stephen Cohen (A*STAR institute of molecular and cell biology). Since this time, the stocks will have been self-propagating under the care of the BDSC with a new generation approximately every 2 weeks. With a phenotype of this severity (Chapter 5), there is a possibility that background modifiers will have occurred that help the organism cope with *miR-137* LOF. This could be a random mutation within a gene that alleviates the strain on the individual's biology. If this partially rectified any of the health phenotypes found in Chapter 4, then these flies were more likely to reproduce and after several generations most of the stock would contain this modifier. If these modifiers were the cause of the sleep phenotype then *miR-137* is not likely to be directly responsible, and therefore additional checking was needed to validate the importance of *miR-137*.

The experiments had varying degrees of success in replicating the phenotype of *miR-137^{KO}* strain. The use of heterozygous controls successfully validates the phenotype seen is as a result of the combined UAS/Gal4 construct, as no controls were significantly higher that would look to cause a false representation. When taking into consideration that the *miR-137^{KO}* strain has a complete genomic loss of *miR-137*, the sleep amount seen when the sponge construct downregulates the miRNA in any of the experiments is impressive. The highest amount of sleep per 24-hour period was in the *elav*-Gal4 flies. Expressing the miR-

137-sponge in just the FB and MB also demonstrated a significantly higher sleep per 24-hours. Since these experiments targeted specific regions, this infers these brain regions are part of the mechanism through which *miR-137* regulates proper sleep function. Following post-hoc Bonferroni corrections, all significance tests meet the new conservative p-value with the exception of the *elav/+* control and the wild-type across the 24-hour period. This benefits the experiment by removing the significant difference between controls, further reinforcing the effect shown by the combined UAS and Gal4 constructs.

6.3   Conclusion

The excessive sleep phenotype was also demonstrated in a heterozygous cross of the original knockout stock and another deficiency strain, which reaffirms that *miR-137* LOF is the cause of the behaviour. This outcome enables the rejection of the null hypothesis in objective C hypothesis 1 (Chapter 1.8).

Downregulation of the miRNA through the UAS/Gal4 construct implicates the function of *miR-137* as within the CNS, and even more specifically related to the homeostatic sleep control by the FB and MB brain regions. These are regions of interest related to many higher cognitive functions like learning and memory in flies, and speech in humans. These results enable the rejection of the null hypotheses in objective C hypotheses 2, 3, and 4 (Chapter 1.8). Altogether, these results provide a putative mechanism and brain region for the sleep phenotype in the *miR-137^{KO}* strain.

Unfortunately, the qPCR was not successful in producing robust gene expression data, though it consistently confirmed the lack of *miR-137* expression in the original knockout strain. Further investigation of methods is needed. Ultimately qPCR data would have been useful to establish the quantitative effect of the sponge knockdown, however the use of

heterozygous construct controls and phenotype changes validates that the sponge system

causes a behavioural change through *miR-137*.

7   *Recovery of normal sleep in miR-137^{KO}*

To better evaluate the mechanisms behind the loss of *miR-137* leading to the phenotype exhibited by the flies, attempts to rectify the underlying cause of the behaviour were undertaken. By attempting to reintroduce *miR-137* expression, in specific regions of the CNS or utilising pharmaceuticals to cure the symptoms this chapter aims to provide additional information through which to hypothesise on the mechanisms linking *miR-137* to sleep and locomotor defects.

7.1   *Reintroduction of miR-137 expression*

The genomic excision of the *miR-137* gene led to a complete loss of the gene throughout the flies, which were then shown to have sleep and locomotor defects (Chapter 5). To further investigate this, reintroduction of *miR-137* gene expression through use of the UAS/Gal4 system to the *miR-137^{KO}* stock was conducted. As the normal expression pattern for *miR-137* was in the brain, one experiment aimed to reintroduce expression within the CNS in a hope to reverse the phenotype. As the main phenotype expressed was a severe sleep defect, another set of experiments were to reintroduce expression within the fan-shaped body, a specific sub region of the brain related to sleep regulation (Donlea *et al.* 2011; Liu *et al.* 2012; Pimental *et al.* 2016).

7.1.1.1 CNS expression

Utilising *elav*-GAL4 driven UAS-*miR-137*, an attempt to reintroduce *miR-137*
expression to the CNS of *D. melanogaster* with *miR-137^{KO}* genotype. The *embryonic lethal
abnormal vision* gene (*elav*) is heavily expressed in the larval and adult CNS, and with it will
also express the UAS insertion of the coding sequence for *miR-137*. Data available from
FlyAtlas (available online: http://flyatlas.gla.ac.uk/) shows the Fragments Per Kilobase of
transcript per Million mapped reads (FPKM) value of the gene expression and the enrichment
score. The enrichment is a measure of the abundance of a gene in a particular tissue relative
to that in the whole fly and can be used to determine whether the expression of a particular
gene is tissue-specific or not. The highlighted figures are the highest enrichment levels which
show a particularly specific *elav* expression across the CNS, head, eye, and Thoracico-
abdominal ganglion (table 7.1).

Table 7.1: *Elav* **gene expression location and enrichment.**

Gene: *elav*	Adult Male			Larval		
Tissue	FPKM	SD	Enrichment	FPKM	SD	Enrichment
Whole body	5.62	0.38	1	2.15	0.36	1
Head	21.5	1.49	3.83	-	-	-
Eye	14.2	3.22	2.53	-	-	-
Brain / CNS	25.48	0.86	4.53	42.54	9.92	19.79
Thoracico-abdominal ganglion	14.36	2.35	2.56	-	-	-
Crop	4.87	0.09	0.87	-	-	-
Midgut	2.34	0.09	0.42	0.86	0.23	0.4
Hindgut	3.91	0.41	0.7	1.21	0.24	0.56
Malpighian Tubules	2.35	0.4	0.42	1.31	0.36	0.61
Fat body	1.81	0.52	0.32	2.1	0.35	0.98
Salivary gland	2.46	0.27	0.44	0.71	0.17	0.33
Heart				-	-	-
Trachea	-	-	-	1.98	0.25	0.92
Ovary	-	-	-	-	-	-
Virgin Spermatheca	-	-	-	-	-	-
Mated Spermatheca	-	-	-	-	-	-
Testis	0.76	0.31	0.14	-	-	-
Accessory glands	2.11	0.35	0.38	-	-	-
Carcass	2.66	0.52	0.47	3.62	0.51	1.68
Rectal pad	2.66	0.34	0.47	-	-	-

7.1.1.2 Fan-shaped body brain expression

The stock used was GMR23E10-Gal4 (#49032) from the BDRC donated by Janelia farms. GMR23E10 causes expression near astA-R1 proteins which are highly localised to the fan-shaped body and nearby dopaminergic neurons (Jenett *et al.* 2012). Data available from FlyAtlas (available online: http://flyatlas.gla.ac.uk/) highlights the highest enrichment levels, which show a particularly specific *astA-R1* expression across the CNS, head, eye, and Thoracico-abdominal ganglion (table 7.2). Reports from Janelia demonstrate, through expression of a fluorescence protein, that the Gal4 construct drives expression specifically in the fan-shaped body and some projecting neurons (figure 7.1).
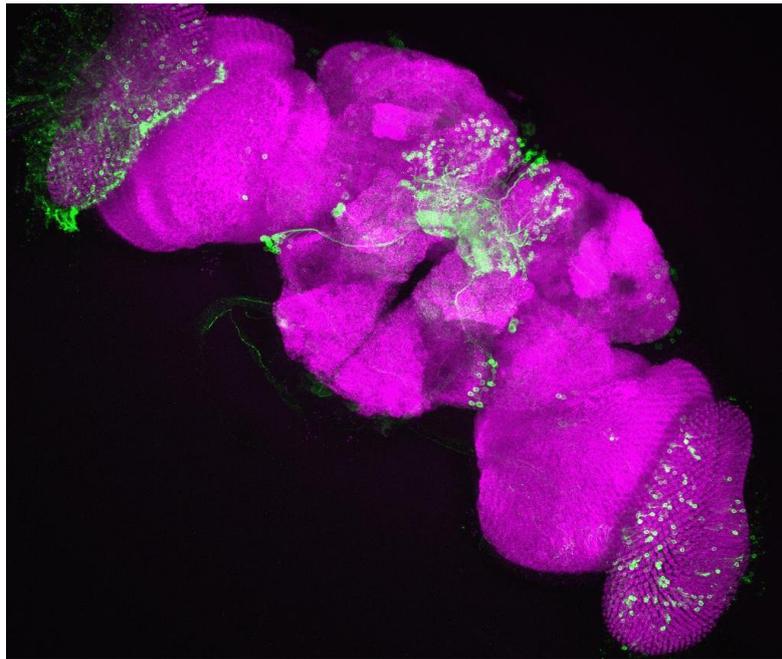
Figure 7.1: Expression of GMR23E10-Gal4 in the *D. melanogaster* brain.

Image obtained from Janelia research centre online reports (available from: https://www.janelia.org/). Image shows the results of a fluorescence protein UAS insertion driven by GMR-23E10-Gal4 construct. The brain is shown in magenta, and the fan-shaped body and projecting neurons in green (Jenett et al. 2012).

Table 7.2: *AstA-R1* gene expression location and enrichment.

Gene: astA-R1	Adult Male			Larval		
Tissue	FPKM	SD	Enrichment	FPKM	SD	Enrichment
Whole body	1.45	0.26	-1	0.09	0.04	-1
Head	8.58	0.86	4.29	-	-	-
Eye	8.61	1.43	4.3	-	-	-
Brain / CNS	9.99	1.53	5	2.26	0.26	1.13
Thoracico-abdominal ganglion	5.16	0.05	2.58	-	-	-
Crop	0.08	0.04	-1	-	-	-
Midgut	0.07	0.03	-1	0.01	0.02	-1
Hindgut	0.02	0.03	-1	0.01	0.02	-1
Malpighian Tubules	0.05	0.06	-1	0	0	-1
Fat body	0	0	-1	0.13	0.13	-1
Salivary gland	0.12	0.09	-1	0.05	0.07	-1
Heart				-	-	-
Trachea	-	-	-	0.07	0.05	-1
Ovary	-	-	-	-	-	-
Virgin Spermatheca	-	-	-	-	-	-
Mated Spermatheca	-	-	-	-	-	-
Testis	0.04	0.06	-1	-	-	-
Accessory glands	0.02	0.03	-1	-	-	-
Carcass	0.55	0.23	-1	0.34	0.15	-1
Rectal pad	0.02	0.02	-1	-	-	-

7.1.2 Method

Two stocks of heterozygous *miR-137^{KO}* with either UAS or Gal4 construct were crossed and the offspring were selected against genetic balancers to identify flies which have both endogenous *miR-137* knockout and both parts of the UAS/Gal4 system. These flies should therefore have the inserted copy of *miR-137* expressed in the regions driven by the gene with Gal4 attached to it. Control strains with heterozygous copies of all insertions individually were also assessed in case any phenotype shown is a result of the insertion not the expression of the construct (table 7.3). Once these adults were collected, they were put through the DAMS system and the data was analysed as described in Chapter 2.7.

Table 7.3: Genotypes used in genetic recovery experiment.

Genotype	Description
w^{1118}	Control strain
Mir-137^{KO}	*miR-137* knock out strain
elav Expressed	Elav-Gal4 driven expression of UAS-miR-137 in the CNS
FB Expressed	GMR23E10-Gal4 driven expression of UAS-miR-137 in the fan-shaped body and projected neurons
FB/+	Heterozygous GMR23E10-Gal4 control strain
elav/+	Heterozygous elav-Gal4 control strain
UAS-miR-137/+	Heterozygous UAS-miR-137 control strain

7.1.3   Results

The total amount of sleep (minutes) across a full 24-hour period show a slight decrease in both recovery attempts (figure 7.2, table 7.4), but only the FB expressed is significantly different from the *miR-137^{KO}* strain (p-value < 0.05). All controls are similar to the control *w^{1118}* strain, with the exception of *elav/+* which is significantly lower (p-value < 0.01). During the 12-hour day period, neither of the recoveries show a significant difference to the *miR-137^{KO}* stock but the controls *elav/+* and *UAS-miR-137/+* are different from the *w^{1118}* strain (p-values of < 0.01 and < 0.05 respectively). During the night phase, both recovery strains slept for significantly less time compared to the *miR-137^{KO}* (p-values of < 0.01). However, they were not significantly different from the control *w^{1118}* flies. The full statistics can be found in the supplementary information.
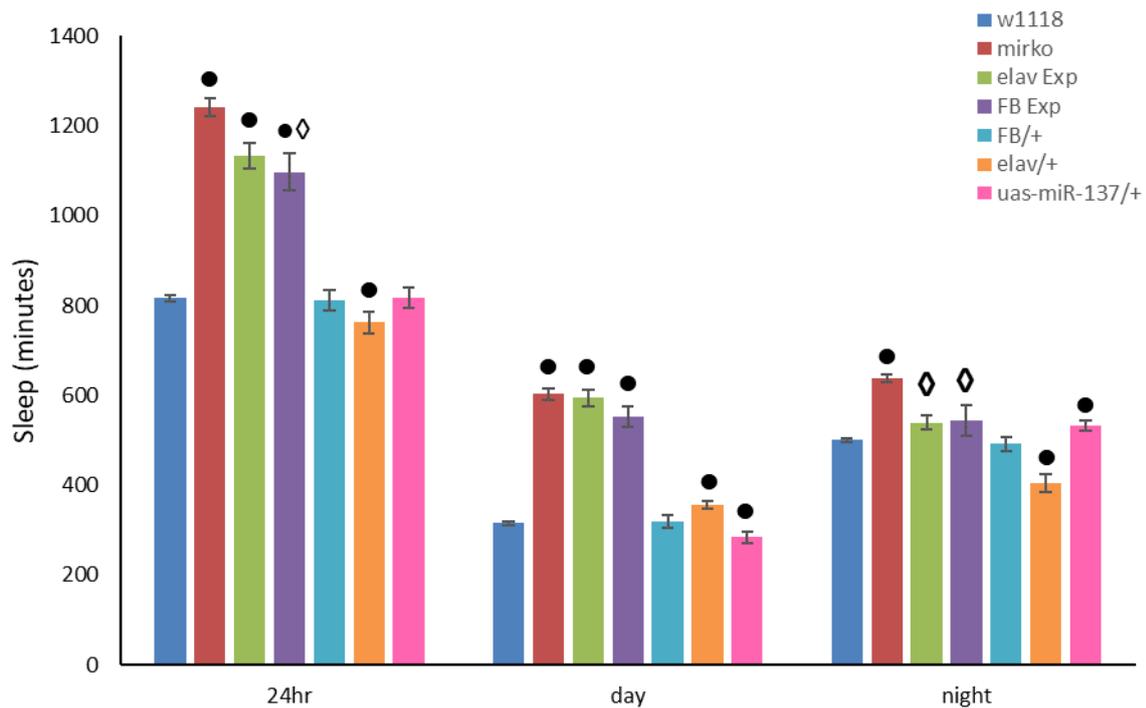
Figure 7.2: Total sleep amount in *miR-137* recovery experiments.

The graph shows both the total 24-hour period, and the separation of the two 12-hour day/night phases. w[1118] (n=324), miR-137[KO] (n=77), elav expressed (n= 4), FB expressed (n = 5), elav/+ (n= 15), FB/+ (n = 16), and UAS-miR-137/+ (n = 16) flies were put into a Drosophila activity monitoring system (DAMS) by Trikinetics and the data was analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings taken across 5 consecutive days and averaged. Genotypes with sample sizes of less than 30 were subjected to the Shapiro-Wilk test for normality, and then all genotypes were compared to the w[1118] strain individually using equal or unequal variance t-tests, with significant differences (p-value < 0.05) signified by ●. Both recovery strains were compared to the miR-137[KO] strain in the same manner and significant differences are signified by ◊. Post-Hoc Bonferroni corrections reduced the critical threshold p-value to 0.000794, and several comparisons did not meet this new value. These were: the FBrecov strain compared to the KO strain across 24 hours and the night period, the elavrecov strain compared to the KO in the night period, the *uasmir137/+* control against the wildtype across the day and night periods, the *elav/+* and *mb/+* controls against the wildtype across the 24 hours period. Error bars are standard error.

Table 7.4: Total sleep amount from *miR-137* recovery experiments.

Genotype	24hr sleep (min)	Day sleep (min)	Night sleep (min)
w[1118]	816.7205	316.3602	500.3602
Mir-137[KO]	1240.403	602.9091	637.4935
elav Expressed	1133.25	594.5	538.75
FB Expressed	1097.2	553.4	543.8
FB/+	811.4375	319.5	491.9375
elav/+	762.0667	356.5333	405.5333
UAS-miR-137/+	817.625	284.375	533.25

150

7.1.4 Discussion

Assumed reintroduction of *miR-137* expression within the CNS (*elav* recovery) appears to have little effect recovering the hyper-sleep phenotype demonstrated in the *miR-137^{KO}* strain (figure 7.2). However, expression in the fan-shaped body shows a significant decrease in the total sleep amount exhibited in 24-hour period (p-value < 0.05). Furthermore, this was due to a slight decrease during the daytime total sleep and a significant decrease in the night period, which was not significantly different to night-time sleep in the *w^{1118}* controls. The most exaggerated phenotype change in the knockout flies is an increase in sleep during the day phase. This could be due to a saturation effect at night, where there is a finite number of minutes available in the timeframe, so it is not possible to physically sleep more at this time. In this experiment, the daytime sleep for both recovery strains were not significantly different from the knockout strain indicating that either the recovery was not successful or that reintroduction of the miRNA into just these regions is insufficient for normal behaviour. There was an effect in the night phase exhibited by both recovery strains, with significant decreases compared to the knockout flies and no significant difference when compared to the controls.

The fan-body is a region of the *D. melanogaster* brain, which is linked to sleep regulation, specifically that of the homeostatic control (Donlea *et al.* 2011; Liu *et al.* 2012; Pimental *et al.* 2016). Activation of the region induces a "need to sleep" and is signalled by the lack of dopamine expression (Liu *et al.* 2012). Interestingly, the method of activation involves function of potassium voltage gated channel *shaker,* and subunits *hyperkinetic* and *shab* in the fly, orthologs of human KCNA/B protein channels. These channels are well documented to have implications in signal transduction and development of neurological symptoms (Somers *et al.* 2010; Askland *et al.* 2012; Imbrici *et al.* 2013). In addition, several of the *shaker* related genes are predicted binding targets for *miR-137* (Agarwal *et al.* 2015).

Hyperactivation of the fan-body, utilising the same GMR23E10-Gal4 construct to drive increased expression of the depolarizing Na+ channels, resulted in an exaggerated hyper-sleep phenotype (Ni *et al.* 2019), like that displayed by the *miR-137^{KO}* strain. Cumulatively, these results provide a potential hypothesis for the relevance of *miR-137* regulation of sleep and the onset of neurological disorders.

Correction using the post-hoc Bonferroni method (described in Chapter 2.10) produced a corrected p-value of 0.000794 which meant several comparisons were no longer significant. Reinforcing the project results was loss of significant differences between several of the control strains and the wildtype flies: *elav/+* and *mb/+* across the 24-hour period, and *uasmir137/+* across both day and night periods. The new threshold also meant that FB-recovery during the 24-hour period and the night period, and the elav-recovery during the night period are not significantly different from the KO strain. This outcome removes the success of the recovery attempts, though as discussed in Chapter 2.10, the Bonferroni correction method is highly conservative. Visual representation demonstrates a difference between the original KO strain and these recovery attempts (figure 7.2), presenting an argument for further evaluation into false negatives (Type II errors) at such a stringent level. Furthermore, due to the random chance of offspring having the right genetic insertions (identified through visual lack of phenotype markers), the numbers of offspring available for the DAMS at the correct time was very limited. This does provide a degree of error when assessing phenotype behaviour as this can be a very flexible measure. Additionally, the expression level of the insertion is unknown due to lack of ability to perform the tests in these areas. The inclusion of heterozygous controls, as well as the original knockout and control strains, work well to evaluate the likelihood of the phenotype changes being as a result of the combined UAS/Gal4 constructs.

7.2   *Pharmacological treatment*

In humans, sleep abnormalities are usually treated through pharmacological chemicals that influence parts of the brain. These medicines treat only the symptom, such as too much or too little sleepiness, and not the underlying cause. As the *miR-137^{KO}* strain exhibits a hyper sleep phenotype, experiments using some pharmacological chemicals were conducted to see if the symptom of *miR-137* loss can be treated in *D. melanogaster*. Furthermore, relatable phenotype changes due to the treatment can help identify abnormal pathway functionality.

7.2.1   *Caffeine*

Caffeine is an example of a widely used substance which is taken to alleviate feelings of tiredness and energise the individual. Commonly taken as coffee or within energy drinks for its stimulating effect, the compound acts on the CNS competing with adenosine molecules (Wu *et al.* 2009). Adenosine binding in the brain results in a depressive effect and subsequent drowsiness or fatigue. In the presence of caffeine molecules, the binding sites for adenosine are inhabited by caffeine compounds with an effect that ultimately leads to a stimulating effect (Lehnig *et al.* 1992; Wu *et al.* 2009). Caffeine has been linked to hyperactivity, sleeplessness, and over time chronic administration can cause an increase in adenosine receptors (Lehnig *et al.* 1992).

As *miR-137^{KO}* flies exhibited a hyper-sleep phenotype, an experiment of adding caffeine to their food was conducted to evaluate the recovery of normal behaviour. Initially the dose of caffeine was 0.5mg/mL, as was done in a previous study (Wu *et al.* 2009). However, this concentration produced little to no effect and after further investigation the study used a different food type that flies tend to eat in larger volumes and therefore the concentration was doubled to 1mg/mL.

7.2.1.1 Method

Caffeine was administered for 5-7 days to 8 w^{1118} and 8 *miR-137^{KO}* adult flies before being put into the DAMS system alongside 8 controls for each genotype. Pupa were removed from a normal food vial and moved into a vial containing food containing either caffeine dosed at 1mg/mL. After the flies eclosed and reached the correct age, they were added to the DAMS with drugged food. The DAMS and subsequent analysis were conducted as described in Chapter 2.7.

7.2.1.2 Results

Addition of 1mg/mL caffeine to food of the flies caused a significant decrease in total sleep amount for the w^{1118} strain compared to controls across the 24-hour period (p-value $<$ 0.05), which is the combination of no day time effect (p-value $>$ 0.05) and a significant decrease in the night period (p-value $<$ 0.01). The *miR-137^{KO}* strain exhibited contradictory results of a significant increase during the day phase (p-value $<$ 0.05). Like the w^{1118}, the knockout strain displayed a significant reduction of sleep in the night period (p-value $<$ 0.05), however the overall change was an insignificant slight increase across the 24-hour period (table 7.5; figure 7.3).

Table 7.5: Total sleep amount from chronic caffeine recovery experiment

	Average sleep amount (minutes [1dp])		
Experiment	**24-hour**	**Day period**	**Night period**
Normal: W^{1118}	784.1	251.0	533.1
Normal: *miR-137^{KO}*	1257.6	588.0	669.6
Caffeine: W^{1118}	593.3	215.1	378.1
Caffeine: *miR-137^{KO}*	1286.3	643.3	643.0
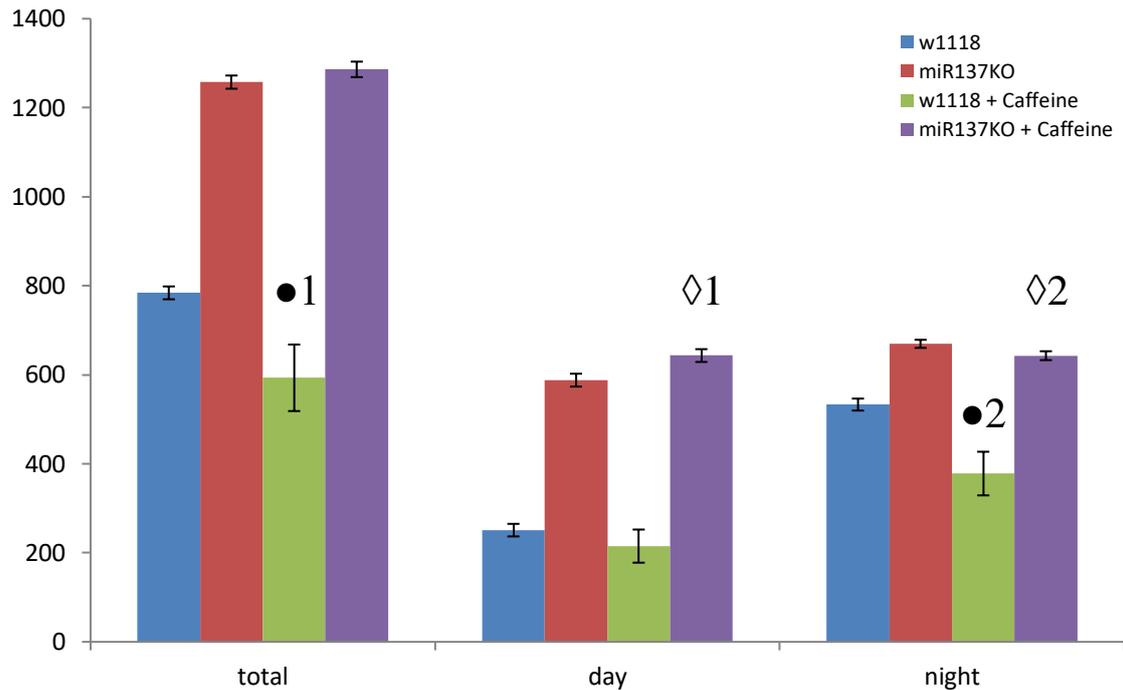
Figure 7.3: The effect of chronic caffeine treatment on total sleep amount

The graph shows both the total 24-hour period, and the separation of the two 12-hour day/night phases. w^{1118} (n=8), $miR\text{-}137^{KO}$ (n=8), w^{1118} with caffeine (n=8), and $miR\text{-}137^{KO}$ with caffeine (n=8) flies were put into a Drosophila activity monitoring system (DAMS) by Trikinetics and the data was analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings taken across 5 consecutive days and averaged. Experiments were compared between the treated and non-treated flies individually using t-tests. In the 24-hour period the w^{1118} flies showed a significant decrease in total sleep ($\bullet$1: p-value < 0.05), while the $miR\text{-}137^{KO}$ flies showed a significant increase during the day period ($\Diamond$1: p-value < 0.05. Both genotypes showed a significant decrease in total night-time sleep amount ($\bullet$2: p-value < 0.01, $\Diamond$2 p-value < 0.05). Following post-hoc correcting for Type I errors using the Bonferroni method as described in chapter 2.10, the caffeine difference exhibited in the wildtype flies across the 24-hour period and the difference in the KO strain during the day and night periods did not meet the corrected p-value of 0.00833. Error bars are standard error.

7.2.1.3   *Discussion*

The use of caffeine as a stimulant to promote wakefulness is a normal occurrence in human daily life, and as such the mechanism of its effect is largely understood. The control flies demonstrated an overall decrease in sleep amount, particularly within the night phase when there is more average sleep volume to be disrupted. This data reinforces numerous other studies which show caffeine intake has an immediate sleep deprivation effect in *Drosophila melanogaster* controls. Potdar *et al.* (2018) fed caffeine to control strain w^{1118} in

the day or night exclusively at concentrations of 0.5mg/ml and 1mg/ml as was in this study, however only for acute periods of time. They reported an immediate decrease in average sleep amount by ~50% in response to 0.5mg/ml treatment, and ~20% in response to 1mg/ml treatment which they explained as likely due to less food consumption because of the high caffeine content. In 2016, Nall *et al.* chronically treated flies with a range of caffeine concentrations; 0.2mg/mL, 0.5mg/mL, and 1mg/mL mixed into sucrose-agar food. They discovered that the increasing concentrations positively correlated with a decrease in night-time sleep, with the highest concentration reducing sleep amount by 30% which is comparable to this study. The increase of sleep deprivation correlating with the increase of caffeine concentration was also shown in a study by Andretic *et al.* (2008). The sleep depriving effect of caffeine was not displayed in the *miR-137^{KO}* strain, which showed an overall decrease that is the reverse effect shown in the control genotype. It is important to note that this experiment used solely male flies, and the effect of chronic caffeine treatment on females remains to be confirmed.

Using caffeine to treat flies with excessive sleep amounts has provided an interesting outcome in flies lacking *miR-137*, which subsequently creates a putative hypothesis of adenosine pathway dysfunction or downstream mechanisms regulating requirement. The expected decrease in sleep in the night period, while not reversing the knockout phenotype completely, does suggest that the caffeine may be having the expected effect. However, this is offset by an increase within the day period that is in contradiction to what should happen upon introduction of a stimulant and infers a potential overpowering sleep drive under the homeostatic control. This sleep-rebound effect was exhibited by flies in the Potdar et al. (2018) study, where removal of caffeine stimulation by moving to non-drugged food produced increased daytime sleep. Correcting for Type I errors produced by multiple t-tests meant that the effect exhibited by the KO strain in the presence of caffeine did not meet the

new threshold level calculated using the Bonferroni method as described in Chapter 2.10. Nonetheless, this result still demonstrates that chronic caffeine dosage did not cause the expected recovery of the KO sleep phenotype. Similarly, the overall 24-hour effect of caffeine on the wildtype flies does not meet the new 0.00833 significance cut-off, though the night period effect remains significant even after such conservative corrections.

7.2.2   3-iodo-tyrosine and lithium chloride

Dopamine is a neurotransmitter responsible for promoting wakefulness and is responsible for the morning peak of activity exhibited in *D. melanogaster* (Kume *et al.* 2005). The phenotype displayed by the *miR-137^{KO}* strain exhibits a somewhat stunted morning peak compared to controls (Chapter 5.1), along with the excessive sleep and decrease motility, these are symptoms synonymous with Drosophila research into dopamine deficiencies (Reimensperger *et al.* 2011). 3-iodo-tyrosine (3IY) is an inhibitor of dopamine synthesis and has been found in fruit flies to promote sleep by lowering dopamine levels (Catterson *et al.* 2010). While counterproductive in reversing the hyper-sleep phenotype, an initial hypothesis for the action of *miR-137* knockout promoting sleep is the desensitization of the dopamine signalling pathway due to increases in *miR-137* predicted targets such as dopamine receptors (*Dop1r1, Dop1r2, d2r*), and tyrosine hydroxylase ortholog *pale* (*ple*) (Jia *et al.* 2016; Agarwal *et al.* 2018). An increase in these genes could increase the amount of dopamine synthesized and the number of receptors to pick up the molecules, resulting in overexcited neurons and a resulting desensitization as the larvae develop. To test this hypothesis, 3IY was given to flies throughout development to try and prevent this overload. A chosen dosage of 0.5mg/mL was used as this concentration proved effective in a study by Catterson *et al.* (2010).

Lithium is well known for its use as a treatment for psychological issues in earlier times. It was administered as pure tablets, but also eventually as lithium chloride (LiCl) salts or lithium carbonate for treatment of acute mania and BD. It has toxic effects on the body in high concentrations which results in renal failure and several other issues. While the exact mechanism of the neuroprotective effects of lithium is not understood, the chemical is known to interact with glycogen-synthase-kinase-3 (GSK3) which is a component of several neuroprotective genetic pathways (Klein *et al.* 1996; Jia *et al.* 2013). *miR-137* has a risk allele associated with SZ and several of its predicted targets are risk genes for many psychiatric disorders such as *CACNA1C* and *TCF4*. Furthermore, *GSK3* itself is a predicted target of *miR-137* downregulation (Agarwal *et al.* 2018). LiCl was introduced to the food of *D. melanogaster* to investigate if this medication could reduce the phenotype caused by *miR-137* LOF. A dosage of 15mM was chosen as this successfully induced an effect in a study by Jia *et al.* (2013).

7.2.2.1   *Method*

For chronic treatment, LiCl or 3IY were administered for 5-7 days to adult flies before being put into the DAMS system. Pupa were removed from a normal food vial and moved into a vial containing food containing either LiCl dosed at 15mM or 3IY mg/mL. After the flies eclosed and reached the correct age, they were added to the DAMS with drugged food. The DAMS and subsequent analysis were conducted as described in Chapter 2.7.

For acute treatment, 16 w^{1118} and 16 $miR-137^{KO}$ male flies were raised normally and put into the DAMS as described in Chapter 2.7. At circa 7am on the 4th day in the DAMS, half of each genotype was moved to new vials containing drugged food containing either LiCl dosed at 15mM or 3IY at 5mg/mL, and the other half was transferred to fresh normal food.

Data was collected and analysed in Microsoft Excel, producing total sleep amount per 24-hours of the 3 days before and 4 days after exposure to the drugged food. The 4th day was excluded to allow the flies to reacclimatise to the vials after disturbance.

7.2.2.2 Chronic treatment results

Chronic exposure of flies with 3IY produced an overall decrease in total sleep amount across a 24-hour period for the $miR\text{-}137^{KO}$ strain (p-value < 0.05), but not a significant change in the w^{1118} controls (p-value > 0.05) (figure 7.4; table 7.6). Most of the change in the knockout flies was a significant decrease of sleep amount during the day period (p-value < 0.01) and a small but not statistically relevant decrease in the night period (p-value > 0.05). No change within any period was statistically significant (p-value < 0.05) for the w^{1118} flies. The full statistics can be found in the supplementary information.

Table 7.6: Total sleep amount from chronic 3IY recovery experiment

	Average sleep amount (minutes [1dp])		
Experiment	24-hour	Day period	Night period
Normal: W^{1118}	841.5	321.5	520.0
Normal: $miR\text{-}137^{KO}$	1239.6	647.0	592.8
3IY: W^{1118}	821.6	308.4	513.3
3IY: $miR\text{-}137^{KO}$	1105.0	544.4	560.6
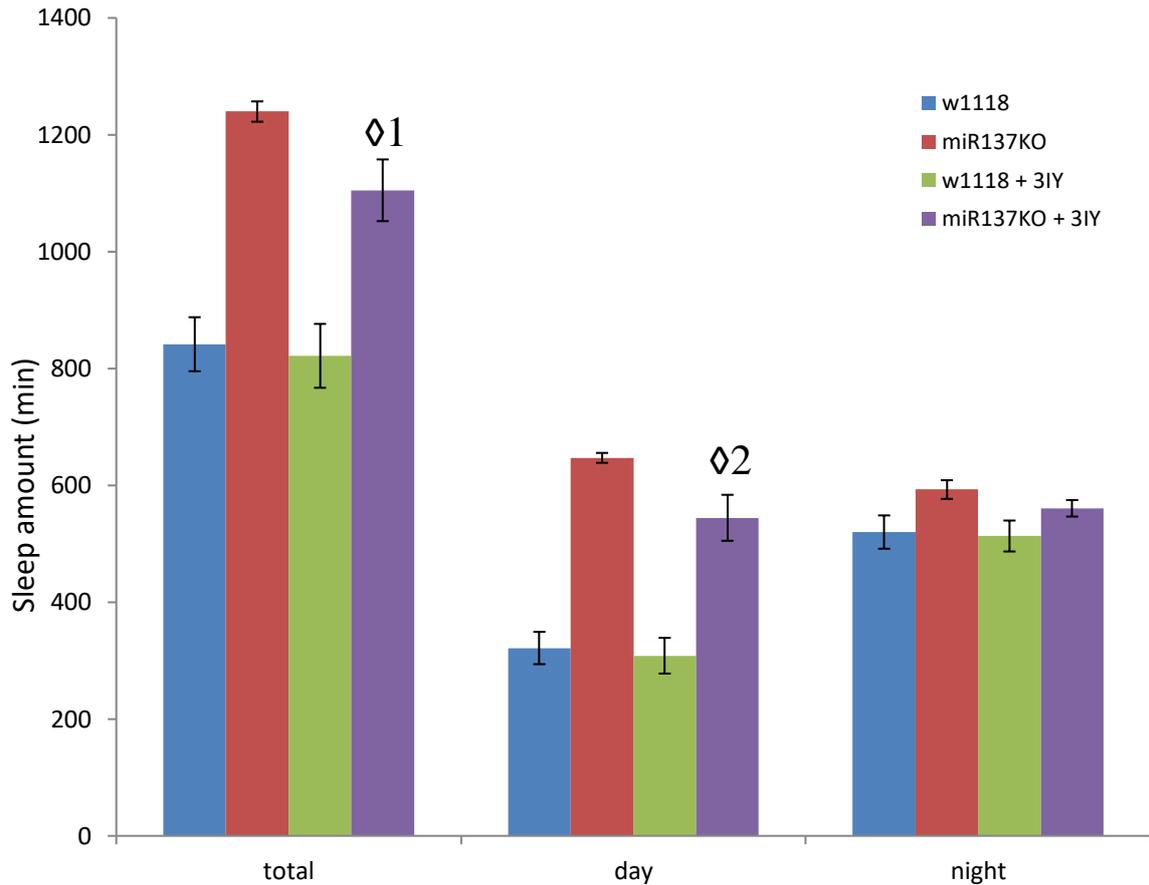
Figure 7.4: The effect of chronic 3IY treatment on total sleep amount

The graph shows both the total 24-hour period, and the separation of the two 12-hour day/night phases. w^{1118} (n=8), $miR\text{-}137^{KO}$ (n=8), w^{1118} with 3IY (n=8), and $miR\text{-}137^{KO}$ with 3IY (n=5) flies were put into a Drosophila activity monitoring system (DAMS) by Trikinetics and the data was analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings taken across 5 consecutive days and averaged. Experiments were compared between the treated and non-treated flies individually using t-tests. In the 24-hour period and the day period the $miR\text{-}137^{KO}$ showed a significant decrease in total sleep (◊1: p-value < 0.05, ◊2: p-value < 0.01). The w^{1118} strain shows no significant difference between treated and non-treated groups. Correcting for Type I error using the Bonferroni method (described in ch2.10) resulted in the 3IY effect in the KO strain for the 24-hour period is no longer significant. Error bars are standard error.

Treatment with 15 mM LiCl did not produce any significant changes in total sleep between the treated and non-treated $miR\text{-}137^{KO}$ flies (figure 7.5; table 7.7), however the controls demonstrated a significant decrease in total sleep across the 24-hour period and both the day and night phases separately (P-values < 0.01).
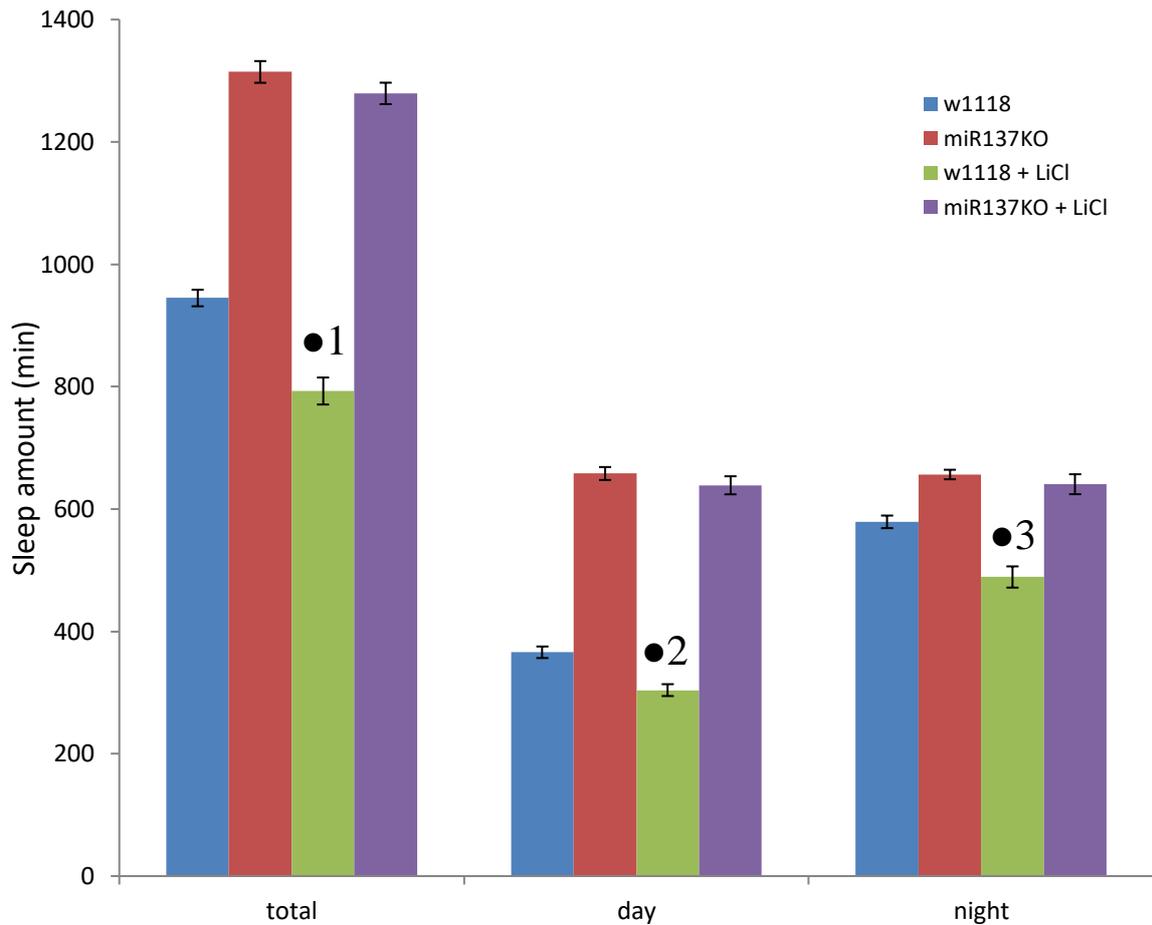
Figure 7.5: The effect of chronic LiCl treatment on total sleep amount

The graph shows both the total 24-hour period, and the separation of the two 12-hour day/night phases. w^{1118} (n=7), miR-137^{KO} (n=8), w^{1118} with 3IY (n=8), and miR-137^{KO} with 3IY (n=8) flies were put into a Drosophila activity monitoring system (DAMS) by Trikinetics and the data was analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings taken across 5 consecutive days and averaged. Experiments were compared between the treated and non-treated flies individually using t-tests. The w^{1118} showed a significant decrease in total sleep across both periods and in total (●1: p-value < 0.01, ●2: p-value < 0.01, ●3: p-value < 0.01). The miR-137^{KO} strain shows no significant difference between treated and non-treated groups. Error bars are standard error.

Table 7.7: Total sleep amount from chronic LiCl recovery experiment

	Average sleep amount (minutes [1dp])		
Experiment	**24-hour**	**Day period**	**Night period**
Normal: W^{1118}	944.8571	365.8571	579
Normal: miR-137^{KO}	1314.375	658	656.375
LiCl: W^{1118}	792.875	304	488.875
LiCl: miR-137^{KO}	1279.25	638.75	640.5

7.2.2.3 Acute treatment results

Treatment of 3IY shows an obvious increase in total sleep amount per 24 hours upon introduction of the drug in w^{1118} flies compared to those not treated (figure 7.6, table 7.8). There is no change in $miR-137^{KO}$ flies compared to non-treated controls.
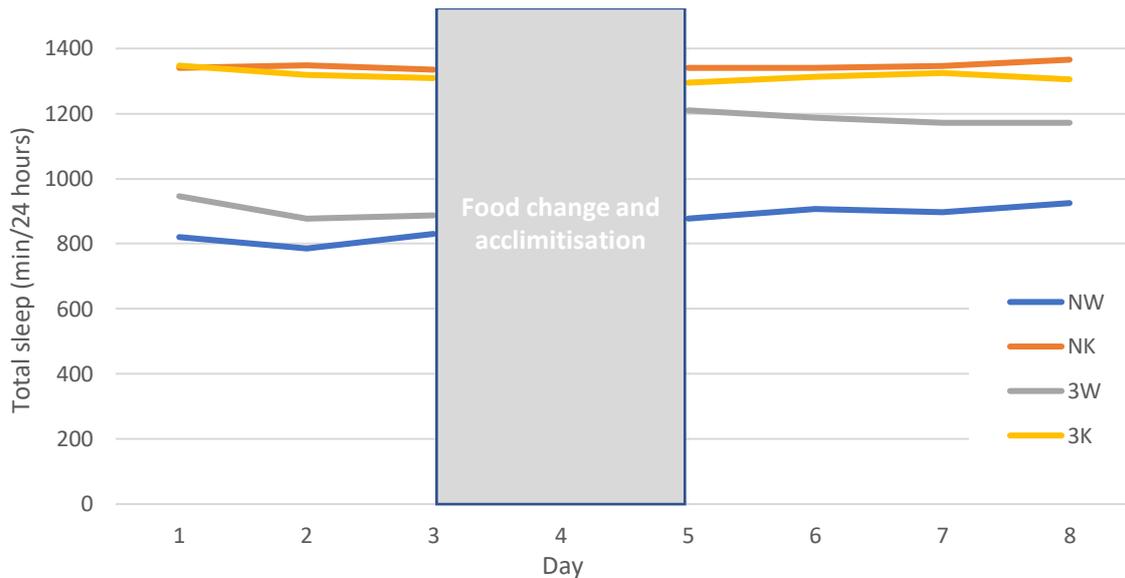
Figure 7.6: The effect of acute 3IY treatment on total sleep amount

The graph shows the 3 days on normal food, followed by a period of acclimatisation following switching of food. w^{1118} (NW, n=8), $miR-137^{KO}$ (NK, n=8), w^{1118} with 3IY (3W, n=7), and $miR-137^{KO}$ with 3IY (3K, n=8) flies were put into a Drosophila activity monitoring system (DAMS) by Trikinetics and the data was analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings for all flies averaged. The w^{1118} strain shows an obvious increase in total sleep amount following exposure to 3IY, while the $miR-137^{KO}$ flies show no difference between treated and non-treated groups.

Table 7.8: Total sleep amount from acute 3IY recovery experiment

Experiment	Average sleep amount (minutes/24 hours[1dp])							
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8
Normal W^{1118}	820.6	785.6	830.6	-	876.9	906.25	896.25	925.0
Normal $miR-137^{KO}$	1340.6	1347.5	1335.0	-	1340.6	1341.2	1345.6	1365.6
3IY W^{1118}	946.4	877.1	887.9	-	1210.0	1187.1	1172.1	1172.1
3IY $miR-137^{KO}$	1347.5	1318.8	1308.8	-	1295.0	1313.1	1324.3	1305.6

Treatment of LiCl shows no change in either w^{1118} or $miR\text{-}137^{KO}$ flies compared to non-treated controls (figure 7.7; table 7.9).
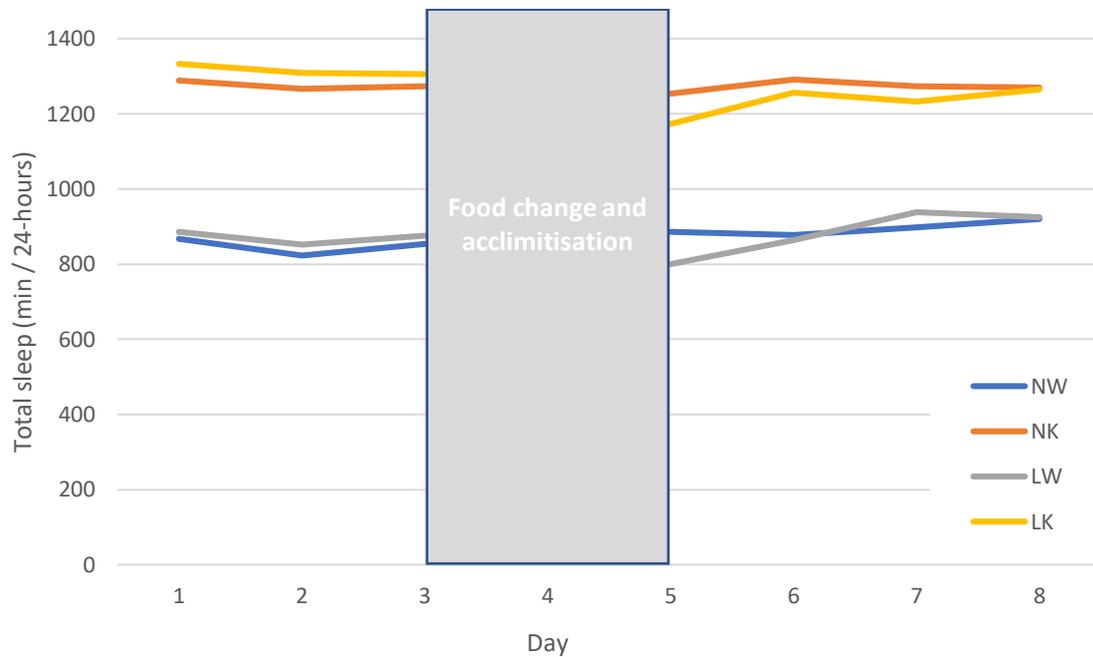
Figure 7.7 The effect of acute LiCl treatment on total sleep amount

The graph shows the 3 days on normal food, followed by a period of acclimatisation following switching of food. w^{1118} (NW, n=8), $miR\text{-}137^{KO}$ (NK, n=8), w^{1118} with LiCl (LW, n=7), and $miR\text{-}137^{KO}$ with LiCl (LK, n=8) flies were put into a Drosophila activity monitoring system (DAMS) by Trikinetics and the data was analysed in Microsoft Excel. One sleep period is counted as 5 minutes of no activity recorded in the DAMS, with recordings for all flies averaged. Both the w^{1118} strain and the $miR\text{-}137^{KO}$ flies show no obvious difference between treated and non-treated groups.

Table 7.9: Total sleep amount from acute LiCl recovery experiment

Experiment	Average sleep amount (minutes/24 hours[1dp])							
	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8
Normal W^{1118}	866.4	822.9	854.2	-	886.4	877.1	897.1	920.0
Normal $miR\text{-}137^{KO}$	1288.8	1266.9	1272.5	-	1253.8	1291.3	1273.8	1270.0
LiCl W^{1118}	885.6	851.9	875.0	-	800.0	863.8	938.1	925.6
LiCl $miR\text{-}137^{KO}$	1333.1	1308.8	1306.3	-	1173.1	1256.3	1232.5	1265.6

7.2.2.4   Discussion

Chronic treatment of 3IY affected the two genotypes differently. The w^{1118} strain showed no real change in sleep amount comparing treated to untreated flies, but the *miR-137^{KO}* flies exhibited a contradictory effect of reduced sleep across the 24-hour period. 3IY is a molecule which inhibits the production of dopamine by limiting the tyrosine hydroxylase (*ple* in *D. melanogaster*), therefore promoting sleep (Catterson *et al.* 2010; Reimensperger *et al.* 2011). Catterson *et al.* (2010) reported a 4.5-fold increase in average sleep bout length, while Tomita *et al.* (2015) demonstrated an increase of circa 20%. These studies used differing food mediums and drug concentrations but produced an increase in control sleep. Furthermore, a higher dose of 10mg/mL 3IY administered for 48 hours produced a tenfold decrease in dopamine levels and had no effect on locomotor climbing ability in male flies (Bainton *et al.* 2000). This dosage was lethal when given for extended periods of time or during larval development. Initially, the data from the chronic treatment was thought to be erroneous because of improperly prepared food/drug concentrations, but acute exposure to the same 3IY food produced the expected increased sleep behaviour within the w^{1118} strain. Potentially the prolonged exposure to 3IY has allowed for some sort of tolerance which prevents the 5-day average sleep value being different in the control flies, but these results reinforce the dopamine desensitisation hypothesis of increased dopamine receptor activation in the absence of *miR-137* creating a tolerance throughout development of these flies. When the Bonferroni method was used to correct for Type I errors, the corrected significance threshold meant that the effect of 3IY on the KO strain across the 24-hour period was no longer deemed significant. Nonetheless, even with the conservative corrected p-value of 0.00833, the difference exhibited by the KO strain in the day was still significant and therefore contradictory to the lack of effect seen in the wildtype controls.

Lithium chloride is known to produce therapeutic effects in *D. melanogaster* models of Huntington's disease and Alzheimer's disease (Berger *et al.* 2005; Sofola *et al.* 2010), thought to be due to inhibition of GSK3B and subsequently preventing the toxic aggregation of proteins characteristic of these diseases (Jia *et al.* 2013). Furthermore, an MDD model in flies (induced by repeated and unavoidable stressors) showed a locomotor deficit which was then reversed upon LiCl treatment (Ries *et al.* 2017). With LiCl's use as an effective neurological treatment, it was hypothesized that LiCl may alleviate some of the sleep symptom exhibited within *miR-137* deficient flies, however no such study appears to have been done previously. Acute exposure to lithium produced little results for either strain, however chronic treatment severely decreased the sleep in w^{1118} flies across both periods of the day. With many medications prescribed for psychological symptoms there is a period required before the level of the drug within the body is enough to produce the therapeutic effect, and the body can adapt to it. This is a potential explanation as to why the acute treatment did not seem to cause any change in any of the flies, but the concentration of 15mM also does not restore the normal sleep phenotype in $miR\text{-}137^{KO}$ flies. There is a small decrease across both periods, but nothing calculated as significant (p-value < 0.05).

7.3　Conclusion

Genetically, the statistically significant decrease in sleep amount shown in the FB-recovery, along with the comparison of this region to the mammalian hypothalamic ventrolateral preoptic nuclei, offers promising options as to the mechanism resulting in exaggerated sleep and links to psychiatric disorders. The Bonferroni corrected p-values ultimately resulted in no significance found with reintroduction of *miR-137* to the CNS as a whole or the fan-shaped body by itself. This ultimately means that the null hypotheses in objective D hypotheses 1 and 2 cannot be rejected (Chapter 1.8).

Exposure to pharmaceuticals produced some confounding results that could infer further neurological abnormalities. The lack of behavioural change when treated with anti-psychotic medicine LiCl, both chronically and acutely, in the knockout flies demonstrates another potential molecular anomaly that has links to psychiatric diseases. The decrease in sleep when chronically exposed to 3IY, but no effect upon acute treatment lends further evidence to the desensitisation to dopamine hypothesis explained in Chapter 7.2.2. The significant effect of 3IY across the day period allows for rejection of the null hypothesis 4 in objective D (Chapter 1.8). The increase of daytime sleep within the knockout flies upon exposure to caffeine and loss of some night-time sleep implicates *miR-137* with the sleep homeostat as was also concluded in Chapter 5. Following Bonferroni corrections, the lack of significant effect overall from caffeine and LiCl treatment means that the null hypotheses for objective D hypotheses 3 and 4 cannot be rejected (Chapter 1.8).

It is worth mentioning that these pilot experiments were conducted on numbers that while enough for comparison, could be amplified for greater experimental validity. Further work could also use a range of drug concentrations and assess food ingestion rates as proof of drug intake. Furthermore, mainly due to limitations with the DAMS and experimental

continuity, these experiments were conducted solely on male *Drosophila* and further work to confirm the effect in females still needs to be completed.

In conclusion, none of the recovery attempts produced results that could be deemed as a full recovery of the phenotype behaviour shown in the *miR-137^{KO}* flies. However, all experiments provided a novel vantage point and research direction for *miR-137* dysfunction that can be assessed in the future. Of interest is the hypothesis of dopamine desensitisation, which is evidenced with the results from the 3IY chronically treatment flies and the dopamine driven fan-body region recovery.

8   Bioinformatics

The experimental results presented in the previous chapters, showed that *miR-137* is important for proper locomotor and sleep behaviours in flies. These results obviously represent the phenotypic effects of sequestering that gene in *Drosophila melanogaster*. However, this project began with a statistical association of human *MIR-137* in a SZ GWAS; a different phenotype. The crossover between many documented psychiatric symptoms and the phenotypes displayed by the *miR-137^{KO}* strain implied potential effects of human *MIR137* in psychiatric disorders and the severe sleep phenotype is also a symptom of many neurological diseases. Therefore, this phenotype appears to be broad based. Unfortunately, the *miR-137^{KO}* phenotype alone does not offer an explanation as to the mechanism with which the microRNA creates this effect, or the many hypothesised genes downstream in the pathways. A complete transcriptome of the fly stock with *miR-137* gene removed could be used to identify any genes where the expression was significantly different compared to control strains. This data could also then be compared with bioinformatic analysis to try and identify genetic pathways where there are multiple expression changes as a result of *miR-137* loss. Additionally, published data from a study of *miR*-137 inhibition in progenitor neural stem cells (Hill *et al.* 2014), provided the opportunity to investigate if similar genetic expression changes matched the *D. melanogaster* data. Furthermore, analysis of other human GWAS data could provide links to differentially expressed genes and other psychiatric or neurological disorders. Ideally pooling all data together might provide synergies between gene sets leading to hypotheses of tangible mechanisms through which the altered expression of *miR-137* acts, resulting in the phenotypes displayed.

8.1 Introduction

Bioinformatics is the modern practice of computer aided analysis and interpretation of complex biological data such as that from gene expression, GWAS, proteomic and *in vivo* experiments. Historically, as research results contained numerous genes or SNPs, researchers would have to methodically accumulate information on each component and then assess and hypothesise pathways and interactions based upon literature. With advances in technology and the wealth of open sourced publicly deposited data, there has been a huge number of bioinformatic programmes developed to help non-bioinformatic skilled researchers quickly access and interpret gene lists for pathways, expression data, sample enrichment, genome-wide significance and more. As these resources have become increasingly used by non-specialists, many online programmes now adopt very simple interfaces to enable accessibility to a greater number of researchers.

This section details my efforts to establish a selection of evidence based and biologically plausible genes affected by *miR-137* dysregulation. Utilising a service for *Drosophila* transcriptomics offered by Cambridge University (FlyChip), the expression of all genes was compared between adult control strain flies and the flies with genomic *miR-137* removed. Using these results alongside other published 'wet-lab' experimental data, I show how I assess cross-species links and produce a select group of matched genes. In addition, I then used publicly available bioinformatic tools and datasets (DAVID, UKbiobank, Neale Lab data, FUMA CTG lab), as well as data received from Dr Richard Anney of Cardiff University to evaluate the ontology and potential links to sleep and locomotion behavioural phenotypes.

8.2   *Drosophila* Transcriptome

To assess the effect of *miR-137* ablation on a genetic level in the fruit fly, a FlyChip gene expression chip was conducted utilising a service at Cambridge University. The service performed a full transcriptome analysis using a custom long oligonucleotide probe array to measure comparative expression values between a test and a control sample, then performed simple statistical analysis to evaluate the significance of the expression differences. The results would give comparative gene expression changes of all identified genes in the *miR-137^{KO}* strain against the control *w^{1118}* flies, which would then be useful for assessing pathways and biological mechanisms affected by *miR-137* dysfunction.

8.2.1   *Preparation of FlyChip samples*

Samples of homogenised *Drosophila* heads in Trizol solution (Thermofisher) were collected and delivered to Cambridge University in dry ice for FlyChip analysis. The FlyChip utilised two biological samples to produce two readings of the comparative expression amount of test sample against control sample. Cambridge University also applied use of the Bioconductor Linear Models for Microarray Data (LIMMA) programme to perform statistical analysis producing a standardised p-value and log2 change factor. The log2 factor represents a fold change between test and control (eg: log2 of 2 signifies a doubling of expression).

Adult flies aged 4 – 7 days post eclosion were put into -20°C for approximately 30 minutes before being transferred to a dissection plate. A total of 50 *Drosophila* of each *w^{1118}* and *miR-137^{KO}* were dissected under a microscope in Phosphate-buffered Saline solution. The heads were collected whole and put directly into a 1.5mL ependorf containing 300uL of Trizol solution. While the samples were being collected the ependorf was stored on ice, and once the full amount was transferred, they were then homogenised using a sterile micropestle.

Following homogenisation, the samples were stored at -80°C before being transported in dry ice. Once delivered, Cambridge University performed RNA extraction then a Chip-style expression array. The time between sample collection and RNA extraction was kept to under 36 hours to minimise RNA degradation as much as possible. The sample size and trizol volume used were suggested amounts by Cambridge University.

8.2.2   *Differentially expressed genes (DEG)*

The pre-analysed results from Cambridge University indicated that in the *miR-137^{KO}* strain there were a total of 7450 genes with a decrease of expression compared to controls, and 6990 genes which were increased compared to controls. Of these, only 30 gene expression changes were significant to a p-value of less than 0.05 (tables 8.1 and 8.2).

Table 8.1: Significantly upregulated genes in the FlyChip.

Gene Symbol	Gene Name	Log2 Change (3dp)	P-Value (3dp)
FBgn0033631	Sod3	4.232	0.002
FBgn0003996	W	3.830	0.002
FBgn0260793	2mit	2.970	0.009
FBgn0031228	ND-15	2.573	0.009
FBgn0010473	Tutl	2.589	0.013
FBgn0001217	Hsc70-2	2.300	0.013
FBgn0027584	CG4757	2.281	0.014
FBgn0051157	CG31157	2.277	0.017
FBgn0031560	CG16713	2.038	0.018

Table 8.2: Significantly downregulated genes in the FlyChip.

Gene Symbol	Gene Name	Log2 Change (3dp)	P-Value (3dp)
FBgn0040074	*Retinin*	-3.753	0.002
FBgn0033065	*Cyp6w1*	-3.377	0.005
FBgn0023171	*rnh1*	-2.632	0.009
FBgn0034330	*CG18107*	-2.610	0.009
FBgn0261477	*Slim*	-2.357	0.013
FBgn0038237	*Pde6*	-2.521	0.017
FBgn0086348	*Se*	-2.190	0.017
FBgn0069973	*CG40485*	-2.090	0.017
FBgn0038343	*Trissin*	-2.066	0.018
FBgn0028371	*Jbug*	-2.070	0.019
FBgn0039459	*IntS12*	-2.042	0.019
FBgn0034860	*CG9812*	-1.951	0.021
FBgn0040256	*Ugt86Dd*	-2.016	0.031
FBgn0035676	*ssp6*	-1.826	0.031
FBgn0050456	*CG30456*	-2.492	0.041
FBgn0267435	*Chp*	-1.852	0.042
FBgn0036183	*CG6083*	-1.788	0.042
FBgn0260746	*Ect3*	-1.774	0.043
FBgn0011693	*Pdh*	-2.023	0.046
FBgn0259229	*CG42329*	-1.814	0.046
FBgn0043534	*Obp57b*	-1.876	0.047

8.2.2.1 Discussion

The 30 Drosophila melanogaster genes listed (P<0.05) consist of nine over-expressed and 21 under-expressed (Tables 8.1 and 8.2) genes. According to Flybase (available at: www.flybase.org), these genes are involved in the following:

Sod3 is a Superoxide dismutase 3 and encodes an extra-cellular copper-zinc superoxide dismutase that can be produced as secreted and membrane bound forms. It converts oxygen free radicals to hydrogen peroxide and is involved in free radical defence and redox balance.

White (*w*) encodes a member of the ABCG2 class of transporters, transporting molecules such as cyclic GMP, biogenic amines and pigments. This is an expected change within the transcriptome as the genetic marker used to differentiate between *miR-137^{KO}* flies and control was a recessive allele of a truncated version of the *w* gene causing white/colourless eyes in the controls.

2mit encodes a putative leucine-rich repeat transmembrane protein that is mainly expressed in the developing central nervous system. In the adult, it has a role in short-term memory and learned courtship behaviour (Baggio *et al.* 2013).

ND-15 NADH dehydrogenase (*ubiquinone*) is likely to be involved with mitochondrial electron transport, and downregulation of the gene causes body size defects and swollen mitochondria in third instar larva (Zhou *et al.* 2019).

Turtle (*tutl*) encodes an Ig-superfamily transmembrane protein. Its roles include axonal tiling, dendrite self-avoidance, axonal pathfinding and coordinated motor control. Loss of function in flies causes locomotor and behavioural defects, plus partial lethality during pupation (Bodily *et al.* 2001).

Heat shock protein cognate 2 (*Hsc70-2*) encodes a member of the heat shock protein family. It functions as a chaperone to promote proper post-transcriptional folding of proteins. In flies with RNAi driven knockdown of *hsc70-2* through development, the larva develops anatomically smaller and feature fragmented mitochondria (Zhou *et al.* 2019).

Retinin is a structural component of the eye, expressed in the extracellular space and around pigment cells. This is likely to be related to the white eye phenotype mentioned under the *white* gene.

Cyp6w1 is predicted to be part of the cytochrome P450 family. It is likely to have heme-binding and oxidoreductase properties and has a peak expression level observed in third instar stages and in adult males.

Rnh1 ribonuclease H1 (*rnh1*) encodes an RNA-DNA hybrid ribonuclease. Partial deletion of the genomic sequence causes developmental defects including a delay in developmental time, pre-pupal and pupal developmental lethality, and melanotic tumours (Filippov *et al.* 2001).

CG18107 is also known as *Bomanin Short 4,* part of the Bomanin family. Bomanins are small secreted immune-response peptides induced by *Toll* signalling, likely responsible for bacterial or infection resistance.

Slim is called '*scruin like at the midline'* and is involved in the stress-activated MAPK signalling cascade. RNAi driven downregulation in the CNS causes sleep dysfunction in both males and females, and anatomical muscle defects in the larval wall (Schnorrer *et al.* 2010; Lobell *et al.* 2017).

Pde6 is a 3',5'-cyclic-GMP phosphodiesterase, which is expressed in the adult eyes and its product is involved in drosopterin (red eye pigment) synthesis. This is another likely genetic change likely caused by the *white* gene change.

Sepia (*se*) is a Glutathione dehydrogenase/transferase, which produces a product used in eye cell pigments. This change is likely as a result of the *white* gene modification.

CG40485 is predicted to be a steroid dehydrogenase, which are NAD or NADP dependent oxidoreductases with steroid as a substrate. RNAi downregulation causes early larval lethality and neuroanatomy defects with less daughter cell neuroblasts (Schnorrer *et al.* 2010; Neumuller *et al.* 2011).

Jitterbug (*jbug*) encodes a Filamin-type protein. Filamins are actin crosslinkers that organize actin filaments into networks and connect them to membrane receptors. The product of *jbug* is required in photoreceptor cells for axon targeting and in tendon cells to maintain its shape during muscle-tendon interaction.

Integrator S12 is predicted to localise with the integrator complex and have involvement with the processing of small nucleolar RNA molecules. RNAi mediated downregulation causes early larval lethality and neuroanatomy defects with smaller neuroblasts (Schnorrer *et al.* 2010; Neumuller *et al.* 2011).

Ugt86Dd is "*UDP-glycosyltransferase family 35 member C1*", part of the glycosyltransferase family that plays a major role in detoxification of substrates, formation of pigments, and olfaction.

Chaoptin (*chp*) is involved in rhabdomere and microvilli organisation and anchoring to the plasma membrane adhesion molecules. Loss of function results in irregular photoreceptor and eye morphology (Pollock *et al.* 1990; Gurudev *et al.* 2014).

CG6083 is involved in Alcohol dehydrogenase (NADP (+)) and Aldehyde reductase. Expression of an RNA-inhibition molecule in *Drosophila* mushroom bodies demonstrated significant neuroanatomical defects, characterised by lobe fusion in approximately 35% of cases (Koboyashi *et al.* 2006).

175

Ect3 is predicted to have beta-galactosidase activity catalysing the hydrolysis of β-D-galactosyl residues in the presence of water to galactose and alcohol. This forms part of the larger carbohydrate metabolic process.

Photoreceptor dehydrogenase (*Pdh*) encodes a retinal pigment cell dehydrogenase involved in retinol metabolism. Loss of function in flies results in progressive degeneration of the rhabdomeres in normal light/dark cycle, and RNAi downregulation in neuroblasts result in higher lethality rate compared to controls when placed in higher temperatures (Neely *et al.* 2010; Wang *et al.* 2010)

Genes *CG42329, Obp57b, ssp6, CG30456, Trissin, CG4757, CG31157, CG9812,* and *CG16713* are of unknown function with little or no information available to assess involvement with the rest of the gene set.

Genes *2mit* and *pde6* are predicted direct targets of *miR-137* on Targetscan, however *2mit* is upregulated in the *miR-137^{KO}* strain which is contradictive of the predicted downregulatory effect. The *pde6* gene does not have a lot of information regarding its function or effect however, the available information implies the change is likely exasperated by the differing eye phenotypes as mentioned previously. In addition to this, none of the human orthologs of these genes calculated by DIOPT are predicted targets in humans and therefore unlikely to be of direct relevance.

Overall, the list includes several photoreceptor/eye morphology related genes which are likely to be due to the truncated white gene marker used extensively across *D. melanogaster* studies for its recessive and easily observable phenotype. Despite this, one of these genes (*pde6*) is a putative direct target of *miR-137* which may be worth investigating in the future. Several genes are implicated within metabolism and energy transfer. There are also a few genes linked to neuroanatomy, be it brain regions, axon guidance, or signalling

between the CNS and muscle. Of interest is gene *CG8063*, which has experimental links to mushroom body development when downregulated. In the FlyChip assay this had a Log2 change of -1.788 (3dp), a significant decrease in expression. This region of the fly brain has been functionally compared to the mammalian cerebral cortex, which is subsequently one of the key regions related to psychiatric disorders in humans (Nesvag *et al.* 2007; Iritani 2013; Sitara 2015). There is also a gene linked to sleep dysfunction, *slim*, which may prove to be a link to the severe phenotype displayed in the flies (Chapter 5.4). The sleep dysfunction phenotype was exhibited in flies with RNAi mediated downregulation of the gene, and this matches the FlyChip data which had a Log2 factor of -2.357 (3dp). There are many genes which do not have any or much information available to assess function or potential phenotypical importance within this project, as the understanding of these genes develops, they may become of more importance.

8.2.3   *Significant gene ontology*

As identification of significantly expressed genes above shows, using Flybase does not provide adequate indication of what pathways they are likely to be involved with collectively. To investigate this, the online platform DAVID was used to calculate molecular interactions and pathways. DAVID is the "database for annotation, visualization, and integrated discovery" (accessible from: david.ncifcrf.gov/), a publicly available web interface for assessing gene lists for enrichment, disease relevance, and biological categories amongst other features. The database is run and updated constantly by the Laboratory of Human Retrovirology and Immunoinformatics. The output calculates enrichment p-values for easy identification of enrichment within the gene set for further analysis.

Visualised pathways were also produced utilising online resource "Gene Mania" (accessible from genemania.org). The online platform has a user-friendly interface and the results are useful for generating hypotheses about gene function, analysing gene lists and prioritising future workload. The algorithm extends the list with functionally similar genes that it identifies using available genomics and proteomics data (Warde-Farley *et al.* 2010). The visual network features lines of different weights indicative of the predicted value of each selected data set for the query. Gene Mania relies upon data sources including; GeneMANIA relies on the following data sources: GEO, BioGRID, EMBL-EBI, Pfam, Ensembl, NCBI, MGI, I2D, InParanoid, Pathway Commons, and also smaller data sets.

8.2.3.1   *Method*

The 30 genes were inputted to the online resource DAVID to look for gene ontology and pathway enrichment using the preloaded algorithms and databases. The significant genes from the FlyChip were put as FlyBase IDs with the *D. melanogaster* background selected. Any results from DAVID analysis were to be assessed for relevant key terms within both functional clusters or individual functional terms, the degree of enrichment, and automatically calculated p-value. This was then used as a basis for further reading to build information for potential methods of action of *miR-137* manipulation on the phenotype.

Additionally, the significant orthologs were created by putting the fly genes into the online DIOPT programme (accessible from: www.flyrnai.org/diopt). These orthologs were then put into the online programme Gene Mania, which created a visual network map of proven or predicted interactions via similar protein regions and localization. To do this, the human ortholog list was copied into the Gene Mania search bar, and the organism type was changed to *Homo sapiens* before selecting the search button.

8.2.3.2 Results

DAVID database did not produce any significant results when the 30 significant *D. melanogaster* genes. 6.7% of the input gene list (2/30) were involved in the Galactose Metabolism pathway (p-value > 0,05). The two genes involved with the pathway were *ect3* and *CG6083*.

Table 8.3: DAVID functional annotation of significant FlyChip genes.

Database	Pathway	Gene Count	Percent of input	Enrichment P-value	Gene Name
KEGG_PATHWAY	Galactose metabolism	2	6.7%	0.095	*CG6083*
					Ect3

Gene Mania produced network maps clearly showing three main clusters around UGT, HSAP, and AKR1 protein families indicated by lots of linkage through "shared protein domains" (figure 8.1). The programme also identified additional links to genes outside of the significant list that was inputted. These include; *KCNAB1, CCS,* and *ACHE* which are linked to the gene set through direct protein interactions or shared protein domains.
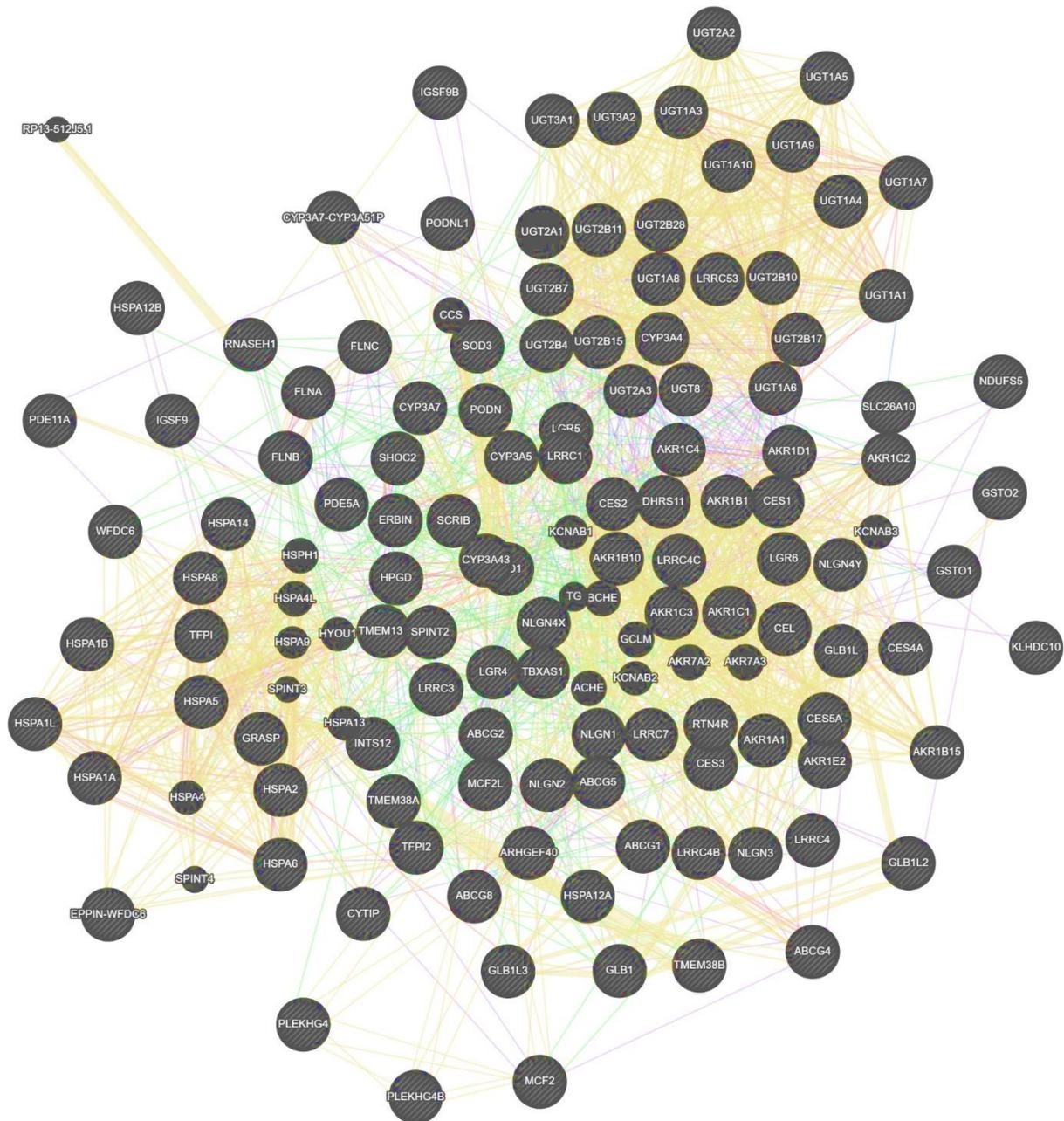
Figure 8.1: Human ortholog gene network of the DEGs from the FlyChip.

The 30 differentially expressed genes (p-value <0.05) were put into DIOPT to produce 119 human orthologs. The analysis and visualisation were created by Gene Mania (accessible from: genemania.org). Additional genes are in full black while inputted genes are in stripes. The colours of lines represent different links: Beige = Shared protein domains, Purple = co-localisation, Red = Physical interactions, Blue = Pathways, Green = Shared protein domains, Orange = Predicted interactions.

8.2.3.3 *Discussion*

DAVID produced one association of the inputted gene list, with 2 genes implicated as part of the galactose metabolism pathway. While the association was not deemed a significant enrichment (P-value > 0.05), galactose metabolism is part of a larger carbohydrate metabolism process where dysfunction within can cause severe side effects ranging from lethargy to death in humans (Sehgal 1998). The gene *CG6083* has been implicated in flies with proper mushroom-body brain development (Kobayashi *et al.* 2006), the fly brain region which shares functional similarities to the mammalian cerebral cortex (Sitara 2015). This region is linked to neurological and psychiatric disorders in humans.

Of the additional genes identified by Gene Mania, *KCNAB1* is of interest as it is part of the potassium voltage-gated channel family known to be involved with synaptic signal transmissions. It is an ortholog of *D. melanogaster* gene *hyperkinetic* (*hk*), which is a subunit of *shaker* channels responsible for action potentials in neurons and transmitter release. It specifically is responsible for activation of the fan-body brain region of the fruit fly which regulates sleep (Bushey *et al.* 2010; Pimental *et al.* 2016). Gene Mania links KCNAB1 via shared protein domains to AKR1A1, the human ortholog of fly gene *CG6083*. *CG6083* is involved in Alcohol dehydrogenase (NADP (+)) and Aldehyde reductase pathways. CG6083 is downregulated by a log2 factor of 1.8 in the FlyChip, and inhibition of *CG6083* in *D. melanogaster* mushroom bodies caused neuroanatomical defects such as lobe fusion in approximately 35% of cases (Koboyashi *et al.* 2006).

8.3 Comparison to Human stem cell transcriptome

The effect of *miR-137* inhibition on the transcriptome of human progenitor human stem cells has been investigated by Hill *et al.* (2014). The published work identifies over 1000 genes which had a difference in gene expression to a significance value of less than 5%. As there is clear evidence of evolutionary conservation in the *miR-137* gene, it is plausible that the mechanisms through which the miRNA works being similar across species too.

To test this hypothesis, the significant *D. melanogaster* genes from the FlyChip were compared with the genes identified in the results of the stem cell culture experiment.

8.3.1 Method: DIOPT Orthologs

To compare fruit fly genes with human genes, they must be converted into the human orthologs. The list of 30 significant *Drosophila* genes was put into the online platform DIOPT (DRSC Integrative Ortholog Prediction Tool) (accessible from www.flyrnai.org/diopt). This identified a list of 119 human genes (table 8.4).

Table 8.4: Human orthologs of the DEGs from the FlyChip

Human orthologs of the Significantly changed Drosophila genes in the FlyChip Data					
ABCG1	CES4A	GSTO2	LRRC3	PLEKHG4B	UGT1A4
ABCG2	CES5A	HPGD	LRRC4	PODN	UGT1A5
ABCG4	CYP3A4	HSPA12A	LRRC4	PODNL1	UGT1A6
ABCG5	CYP3A43	HSPA12B	LRRC4B	RNASEH1	UGT1A7
ABCG8	CYP3A5	HSPA14	LRRC4B	RTN4R	UGT1A8
AKR1A1	CYP3A7	HSPA1A	LRRC4C	SCRIB	UGT1A9
AKR1B1	CYP3A7-CYP3A51P	HSPA1B	LRRC4C	SHOC2	UGT2A2
AKR1B10	CYTIP	HSPA1L	LRRC53	SLC26A10	UGT2A3
AKR1B15	DHRS11	HSPA2	LRRC7	SOD1	UGT2B10
AKR1C1	EPPIN-WFDC6	HSPA5	MCF2	SOD3	UGT2B11
AKR1C2	ERBIN	HSPA6	MCF2L	SPINT2	UGT2B15
AKR1C3	FLNA	HSPA8	NDUFS5	TBXAS1	UGT2B17
AKR1C4	FLNB	IGSF9	NLGN1	TFPI	UGT2B28
AKR1D1	FLNC	IGSF9B	NLGN2	TFPI2	UGT2B4
AKR1E2	GLB1	INTS12	NLGN3	TMEM135	UGT2B7
ARHGEF40	GLB1L	KLHDC10	NLGN4X	TMEM38A	UGT3A1
CEL	GLB1L2	LGR4	NLGN4Y	TMEM38B	UGT3A2
CES1	GLB1L3	LGR5	PDE11A	UGT1A1	UGT8
CES2	GRASP	LGR6	PDE5A	UGT1A10	WFDC6
CES3	GSTO1	LRRC1	PLEKHG4	UGT1A3	------------

8.3.2   *Method: Matched gene expression change direction*

Hill *et al.* 2014 published a list of 1025 genes which were found to be differentially expressed in human neuro stem cells (CTXOE03) when *MIR-137* was inhibited using a Mirvarna Mir-inhibitor. Comparison of the FlyChip DIOPT-calculated orthologs and the Hill data produced a list of genes which had different expression levels in both flies and human cells when *miR-137* function is decreased. Additionally, this list was further filtered for genes who had matching expression change direction, up or down, to keep only genes which behaved in the same manner across both organisms. All comparisons and conversions were performed in Microsoft Excel. Two lists were created, one of the Fly orthologs and one of the results from Hill *et al.* and then the MATCH function was used to check if the same gene

name could be found in both data sets. The matches were then separated into another list and the expression change values were associated with the gene names using VLOOKUP function to search for the right cell in the original data table. Using the IF function, another list was created if both values for the gene (one from FlyChip and one from Hill *et al.*) were either positive or negative (figure 8.2).
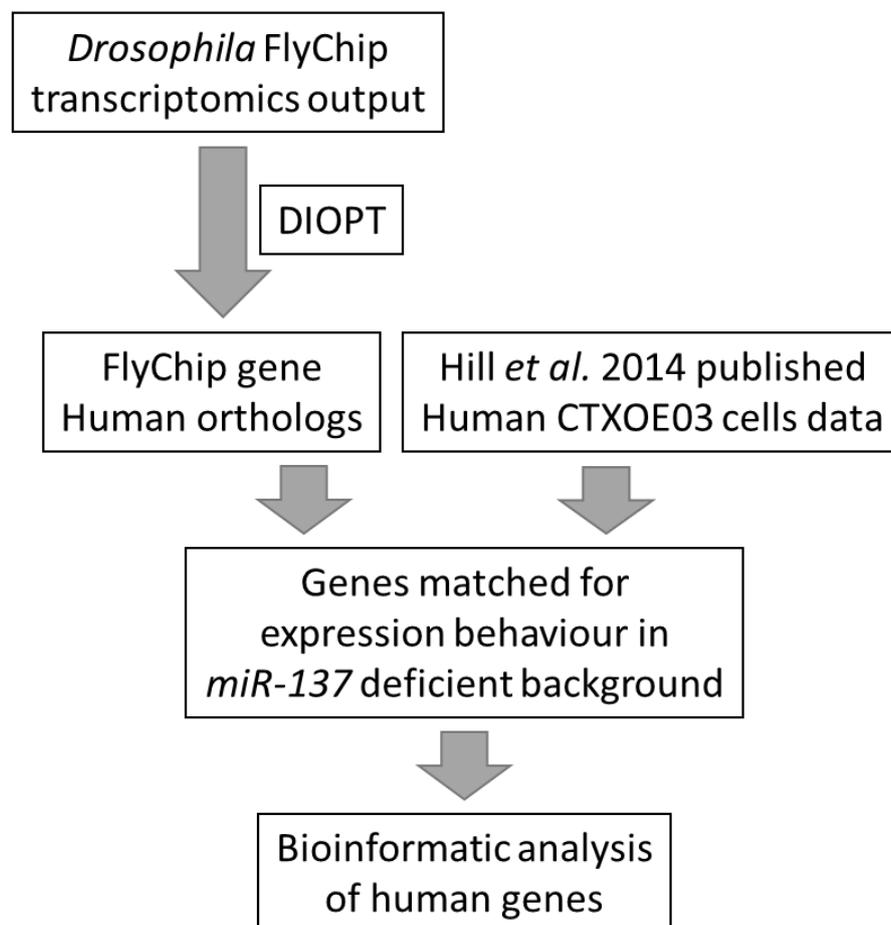
Figure 8.2: Method of creating matched-expression candidate gene set

Representation of method behind the creation of the candidate gene set from both human cells and Drosophila melanogaster transcriptome data. Two data sets were compared using Microsoft Excel to produce a list of genes which displayed similar genetic expression change behaviour when miR-137 expression was lost, or function was inhibited. One data set was a total transcriptome comparison of flies with endogenous miR-137 region removed and controls, while the second data set was the transcriptome effects of miR-137 inhibition in CTXOE03 cells compared to controls (Hill et al. 2014). The genes from the FlyChip were first converted to human orthologs utilising the DIOPT (DRSC Integrative Ortholog Prediction Tool) to allow comparison with the CTXOE03 data. The resultant multi-organism matched gene set was then used as one of the lists in further bioinformatic analysis.

8.3.2.1 Results

Comparing the total FlyChip orthologs with the significant genes from the Hill *et al.* (2014) results separated a list of 403 individual genes that behaved in the same manner across two organisms when in an environment lacking *miR-137* (table 8.5).

Table 8.5: Genes directionally matched in FlyChip and CTXOE03

flyCHIP-Hill genes							
AACS	CENTG2	EDEM1	HSGT1	MARK1	PSMA5	SFRP4	TRIM33
ABCB9	CFL2	EFEMP1	HSP90AA1	MAX	PSMB8	SFRS15	TROVE2
ABHD7	CFLAR	EIF4A1	HSP90AB1	MSI2	PTGR1	SGCE	TRPM3
ACAD10	CHCHD9	ELP3	HSPA2	MYH9	PTGS1	SH3PXD2A	TSC22D3
ACOT9	CIB1	EML1	HSPE1	MYNN	PTPN3	SHC2	TSPAN1
ACSBG1	CLSTN1	ENO3	HTR1E	N4BP2L1	PUS3	SHROOM2	TSPAN5
ADAT1	CMPK1	ENTPD6	IDS	NACA	PXDN	SIGLEC6	TSPAN9
ADRA2B	CMYA5	EPHB2	IFIT1	NANS	RAB2A	SLC15A4	TULP4
ADRM1	CNBP	ETNK2	IGFBP3	NARS	RAB33B	SLC19A1	TXNL1
AGFG1	CNN2	EXOC2	INA	NBPF14	RAB5C	SLC25A11	UBE2F
AGK	CNNM4	EXT1	ING4	NCSTN	RAB7L1	SLC25A14	UBE3C
AIFM1	CNOT4	FAHD1	INTS4	NDUFB6	RAC1	SLC25A20	UBQLN1
AK2	COL11A1	FAM177A1	IQCB1	NDUFV1	RAD21	SLC25A22	UCK2
AKR1B1	CPSF4L	FAM179B	IRS1	NFS1	RAG1AP1	SLC25A25	UPF3A
AKT3	CPT1A	FAM86B1	ITGA10	NHP2L1	RANBP6	SLC26A2	USF2
ALDH3A2	CPT1C	FBXL12	ITGA2	NHS	RANBP9	SLC26A7	USP14
ALDH9A1	CRIM1	FCGR3A	JARID1D	NKTR	RASD1	SLC35A2	USP2
ALG14	CRKL	FGF13	KBTBD7	NOP16	RBBP5	SLCO3A1	USP4
ALK	CSNK1E	FMNL1	KCNIP1	NOTCH2NL	RBED1	SLCO4A1	VAV3
ANAPC4	CSNK1G1	FOXH1	KCNK1	NP	RBM12	SLCO5A1	VPS25
ANKS1A	CTCF	FOXO3	KCNK3	NPC1	RBM28	SLK	VPS37B
ANXA6	CTDSP2	FRRS1	KCNS1	NRD1	RBM45	SMAP1	WASF3
AP1S2	CTSB	FRZB	KIAA0528	NUP155	RDH11	SNRPA	WNT7B
ARAP3	CTSC	FTL	KIAA0564	NUP54	RDH14	SNX32	XRN1
ARL5A	CUL4A	FTSJ1	KIF21A	NUP62	REEP1	snx6	YBX1
ATP5G2	CUX1	FYN	KIF3B	NUP93	REEP5	SOCS3	YWHAB
ATP6V0A2	CYB561D2	FZD3	KIF5C	NXF1	RFWD2	SOD2	ZDHHC24
ATP6V0E1	CYB5A	GADD45A	KLF11	OASL	RGS17	SORL1	ZDHHC8
AUTS2	CYFIP1	GALNTL4	KPNA2	OGFOD1	RNASEH1	SPATA2L	ZMYM5
B3GALNT1	CYTSA	GGNBP2	KRAS	OLFML2B	RNF112	SPATA6	ZNF146
BBOX1	DACH1	GPHN	KRT80	P4HA2	RNF114	SRPK2	ZNF197
BIRC3	DDB1	GPKOW	KTELC1	PAICS	RNF122	SRRM1	ZNF281
BIRC6	DDR2	GPR98	LAMA4	PATE1	RNF19A	STK35	ZNF304
BRP44L	DEAF1	GPT2	LAPTM4B	PCNA	ROCK2	STK38	ZNF33B
BTBD7	DHFR	HADHB	LGALS8	PCOLCE2	RPL18	STX5	ZNF343
BTG3	DMXL2	HBA2	LGALS9B	PELI1	RPL19	SULF2	ZNF37A
C14orf118	DNAJC12	HERC6	LIPA	PEPD	RPL24	SURF4	ZNF397
C14orf143	DNAJC18	HES2	LOC10013153	PHF6	RPS6	SV2A	ZNF407
C1orf83	DNAJC7	HIP1	LOC641298	PIGF	RPS7	SYF2	ZNF485
C1RL	DNAL1	HISPPD2A	LONP2	PIGN	RPSA	TACC1	ZNF512B
C21orf55	DNALI1	HIST1H4C	LOX	PKNOX1	RSU1	TAF1	ZNF626
C6orf162	DNM3	HIVEP1	LRBA	PLAA	RUNX1T1	TBC1D23	ZNF630
C8orf33	DPH2	HK1	LRRC8D	PMS2L1	SDAD1	TEX10	ZNF674
CANX	DPP3	HK2	LRRC8E	POLA1	SDHC	TJP2	ZNF75A
CAPRIN1	DSC2	HMGB1	LTBP3	POLR2D	SEC11A	TM2D1	ZNF768
CASKIN2	DUS4L	HNRNPAB	LTF	POLR2L	SEMA3C	TM9SF1	
CCDC127	DUSP4	HNRNPD	LYL1	PPAPDC3	SEMA6C	TMEM164	
CCT2	DYM	HNRNPL	LYPD6B	PPIC	SENP3	TMEM185A	
CD44	DYNC1H1	HNRNPUL2	MAGED1	PPIE	SERINC3	TMEM19	
CDC2L5	ECAT8	HNRPA2B1	MAP2	PRKAG2	SETDB2	TRAF7	
CDKN2C	ECE1	HSDL1	MAP3K2	PSMA3	SF3B14	TRIM25	

8.3.2.2 Discussion

The number of matches found across the two datasets was extensive and prompted further assessment of the method going forward. In discussion with Dr Richard Anney from Cardiff University, it was noted that the process of comparing the entire genome of the fruit fly (around 14400 genes) to 1000 significant DEGs was very likely to produce matches due to sheer volume. Furthermore, filtering of upregulation or downregulation creates a statistical chance of a 50/50 for a positive match. Overall this is likely to present false positive results which could consequently lead to misrepresentation of the importance of certain genes.

The DIOPT programme automatically calculates the orthologs of genes in an alternative organism. When converting from human to *Drosophila*, the list is likely to get smaller due to the overall genome size and likelihood of specific sub-genes in a simpler organism. In contrast, from *Drosophila* to humans produces multiple orthologs in the same family (eg: *D. melanogaster* gene *CG6083* is an ortholog of genes *AKR1A1, AKR1C1, AKR1C2, AKR1C3, AKR1C4,* and *AKR1E2*). With the conversion of human to fly being a degenerative method, you might lose potential genes but alternatively the conversion of fly to human includes a wider range of genes which might not be indicative of the genetic function in the fly.

8.3.3   Matched significant genes

Following the production of the match expression direction list and encountering potential false positives, a more specific comparison of the two datasets was conducted. To identify higher value gene matches, only the significant differentially expressed genes from the FlyChip were matched to the data from Hill *et al.* (2014).

8.3.3.1 Method

The top genes with a p-value of less than 0.05 were selected from the FlyChip orthologs from DIOPT. This list was then compared against the data from Hill *et al.* 2014 for matches in excel using the MATCH function to look for name matches in both lists. The genes identified from matching across the two organisms' experiments were put into DAVID and Gene Mania as previously done in section 8.1.3 to assess any enriched pathways and ontology for information on how they might relate to the fly phenotype.

8.3.3.2 Results

Comparison of the FlyChip orthologs and the Hill *et al.* (2014) data identified 3 genes of interest which were differentially expressed in both data sets (table 8.6).

Table 8.6: Significant DEGs in both FlyChip and CTXOE03

Drosophila gene	Human ortholog	Information
rnh1	*RNASEH1*	Both the fly gene and human gene were significantly downregulated, with p-values < 0.01
hsc70-2	*HSPA2*	The fly gene was significantly upregulated while the human gene was significantly downregulated, with p-values < 0.05 respectively.
CG6083	*AKR1B1*	Both the fly gene and human gene were significantly downregulated, with p-values < 0.05

DAVID did not produce any enriched pathways from the 3 candidate genes; however, Gene Mania added several genes to the pathway including members of the *TOR* family. The network appears to be mostly separated into 3 distinct sections, indicating little overlap between the mechanisms of the 3 candidate genes (figure 8.3).
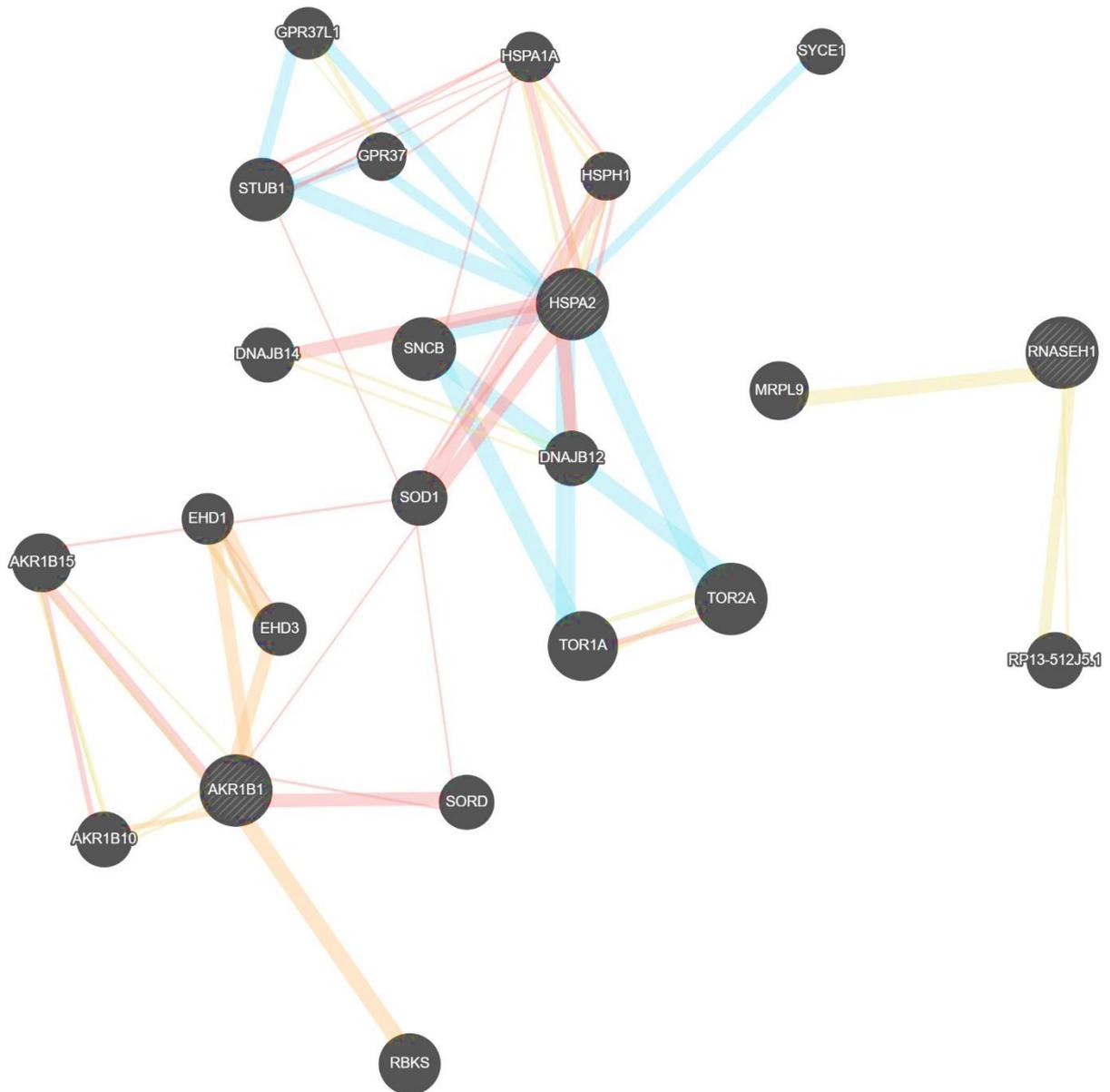
Figure 8.3: Genetic network of genes; HSPA2, AKR1B1, and RNASEH1

Created by Gene Mania (accessed from: genemania.org/). Additional genes are in full black while inputted genes are in stripes. The colours of lines represent different links: Beige = Shared protein domains, Purple = co-localisation, Red = Physical interactions, Blue = Pathways, Green = Shared protein domains, Orange = Predicted interactions. The weight of the lines is in proportion to biological process involvement with the 3 inputted genes.

8.3.3.3 Discussion

The comparison of two very different experiment results has shown a lack of crossover which is explained by substantial differences between the two studies. One of the more obvious differences is the species or tissue type used to evaluate *miR-137* effect on the transcriptome. The Hill study utilised human neural stem cells in a culture, while the FlyChip was completed using whole *D. melanogaster* heads. Further differences are within the experimental approach to *MIR-137* expression change. The CTXOE03 cells were cultured as normal and then exposed to a chemical miRNA inhibitor, preventing the expressed miRNA to function properly. In the flies, the genomic region containing the *miR-137* gene was substituted for another gene. This causes complete loss of miRNA expression throughout development and adulthood. These differences create many incomparable factors like level of *MIR-137* expression remaining and potentially biological coping mechanisms from lack of miRNA in development.

AKR1B1 is the ortholog of *Drosophila* gene *CG6083*. Aldo-keto reductase family 1 member B1, prevalent in the adrenal glands and responsible for glucose metabolism and osmoregulation (Shah *et al.* 1997). In humans, dysfunction within the gene has been implicated with increased risk of nephropathy, neuropathy, and retinopathy in insulin-dependent diabetics. The gene encodes an enzyme responsible for the metabolism of glucose aldehydes to sorbitol which is then slowly broken down by sorbitol dehydrogenase. In hyperglycaemic conditions, there is a dramatic increase in sorbitol concentration which can cause additional cellular hyperosmotic stresses and subsequently may increase damage caused by diabetes (Chung and Lamendola 1989). Expression of an RNA-inhibition molecule in *Drosophila* mushroom bodies demonstrated significant neuroanatomical defects, characterised by lobe fusion in approximately 35% of cases (Koboyashi *et al.* 2006). The mushroom body is functional comparable to the mammalian cerebral cortex (Sitara *et al.*

190

2015), which is responsible for higher cognitive thought processes such as memory, thinking, perceiving or language. Neuroanatomical changes in this region; such as decreased cortical thickness, decreased dendrite length, and overall decrease in brain volume has been associated with psychiatric disorders like SZ (Nesvag *et al.* 2007; Iritani 2013).

HSPA2 is the ortholog of *Drosophila* gene *hsc70-2*. The gene encodes heat shock 70d protein 2, which is expressed abundantly in the brain, heart, and other muscles (Roux *et al.* 1994). The protein has been implicated with sperm maturation, with cells not expressing sufficient amounts of HSPA2 showing cytoplasmic retention (Huszar *et al.* 2000). The protein was also found in elevated amounts in breast cancer tissues, both primary and metastatic (Rohde *et al.* 2005). The family of heat shock proteins have been linked to SZ and are thought to play a protective role for the CNS (Kim *et al.* 2008). Furthermore, these proteins are chaperones responsible for repair, growth, and maintenance of cellular structures that may be damaged from stress conditions (Welch 1991; Kim *et al.* 2008). It is thought that there is a potential for such cellular stresses to increase the likelihood of developing SZ and associated disorders, as expression of oxidative stress and metabolism genes differentiated in patients suffering from SZ (Prabakoran *et al.* 2004). Reducing the expression of both HSPA1 and HSPA2 in cancer cells produced an antiproliferative effect on the cells and, when downregulated individually, produced growths with different morphologies. In mice, knockout of the gene has produced partial sterility of the males without affecting the females (Dix *et al.* 1996). In the fruit fly, the enzyme is known to function as a chaperone to promote proper post-transcriptional folding of proteins. In flies with RNAi driven knockdown of *hsc70-2* through development, the larva develops anatomically smaller and feature fragmented mitochondria (Zhou *et al.* 2019).

RNASEH1 is the human ortholog of *Drosophila* gene *rnh1*. The gene encodes an endonuclease expressed in both the nucleus and mitochondria, which digests the RNA

component of RNA-DNA double-stranded hybrids (Reyes *et al.* 2015). Heterozygous double mutants of the *RNASEH1* gene was found in some individuals who developed muscle paralysis around the eyes as a result of loss of *RNASEH1* function. The cells were shown to have fragmented mitochondria and lower action potential across the mitochondrial membrane, synonymous with mitochondrial dysfunction (Reyes *et al.* 2015). In mice, deletion of the *rnaseh1* ortholog caused developmental arrest at around 8 days, due to a decrease in mitochondrial DNA and subsequent apoptotic cell death (Ceritelli *et al.* 2003). Partial deletion of the genomic sequence of *rnh1* in *D. melanogaster* causes severe developmental defects including a delay in developmental time, pre-pupal and pupal developmental lethality, and melanotic tumours (Filippov *et al.* 2001).

Putting either the human genes or the fly genes into DAVID did not produce any significant results for pathway enrichment, however Gene Mania produced some additional genes which are linked to the 3 candidates. These include members of the TOR family in a pathway with *HSPA2* (Wu *et al.* 2010), predicted links between *EDH1/3* and *AKR1B1* (Stuart *et al.* 2013), and pathways between *SYCE1* and *HSPA2* (Jassal *et al.* 2020). Members of the TOR family have been linked to dystonia, an involuntary muscle movement disorder (Charlesworth *et al.* 2017). *EDH3* is linked to MDD in Chinese population studies (Shi *et al.* 2012), and dysfunction of the fly ortholog *past1* creates neuromuscular junction synapse abnormalities and developmental delay of ~20% (Olswang-Kutz *et al.* 2008; Koles *et al.* 2015). The candidate genes are separated in the gene network map, indicating little overlap in the mechanisms that they are involved in. This could be indicative of the complexity of miRNA regulation networks and the difficulty in comparing two organisms.

8.3.4   Conclusion: FlyChip vs Stem cell data

In conclusion, the attempt to find overlap between the genetic changes in human CTXOE03 and *D. melanogaster* heads demonstrate did not produce significant results, meaning the acceptance of the null hypothesis from Objective E hypothesis 1 (Chapter 1.8). The initial comparison of the entire FlyChip to the significant genes produced from *miR-137* analysis in CTXOE03 cells had a high potential to produce false results and therefore the method was changed to look at higher significance target genes only. Overall, there were not many similarities between the significant data sets, with a calculated 0.25% correlation coefficient. This produced a select subset of just 3 genes across the ~14000 FlyChip genes (30 with a p-value <0.05) and over 1000 genes from Hill *et al.* study (2014). Arguably, there are many differences within the experimental designs of the studies producing the datasets that this result is not surprising. However, the results interestingly demonstrate a heat shock protein which is significantly expressed differently across both experiments but in different expression directions. This could be evidence of a coping mechanism within the flies, as heat shock proteins are known to have protective properties for the nervous system (Kim *et al.* 2008). The three genes identified as significant across both data sets shows some potential implications for *miR-137* expression regulating pathways particularly *HSPA2*, and its apparent involvement with *TOR* and *EDH* genes. The phenotypes associated with dysfunction of these genes are synonymous with some symptoms of degenerative neurological diseases and psychiatric disorders, creating a putative method of action for the locomotor defect exhibited in Chapter 5.

8.4   Comparison to human GWAS

With many of *miR-137* target genes already implicated with neurological diseases and psychiatric disorders, further comparison of the significant FlyChip genes with phenotypically relevant GWAS could bring new links between *miR-137* and these human diseases.

8.4.1   UKBiobank/FINNGEN Neale Lab published datasets

The UK biobank is an ongoing project to provide an open database of some 500000 people who were subjected to a series of medical tests and measurements alongside a self-answered questionnaire. The project was started in 2006 and aims to follow many of the individuals for up to 30 years accumulating additional data to be added to the biobank. Currently there is a wealth of data alongside the genomes including body measurements, medical diagnoses, activity monitoring, MRI scans, and more. The UK biobank data has some obvious flaws with the method of self-reported and online questionnaire style of data gathering, however some parameters are much more robust and useful for studying. These can be particularly useful alongside more specific databases searching for correlations.

Finngen is a Finnish project which is accumulating a holistic dataset from novel data and the existing biobanks in collaboration with hospitals, universities, and pharmaceutical companies. The project ultimately aims to produce a similar database to the UKBiobank with 500000 participants, each with genomic data and health registry diagnoses information. Unlike the UKbiobank, the diagnoses are clinical based and therefore likely to be more reliable than self-reported participant measures. However, the Finngen project started in 2016 and is still in early stages of collecting data which makes it a smaller database currently.

While many studies have been conducted utilising the data put out, researchers within the Ben Neale lab have outputted cleaned and imputed data from their GWAS research that are open source to anybody that wants to download and use them. This data is accessible from www.nealelab.is/. The data has been separated into the individual questions/parameters that were put into the database, such as "self-reported schizophrenia" or "smokes tobacco on most days".

GWAS data was selected for parameters associated with neurological or psychiatric disorders (table 8.7). This data was further edited and then put through an online platform to calculate significant SNPs and produce Manhattan plots of genes with genome wide significance for each parameter. The resulting SNPs and genes were then cross referenced again the candidate gene list (table 8.5).

Table 8.7: Selected neurological UKBiobank and FINNGEN parameters.

Database	Identifier	Category
UKBiobank	20002_1262	Self-reported: Parkinson's disease
	20002_1286	Self-reported: depression
	20002_1289	Self-reported: schizophrenia
	20002_1291	Self-reported: mania/bipolar disorder/manic depression
	20002_1482	Self-reported: chronic fatigue syndrome
	20002_1614	Self-reported: stress
	20003_1140867490	Medicine: lithium product
	20126_1	Bipolar I Disorder
	20126_3	Probable Recurrent major depression (severe)
	20544_1	Diagnosed: Social anxiety or phobia
	20544_10	Diagnosed: mania/bipolar disorder/manic depression
	20544_11	Diagnosed: Depression
	20544_14	Diagnosed: Autism, Asperger's
	20544_15	Diagnosed: Anxiety
	20544_2	Diagnosed: schizophrenia
	23104_irnt	Body mass index (BMI)
	G20	Diagnoses - main ICD10: G20 Parkinson's disease
	G47	Diagnoses - main ICD10: G47 Sleep disorders
	AD	Alzheimer's disease
	F5_ALLANXIOUS	All anxiety disorders
FINNGEN	F5_DEMENTIA	Dementia
	F5_DEPRESSIO	Depression
	F5_SCHIZO	Schizophrenia, schizotypal and delusional disorders

8.4.1.1 *Python editing of GWAS results data*

Editing of Neale's public access clean and imputed GWAS data was done in python 3 programming language using Anaconda PC platform (See supplementary information for script). The original download files are not compatible with FUMA as the site requires separated chromosome and case-pair position columns, which in Neale's download files were conjugated as "chromosome:position:amino:substitute" in a single column.

The code was written to produce the minimum required for the FUMA input, which was stated as chromosome, position, and p-value only. The code mainly utilised the "pandas" library ability to manipulate dataframes. It was written to read and decrypt the .bgz file with the gzip library, then extract and separate the variant column, and then create a new dataframe with only the three necessary columns (figure). This dataframe was then outputted to a .csv file which was compatible with FUMA in both size and filetype.

The original python code extracted the "variant" column and created a new dataframe separating the column at the ":" separator and then joined them to the original dataframe. This created 4 new columns at the beginning named "Chromosome", "position", "amino1", and "amino2". This dataframe was then exported to .csv file and then zipped with GZIP function to be used with FUMA. Unfortunately, the filesize even when zipped was too large for the website as it takes a maximum of 600mb per upload.

The next code created the same end dataframe with the new columns but removed rows which had a p-value of over 0.01 and were therefore not of interest. As a result, the dataframe was significantly shorter and when zipped released a file of a size compatible with the FUMA site. However, it was unclear if the selective deletion of GWAS data had implications for the average p-value calculations conducted within FUMA that would have given false results. Upon development of the final code and following testing it was found that selective deletion of the GWAS data had no effect on the significant output results.

The final code which simply created a new dataframe of just chromosome, position, and p-value, was found to provide a small enough filesize while also not selectively deleting non-significant SNPs.

8.4.1.2 *Calculating significant SNP and genes*

The Complex Genetic Traits lab at VA university in Amsterdam runs an online Functional mapping and annotation (FUMA) platform available for public use at fuma.ctglab.nl/. The platform is primarily split into 3 sections: GENE2FUNC, Cell Type, and SNP2GENE. FUMA utilises several different pre-existing data sets and tools for these analyses including PLINK and the 1000 genomes project for calculating r2 and MAF, MAGMA for gene set analyses, GTEX for tissue-based gene expression, and MsigDB for gene sets grouping (Purcell *et al.* 2007; Liberzon *et al.* 2011; 1000 genomes project consortium 2015; Leeuw *et al.* 2015; GTEX consortium 2015).

SNP2GENE takes GWAS results and analyses them for significant SNPs and subsequent related genes. The output includes a gene-based Manhattan plot, and easily downloadable lists of significant SNPs and associated information like genomic location. The minimalist GWAS files created from python manipulation of the Neale data were uploaded into SNP2GENE. The number of samples was added to the settings (361141) but no other parameters were changed. The total results were downloaded from the download tab and the "snps.txt" file was opened and copied to excel. The SNPS and correlating nearest genes of all significant hits were checked for matches within the target gene list created from the directional matched CTXOE03 and FlyChip data.

8.4.1.3 Results

The SNP2GENE programme produced a total of 2508 significant SNPs or genes across the selected parameters. The outcomes which have a matched gene against the candidate gene list (directionally matched FlyChip and Hill genes) are listed here, while the others are in supplementary materials.

Parameter 20002_1289 refers to the initial UKBiobank assessment centre whereupon the participant was asked to disclose any history of medical disorders to a qualified nurse or doctor in a verbal interview. Participants who were unable to remember, or were uncertain of, the specific disorder were able to describe the symptoms, and this was later classified by the medical professional. There were 642 people who reported a history of SZ out of 383782 participants. This data was then cleaned and imputed by Neale *et al.* into 361194 samples, of which 391 cases against 360750 controls. Genome wide significance was defined at P = 0.05/19712 = 2.537e-6, identifying significant genesTCEAL1, VHLL, and CSF3L (figure 8.4).
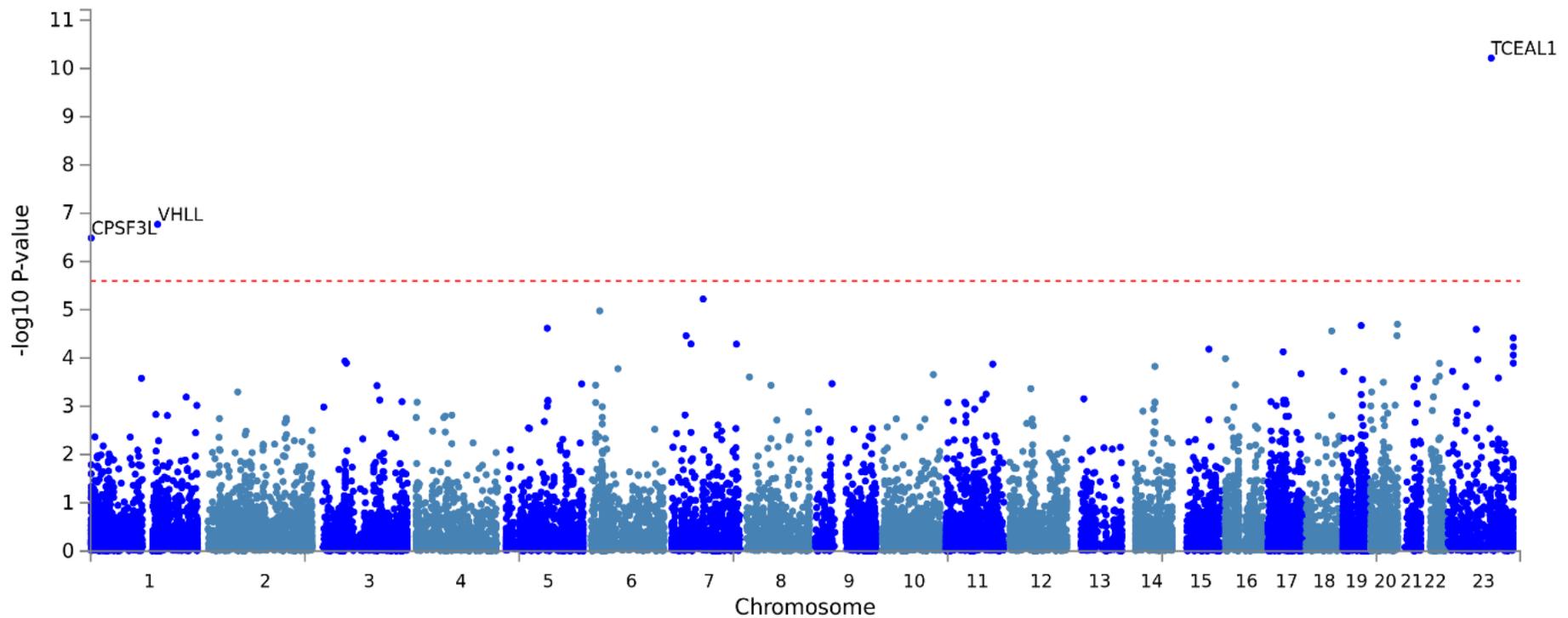
Figure 8.4: Manhattan plot for UKBiobank 20002_1289; Reported schizophrenia

The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: TCEAL1, VHLL, and CSF3L.

Parameter 20544_11 refers to participants disclosing if they had any current or historical diagnoses for depression. The question was part of an online follow-up questionnaire revolving around mental health. There were 50099 participants, of which 33422 claimed to have been diagnosed at some point in their life. This data was cleaned and imputed by Neale *et al.* and a total of 117782 was analysed, with 25087 cases against 92695 controls. Significant genes above the determined genome wide significant level were: *SLC25A20, CCDC71, C3orf84, CCDC36, BTN3A2, BTN1A1, HIST1H2BL, HIST1H2AJ, HIST1H2BN, OR2B2, ZKSCAN4, PGBD1, TMEM106B, RHOF, SETD1B, METTL9, SLC47A1,* and *SUPT4H1* (figure 8.5).

Parameter 20544_2 refers to participants disclosing if they had any current or historical diagnoses for SZ. The question was part of an online follow-up questionnaire revolving around mental health. There were 50099 participants, of which 157 claimed to have been diagnosed at some point in their life. This data was cleaned and imputed by Neale *et al.* and a total of 117716 was analysed, with 103 cases against 117603 controls. Significant genes above the determined genome wide significant level were: *AL645608.2, RYBP, DHFRL1, RGS14, RING1, RP11-89N17.1, OR1L3, SETD1B, PPP4R4, SERPINA10, RP11-17M16.1, FAM187B, MAOB,* and *CDK16* (figure 8.6)
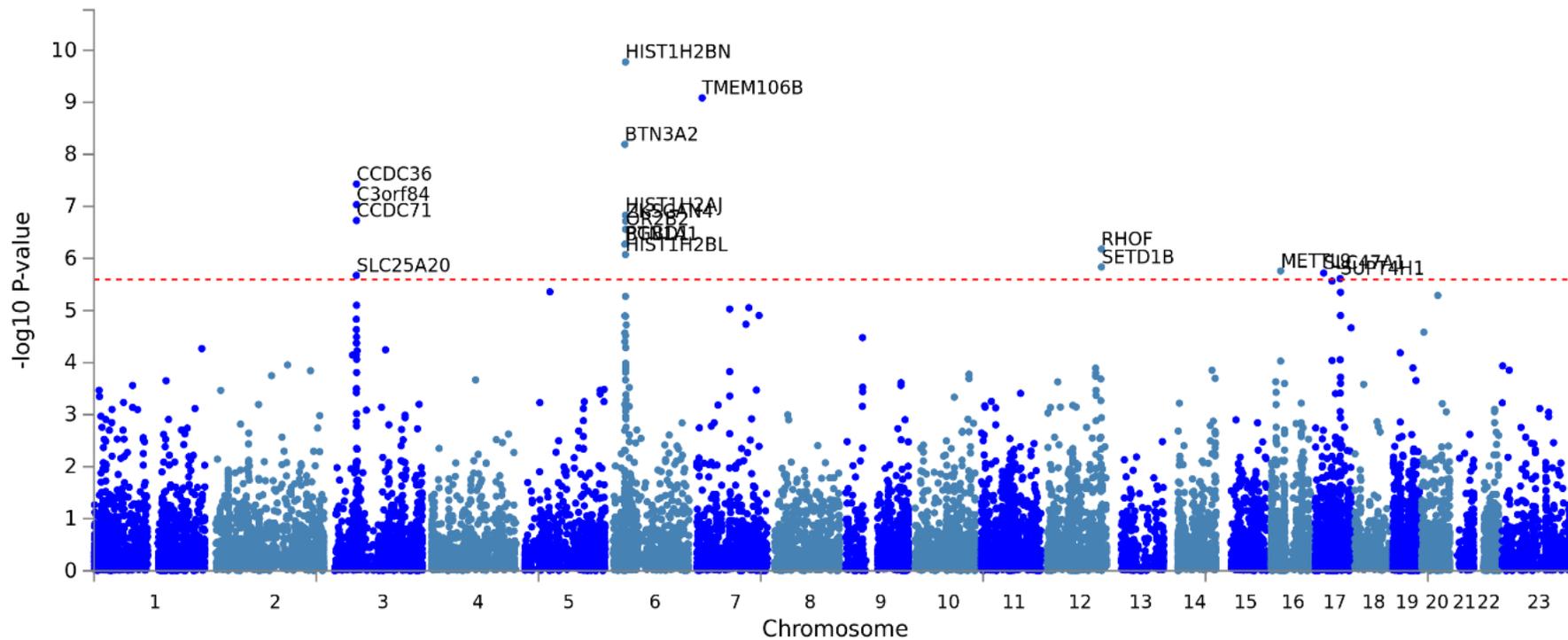
Figure 8.5: Manhattan plot for UKBiobank 20544_11; Diagnosed Depression.

The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: SLC25A20, CCDC71, C3orf84, CCDC36, BTN3A2, BTN1A1, HIST1H2BL, HIST1H2AJ, HIST1H2BN, OR2B2, ZKSCAN4, PGBD1, TMEM106B, RHOF, SETD1B, METTL9, SLC47A1, and SUPT4H1.

Figure 8.6: Manhattan plot for UKBiobank 20544_2; Diagnosed schizophrenia.

The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: AL645608.2, RYBP, DHFRL1, RGS14, RING1, RP11-89N17.1, OR1L3, SETD1B, PPP4R4, SERPINA10, RP11-17M16.1, FAM187B, MAOB, and CDK16.

Comparison of the matched expression direction candidate gene list and the results from FUMA analysis revealed a select few genes present in both (table 8.8). The FUMA output also included SNP locations within the genes, of which two (rs7607527 and rs114534142) are exonic while rs79253132 is intergenic and rs149013672 is intronic with no likely subsequent gene effect. Of the two exonic SNPs, only rs114534142 was reported to influence the gene where it would cause a missense variant and subsequent coding sequence change through dbSNP.

Table 8.8: Genes in neurological GWAS and candidate list

GWAS group	GWAS meaning	Associated gene
20002_1289	Self-reported: Schizophrenia	*HADHB*
20544_11	Diagnosed: Depression	*SLC25A20*
20544_2	Diagnosed: Schizophrenia	*SDAD1*
20544_2	Diagnosed: Schizophrenia	*NUP54*
G20	Diagnosis: ICD10 Parkinson's disease	*HIST1H4C*

8.4.1.4 Discussion

Comparison of the candidate gene list and outputs from the FUMA SNP2GENE function (both Manhattan plots and significant SNPs) offered a select list of 5 human genes. Four of these genes are from psychiatric disorder related parameters (SZ and Depression), with the other being neurological diseases (Parkinson's disease). The SNPs do not currently have reported clinical significance on ClinVar, but further investigation of the genes provided some options for mechanism through which *miR-137* may cause the locomotor or sleep defects.

HADHB, or *Hydroxyacyl-CoA Dehydrogenase Trifunctional Multienzyme Complex Subunit Beta*, is the beta subunit of the mitochondrial trifunctional protein (MTP), which catalyses the last three steps of mitochondrial beta-oxidation of long chain fatty acids (Das *et al.* 2006). This oxidation is the major metabolic system to catabolize fatty acids producing energy for many cellular processes (Das *et al.* 2006; Kishita *et al.* 2012). In humans, problems with the MTP is diagnosed as mitochondrial trifunctional protein deficiency, characterised with symptoms of lethargy and hypoglycaemia (Das *et al.* 2006; Choi *et al.* 2007). Furthermore, if not treated the build-up of fatty acid metabolites can cause tissues damage to major organs. Mutations specifically within the *HADHB* gene can cause similar problems to MTP-deficiency, alongside hypotonia, neuropathy, myopathy, liver problems, and even sudden death in infant sufferers (Das *et al.* 2006; Choi *et al.* 2007). The gene was identified in relation to SNP rs7607527, which was found to be significant in the GWAS parameter "self-reported schizophrenia" from the UKBiobank in Section 8.3.5.2 of this project. HADHB proteins were found to bind to dysbindin-1 in mice skeletal muscle, a known SZ related protein (Oyama *et al.* 2009). In both the FlyChip experiment and Hill *et al.* study the expression levels of *HADHB* (or fly ortholog FBgn0025352 *thiolase*) decreased when *miR-137* levels were decreased/removed. The *thiolase* gene shares a 66.2% identity with the human *HADHB* gene and removal of genomic *thiolase* has been seen to negatively influence *Drosophila* lifespan and locomotor ability (Kishita *et al.* 2012). The phenotype and expression change direction match up with *miR-137* deficient flies in this project, providing a possible mechanism for *miR-137* involvement in the locomotor defect.

The *SLC25A20* gene encodes *carnitine-acylcarnitine translocase* (*CACT*), a component of the carnitine cycle, which is necessary to shuttle long-chain fatty acids from the cytosol into the intramitochondrial space for mitochondrial beta-oxidation of fatty acids. In humans, problems within the *CACT* gene can lead to CACT-deficiency-disease, which is characterised

with symptoms of drowsiness, skeletal muscle weakness, hepatic abnormalities, increased

ammonia in the blood and subsequent death from multiorgan failure and neuropathy (Huizing

et al. 1998; Rubio-Gozalbo *et al.* 2004; Yan *et al.* 2017). The *SLC25A20* gene expression has

also been investigated as a blood marker for Parkinson's disease, and furthermore was

significantly enriched within a sleep regulating pathway (Jiang *et al.* 2019). The *SLC25A20*

gene has multiple likely orthologs in Drosophila; FBgn0031881 *mme1*, FBgn0032219

CG4995, and FBgn0039525 *CG5646*. In both the Flychip results and the data from Hill *et al.*

(2014) these genes showed an increase in expression with removal/inhibition of genomic

miR-137 expression. In drosophila, knockdown of *mme1* gene via RNAi based silencing

caused a decrease in lifespan of approximately 30% (Cui *et al.* 2016). The decrease in

lifespan matches phenotype data from this study of *miR-137* KO, however the phenotype was

exhibited with a decrease in drosophila *mme1* while the Flychip/hill data suggests the gene

expression would have been higher in the *miR-137^{KO}* strain.

SDAD1 or *SDA containing domain 1* gene belongs to the SDAD family, which has been

recently discovered (Zeng *et al.* 2017; Jing *et al* 2019). In humans, *SDAD1* gene expression

has been found to positively correlate with the proliferation of colon cancer cells (Zeng *et al.*

2017) and in mice the gene was shown to influence cardiac hypertrophy (Jing *et al* 2019). In

drosophila FBgn0033379 *mys45a*, both flychip and Hill *et al*. data showed an increase in

expression when *miR-137* was removed. In *Drosophila*, manipulating the expression of

mys45a gene with the UAS/Gal4 constructs results in neuroanatomy defects and shorter

lifespans (Neely *et al.* 2010; Neumuller *et al.* 2011). However, these phenotypes were

displayed when the expression of *mys45a* was decreased, which contrasts with the gene

expression and phenotype behaviour of this study.

NUP54 encodes nucleoporin54, part of the complex protein structures which create pores in the nuclear membrane in eukaryotic cells. This subunit is known to bind directly to the nuclear import factors NUP97 and NTF2 (Hu *et al.* 1996). Loss of genomic *NUP54* increased DNA ionising radiation susceptibility, resulting in an increase in cell death and an increase in chromosome aberrations (Rodriguez-Berriguete *et al.* 2018). Furthermore, NUP54 is directly involved in DNA repair during homologous recombination, as demonstrated by a lack of RAD51 activity in *NUP54*-deficient HeLa (human cervical cancer) cell cultures (Rodriguez-Berriguete *et al.* 2018). In both humans and *C. elegans* the NUP54 proteins have been found to directly recruit polo-like kinase 1 (PLK1) proteins in breakdown of the nuclear envelope during mitosis, a process vital for proper chromosomal separation and ultimately successful cell replication (Martino *et al.* 2017). This project identified the *NUP54* gene from its association with SNP rs79253132 in the GWAS data for UKBiobank participants with a medical diagnosis for SZ. In drosophila, the *NUP54* gene ortholog is FBgn0033737 *nup54*, and in both flychip and Hill *et al.* data there was an increase in expression when *miR-137* was downregulated/ablated. Studies into *Drosophila nup54* have shown that RNAi mediated knockdown of expression in the CNS does not affect stock viability or behavioural response to high temperatures (Neely *et al.* 2010). At present, although *NUP54* appears to be involved in vital cellular processes there are no phenotype matches to this project or those of neurological diseases and psychiatric disorders

HIST1H4C is a member of the histone family. The gene encodes a protein which is involved in the wrapping of DNA in eukaryotic chromatin during cell division. Methylation and acetylation of specific parts of the HIST1H4C protein are common epigenetic features in cancer cells (Fraga *et al.* 2005; Savage *et al.* 2008). Dang *et al.* (2009) reported a correlation with *HIST1H4C* expression and age, along with SIRT1 protein abundance, resulting in compromised transcriptional silencing in older cells. The *HIST1H4C* gene was highlighted as

significant within the Parkinson's disease GWAS data from the UKBiobank. In both the human cells in Hill *et al.* and in the Flychip from this study, the gene expression (in the human or *Drosophila* gene ortholog FBgn0013981 *his4r*) increased with the removal of *MIR-137*. Studies into *his4r* gene expression in *Drosophila* has shown that introducing a hypomorphic allele has a partially lethal effect on stocks and complete sterility (McKay *et al.* 2015).

8.4.1.5   *UKBiobank GWAS limitations and data set refinement*

Through communications with Dr Richard Anney of Cardiff University, discrepencies were identified within the datasets used from the Neale lab releases of the UKBiobank data. This was likely due to incomplete processing of the publicly available data, leading to potential false positive results. Alongside this, the UKBiobank is potentially flawed in terms of data collection metric. Enrolment into the database was on a voluntary basis, and if you have illnesses you are unlikely to participate in such a study, producing a skew in the population spread. Therefore, parameters such as "diagnosed with schizophrenia" may not truly be representative, and selection of parameters should be less specific to disease and diagnoses.

8.4.2   *UKBiobank Sleep Parameters*

Through communication with Richard Anney of Cardiff University, lists of top genes from some pre-analysed UKBiobank GWAS data were obtained (table 8.9). The parameters were in relation to sleep based characteristics, rather than specifically searching for disease diagnoses as identified as problematic in Chapter 8.4.1. The genes were pre-filtered to only those which have minimum reported p-value of 0.00001. These were then cross-referenced to

the significant FlyChip gene list to identify other potential links with the FlyChip information and other human parameters/diseases/disorders.

Table 8.9: Sleep parameters used from UKBiobank.

UKBioBank ID	Parameter
1160	Sleep duration
1170	Getting up in the morning
1180	Morning/evening person
1190	Nap during day
1200	Sleeplessness / insomnia
1220	Daytime dozing / sleeping

8.4.2.1 *Method*

The lists of top GWAS significant genes were copied into Microsoft Excel and compared to the human orthologs calculated by putting the significantly DEGs from the FlyChip into DIOPT. If the gene name was found in both list then this constituted as a match, and the list of these genes were separated for further analysis.

Dr Richard Anney conducted gene set enrichment analysis by using the Multi-marker Analysis of Genomic Annotation (MAGMA) programme on the genes which were found in both the top GWAS results and the FlyChip significant orthologs. MAGMA uses a multiple regression approach to properly incorporate linkage disequilibrium between markers and to detect multi-marker effects, therefore producing better statistics for the association of a gene set to the GWAS SNPs.

The fly genes identified in both the GWAS results and the FlyChip orthologs were put into DAVID to assess any enrichment in pathways. The fly genes were used because the conversion of fly to human genes creates multiple related genes which falsely identify as enriched within the gene set. The human genes were used in Gene Mania to identify any additional genes involved in the pathway.

8.4.2.2 Results

Comparison of the UKBiobank sleep parameters and the significant FlyChip gene orthologs have identified a list of 28 human genes from 7 *D. melanogaster* genes which exist in both the top GWAS results and DEGs from the FlyChip (table 8.10).

Table 8.10: Genes in sleep related GWAS and significant candidate list

Human Gene	*Drosophila* ortholog ID	*Drosophila* gene	FlyChip expression change (Log2 3dp)
ABCG5	FBgn0003996	*w*	3.830
ABCG8	FBgn0003996	*w*	3.830
AKR1A1	FBgn0036183	*CG6083*	-1.788
AKR1C1	FBgn0036183	*CG6083*	-1.788
AKR1C2	FBgn0036183	*CG6083*	-1.788
AKR1C3	FBgn0036183	*CG6083*	-1.788
AKR1C4	FBgn0036183	*CG6083*	-1.788
AKR1E2	FBgn0036183	*CG6083*	-1.788
CES2	FBgn0027584	*CG4757*	2.281
CES3	FBgn0027584	*CG4757*	2.281
CES4A	FBgn0027584	*CG4757*	2.281
CES5A	FBgn0027584	*CG4757*	2.281
GLB1L3	FBgn0260746	*Ect3*	-1.774
HSPA1A	FBgn0001217	*Hsc70-2*	2.300
HSPA1B	FBgn0001217	*Hsc70-2*	2.300
HSPA1L	FBgn0001217	*Hsc70-2*	2.300
HSPA2	FBgn0001217	*Hsc70-2*	2.300
HSPA5	FBgn0001217	*Hsc70-2*	2.300
HSPA6	FBgn0001217	*Hsc70-2*	2.300
HSPA8	FBgn0001217	*Hsc70-2*	2.300
LGR6	FBgn0267435	*Chp*	-1.852
LRRC1	FBgn0267435	*Chp*	-1.852
LRRC4	FBgn0267435	*Chp*	-1.852
LRRC7	FBgn0267435	*Chp*	-1.852
MCF2L	FBgn0050456	*CG30456*	-2.492
NLGN1	FBgn0027584	*CG4757*	2.282
PLEKHG4	FBgn0050456	*CG30456*	-2.492
SLC26A10	FBgn0050456	*CG30456*	-2.492

The MAGMA analysis conducted at Cardiff University shows that there was not a significant enrichment of matched genes found in any of the GWAS results (table 8.11).

Table 8.11: MAGMA enrichment analysis of sleep related GWAS.

GWAS	P-value (3dp)	Standard Error (3dp)
1160-Sleep duration	0.539	0.097
1170-Getting up in the morning	0.763	0.098
1180-Morning/evening person	0.456	0.11
1190-Nap during day	0.798	0.098
1200-Sleeplessness / insomnia	0.964	0.094
1220-Daytime dozing / sleeping	0.403	0.089

DAVID did not produce any enriched pathways from the 7 candidate genes, however Gene Mania identified additional genes including *KCNAB* proteins, *HSPA* proteins, and *AKR* proteins. The network map shows 3 main clusters of interactions around similar protein domains, as expected from multiple orthologs coming from one fly gene (figure 8.7).
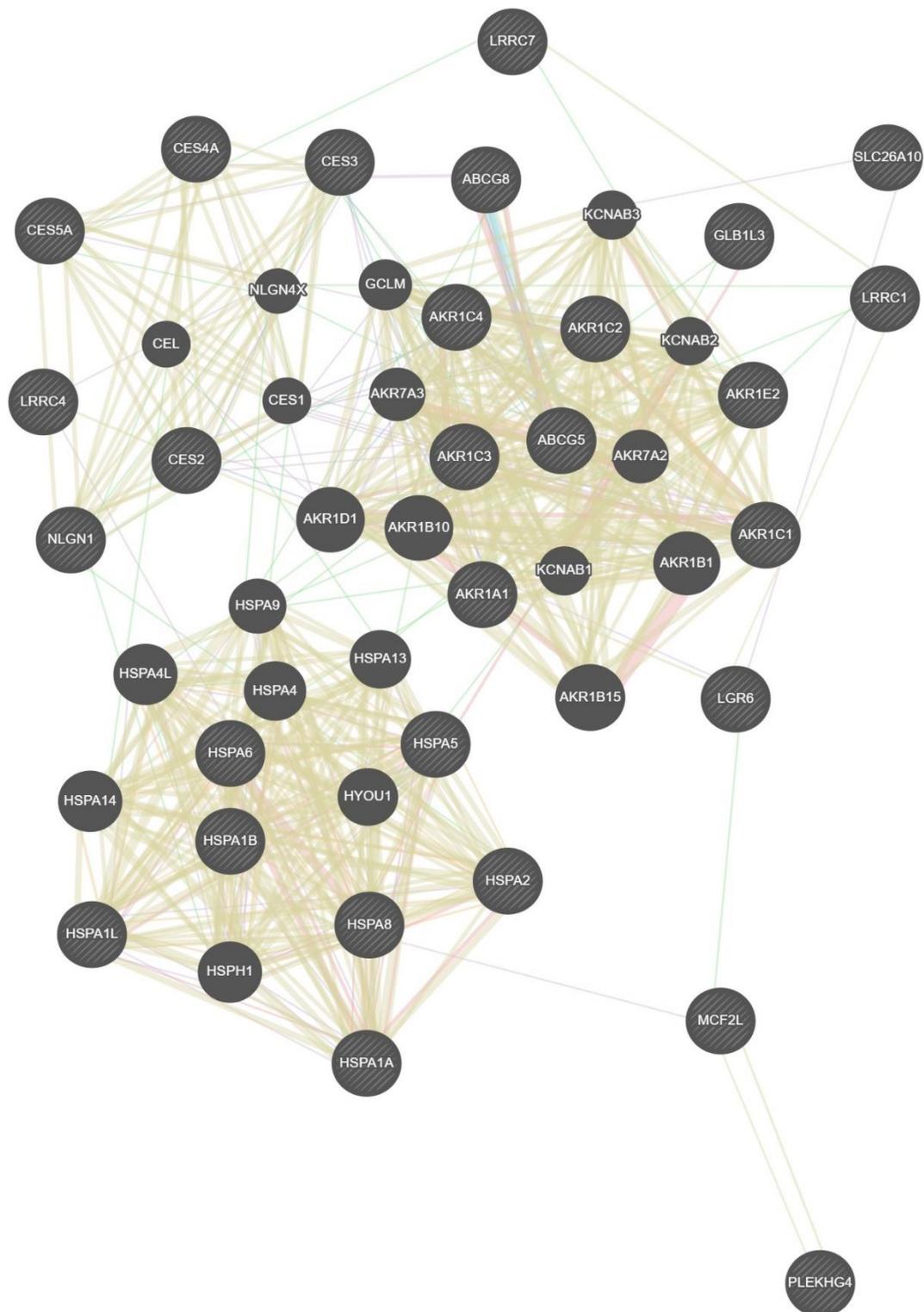
Figure 8.7: Genetic network of FlyChip DEGs in sleep related GWAS.

Genetic network of the 28 human orthologs of the 7 significant DEGs in the FlyChip and UKBiobank Sleep related GWAS results. The analysis and visualisation were created by Gene Mania (accessible from: genemania.org). Additional genes are in full black while inputted genes are in stripes. The colours of lines represent different links: Beige = Shared protein domains, Purple = co-localisation, Red = Physical interactions, Blue = Pathways, Green = Shared protein domains, Orange = Predicted interactions. The weight of the lines is in proportion to biological process involvement with the 3 inputted genes.

8.4.3   *Schizophrenia and major depressive disorder*

Through communication with Dr Richard Anney at Cardiff University, a list of top genes was obtained for GWAS studies into MDD and SZ (Pardinas *et al.* 2016; Ripke *et al.* 2018). The genes were pre-filtered to only those which have minimum reported p-value of 0.00001. These were then cross-referenced to the significant FlyChip gene ortholog list to identify genes of interest.

8.4.3.1   *Method*

The lists of top GWAS significant genes were copied into Microsoft Excel and compared to the human orthologs calculated by putting the significantly DEGs from the FlyChip into DIOPT. If the gene name was found in both list then this constituted as a match, and the list of these genes were separated for further analysis.

Dr Richard Anney conducted gene set enrichment analysis by using the Multi-marker Analysis of Genomic Annotation (MAGMA) programme on the genes which were found in both the top GWAS results and the FlyChip significant orthologs. MAGMA uses a multiple regression approach to properly incorporate linkage disequilibrium between markers and to detect multi-marker effects, therefore producing better statistics for the association of a gene set to the GWAS SNPs (De Leeuw *et al.* 2014).

The fly genes identified in both the GWAS results and the FlyChip orthologs were put into DAVID to assess any enrichment in pathways. The fly genes were used because the conversion of fly to human genes creates multiple related genes which falsely identify as enriched within the gene set. The human genes were used in Gene Mania to identify any additional genes involved in the pathway.

Comparison of the MDD and SZ GWAS results and the significant FlyChip gene

orthologs have identified a list of 13 human genes from 7 *D. melanogaster* genes which exist

in both the top GWAS results and DEGs from the FlyChip (table 8.12).

Table 8.12: Genes in psychiatric disorder GWAS and significant candidate list

Human Gene	*Drosophila* ortholog ID	*Drosophila* gene	FlyChip expression change (Log2 3dp)
GLB1L	FBgn0260746	*Ect3*	-1.774
GLB1L2	FBgn0260746	*Ect3*	-1.774
GLB1L3	FBgn0260746	*Ect3*	-1.774
HSPA1A	FBgn0001217	*Hsc70-2*	2.300
HSPA1B	FBgn0001217	*Hsc70-2*	2.300
HSPA1L	FBgn0001217	*Hsc70-2*	2.300
IGSF9B	FBgn0010473	*Tutl*	0.603
LRRC4	FBgn0267435	*Chp*	-1.852
NDUFS5	FBgn0031228	*ND-15*	2.573
NLGN4X	FBgn0027584	*CG4757*	2.281
PLEKHG4	FBgn0050456	*CG30456*	-2.492
SCRIB	FBgn0267435	*chp*	-1.852
SLC26A10	FBgn0050456	*CG30456*	-2.492

The MAGMA analysis conducted at Cardiff University shows that there was not a

significant enrichment of matched genes found in either of the GWAS results (table 8.13).

Table 8.13 MAGMA enrichment analysis of psychiatric disorder GWAS.

GWAS	P-value (3dp)	Standard Error (3dp)
Major depressive disorder	0.826	0.092
Schizophrenia	0.608	0.11

DAVID did not produce any enriched pathways from the 7 fly genes; however, Gene Mania added several genes to the pathway including *GLB1*, *SLC26* proteins, and *HSPA* proteins. The network map shows 3 main clusters of interactions around similar protein domains, as expected from multiple orthologs coming from one fly gene (figure 8.8).
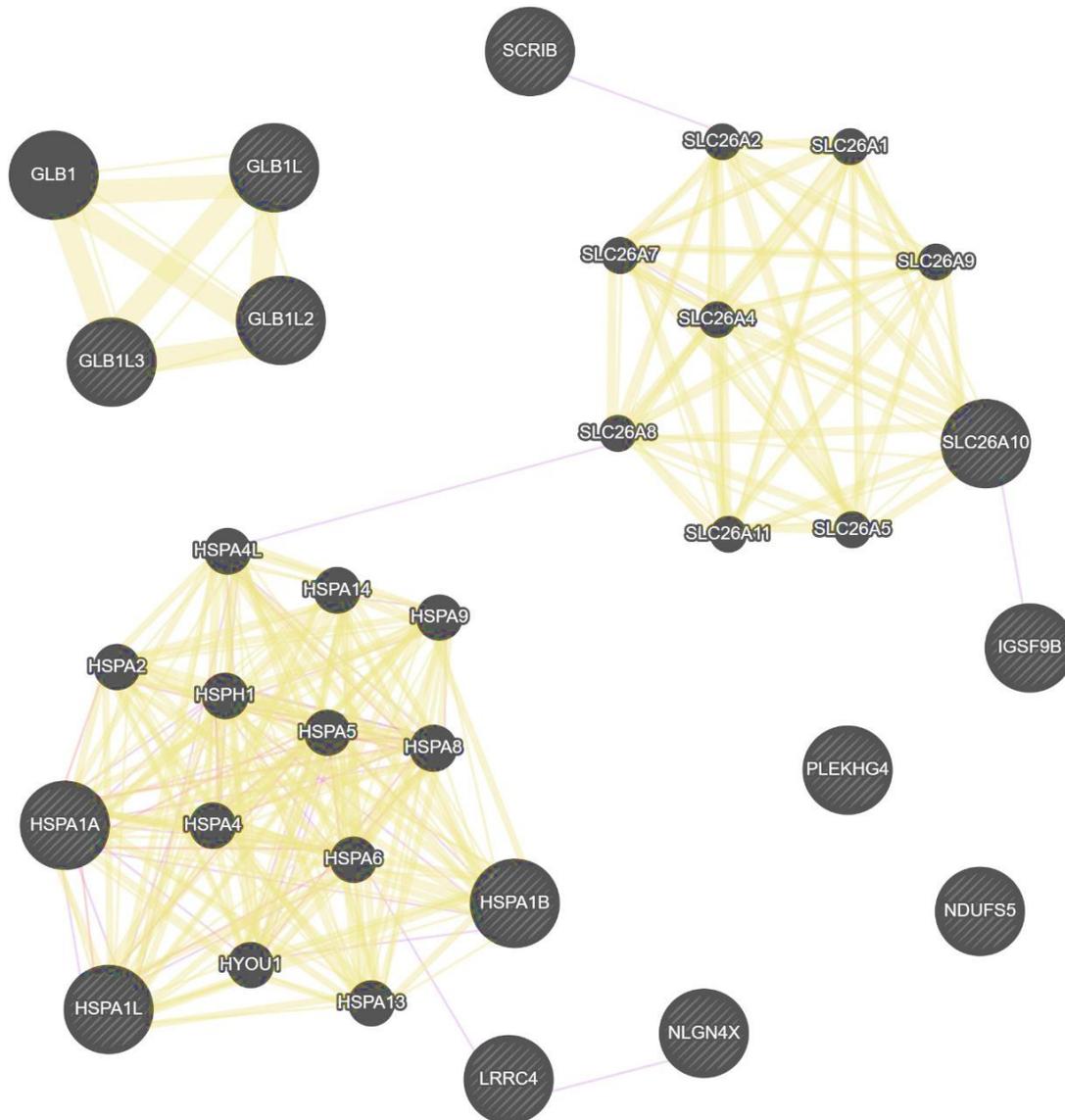
Figure 8.8: Genetic network of FlyChip DEGs in psychiatric disorder GWAS.

Genetic network of the 13 human orthologs of the 7 significant DEGs in the FlyChip and UKBiobank Sleep related GWAS results. The analysis and visualisation were created by Gene Mania (accessible from: genemania.org). Additional genes are in full black while inputted genes are in stripes. The colours of lines represent different links: Beige = Shared protein domains, Purple = co-localisation, Red = Physical interactions, Blue = Pathways, Green = Shared protein domains, Orange = Predicted interactions. The weight of the lines is in proportion to biological process involvement with the 3 inputted genes.

8.5   *Discussion*

Comparison of the significant FlyChip orthologs to GWAS data proved fruitful in identifying genes which should be prioritized in future work. There are several genes which exist in both the UKBiobank sleep parameters and the MDD and SZ studies, likely due to the overlap of sleep and psychiatric disorders causation or symptom. Of the significant DEGs isolated by comparing the top FlyChip results and the data from Hill *et al.*, both *CG6083* and *hsc70-2* (or the human orthologs) were found to be in the top sleep related GWAS gene list. Furthermore, *hsc70-2* and it's orthologs were found in both the sleep and the psychiatric studies GWAS results. Gene Mania produced a connection between *AKR1A1* and *KCNAB1*, which is a member of the potassium voltage gated ion channels linked with psychiatric disorders. Alongside the published data implicating *CG6083* with mushroom body development and the function of *HSPA2* to prevent oxidative stress and subsequently psychiatric disorders, these genes are clearly putative causes of the phenotype displayed in Chapter 5.

Many of the genes found to correlate between the FlyChip and the GWAS data appear to exist in quite separate clusters on the network maps created by Gene Mania and did not produce significant enrichment in DAVID. This reinforces the notion that there is not a single network affected by the loss of genomic *miR-137*, but instead the miRNA is involved in regulating multiple pathways. This also provides some explanation as to the lack of results comparing the two wet lab studies, as the complexity of such networks must become quite confounding and ultimately difficult to determine.

None of the 5 genes identified as part of the psychiatric disorders UKBiobank GWAS were found in the GWAS from the SZ or MDD studies, adding weight to the argument of the

UKBiobank data having a population bias. However, with such a small subset of the results being compared there is not enough evidence to say this with certainty.

8.6   *Conclusion*

Overall, the use of bioinformatical analysis has provided a new aspect to his project which has demonstrated the difficulties of cross-species data comparisons, the overlap of sleep and psychiatric disorder significant GWAS results and provided a set of candidate genes with plausible biological importance in either psychiatric disorders or specifically the phenotypes exhibited in the *miR-137^{KO}* flies. The lack of significant enrichment of the Flychip DEGs and any of the GWAS gene sets ultimately means that the null hypotheses 1, 2, and 3 from objective F must be accepted (Chapter 1.8).

Through the repeated identification of *D. melanogaster* genes *hsc70-2* and *CG6083* in several of the GWAS results as well as significant DEGs in both studies, this project selects these as genes of high importance to assess as part of a follow-on project.

9 The link between miR-137, psychiatric disorders, and sleep

The focus of this project was to investigate the link between *miR-137* and psychiatric disorders, specifically SZ which is already associated with the miRNA through GWAS studies (Ripke *et al.* 2011; Duan *et al.* 2014). There are already several experimental studies connecting the miRNA to psychiatrically relevant phenotypes like brain morphology, cognitive function, and synaptic plasticity (Cummings *et al.* 2013; Green *et al.* 2013; Kuswanto *et al*. 2015; Cosgrove *et al.* 2017). *miR-137* is also highly localised within the brain, further reinforcing the contribution of the miRNA to the aetiology of psychiatric disorders. There is a substantial overlap between the symptoms of psychiatric disorders, such as SZ, and those of neurological diseases like Parkinson's disease. This overlap could also lead to linking *miR-137* to a wider range of human illnesses.

The model organism, *D. melanogaster*, was chosen as a suitable research subject due to the comparable genetics and already published phenotypes synonymous with psychiatric and neurological symptoms. Firstly, the purpose of *miR-137* had to be established, which was achieved by investigating the effect of total removal of genomic expression from the flies. Further research then implicated specific genes and mechanisms through which these phenotypes may be caused.

This project investigated *miR-137* with a holistic organism-level aspect, providing a wider picture of the functional relevance of *miR-137* in relation to overall health and behaviour.

9.1 The effect of miR-137 knockout

The simplest starting point was to associate the loss of genomic *miR-137* with any behavioural phenotypes to evaluate if the miRNA is significant enough to produce any observable behaviour changes. The fly strain obtained from the BDSC had the *miR-137* locus substituted for an insertion which presented a red-eye phenotype. This information was successfully verified through presence/absence PCR genotyping, DNA sequencing, and local alignment against the expected sequences (Chapter 3). QPCR attempts also found zero *miR-137* expression within the *miR-137^{KO}* strain (Chapter 6), further proving complete *miR-137* gene LOF.

While the *miR-137^{KO}* stock lacked *miR-137* expression, it was not guaranteed that this was the cause of the behavioural phenotypes. Ablation of a miRNA molecule was unlikely to not have had any consequences and, with the constant interbreeding of the strain at the BDSC, background modifiers could have accumulated to help cope with the effects. The modifiers would be stock specific, and through a set of crosses the total sleep phenotype was attributed to *miR-137* LOF (chapter 6.1). Furthermore, the heterozygous loss of *miR-137* produced a phenotype of approximately 50% severity to the full knockout.

In this research, the complete removal of the *miR-137* gene presented significant phenotypes across several aspects of general fly health and behaviour. These phenotypes include a significant developmental delay at the pupal stage and severe overall hypoactivity.

9.1.1 Developmental phenotype

There was a developmental delay identified in the pupal stage of the fly life cycle (Chapter 4.2), where $miR\text{-}137^{KO}$ flies reached adulthood slower than controls after spending an average of 0.8 days longer in chrysalis. Furthermore, there was also evidence that this delay was accompanied by an increase in mortality. During the pupal stage, the larva undergo metamorphosis that transforms the entire anatomy into that of an adult fly. It is known that the CNS is dramatically remodelled at this time, converting and creating neurons to function within the adult nervous system (Robertson 1936; Technau and Heisenberg 1982; Truman 1990; Tully *et al.* 1994; Hartenstein *et al.* 2008; Veverytsa and Allen 2013). These neurons require a high degree of synaptic plasticity to enable complex processes like axon and dendritic remodelling or integration into adult neuron circuits (Hartenstein *et al.* 2008; Veverytsa and Allen 2013). The apparent importance of *miR-137* at this stage of the fly reaffirms the role of *miR-137* and its targets in neuroanatomical development and formation as discussed in Chapter 1.

Two of the DEGs identified by transcriptome analysis of the $miR\text{-}137^{KO}$ flies already have developmental phenotypes associated with them. *Turtle* LOF in flies causes partial lethality during pupation (Bodily et al. 2001), and partial deletion of *rnh1* causes a delay in developmental time and pupal stage lethality (Filippov et al. 2001). Of these, *rnh1* (ribonuclease inhibitor 1) was significantly downregulated with a log2 factor of approximately -2.6, which correlates with a deletion causing LOF. While exact method of *rnh1* effecting development is not clear, this provides a logical hypothesis for the mechanism of *miR-137* on pupal-stage lethality.

9.1.2   *Behavioural phenotypes*

Behaviourally, the *miR-137^KO* flies had a severely hypoactive phenotype when compared to *w^1118* controls. The actogram produced from DAMS analysis shows that throughout the knockout flies have a significantly lower activity across the entire 24-hour period, most exaggerated at the morning peak, compared to controls (chapter 5). Separation of sleep and locomotor phenotypes was largely achieved by directly assessing movement through video recording and tracking individual flies. This identified periods of quiescence in the *miR-137^KO* flies alongside locomotor defects (chapter 5), indicating that the actogram result was a combined effect of both locomotor and sleep defects.

9.1.2.1   *Locomotor ability*

Overall, the *miR-137^KO* flies had a slower velocity at peak activity times than the *w^1118* controls. The knockout strain also performed worse in a negative geotaxis assay, another simple assay to evaluate movement ability. At this point it is not clear as to how *miR-137* LOF causes this phenotype as this project did not investigate further. However, the localisation of *miR-137* expression at the synapses and importance for synaptic plasticity would infer that the phenotype is related to neuron signalling rather than muscular problems. However, several of the DEGs in the *miR-137^KO* stock were functionally relevant to metabolism and mitochondria function. Of these genes, four were found to correlate with human orthologs identified as significant within SZ and MDD GWAS (*ND-15, Hsc70-2, CG6083*, and *Ect3)*. As some psychiatric disorders also have aspects of locomotor defects, these could potentially be genes affected by *miR-137* and subsequently causing the phenotype.

9.1.2.2 *Sleep-like behaviour*

The change in sleep-like behaviour appeared to be the most severe of the phenotypes found in the *miR-137^{KO}* strain and became the focus of this project. The DAMS identified total sleep amount in the *miR-137^{KO}* flies to be approximately 86% of the 24-hour period, which is a significant increase when compared to *w^{1118}* which sleep circa 57%. Sleep governed by genetics and is regulated by two largely independent mechanisms; the circadian and the homeostat pathways (Cirelli 2009; Crocker and Sehgal 2010; Sehgal and Mignot 2011; Donlea *et al.* 2013). Initial investigation was to establish which mechanism was affected by *miR-137* LOF.

The actogram (chapter 5.1) and sleep graph (chapter 5.4) demonstrate that despite the obvious different in overall sleep/activity, both the strains exhibit a similar behavioural pattern and therefore do not have a circadian defect. The flies were also put into constant darkness to investigate if the circadian clock could keep the entrained 12-hour rhythm, though with the overall activity count in the *miR-137^{KO}* this proved difficult to assess.

Lack of a circadian phenotype indicates that the role of *miR-137* is within the homeostatic regulation of sleep. This is responsible for the need to sleep following periods of wakefulness and the sleep rebound effect after sleep-deprivation (Donlea *et al.* 2014). In *D. melanogaster,* the fan-shaped body is responsible for homeostatic sleep regulation. This region is functionally comparable to the mammalian hypothalamus and hypothalamic ventrolateral preoptic nuclei (Donlea *et al.* 2014). Artificial stimulation of the fan-shaped body induces sleep on command and the region is negatively modulated by dopamine signalling (Lui *et al.*2012).

The DEGs in the FlyChip included a gene, *slim*, which showed a sleep defect in flies with RNAi mediated downregulation of the gene in the CNS (Lobell *et al.* 2017). This

matches the *miR-137^{KO}* data which had a Log2 factor of -2.357 (3dp). The gene functions to positively regulate the stress-activated MAPK cascade, an evolutionarily conserved pathway vital for transduction pathways responsible for cell growth and proliferation (Morrison 2012; Sekine *et al.* 2012).

9.1.3   Genetic pathways

The high number of predicted and proven targets of direct *miR-137* regulation provides many logical options for the genes and pathways that could result in the phenotypes identified in Chapters 4 and 5. It is obvious that *miR-137* is part of a complex genetic network and therefore it is unreasonable to assume there is a singular pathway responsible for the behaviour changes in *miR-137^{KO}* flies. Having multiple phenotypes also reinforces the likelihood of several factors contributing to the overall effect.

Transcriptome analysis of *miR-137^{KO}* flies identified several genes which already have phenotype associations relatable to locomotor, neuroanatomy, and sleep (chapter 8.2). Adding these genes to DAVID for ontology did not produce any significantly enriched pathways, demonstrating that the identified genes are not related in terms of function.

The sleep phenotype is associated with the homeostatic control pathway (chapter 5.5 and 9.1.2), which is regulated by the neurotransmitter dopamine. This is a sleep regulating molecule that is known to promote wakefulness. The hypoactive phenotype specifically in the morning peak in the actogram (chapter 5.1) suggests a dopamine pathway problem, and this is known to be regulated by dopamine (Lui *et al.* 2012). Dopamine regulates the activation of the fan-shaped body and the mushroom body through upward or downward dopamine neurons (Sitaraman *et al.* 2015). The FB promotes sleep which upon polarization sends

signals to sleep (Donlea *et al.* 2011; Lui *et al.* 2012). The MB exists in multiple subregions, some of which promote sleep, and some promote wakefulness (Sitaraman *et al.* 2015).

Pharmaceutical treatments were used to try recover the sleep phenotype, as the effect of the drug can be used as test for normal action of the related pathway. Caffeine was used as a stimulant to reduce sleep but produced an unexpected increase in total sleep amount across the 24-hour period (chapter 7.2.1). Dopamine is mostly known for its role in promoting wakefulness, and *miR-137* is predicted to downregulate genes in the dopamine pathway such as the *pale* gene which is the rate-limiting enzyme producing dopamine. It was hypothesized that without correct *miR-137* expression, the flies would be exposed to increased dopamine signalling throughout development leading to dopamine desensitisation. A *pale* inhibitor, 3IY, was administered to w^{1118} and *miR-137^{KO}* flies to test this hypothesis. 3IY is known to promote sleep (Catterson *et al.* 2010), an effect that was observed in control flies when treated acutely (chapter 7.2.2). Chronic treatment of 3IY presented no significant change in the total sleep of controls, but it significantly decreased the sleep in *miR-137^{KO}* flies during the day and total 24-hour period. This counter-intuitive change reinforces the hypothesis of dopamine tolerance due to *miR-137* LOF.

9.1.4   *Specific brain regions*

miR-137 is known to be widely expressed within the brain, making this the likely location for the miRNA to function (Willemsen et al. 2011; Yin et al. 2014). The *miR-137^{KO}* stock used has a genomic substitution resulting in complete LOF across the entire organism, which does not associate the sleep phenotype to just the brain. Specific downregulation of *miR-137* expression within the CNS caused a significant increase in sleep to circa 80% of the 24-hour period (chapter 6.2). A decrease of *miR-137* expression in the brain of these flies was

also verified through qPCR, confirming that *miR-137* function is associated with the CNS. Although the phenotype was somewhat less than that of the *miR-137^{KO}*, this can likely be attributed to the difference between full knockout and downregulation in this region.

Investigation into the sleep behaviour and the genetic pathways of the phenotype implicated *miR-137* function with the fan-shaped and mushroom shaped bodies (chapter 9.1.3). Selective downregulation of *miR-137* in the fan-shaped body and the mushroom body both caused a significant increase in total sleep amount (~73% of the 24-hour period). While not as severe as the knockout phenotype, this increase was caused by downregulation in a very specific area, indicating that the FB and MB are regions which are relevant to sleep in the *miR-137^{KO}* flies. Unfortunately, the level of downregulation was not possible to measure due to the lack of ability and equipment to perform such dissections and RNA expression analysis.

Reintroduction of *miR-137* to the FB only within the *miR-137^{KO}* background further reinforced the importance of the FB region in the regulation of sleep by *miR-137* (Chapter 7.1). The excessive sleep phenotype was significantly reduced by ~12% (p-value < 0.05), while reintroduction of *miR-137* in the whole CNS did not produce a significant result. Similarly, to the targeted downregulation, it is not possible to assess the level of gene reintroduction in the FB region alone. However, to produce a significant reduction in sleep within the knockout background further validates the regulation of sleep by *miR-137* through the fan-shaped body.

9.2   Human relevance

The purpose of a model organism in genetic research is to be able to relate the results back to humans, most often for the purpose of investigating the mechanisms behind diseases. The fruit fly is a proven model for research into the genetics behind many diseases ranging from cardiovascular disease to neurological symptoms. The aim of this project was originally built upon GWAS results statistically linking *miR-137* to SZ in humans (Ripke *et al.* 2011; Duan *et al.* 2014), and the final area of research cross referenced the differentially expressed genes in the *miR-137^{KO}* flies to other GWAS (Chapter 8). The phenotypes observed in the *miR-137^{KO}* strain are all comparable to human symptoms within psychiatric disorders, potentially implying these are associated with *miR-137* in humans.

Due to the nature of GWAS based studies, there is a rapidly increasing volume of data quickly becoming available to researchers. At the time of this research, data from the GWAS studies regarding MDD and SZ, by Pardinas *et al.* (2016) and Ripke *et al.* (2018), and the UKBiobank was used because of availability. There are many more studies regarding neurological diseases, psychiatric disorders, and sleep parameters which could provide alternate data sets worth cross-comparing to the FlyChip output. Some examples of such include: Jones *et al.* (2019) study into circadian rhythm using the UKBiobank plus 23andMe datasets, Howard *et al.* (2019) study into depression using an accumulation of 3 major depression studies, and studies by Doherty *et al.* (2018) and Dashti *et al.* (2019) study physical activity and sleep duration using fitness tracker data and the UKBiobank.

9.2.1 Psychiatric disorders

It is known that the aetiology of psychiatric disorders has a genetic basis and are often a result of a complex network of many risk factors cumulatively creating the symptoms. The results from the bioinformatics (chapter 8) identified some key genes which have a plausible link between *miR-137* expression knockdown/knockout and human disease statistics. The *miR-137* gene is highly conserved across many species, with the mature miRNA construct being just one nucleotide different between *D. melanogaster* and humans. The high sequence homology implies that the function of *miR-137* is comparable across species.

Comparison of DEGs in the *miR-137^{KO}* flies to human SZ and MDD GWAS produced a list of 7 genes: *ect3, hsc70-2, tutl, ND-15, CG4757, CG30456,* and *chp*. *Turtle* (*tutl*) functions roles include axonal tiling, dendrite self-avoidance, axonal pathfinding and coordinated motor control. Loss of function in flies causes locomotor and behavioural defects, plus partial lethality during pupation (Bodily et al. 2001). Issues with movement and fatigue are symptoms of psychiatric disorders such as SZ patients demonstrating slow movement (Jakubowski et al. 2012; Lee et al. 2013).

An alternative method of researching the genetics behind psychiatric disorders is the sequencing of genomes for the identification of rare risk variants. GWAS identify common risk alleles from large groups, however the aetiology of psychiatric disorders is known to exist as an accumulation of genetic variations increasing susceptibility. The combined polygenic effect of the common risk alleles identified in GWAS studies accounted for a small portion of disease liability and therefore this network of risk alleles must also include rare variants (Loh *et al.* 2015; Singh *et al.* 2017). Studies on rare variants have reported some have a far larger effect on individual risk than common variants and both rare and common variants can occur in the same genes or pathways (Zhu *et al.* 2014; Pocklington *et al.* 2014; Singh *et al.* 2016; Singh *et al.* 2017). These studies sequence the exome or genome of cases

and controls to identify low-frequency and rare variants with a minor allele frequency of less than 0.5% (Panoutsopoulou *et al.* 2013; Duan *et al.* 2014; Pocklington *et al.* 2014; Rees *et al.* 2014; Sidore *et al.* 2015; Singh *et al.* 2017). There are studies identifying rare *miR-137* risk variants for SZ. Duan *et al.* (2014) sequenced the miR-137 locus of ~2600 SZ patients and control cases and identified 133 variants with minor allele frequency of <0.5%. One of these, 1: g.98515539 A>T, was found in just 11 of the SZ cases and influenced the enhancer region to ultimately decrease *miR-137* expression. This differing approach could likely provide a more directly relevant dataset to compare against. For example, the flies in this project had genomic ablation of the miR-137 gene and this was used to evaluate the genes overall importance and function, and then any transcriptome changes were compared to other genes in the GWAS data. However, a comparison of the fly results to the transcriptome of a human dataset with any rare LOF variant in the *miR-137* locus could provide a stronger link between fly and human. Furthermore, rare variant functional study data could be cross-referenced to the FlyChip results in this study and identify new genes of interest in addition to what has already been identified in this project.

9.2.2 Sleep

Sleep is a vital process required for life. In terms of fundamental survival, the action of sleep contradicts common sense by voluntarily becoming less environmentally aware and generally vulnerable for an extended period. A sleep-like state is found across many species, even insects like *D. melanogaster* further emphasising its importance. The fruit fly has a documented sleep state that is comparable to human sleep in many aspects. During sleep, both fruit flies and humans have increased arousal thresholds and decreased environmental awareness. Furthermore, both species have homeostatic and circadian regulation mechanisms

governing sleep, which implement similar molecular pathways like dopamine signalling (Shaw et al 2000; Bushey et al 2010; Tomita et al. 2015).

Taking the DEGs from the FlyChip results and comparing them to the top genes from sleep-phenotype parameters of the UKBiobank resulted in 6 genes: *w, ect3, hsc70-2, chp, CG30456,* and *CG4757.* None of these genes currently have experimental links to a sleep-related phenotype reported on Flybase but provide direction for future work.

Sleep is a symptom of psychiatric disorders such as SZ and BD. These exhibit differences in sleep behaviour than would be expected of "normal" sleep (Benca *et al.* 1992; Monti and Monti 2004; Lee and Douglass 2010; Krystal 2012). Alongside differences in sleep/wake cycle, symptoms such as catatonia are comparable to being sleepy and fatigued.

The implication of *miR-137* regulating sleep also relates to not just psychiatric disorders, but also many other diseases. Neurological diseases like Parkinson's disease, Dementia, and Huntington's disease also include changes in sleep pattern such as excessive daytime sleepiness or insomnia (Happe *et al.* 2002; Garcia-Borreguero *et al.* 2003; Rose *et al.* 2010; Herzog-Krzywoszanska and Krzywoszansk 2019). While sleep problems can be symptoms, sleep dysregulation is also associated with increased risk of obesity, diabetes, heart disease, and strokes (Spiegel *et al.* 2005; Knutson and Cauter 2008; Khandelwal *et al.* 2017; Young *et al.* 2019). Chronic sleep deprivation has even been statistically correlated with increased suicide ideation in teenagers (Whitmore and Smith 2018).

9.3   Conclusion

The complex nature of miRNA regulation dictates that biological link between *miR-137* and psychiatric disorders will include many genetic pathways. This complicates the research into the miRNA function, particularly in genetic recovery experiments, as the overall phenotype defect will not be contributable to any single gene. This project identified behavioural and developmental defects in *miR-137^{KO}* flies and found that the activity phenotype was a compound effect of locomotor and sleep-like behaviour abnormalities.

Results of investigating the development in *miR-137^{KO}* flies infer importance of *miR-137* in the pupal stage, which is when the larva undergo metamorphosis into an adult fly. This period is associated with dramatic remodelling of the CNS, where a high degree of synaptic plasticity is required for conversion of existing neurons to function in adult fly CNS. Identifying the importance of *miR-137* in this process reaffirms the miRNAs role synaptic plasticity regulation and neurogenesis.

The data collected in this project suggests that *miR-137* is required for proper function of the sleep regulation by the *Drosophila melanogaster* fan-shaped body and mushroom body, brain regions functionally comparable to the human cerebral cortex, hypothalamic ventrolateral preoptic nuclei, and hypothalamus. Furthermore, in the absence of the miRNA, this defect is due to dopamine desensitisation and an increase in mental pressure by the homeostatic control to sleep. Chronic treatment with a dopamine inhibitor decreased the sleep amount in the *miR-137^{KO}* flies, while dopamine inhibition promotes sleepiness in the controls. In normal conditions, voltage-gated potassium ion channels are regulated by dopamine when controlling the activation of the *D. melanogaster* fan-shaped body, and subsequently the sleep signals sent from this brain region. These potassium channels and

dopamine receptors and transporters are also putative targets of *miR-137* and linked via human AKR1 proteins to sleep-phenotype UKBiobank GWAS data in this work.

Comparing the phenotype of *miR-137* LOF in *D. melanogaster* to the symptoms of psychiatric disorders and neurological diseases in humans results in some obvious overlap, as many diseases are associated with sleep or locomotor defects. In terms of human genetics, the statistical evidence of the GWAS implicated a select group of genes vital for general life such as heat shock proteins and potassium channels. Furthermore, several DEGs in the *miR-137^{KO}* strain were also identified as statistically significant in sleep phenotype and psychiatric disorder GWAS.

In conclusion, identifying the link between *miR-137* and psychiatric disorders is a difficult task due to the complex nature of miRNA regulation, the lack of clear psychiatric diagnostic criteria, and the overall intricacy of the brain. Nonetheless, this project identified several genes and mechanisms behind phenotypes in the fly synonymous to those of psychiatric disorders. Furthermore, the project then focused on investigating the excessive sleep amount in *miR-137^{KO}* flies and discovered that *miR-137* regulates proper sleep through maintaining the balance of dopamine within the homeostatic control pathways, a completely novel finding. Additionally, this regulation is likely through interacting with voltage-gated potassium ion channels and dopamine transporters in two *D. melanogaster* brain regions; the fan-shaped body and the mushroom body. These regions are functionally comparable to human brain regions associated with the development of neurological symptoms, reinforcing the relevance of *miR-137* with sleep and psychiatric disorders.

10 Future work

This project identified a novel role for *miR-137* in regulating sleep. However, there is a lot more research that can be done to further refine the mechanisms involved. While this work focused on association between sleep and psychiatric disorders, sleep dysfunction is also linked with increased risk of obesity, diabetes, heart disease, and strokes. There is potential for *miR-137* to have implications with a much wider range of human ailments if these links can be further studied.

Unsuccessful gene expression experiments for the sponge-mediated *miR-137* knockdown and *miR-137* overexpression is an unfortunate outcome of this thesis (chapter 6.2). Future work would include repeats of these experiments to evaluate the level of expression change caused by the sponge construct and correlate the results with the behavioural data for a better understanding of enough *miR-137* for "normal" sleep. It is obvious that the method used was flawed, and future work should consider improvements mentioned in Chapter 6.2.

The hypothesis of the dopamine desensitisation in the flies through increased activation of the fan-shaped body requires more work to fully establish. The measurement of sleep in this study was primarily through associating 5-minutes of no activity to 1 "count of sleep". Better methods of measuring sleep could help elucidate this phenotype. The dopamine component could be strengthened through quantification of dopamine levels, or genetic dopamine pathway inhibition in the KO background to rescue the phenotype similarly to 3IY treatment in chapter 7.2.

The bioinformatics section intended to help bridge the gap between *D. melanogaster* genetics and human diseases, identifying several genes of interest as a result of novel transcriptome data and comparison to GWAS results. More investigation is required into the

DEGs found, especially CG6083 and hsc70-2 which were identified in multiple experimental data sets and psychiatric GWAS results. Additionally, as mentioned in Ch9.2, there are many other GWAS study datasets that could be retrieved to cross compare with the FlyChip results, potentially indicating additional genes of interest.

Ultimately, I hope this work will prove useful for the continued efforts into deciphering the complex genetic aetiology behind psychiatric disorders. There is novel data produced within this project, particularly the sleep phenotype exhibited with complete *miR-137* ablation, which can be further developed and used as part of the wider picture in neurological research.

11 References

Adams, M. D. and Sekelsky, J. J., 2002. From Sequence To Phenotype: Reverse Genetics in Drosophila Melanogaster. *Nature Reviews Genetics*, 3 (3), 189–198.

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P., 2015. Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4, 1–38.

Agarwal, V., Subtelny, A. O., Thiru, P., Ulitsky, I., and Bartel, D. P., 2018. Predicting microRNA targeting efficacy in Drosophila. *Genome Biology*, 19 (1), 1–23.

Aksoy, A., 2019. Integrated model for renewable energy planning in Turkey. *International Journal of Green Energy*, 16 (1), 34–48.

Allebrandt, K. V., Teder-Laving, M., Cusumano, P., Frishman, G., Levandovski, R., Ruepp, A., Hidalgo, M. P. L., Costa, R., Metspalu, A., Roenneberg, T., and De Pittà, C., 2017. Identifying pathways modulating sleep duration: From genomics to transcriptomics. *Scientific Reports*, 7 (1), 1–11.

Alles, J., Fehlmann, T., Fischer, U., Backes, C., Galata, V., Minet, M., Hart, M., Abu-Halima, M., Grässer, F. A., Lenhof, H. P., Keller, A., and Meese, E., 2019. An estimate of the total number of true human miRNAs. *Nucleic Acids Research*, 47 (7), 3353–3364.

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., Dreyfuss, G., Eddy, S. R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., and Tuschl, T., 2003. A uniform system for microRNA annotation. *RNA (New York, N.Y.)*, 9 (3), 277–9.

Andretic, R., Kim, Y.C., Jones, F.S., Han, K.A., and Greenspan, R.J., 2008. Drosophila D1 dopamine receptor mediates caffeine-induced arousal. *PNAS*. 105 (51), 20392 – 20397.

Aradska, J., Bulat, T., Sialana, F. J., Birner-Gruenberger, R., Erich, B., and Lubec, G., 2015. Gel-free mass spectrometry analysis of *Drosophila melanogaster* heads. *Proteomics*, 15 (19), 3356–3360.

Arakawa, Y., Yokoyama, K., Tasaki, S., Kato, J., Nakashima, K., Takeyama, M., Nakatani, A., and Suzuki, M., 2019. Transgenic mice overexpressing miR-137 in the brain show SZ-associated behavioral deficits and transcriptome profiles. *PLoS ONE*, 14 (7), 1–15.

Argyraki, A., Markvart, M., Stavnsbjerg, C., Kragh, K. N., Ou, Y., Bjørndal, L., Bjarnsholt, T., Petersen, P. M., Muramoto, Y., Kimura, M., Nouda, S., Solution, I. N., Takayoshi, T., Takuya, M., Jun, S., Norimichi, N., Kenji, T., Hideki, H., Germicidal, U., Handbook, I., Science, F., Science, F., Science, N., Scholarship, T., Note, A., Note, A., Welch, D., Buonanno, M., Shuryak, I., Randers-Pehrson, G., Spotnitz, H. M., Brenner, D. J., Huo, Y., Aboud, K., Kang, H., Cutting, L. E., Bennett, A., De Melo, W. C. M. A., Avci, P., De Oliveira, M. N., Gupta, A., Vecchio, D., Sadasivam, M., Chandran, R., Huang, Y. Y., Yin, R., Perussi, L. R., Tegos, G. P., Perussi, J. R., Dai, T., Hamblin, M. R., Connolly, E., Lee, M. J., Kwon, J. S., Jiang, H. B., Choi, E. H., Park, G., Kim, K. M.,

Shin, J., Kim, S., Kim, D., and Kang, D., 2018. （生物用）HHS Public Access. *2015 IEEE Summer Topicals Meeting Series, SUM 2015*, 10 (1), 1–13.

Armstrong, J. D., Texada, M. J., Munjaal, R., Baker, D. A., and Beckingham, K. M., 2006. Gravitaxis in Drosophila melanogaster: A forward genetic screen. *Genes, Brain and Behavior*, 5 (3), 222–239.

Assimakopoulos, K., Karaivazoglou, K., Skokou, M., Kalogeropoulou, M., Kolios, P., Gourzis, P., Patrinos, G., and Tsermpini, E., 2018. Genetic Variations Associated with Sleep Disorders in Patients with SZ: A Systematic Review. *Medicines*, 5 (2), 27.

Atilano, M.L., Gittenberg, M., Monteiro, A., Copely, R.R., and Ligoxygakis, P., 2017. MicroRNAs That Contribute to Coordinating the Immune Response in Drosophila melanogaster. *Genetics*. 207, 163 – 178.

Aydin O(1), Unal Aydin P(1), Arslan A(2). Development of Neuroimaging-Based Biomarkers in Psychiatry. *Adv Exp Med Biol*. 2019;1192:159-195. doi: 10.1007/978-981-32-9721-0_9.

Azizan, E. A., Poulsen, H., Tuluc, P., Zhou, J., Clausen, M. V, Lieb, A., Maniero, C., Garg, S., Bochukova, E. G., Zhao, W., Shaikh, L. H., Brighton, C. A., Teo, A. E., Davenport, A. P., Dekkers, T., Tops, B., Küsters, B., Ceral, J., Yeo, G. S., Neogi, S. G., McFarlane, I., Rosenfeld, N., Marass, F., Hadfield, J., Margas, W., Chaggar, K., Solar, M., Deinum, J., Dolphin, A. C., Farooqi, I. S., Striessnig, J., Nissen, P., and Brown, M. J., 2013. Somatic mutations in ATP1A1 and CACNA1D underlie a common subtype of adrenal hypertension. *Nat Genet*, 45 (9), 1055–1060.

Badalà, F., Nouri-mahdavi, K., and Raoof, D. A., 2008. NIH Public Access. *Computer*, 144 (5), 724–732.

Bainton, R.J., Tsai, L.T., Singh, C., Moore, M.S., Neckameyer, W.S., and Heberlein, U., 2000. Dopamine modulates acute respponses to cocaine, nicotine and ethanol in Drosophila. *Current Biology*. 10 (4), 187 – 194.

Baggio, F., Bozzato, A., Benna, C., Leonardi, E., Romoli, O., Cognolato, M., Tosatto, S. C. E., Costa, R., and Sandrelli, F., 2013. 2mit, an Intronic Gene of Drosophila melanogaster timeless2, Is Involved in Behavioral Plasticity. *PLoS ONE*, 8 (9), 1–19.

Barbeira, A. N., Pividori, M. D., Zheng, J., Wheeler, H. E., Nicolae, D. L., and Im, H. K., 2019. Integrating predicted transcriptome from multiple tissues improves association detection. *PLoS Genetics*, 15 (1), 1–20.

Barešić, A., Nash, A. J., Dahoun, T., Howes, O., and Lenhard, B., 2019. Understanding the genetics of neuropsychiatric disorders: the potential role of genomic regulatory blocks. *Molecular Psychiatry*, 6–18.

Barsotti, C. and Ipata, P. L., 2004. Metabolic regulation of ATP breakdown and of adenosine production in rat brain extracts. *International Journal of Biochemistry and Cell Biology*, 36 (11), 2214–2225.

Bartel, D. P., 2004. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell*, 116 (2), 281–297.

Bass, T. M., Weinkove, D., Houthoofd, K., Gems, D., and Partridge, L., 2007. Effects of resveratrol on lifespan in Drosophila melanogaster and Caenorhabditis elegans. *Mechanisms of Ageing and Development*, 128 (10), 546–552.

Battle, D. J., Kasim, M., Yong, J., Lotti, F., Lau, C. K., Mouaikel, J., Zhang, Z., Han, K., Wan, L., and Dreyfuss, G., 2006. The SMN complex: An assembly machine for RNPs. *Cold Spring Harbor Symposia on Quantitative Biology*, 71, 313–320.

Behrendt, R. and Young, C., 2004. Behrendt2004Hallucinations in SZ, 771–830.

Benca, R. M., Obermeyer, W. H., Thisted, R. A., Gillin, ; J Christian, and Obermeyer, M., 2015. Sleep and Psychiatric Disorders A Meta-analysis From the Departments of Psychiatry (Dr Benca).

Benjamini, Y., and Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology,* 57(1), 289–300.

Berger, S. M. and Bartsch, D., 2014. The role of L-type voltage-gated calcium channels Cav1.2 and Cav1.3 in normal and pathological brain function. *Cell and Tissue Research*, 357 (2), 463–476.

Berger, Z., Ttofi, E. K., Michel, C. H., Pasco, M. Y., Tenant, S., Rubinsztein, D. C., and O'Kane, C. J., 2005. Lithium rescues toxicity of aggregate-prone proteins in Drosophila by perturbing Wnt pathway. *Human Molecular Genetics*, 14 (20), 3003–3011.

Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J., 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409 (6818), 363–366.

Betel, D., Koppal, A., Agius, P., Sander, C., and Leslie, C., 2010. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biology*, 11 (8), R90.

Bidaki, R., Zarei, M., Toosi, A. K., and Shooshtari, M. H., 2012. A review on genetics of sleep disorders. *Iranian Journal of Psychiatry and Behavioral Sciences*, 6 (1), 12–19.

Bodily, K. D., Morrison, C. M., Renden, R. B., and Broadie, K., 2001. A novel member of the Ig superfamily, turtle, is a CNS-specific protein required for coordinated motor control. *Journal of Neuroscience*, 21 (9), 3113–3125.

Bohnsack, M. T., Czaplinski, K., and Gorlich, D., 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA (New York, N.Y.)*, 10 (2), 185–91.

Bonferroni, C., 1936. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni Del R Istituto Superiore Di Scienze Economiche E Commericiali Di Firenze*, 8, 3–62.

Bosche, Bert; Macdonald, R. L., 2015. Letter To The Editor: "Letter to the Editor". *Stroke*, 46 (2), e126–e127.

Boudreau, R.L., Jiang, P., Gilmore, B.L., Spengler, R.M., Tirabassi, R., Nelons, J.A., Ross, C.A., Xing, Y., and Davidson, B.L., 2015. Transcriptome-wide Discovery of microRNA Binding Sites in Human Brain, *Neuron.* 81, 294–305.

Brand, A. H. and Perrimon, N., 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development (Cambridge, England)*, 118 (2), 401–15.

Brenman-Suttner, D. B., Yost, R. T., Frame, A. K., Robinson, J. W., Moehring, A. J., and Simon, A. F., 2019. Social behavior and aging: A fly model. *Genes, Brain and Behavior*, (October 2018), 1–18.

Brent, J., Werner, K., and McCabe, B. D., 2009. Drosophila larval NMJ immunohistochemistry. *Journal of visualized experiments: JoVE*, 4–5.

Brennecke, J., Hipfner, D. R., Stark, A., Russell, R. B. and Cohen, S. M., 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* 113, 25–36.

Brunner, E., Ahrens, C. H., Mohanty, S., Baetschmann, H., Loevenich, S., Potthast, F., Deutsch, E. W., Panse, C., de Lichtenberg, U., Rinner, O., Lee, H., Pedrioli, P. G. a, Malmstrom, J., Koehler, K., Schrimpf, S., Krijgsveld, J., Kregenow, F., Heck, A. J. R., Hafen, E., Schlapbach, R., and Aebersold, R., 2007. A high-quality catalog of the Drosophila melanogaster proteome. *Nature biotechnology*, 25 (5), 576–583.

Brunner, J. I., Gotter, A. L., Millstein, J., Garson, S., Fox, S. V, Savitz, A. T., Yang, H. S., Fitzpatrick, K., Zhou, L., Owens, R., Webber, A. L., Vitaterna, M. H., Kasarskis, A., Victor, N., Turek, F., Renger, J. J., and Winrow, C. J., 2013. NIH Public Access, 25 (4), 167–181.

Bullock TH. Identifiable and addressed neurons in the invertebrates. Neurobiology of the Mauthner Cell. 1978:1–12.

Bushey, D., Hughes, K. a, Tononi, G., and Cirelli, C., 2010. Sleep, aging, and lifespan in Drosophila. *BMC neuroscience*, 11, 56.

Carbone, M. A., Yamamoto, A., Huang, W., Lyman, R. A., Meadors, T. B., Yamamoto, R., Anholt, R. R. H., and Mackay, T. F. C., 2016. Genetic architecture of natural variation in visual senescence in *Drosophila. Proceedings of the National Academy of Sciences*, 113 (43), E6620–E6629.

Carthew, R.W., Agbu, P., and Giri, R., 2017. MicroRNA function in Drosophila melanogaster. *Cell Developmental Biology.* 65, 29 – 37.

Carvalho, G. B., Ja, W. W., and Benzer, S., n.d. DNA isolation and amplification from Drosophila wings, 4–5.

Catterson, J. H., Knowles-Barley, S., James, K., Heck, M. M. S., Harmar, A. J., and Hartley, P. S., 2010. Dietary modulation of Drosophila sleep-wake behaviour. *PLoS ONE*, 5 (8).

Cavanaugh, D. J., Geratowski, J. D., Wooltorton, J. R. A., Jennifer, M., Hector, C. E., Zheng, X., Johnson, E. C., and James, H., 2015. rhythms in Drosophila, 157 (3), 689–701.

Cerritelli, S. M., Frolova, E. G., Feng, C., Grinberg, A., Love, P. E., and Crouch, R. J., 2003. Failure to Produce Mitochondrial DNA Results in Embryonic Lethality in Rnaseh1 Null Mice strand mode of DNA synthesis and that the RNA-DNA hybrids are RNA primers of DNA replication. The presence of abundant RNA-DNA hybrids in replicating Laboratory of M. *Molecular Cell*, 11, 807–815.

Chan, C. S., Guzman, J. N., Ilijic, E., Mercer, J. N., Rick, C., Tkatch, T., Meredith, G. E., and Surmeier, D. J., 2007. 'Rejuvenation' protects neurons in mouse models of Parkinson's disease. *Nature*, 447 (7148), 1081–1086.

Chanut-Delalande, H., Jung, A. C., Lin, L., Baer, M. M., Bilstein, A., Cabernard, C., Leptin, M., and Affolter, M., 2007. A genetic mosaic analysis with a repressible cell marker screen to identify genes involved in tracheal cell migration during Drosophila air sac morphogenesis. *Genetics*, 176 (4), 2177–2187.

Charlesworth, G., Bhatia, K. P., and Wood, N. W., 2013. The genetics of dystonia: New twists in an old tale. *Brain*, 136 (7), 2017–2037.

Chen, C., Ridzon, D. A., Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., Barbisin, M., Xu, N. L., Mahuvakar, V. R., Andersen, M. R., Lao, K. Q., Livak, K. J., and Guegler, K. J., 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, 33 (20), 1–9.

Chen, D., Hoshi, T., Sehgal, A., and Koh, K., 2010. NIH Public Access, 13 (1), 69–75.

Chen, L., Wang, X., Huang, W., Ying, T., Chen, M., Cao, J., and Wang, M., 2017. MicroRNA-137 and its downstream target LSD1 inversely regulate anesthetics-induced neurotoxicity in dorsal root ganglion neurons. *Brain Res Bull*. 135:1-7.

Chen, W., Du, J., Li, X., Zhi, Z., and Jiang, S., 2020. microRNA-137 downregulates MCL1 in ovarian cancer cells and mediates cisplatin-induced apoptosis. *Pharmacogenomics*. 21(3):195-207.

Chen, Y. W., Song, S., Weng, R., Verma, P., Kugler, J. M., Buescher, M., Rouam, S., and Cohen, S., 2014. Systematic study of Drosophila MicroRNA functions using a collection of targeted knockout mutations. *Developmental Cell*, 31 (6), 784–800.

Cheng, Y., Wang, Z.M., Tan, W., Wang, X., Li, Y., Bai, B., Li, Y., Zhang, S.F., Yan, H.L., Chen, Z.L., Liu, C.M., Mi, T.W., Xia, S., Zhou, Z., Liu, A., Tang, G.B., Liu, C., Dai, Z.J., Wang, Y.Y., Wang, H., Wang, X., Kang, Y., Lin, L., Chen, Z., Xie, N., Sun, Q., Xie, W., Peng, J., Chen, D., Teng, Z.Q., Jin, P., 2018. Partial loss of psychiatric risk gene Mir137 in mice causes repetitive behavior and impairs sociability and learning via increased Pde10a. 1. *Nat Neurosci*. 21(12):1689-1703.

Choi, J. H., Yoon, H. R., Kim, G. H., Park, S. J., Shin, Y. L., and Yoo, H. W., 2007. Identification of novel mutations of the HADHA and HADHB genes in patients with mitochondrial trifunctional protein deficiency. *International Journal of Molecular Medicine*, 19, 81–87.

Chungs, S. and Lamendola, J., 1989. Cloning and Sequence Determinationo f Human Placental Aldose.pdf. *The Journal of biological chemistry*, 264 , 14775–14777.

Cirelli, C. and Bushey, D., 2008. Sleep and Wakefulness in Drosophila melanogaster. *Annals of the New York Academy*, (608), 323–329.

Cirelli, C., 2009. The genetic and molecular regulation of sleep: from fruit flies to humans. *Nature Reviews Neuroscience*, 10, 549–560.

Cirelli, C., Bushey, D., Hill, S., Huber, R., Kreber, R., Ganetzky, B., and Tononi, G., 2005. Reduced sleep in Drosophila Shaker mutants. *Nature*, 434 (7037), 1087–1092.

Collins, a L., Kim, Y., Bloom, R. J., Kelada, S. N., Sethupathy, P., and Sullivan, P. F., 2014. Transcriptional targets of the SZ risk gene MIR137. *Translational psychiatry*, 4, e404.

Cosgrove, D., Harold, D., Mothersill, O., Anney, R., Hill, M. J., Bray, N. J., Blokland, G., Petryshen, T., Richards, A., Mantripragada, K., Owen, M., O'Donovan, M. C., Gill, M., Corvin, A., Morris, D. W., and Donohoe, G., 2017. MiR-137-derived polygenic risk: Effects on cognitive performance in patients with SZ and controls. *Translational Psychiatry*, 7.

Costa, E., Silva, J.A., Steffen, R.E., 2019. Urban environment and psychiatric disorders: a review of the neuroscience and biology. *Metabolism*. 100S:153940.

Crocker, A. and Sehgal, A., 2010. Genetic analysis of sleep. *Genes & Development*, 24 (12), 1220–1235.

Cui, Y., Zhao, S., Wang, X., and Zhou, B., 2016. A novel Drosophila mitochondrial carrier protein acts as a Mg2+ exporter in fine-tuning mitochondrial Mg2+ homeostasis. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1863 , 30–39.

Cummings, E., Donohoe, G., Hargreaves, A., Moore, S., Fahey, C., Dinan, T. G., McDonald, C., O'Callaghan, E., O'Neill, F. A., Waddington, J. L., Murphy, K. C., Morris, D. W., Gill, M., and Corvin, A., 2013. Mood congruent psychotic symptoms and specific cognitive deficits in carriers of the novel SZ risk variant at MIR-137. *Neuroscience Letters*, 532 , 33–38.

Dang, W., Steffen, K. K., Perry, R., Dorsey, J. A., Johnson, F. B., Shilatifard, A., Kaeberlein, M., Kennedy, B. K., and Berger, S. L., 2009. Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature*, 459 (7248), 802–807.

Das, A. M., Illsinger, S., Lücke, T., Hartmann, H., Ruiter, J. P. N., Steuerwald, U., Waterham, H. R., Duran, M., and Wanders, R. J. A., 2006. Isolated mitochondrial long-chain ketoacyl-CoA thiolase deficiency resulting from mutations in the HADHB gene. *Clinical Chemistry*, 52, 530–534.

DeCarolis, N.A., Eisch, A.J., 2010. Hippocampal neurogenesis as a target for the treatment of mental illness: acritical evaluation. *Neuropharmacology*. 58:884-93.

de Leeuw, C. A., Mooij, J. M., Heskes, T., and Posthuma, D., 2015. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Computational Biology*, 11, 1–19.

Dennison, C. A., Legge, S. E., Pardiñas, A. F., and Walters, J. T. R., 2019. Genome-wide association studies in SZ: Recent advances, challenges and future perspective. *SZ Research*.

Devanna, P. and Vernes, S. C., 2014. A direct molecular link between the autism candidate gene RORa and the SZ candidate MIR137. *Scientific reports* , 4 , 3994.

Diegelmann, S., Fiala, A., Leibold, C., Spall, T., and Buchner, E., 2002. Transgenic flies expressing the fluorescence calcium sensor Cameleon 2.1 under UAS control. *Genesis (New York, N.Y. : 2000)* , 34 (1–2), 95–8.

Dix, D. J., Allen, J. W., Collins, B. W., Mori, C., Nakamura, N., Poorman-Allen, P., Goulding, E. H., and Eddy, E. M., 1996. Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis, and male infertility. *Proceedings of the National Academy of Sciences of the United States of America*, 93 , 3264–3268.

Doherty, J.L., Owen, M.J., 2014. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome Med*. 28;6:29.

Doherty, A., Smith-Byrne, K., Forreira, T., Holmes, M.V., Holmes, C., Pulit, S.L., and Lindgren, C.M., 2018. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nature communications*. 9 (5257).

Dokucu, M. E., Yu, L., and Taghert, P. H., 2005. Lithium- and valproate-induced alterations in circadian locomotor behavior in Drosophila. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, 30 (12), 2216–2224.

Dong, P., Xiong, Y., Watari, H., Hanley, S. J. B., Konno, Y., Ihira, K., Yamada, T., Kudo, M., Yue, J., and Sakuragi, N., 2016. MiR-137 and miR-34a directly target Snail and inhibit EMT, invasion and sphere-forming ability of ovarian cancer cells. *Journal of Experimental & Clinical Cancer Research*, 1–9.

Donlea, J. M., Pimentel, D., and Miesenböck, G., 2014. Neuronal machinery of sleep homeostasis in Drosophila. *Neuron*, 81 , 860–872.

Donlea, J. M., Pimentel, D., Talbot, C. B., Kempf, A., Omoto, J. J., Hartenstein, V., and Miesenböck, G., 2018. Recurrent Circuitry for Balancing Sleep Need and Sleep. *Neuron*, 97 , 378-389.

Donlea, J.M., Thimgam, M.S., Suzuki, Y., Gottschalk, L., and Shaw, P.J., 2011. Inducing sleep by remote controlfacilitates memory consolidation in drosophila. *science*. 332 (6037), 1571–1576.

Downen, M., Belkowski, S., Knowles, H., Cardillo, M., and Prystowsky, M. B., 1999. Developmental expression of voltage-gated potassium channel beta subunits. *Brain research Developmental brain research*, 117, 71–80.

Dragicevic, E., Poetschke, C., Duda, J., Schlaudraff, F., Lammel, S., Schiemann, J., Fauler, M., Hetzel, A., Watanabe, M., Lujan, R., Malenka, R. C., Striessnig, J., and Liss, B., 2014. Cav1.3 channels control D2-autoreceptor responses via NCS-1 in substantia nigra dopamine neurons. *Brain*, 137, 2287–2302.

Duan, J., Shi, J., Fiorentino, A., Leites, C., Chen, X., Moy, W., Chen, J., Alexandrov, B. S., Usheva, A., He, D., Freda, J., O'Brien, N. L., McQuillin, A., Sanders, A. R., Gershon, E. S., Delisi, L. E., Bishop, A. R., Gurling, H. M. D., Pato, M. T., Levinson, D. F., Kendler, K. S., Pato, C. N., and Gejman, P. V., 2014. A rare functional noncoding variant at the GWAS-Implicated MIR137/MIR2682 locus might confer risk to SZ and bipolar disorder. *American Journal of Human Genetics*, 95 , 744–753.

Duffy, J. B., 2002. GAL4 system in Drosophila: a fly geneticist's Swiss army knife. *Genesis (New York, N.Y. : 2000)*, 34 (1–2), 1–15.

Encode Consortium, Carolina, N., and Hill, C., 2013. For Junk DNA. *Nature*, 489 (7414), 57–74.

Enright, A. J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D. S., 2003. MicroRNA targets in Drosophila. *Genome biology*, 5.

Equence, C. E. S., Iology, T. O. B., The, C., and Consortium, S., 1998. Genome sequence of the nematode C. elegans: A platform for investigating biology. *Science*, 282 (5396), 2012–2018.

Featherstone, D. E., Chen, K., and Broadie, K., 2009. Harvesting and preparing Drosophila embryos for electrophysiological recording and other procedures. *Journal of Visualized Experiments*, (27), 3–5.

Ferreira, M., O'Donovan, M., Meng, Y., 2008. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 40, 1056–1058.

Fitzgerald, J., Morris, D.W., Donohoe, G., 2020. Cognitive Genomics: Recent Advances and Current Challenges. *Curr Psychiatry Rep*. 22, 2.

Filippov, V., Filippova, M., and Gill, S. S., 2001. Drosophila RNase H1 is essential for development but not for proliferation. *Molecular Genetics and Genomics*, 265, 771–777.

Fischbach KF, Dittrich AP. The optic lobe of Drosophila melanogaster. I. A Golgi analysis of wildtype structure. Cell and Tissue Research. 1989:441–475

Foley, C., Corvin, A., and Nakagome, S., 2017. Genetics of SZ: Ready to Translate? *Curr Psychiatry Rep*. 19, 61.

Fraga, M. F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., Iyer, N. G., Pérez-Rosado, A., Calvo, E., Lopez, J. A., Cano, A., Calasanz, M. J., Colomer, D., Piris, M. Á., Ahn, N., Imhof, A., Caldas, C., Jenuwein, T., and Esteller, M., 2005. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature Genetics*, 37, 391–400.

Frank, C. A., Wang, X., Collins, C. a, Rodal, A. a, Yuan, Q., Verstreken, P., and Dickman, D. K., 2013. New approaches for studying synaptic development, function, and plasticity using Drosophila as a model system. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33 (45), 17560–8.

Freeman, A., Pranski, E., Miller, R. D., Radmard, S., Bernhard, D., Jinnah, H. A., Betarbet, R., Rye, D. B., and Sanyal, S., 2012. Sleep fragmentation and motor restlessness in a drosophila model of restless legs syndrome. *Current Biology*, 22 (12), 1142–1148.

Fries, G. R., Carvalho, A. F., and Quevedo, J., 2018. The miR-Nome of bipolar disorder. *Journal of Affective Disorders*, 233, 110–116.

Furuta, M., Kozaki, K.I., Tanaka, S., Arii, S., Imoto, I., Inazawa, J., 2010. miR-124 and miR-203 are epigenetically silenced tumor-suppressive microRNAs in hepatocellular carcinoma. *Carcinogenesis*. 31, 766-76.

Gallopin, T., Luppi, P. H., Cauli, B., Urade, Y., Rossier, J., Hayaishi, O., Lambolez, B., and Fort, P., 2005. The endogenous somnogen adenosine excites a subset of sleep-promoting neurons via A2A receptors in the ventrolateral preoptic nucleus. *Neuroscience*, 134, 1377–1390.

Ganetzkyl, B., 1982. Indirect Suppression Involving Behavioral Mutants. *Genetics*, 597–614.

Gao, F. B., Brenman, J. E., Lily Yeh Jan, and Yuh Nung Jan, 1999. Erratum: Genes regulating dendritic outgrowth, branching, and routing in Drosophila (Genes and Development 13 (2549-2561)). *Genes and Development*, 13 (23), 3170.

Garcia-Borreguero, D., Larrosa, O., and Bravo, M., 2003. Parkinson's disease and sleep. *Sleep Medicine Reviews*, 7, 115–129.

Gentile, S., Fusco, M.L., 2019. SZ and motherhood. *Psychiatry Clin Neurosci*. 73, 376-385.

Gottesman, I.I., Gould, T.D., 2003. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry*. 160, 636-45.

Gonzalez-Giraldo, Y., Gonzalez-Reyes, R. E., and Forero, D. A., 2016. A functional variant in MIR137, a candidate gene for SZ, affects Stroop test performance in young adults. *Psychiatry Research*, 236, 202–205.

Green, E. K., Grozeva, D., Jones, I., Jones, L., Kirov, G., Caesar, S., Gordon-Smith, K., Fraser, C., Forty, L., Russell, E., Hamshere, M. L., Moskvina, V., Nikolov, I., Farmer, a, McGuffin, P., Holmans, P. a, Owen, M. J., O'Donovan, M. C., and Craddock, N., 2009.

The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of SZ. *Molecular psychiatry*, 1–7.

Green, M. J., Cairns, M. J., Wu, J., Dragovic, M., Jablensky, A., Tooney, P. A., Scott, R. J., and Carr, V. J., 2013. Genome-wide supported variant MIR137 and severe negative symptoms predict membership of an impaired cognitive subtype of SZ. *Molecular Psychiatry*, 18, 774–780.

Guan, F., Li, L., Qiao, C., Chen, G., Yan, T., Li, T., Zhang, T., and Liu, X., 2015. Evaluation of genetic susceptibility of common variants in CACNA1D with SZ in Han Chinese. *Scientific reports*, 5, 12935.

Guan, F., Ni, T., Han, W., Lin, H., Zhang, B., Chen, G., Zhu, L., Liu, D., and Zhang, T., 2019. Evaluation of the relationships of the WBP1L gene with SZ and the general psychopathology scale based on a case–control study. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, (December), 1–8.

Guan, F., Zhang, B., Yan, T., Li, L., Liu, F., Li, T., Feng, Z., Zhang, B., Liu, X., and Li, S., 2014. MIR137 gene and target gene CACNA1C of miR-137 contribute to SZ susceptibility in Han Chinese. *SZ Research*, 152, 97–104.

Gurudev, N., Yuan, M., and Knust, E., 2014. chaoptin, prominin, eyes shut and crumbs form a genetic network controlling the apical compartment of Drosophila photoreceptor cells. *Biology Open*, 3, 332–341.

Guzman, J. N., Sánchez-Padilla, J., Chan, C. S., and Surmeier, D. J., 2009. Robust pacemaking in substantia nigra dopaminergic neurons. *J Neurosci*, 29 (35), 11011–11019.

Ha M., and Kim, V.N., 2014. Regulation of microRNA biogenesis. *Nature Reviews*. 15, 509 – 524.

Happe, S., Lüdemann, P., and Berger, K., 2002. The association between disease severity and sleep-related problems in patients with Parkinson's disease. *Neuropsychobiology*, 46, 90–96.

Harbison, S. T. and Sehgal, A., 2008. Quantitative genetic analysis of sleep in Drosophila melanogaster. *Genetics*, 178, 2341–2360.

Hartenstein, V., Cardona, A., Pereanu, W., and Younossi-Hartenstein, A., 2008. Modeling the Developing Drosophila Brain: Rationale, Technique, and Application. *BioScience*, 58, 823–836.

Hartley, P. S., Motamedchaboki, K., Bodmer, R., and Ocorr, K., 2016. SPARC-dependent cardiomyopathy in drosophila. *Circulation: Cardiovascular Genetics*, 9, 119–129.

Hauberg, M. E., Roussos, P., Grove, J., Børglum, A. D., and Mattheisen, M., 2016. Analyzing the Role of MicroRNAs in SZ in the Context of Common Genetic Risk Variants. *JAMA psychiatry*, 73, 369–77.

He, E., Lozano, M.A.G., Stringer, S., Watanabe, K., Sakamoto, K., den Oudsten, F., Koopmans, F., Giamberardino, S.N., Hammerschlag A, Cornelisse LN, Li KW, van Weering J, Posthuma D, Smit AB, Sullivan PF, Verhage M., 2018. MIR137 SZ-associated locus controls synaptic function by regulating synaptogenesis, synapse maturation and synaptic transmission. *Hum Mol Genet*. 27(11), 1879-1891.

Heckers, S., 2001. Neuroimaging studies of the hippocampus in SZ. *Hippocampus*, 11, 520–528.

Herzog-Krzywoszanska, R. and Krzywoszanski, L., 2019. Sleep disorders in Huntington's disease. *Frontiers in Psychiatry*, 10, 1–9.

Heyes, S., Pratt, W. S., Rees, E., Dahimene, S., Ferron, L., Owen, M. J., and Dolphin, A. C., 2015. Genetic disruption of voltage-gated calcium channels in psychiatric and neurological disorders. *Progress in Neurobiology*, 134, 36–54.

Hill, M. J., Donocik, J. G., Nuamah, R. A., Mein, C. A., Sainz-Fuertes, R., and Bray, N. J., 2014. Transcriptional consequences of SZ candidate miR-137 manipulation in human neural progenitor cells. *SZ Research*, 153 (1–3), 225–230.

Hirtz, J. J., Braun, N., Griesemer, D., Hannes, C., Janz, K., Löhrke, S., Müller, B., and Friauf, E., 2012. Synaptic refinement of an inhibitory topographic map in the auditory brainstem requires functional Cav1.3 calcium channels. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 32 (42), 14602–14616.

Hobert, O., 2007. miRNAs Play a Tune. *Cell*, 131, 22–24.

Horwitz, T., Lam, K., Chen, Y., Xia, Y., and Liu, C., 2019. A decade in psychiatric GWAS research. *Molecular Psychiatry*, 24, 378–389.

Howard, D.M., Adams, M.J., Clarke, T., Hafferty, J.D., Gibson, J., Shirali, M., Coleman, J.R.I., Hagenaars, S.P., and Ward, J., 2019. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience*. 22 (3), 343 – 352.

Howell, K. R. and Law, A. J., 2019. Neurodevelopmental concepts of SZ in the genome-wide association era: AKT/mTOR signaling as a pathological mediator of genetic and environmental programming during development. *SZ Research*.

Howes, O.D., McCutcheon, R., Owen, M.J., Murray, R.M., 2017. The role of genes, stress, and dopamine in the development of SZ. *Biol Psychiatry*. 81, 9–20

Hu, T., Guan, T., and Gerace, L., 1996. Molecular and functional characterization of the p62 complex, an assembly of nuclear pore complex glycoproteins. *Journal of Cell Biology*, 134, 589–601.

Huang, D. W., Sherman, B. T., and Lempicki, R. A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4, 44–57.

Huizing, M., Wendel, U., Ruitenbeek, W., Iacobazzi, V., IJlst, L., Veenhuizen, P., Savelkoul, P., Van Den Heuvel, L. P., Smeitink, J. A. M., Wanders, R. J. A., Trijbels, J. M. F., and Palmieri, F., 1998. Carnitine-acylcarnitine carrier deficiency: Identification of the molecular defect in a patient. *Journal of Inherited Metabolic Disease*, 21, 262–267.

Huszar, G., Stone, K., Dix, D., and Vigue, L., 2000. Putative Creatine Kinase M-Isoform in Human Sperm Is Identifiedas the 70-Kilodalton Heat Shock Protein HspA21. *Biology of Reproduction*, 63, 925–932.

Imbrici, P., Camerino, D. C., and Tricarico, D., 2013. Major channels involved in neuropsychiatric disorders and therapeutic perspectives. *Frontiers in Genetics*, 4, 1–19.

Iritani, S., 2014. What happens in the brain of SZ patients? An investigation from the viewpoint of neuropathology. *Nagoya Journal of Medical Science*, 76 (1–2), 11–28.

Ishimoto, H., Lark, A., and Kitamoto, T., 2012. Factors that differentially affect daytime and nighttime sleep in Drosophila melanogaster. *Frontiers in Neurology*, FEB, 1–5.

Jakubowski, B. R., Longoria, R. A., and Shubeita, G. T., 2012. A high throughput and sensitive method correlate neuronal disorder genotypes to Drosophila larvae crawling phenotypes. *Fly*, 6, 303–308.

Jablensky, A., McNeil, T.F., and Morgan, V.A., 2017. Barbara Fish and a Short History of the Neurodevelopmental Hypothesis of SZ. *Schizophr Bull*. 43, 1158-1163.

James, S. L., Abate, D., Abate, K. H., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R. S., Abebe, Z., Abera, S. F., Abil, O. Z., Abraha, H. N., Abu-Raddad, L. J., Abu-Rmeileh, N. M. E., Accrombessi, M. M. K., Acharya, D., Acharya, P., Ackerman, I. N., Adamu, A. A., Adebayo, O. M., Adekanmbi, V., Adetokunboh, O. O., Adib, M. G., Adsuar, J. C., Afanvi, K. A., Afarideh, M., Afshin, A., Agarwal, G., Agesa, K. M., Aggarwal, R., 2018. Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392 (10159), 1789–1858.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., and D'Eustachio, P., 2019. The reactome pathway knowledgebase. *Nucleic acids research*, 48 (November 2019), 498–503.

Jenett, A., Rubin, G. M., Ngo, T. T. B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B. D., Cavallaro, A., Hall, D., Jeter, J., Iyer, N., Fetter, D., Hausenfluck, J. H., Peng, H., Trautman, E. T., Svirskas, R. R., Myers, E. W., Iwinski, Z. R., Aso, Y., DePasquale, G. M., Enos, A., Hulamm, P., Lam, S. C. B., Li, H. H., Laverty, T. R., Long, F., Qu, L., Murphy, S. D., Rokicki, K., Safford, T., Shaw, K., Simpson, J. H., Sowell, A., Tae, S., Yu, Y., and Zugates, C. T., 2012. A GAL4-Driver Line Resource for Drosophila Neurobiology. *Cell Reports*, 2 , 991–1001.

Jia, D. D., Zhang, L., Chen, Z., Wang, C. R., Huang, F. Z., Duan, R. H., Xia, K., Tang, B. S., and Jiang, H., 2013. Lithium chloride alleviates neurodegeneration partly by inhibiting activity of GSK3â in a SCA3 drosophila model. *Cerebellum*, 12, 892–901.

Jia, X., Wang, F., Han, Y., Geng, X., Li, M., Shi, Y., Lu, L., and Chen, Y., 2016. miR-137 and miR-491 Negatively Regulate Dopamine Transporter Expression and Function in Neural Cells. *Neuroscience Bulletin*, 32, 512–522.

Jiang, F., Wu, Q., Sun, S., Bi, G., and Guo, L., 2019. Identification of potential diagnostic biomarkers for Parkinson's disease. *FEBS Open Bio*, 9, 1460–1468.

Jiang, K., Ren, C., and Nair, V. D., 2013. MicroRNA-137 represses Klf4 and Tbx3 during differentiation of mouse embryonic stem cells. *Stem Cell Research*, 11, 1299–1313.

Jing, L., Li, S., Wang, J., and Zhang, G., 2019. Long non-coding RNA small nucleolar RNA host gene 7 facilitates cardiac hypertrophy via stabilization of SDA1 domain containing 1 mRNA. *Journal of Cellular Biochemistry*, 120, 15089–15097.

Jiren, F., 2007. Bracketing likened to flowers, branches and foliage: Architectural metaphors and conceptualization in tenth to twelfth-century China as reflected in the Yingzao fashi. *T'oung Pao*, 93 (4–5), 369–432.

Jones, S.E., Lane, J.M., Wood, A.R., Hees, V.T., Tyrrell, J., and 28 others, 2019. Genome wide association analyses of chronotype in 697828 individuals provides insight into circadian rhythms. *Nature communications.* 10 (343).

Juźwik, C. A., Drake, S., Lécuyer, M. A., Johnson, R. M., Morquette, B., Zhang, Y., Charabati, M., Sagan, S. M., Bar-Or, A., Prat, A., and Fournier, A. E., 2018. Neuronal microRNA regulation in Experimental Autoimmune Encephalomyelitis. *Scientific Reports*, 8, 1–16.

Kaas, G. A., Kasuya, J., Lansdon, P., Ueda, A., Iyengar, A., Wu, C.-F., and Kitamoto, T., 2016. Lithium-responsive seizure-like hyperexcitability is caused by a mutation in the Drosophila voltage-gated sodium channel gene paralytic. *eNeuro*, 3 (October).

Kandemir, H., Erdal, M. E., Selek, S., Ay, Ö. İ., Karababa, I. F., Kandemir, S. B., Ay, M. E., Yılmaz, Ş. G., Bayazıt, H., and Taşdelen, B., 2014. Evaluation of several micro RNA (miRNA) levels in children and adolescents with attention deficit hyperactivity disorder. *Neuroscience letters*, 580, 158–62.

Kaplan, R.M., 2018. Being Bleuler: the second century of SZ. *Australia Psychiatry*. 16, 305-11.

Kelleher, I., Cederlöf, M., and Lichtenstein, P., 2014. Psychotic experiences as a predictor of the natural course of suicidal ideation: a Swedish cohort study. *World Psychiatry*. 13, 184-8.

Karres, J. S., Hilgers, V., Carrera, I., Treisman, J., and Cohen, S. M., 2007. The Conserved microRNA MiR-8 Tunes Atrophin Levels to Prevent Neurodegeneration in Drosophila. *Cell*, 131, 136–145.

Kasuya, J., Kaas, G., and Kitamoto, T., 2009. Effects of lithium chloride on the gene expression profiles in Drosophila heads. *Neuroscience Research*, 64, 413–420.

Khandelwal, D., Dutta, D., Chittawar, S., & Kalra, S. (2017). Sleep Disorders in Type 2 Diabetes. Indian journal of endocrinology and metabolism, 21, 758–761.

Khvorova, A., Reynolds, A., and Jayasena, S. D., 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell*, 115, 209–216.

Kim, J. J., Mandelli, L., Lim, S., Lim, H. K., Kwon, O. J., Pae, C. U., Serretti, A., Nimgaonkar, V. L., Paik, I. H., and Jun, T. Y., 2008. Association analysis of heat shock protein 70 gene polymorphisms in SZ. *European Archives of Psychiatry and Clinical Neuroscience*, 258, 239–244.

Kim, V.N., Han, J., and Siomi, M.C., 2009. Biogenesis of small RNAs in animals. *Nature Reviews.* 10, 126 – 139.

Kishita, Y., Tsuda, M., and Aigaki, T., 2012. Impaired fatty acid oxidation in a Drosophila model of mitochondrial trifunctional protein (MTP) deficiency. *Biochemical and Biophysical Research Communications*, 419, 344–349.

Klein, P. S. and Melton, D. A., 1996. A molecular mechanism for the effect of lithium on development. *Proceedings of the National Academy of Sciences of the United States of America*, 93 (16), 8455–8459.

Klemenz, R., Weber, U., and Gehring, walter j, 1987. The white gene as a marker in a new p-element vector for gene transfer in Drosophila. *nucleic acids research*, 15 (10), 3947–3959.

Knutson, K. L. and Van Cauter, E., 2008. Associations between sleep loss and increased risk of obesity and diabetes. *Ann N Y Acad Sci*, 1129, 287–304.

Kobayashi, M., Michaut, L., Ino, A., Honjo, K., Nakajima, T., Maruyama, Y., Mochizuki, H., Ando, M., Ghangrekar, I., Takahashi, K., Saigo, K., Ueda, R., Gehring, W. J., and Furukubo-Tokunaga, K., 2006. Differential microarray analysis of Drosophila mushroom body transcripts using chemical ablation. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (39), 14417–14422.

Koles, K., Messelaar, E. M., Feiger, Z., Yu, C. J., Frank, C. A., and Rodal, A. A., 2015. The EHD protein Past1 controls postsynaptic membrane elaboration and synaptic function. *Molecular Biology of the Cell*, 26 (18), 3275–3288.

Kong, Y., Liang, X., Liu, L., Zhang, D., Wan, C., Gan, Z., Yuan, L., and Jiang, B. H., 2015. High throughput sequencing identifies MicroRNAs mediating ??-synuclein toxicity by targeting neuroactive-ligand receptor interaction pathway in early stage of Drosophila Parkinson's disease model. *PLoS ONE*, 10, 1–24.

Kozlowska, E., Krzyzosiak, W. J., and Koscianska, E., 2013. Regulation of Huntingtin gene expression by miRNA-137, -214, -148a, and their respective isomiRs. *International Journal of Molecular Sciences*, 14, 16999–17016.

Kozomara, A., and Griffiths-Jones, S., 2014. Mirbase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research*. 42, 68 – 73.

Krystal, A. D., 2012. Psychiatric Disorders and Sleep. *Neurologic Clinics*, 30, 1389–1413.

Kumar, D., Dedic, N., Flachskamm, C., Deussing, J., and Kimura, M., 2013. CAV1.2 calcium channel is involved in the circadian regulation of sleep. *Sleep Medicine*, 14 (2013), 175.

Kume, K., Kume, S., Park, S. K., Hirsh, J., and Jackson, F. R., 2005. Cellular/Molecular Dopamine Is a Regulator of Arousal in the Fruit Fly. *Journal of Neuroscience*, 25 (32), 7377–7384.

Kuswanto, C. N., Sum, M. Y., Qiu, A., Sitoh, Y. Y., Liu, J., and Sim, K., 2015. The impact of genome wide supported microRNA-137 (MIR137) risk variants on frontal and striatal white matter integrity, neurocognitive functioning, and negative symptoms in SZ. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, 168, 317–326.

Kwon, E., Wang, W., and Tsai, L. H., 2013. Validation of SZ-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. *Molecular Psychiatry*, 18, 11–12. Available from: http://dx.doi.org/10.1038/mp.2011.170.

Lane, J. M., Liang, J., Irma, V., Anderson, S. G., Bechtold, D. A., Bowden, J., Emsley, R., Gill, S., Little, max a, Luik, annemaries i, Loudon, A., Scheer, frank ajl, Purcell, shaun m, Kyle, simon d, Lawlor, deborah a, Zhu, X., Redline, S., Ray, david w, Rutter, marti k, and Saxena, R., 2017. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nature Genetics*, 49, 274–281.

Lecture, K., 1998. Galactosaemia today. *The enigma*, 21, 455–471.

Lee, E. K. and Douglass, A. B., 2010. Sleep in psychiatric disorders: Where are we now? *Canadian Journal of Psychiatry*, 55, 403–412.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., and Kim, V. N., 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425 (6956), 415–419.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S. H., and Kim, V. N., 2004. MicroRNA genes are transcribed by RNA polymerase II. *Embo J*, 23 (20), 4051–4060.

Lee, Y. H., Kim, J.-H., and Song, G. G., 2013. Pathway analysis of a genome-wide association study in SZ. *Gene*, 525 , 107–15.

Lee, K.M., Hawi, Z.H., Parkington, H.C., Parish, C.L., Kumar, P.V., Polo, J.M., Bellgrove, M.A., and Tong, J., 2020. The application of human pluripotent stem cells to model the neuronal and glial components of neurodevelopmental disorders. *Mol Psychiatry*. 25, 368-378.

Lee, S.J., Jeong, J.H., Kang, S.H., Kang, J., Kim, E.A., Lee, J., Jung, J.H., Park, H.Y., and Chae, Y.S., 2019. MicroRNA-137 Inhibits Cancer Progression by Targeting Del-1 in Triple-Negative Breast Cancer Cells. *Int J Mol Sci*, 20(24).

Lee, T.Y., Kim, S.N., Jang, J.H., Shim, G., Jung, W.H., Shin, N.Y., and Kwon, J.S., 2013. Neural correlate of impulsivity in subjects at ultra-high risk for psychosis. *Prog Neuropsychopharmacol Biol Psychiatry*. 45, 165-9.

Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P., and Burge, C. B., 2003. 33-Prediction of Mammalian MicroRNA Targets. *Cell*, 115, 787–798.

Lewis, S.W., and Murray, R.M., 1987. Obstetric complications, neurodevelopmental deviance, and risk of SZ. *J Psychiatr Res*. 21, 413-21.

Liang, M.L., Hsieh, T.H., Ng, K.H., Tsai, Y.N., Tsai, C.F., Chao, M.E., Liu, D.J., Chu, S.S., Chen, W., Liu YR, Liu RS(10)(11)(12), Lin SC(13), Ho DM(10)(13), Wong TT, Yang MH, and Wang, H.W., 2016. Downregulation of miR-137 and miR-6500-3p promotes cell proliferation in pediatric high-grade gliomas. *Oncotarget*. 7(15), 19723-37.

Li, L. L., de Mera, R. M. M. F., Chen, J., Ba, W., Kasri, N. N., Zhang, M., and Courtney, M. J., 2015. Unexpected heterodivalent recruitment of NOS1AP to nNOS reveals multiple sites for pharmacological intervention in neuronal disease models. *Journal of Neuroscience*, 35 (19), 7349–7364.

Li, W., Zhang, X., Zhuang, H., Chen, H. G., Chen, Y., Tian, W., Wu, W., Li, Y., Wang, S., Zhang, L., Chen, Y., Li, L., Zhao, B., Sui, S., Hu, Z., and Feng, D., 2014. MicroRNA-137 is a novel hypoxia-responsive MicroRNA that inhibits mitophagy via regulation of two mitophagy receptors FUNDC1 and NIX. *Journal of Biological Chemistry*, 289 (15), 10691–10701.

Li, Y., Zhang, S., Yan, H., Chen, Z., Liu, C., Mi, T., Xia, S., Zhou, Z., Liu, A., Tang, G., Liu, C., Wang, Y., Wang, H., Wang, X., Kang, Y., Lin, L., Xie, N., Sun, Q., Xie, W., Peng, J., and Chen, D., 2018. Partial loss of psychiatric risk gene miR-137 in mice causes repetitive behavior and impairs sociability and learning via increased Pde10a. *Nature Neuroscience*, 21 (12), 1689–1703.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J. P., 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27 (12), 1739–1740.

Lin, F. J., Pierce, M. M., Sehgal, A., Wu, T., Skipper, D. C., and Chabba, R., 2010. Effect of taurine and caffeine on sleep-wake activity in Drosophila melanogaster. *Nature and Science of Sleep*, 2, 221–231.

Lin, J., Sergeeva, O. A., and Haas, H. L., 2011. Histamine H 3 Receptors and Sleep-Wake Regulation. *The Journal of pharmacology and experimental therapeutics*, 336, 17–23.

Lin, J. C., Ho, W. H., Gurney, A., and Rosenthal, A., 2003. The netrin-G1 ligand NGL-1 promotes the outgrowth of thalamocortical axons. *Nature Neuroscience*, 6 (12), 1270–1276.

Lin, M., Hrabovsky, A., Pedrosa, E., Wang, T., Zheng, D., and Lachman, H. M., 2012. Allele-Biased Expression in Differentiating Human Neurons: Implications for Neuropsychiatric Disorders. *PLoS ONE*, 7, 1–10.

Linford, N. J., Bilgir, C., Ro, J., and Pletcher, S. D., 2013. Measurement of lifespan in Drosophila melanogaster. *Journal of Visualized Experiments*, (71), 1–9.

Lionel, A. C., Vaags, A. K., Sato, D., Gazzellone, M. J., Mitchell, E. B., Chen, H. Y., Costain, G., Walker, S., Egger, G., Thiruvahindrapuram, B., Merico, D., Prasad, A., Anagnostou, E., Fombonne, E., Zwaigenbaum, L., Roberts, W., Szatmari, P., Fernandez, B. A., Georgieva, L., Brzustowicz, L. M., Roetzer, K., Kaschnitz, W., Vincent, J. B., Windpassinger, C., Marshall, C. R., Trifiletti, R. R., Kirmani, S., Kirov, G., Petek, E., Hodge, J. C., Bassett, A. S., and Scherer, S. W., 2013. Rare exonic deletions implicate the synaptic organizer gephyrin (GPHN) in risk for autism, SZ and seizures. *Human Molecular Genetics*, 22 (10), 2055–2066.

Liu, C., Haynes, P. R., Donelson, N. C., Aharon, S., and Griffith, L. C., 2015. Sleep in Populations of Drosophila Melanogaster. *eNeuro*, 2, 1-15.

Liu, Q., Liu, S., Kodama, L., Driscoll, M. R., and Wu, M. N., 2012. Two dopaminergic neurons signal to the dorsal fan-shaped body to promote wakefulness in Drosophila. *Current Biology*, 22 (22), 2114–2123.

Liu, T., Dartevelle, L., Yuan, C., Wei, H., Wang, Y., Ferveur, J. F., and Guo, A., 2009. Reduction of dopamine level enhances the attractiveness of male Drosophila to other males. *PLoS ONE*, 4, 1–5.

Liu, Y. ping, Meng, J. hua, Wu, X., Xu, F. L., Xia, X., Zhang, X. cen, Liu, Y., Yao, J., and Wang, B. jie, 2019. Rs1625579 polymorphism in the MIR137 gene is associated with the risk of SZ: updated meta-analysis. *Neuroscience Letters*, 713 (77), 13453-5.

Liu, Y., Harding, M., Pittman, A., Dore, J., Striessnig, J., Rajadhyaksha, A., Chen, X., Bargas, J., Howe, A., Eberwine, J., Cao, Y., Surmeier, D., Busquet, P., Nguyen, N., Schmid, E., Tanimoto, N., Seeliger, M., Ben-Yosef, T., Mizuno, F., Akopian, A., Striessnig, J., Singewald, N., Cao, J., Covington, H., Friedman, A., Wilkinson, M., Walsh, J., Cooper, D., Nestler, E., Han, M., Cardozo, D., Bean, B., Carlin, K., Jones, K., Jiang, Z., Jordan, L., Brownstone, R., Chan, C., Guzman, J., Ilijic, E., Mercer, J., Rick, C., Striessnig, J., Koschak, A., Striessnig, J., Koschak, A., Sinnegger-Brauns, M., Hetzenauer, A., Nguyen, N., Busquet, P., Pelster, G., Singewald, N., Takada, M., Kang, Y., Imanishi, M., Tobler, P., Fiorillo, C., Schultz, W., Tritsch, N., Sabatini, B., Xu, W., Lipscombe, D., Ye, J., Zhang, J., Xiao, C., Kong, J., Zhang, H., Fu, Y., Altier, C., Platzer, J., Surmeier, D., Bezprozvanny, I., Zhang, J., Berra-Romani, R., Sinnegger-Brauns, M., Striessnig, J., Blaustein, M., Matteson, D., Zhang, L., Liu, Y., and Chen, X., 2014. Cav1.2 and Cav1.3 L-type calcium channels regulate dopaminergic firing activity in the mouse ventral tegmental area. *Journal of neurophysiology*, 112, 1119–30.

Llinás, R. R., Ribary, U., Jeanmonod, D., Kronberg, E., and Mitra, P. P., 1999. Thalamocortical dysrhythmia: A neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *Proceedings of the National Academy of Sciences of the United States of America*, 96 (26), 15222–15227.

Lo, P., Bhatia, G., Gusev, A., Finucane, H.K., and 11 others, 2015. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance components analysis. *Nature genetics*. 47 (12), 1385 – 1392.

Lobell, A. S., Kaspari, R. R., Serrano Negron, Y. L., and Harbison, S. T., 2017. The genetic architecture of Ovariole number in Drosophila melanogaster: Genes with major, quantitative, and pleiotropic effects. *G3: Genes, Genomes, Genetics*, 7, 2391–2403.

Losses, D. and Larger, L., 2016. HHS Public Access, 16, 338–348.

Lu, J., Shen, Y., Qingfu, W., Kumar, S., He, B., Shi, S., Carthew, R.W., Wang, S.M., and Wu, C., 2008. The birth and death of microrna genes in drosophila. *Nature Genetics,* 40 (3), 351 – 355.

Lund, E., Güttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U., 2004. Nuclear export of microRNA precursors. *Science*, 303 (5654), 95–98.

Maas, D.A., Vallès, A., and Martens, G.J.M., 2017. Oxidative stress, prefrontal cortex hypomyelination and cognitive symptoms in SZ. *Transl Psychiatry*. 18, 7.

MacRae, I. J., Ma, E., Zhou, M., Robinson, C. V., and Doudna, J. A., 2008. In vitro reconstitution of the human RISC-loading complex. *Proceedings of the National Academy of Sciences*, 105, 512–517.

Mahmoudi, E. and Cairns, M. J., 2017. MiR-137: An important player in neural development and neoplastic transformation. *Molecular Psychiatry*, 22, 44–55.

Marschallinger, J., Sah, A., Schmuckermair, C., Unger, M., Rotheneichner, P., Kharitonova, M., Waclawiczek, A., Gerner, P., Jaksch-Bogensperger, H., Berger, S., Striessnig, J., Singewald, N., Couillard-Despres, S., and Aigner, L., 2015. The L-type calcium channel Cav1.3 is required for proper hippocampal neurogenesis and cognitive functions. *Cell Calcium*, 58, 606–616.

McGurk, L., Berson, A., and Bonini, N. M., 2015. Drosophila as an in vivo model for human neurodegenerative disease. *Genetics*, 201, 377–402.

McGuffin, P., Owen, M.J., and Farmer, A.E., 1995. Genetic basis of SZ. *Lancet*. 346(8976), 678-82.

McMillan, P.A., and McGuire, T.R, 1992. The homeotic gene, spineless-aristopedia, affects geotaxis in drosophila melanogaster. *Behaviour Genetics*, 22, 557 – 573.

Meister, G., Landthaler, M., Peters, L., Chen, P. Y., Urlaub, H., Lührmann, R., and Tuschl, T., 2005. Identification of novel argonaute-associated proteins. *Current Biology*, 15 (23), 2149–2155.

Mirzoev, A., Bercovici, E., Stewart, L. S., Cortez, M. A., Snead, O. C., and Desrocher, M., 2012. Circadian profiles of focal epileptic seizures: A need for reappraisal. *Seizure*, 21, 412–416.

Mishima, T., Sadovsky, E., Gegick, M. E., and Sadovsky, Y., 2016. Determinants of effective lentivirus-driven microRNA expression in vivo. *Scientific Reports*, 6 (August), 1–11.

Mollon, J., Mathias, S.R., Knowles, E.E.M., Rodrigue A, Koenis MMG, Pearlson GD, Reichenberg, A., Barrett J, Denbow D, Aberizk K, Zatony M, Poldrack RA, Blangero J, Glahn, D.C., 2020. Cognitive impairment from early to middle adulthood in patients with affective and nonaffective psychotic disorders. *Psychol Med*. 50, 48-57.

Monti, J. M. and Monti, D., 2004. Sleep in SZ patients and the effects of antipsychotic drugs. *Sleep Medicine Reviews*, 8, 133–148.

Moore, R., Casale, F. P., Jan Bonder, M., Horta, D., Heijmans, B. T., Peter, P. A., van Meurs, J., Isaacs, A., Jansen, R., Franke, L., Boomsma, D. I., Pool, R., van Dongen, J., Hottenga, J. J., van Greevenbroek, M. M. J., Stehouwer, C. D. A., van der Kallen, C. J. H., Schalkwijk, C. G., Wijmenga, C., Zhernakova, A., Tigchelaar, E. F., Slagboom, P. E., Beekman, M., Deelen, J., van Heemst, D., Veldink, J. H., van den Berg, L. H., van Duijn, C. M., Hofman, B. A., Uitterlinden, A. G., Jhamai, P. M., Verbiest, M., Suchiman, H. E. D., Verkerk, M., van der Breggen, R., van Rooij, J., Lakenberg, N., Mei, H., van Iterson, M., van Galen, M., Bot, J., van't Hof, P., Deelen, P., Nooren, I., Moed, M., Vermaat, M., Zhernakova, D. V., Luijk, R., Jan Bonder, M., van Dijk, F., Arindrarto, W., Kielbasa, S. M., Swertz, M. A., van Zwet, E. W., Franke, L., Barroso, I., and Stegle, O., 2019. A linear mixed-model approach to study multivariate gene–environment interactions. *Nature Genetics*, 51, 180–186.

Morgan, M. D., Pairo-Castineira, E., Rawlik, K., Canela-Xandri, O., Rees, J., Sims, D., Tenesa, A., and Jackson, I. J., 2018. Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability. *Nature Communications*, 9, 1–10.

Morrison, D. K., 2012. MAP kinase pathways. *Cold Spring Harbor Perspectives in Biology*, 4 (11), 1–5.

Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G., 2002. miRNPs: A novel class of ribonucleoproteins containing numerous microRNAs. *Genes and Development*, 16, 720–728.

Murai, K., Sun, G., Ye, P., Tian, E., Yang, S., Cui, Q., Sun, G., Trinh, D., Sun, O., Hong T, Wen Z, Kalkum M, Riggs AD3, Song H5,6,7, Ming GL5,6,7,8, Shi Y., 2016. The TLX-miR-219 cascade regulates neural stem cell proliferation in neurodevelopment and SZ iPSC model. *Nat Commun.* 7, 10965.

Murray, R.M., Bhavsar, V., Tripoli, G., and Howes, O., 2017. 30 Years on: How the Neurodevelopmental Hypothesis of SZ Morphed into the Developmental Risk Factor Model of Psychosis. *Schizophr Bull*. 43, 1190-1196.

Murray, R.M., and Fearon, P., 1999. The developmental 'risk factor' model of SZ. *J Psychiatr Res.* 33:497–499.

Murray RM, Lewis SW., 1987. Is SZ a neurodevelopmental disorder? *Br Med J (Clin Res Ed)*. 295(6600), 681-2.

Murchison, E. P. and Hannon, G. J., 2004. miRNAs on the move: miRNA biogenesis and the RNAi machinery. *Current Opinion in Cell Biology*, 16, 223–229.

Nakahara, K., and Carthew, R.W., 2004. Expanding roles for miRNAs and siRNAs in cell regulation. *Curr Opin Cell Biol*. 16, 127-33.

Nall, A. H., Shakhmantsir, I., Cichewicz, K., Birman, S., Hirsh, J., and Sehgal, A., 2016. Caffeine promotes wakefulness via dopamine signalling in Drosophila. *Scientific reports*, 6, 20938.

Neely, G. G., Hess, A., Costigan, M., Keene, A. C., Goulas, S., Langeslag, M., Griffin, R. S., Belfer, I., Dai, F., Smith, S. B., Diatchenko, L., Gupta, V., Xia, C. ping, Amann, S., Kreitz, S., Heindl-Erdmann, C., Wolz, S., Ly, C. V., Arora, S., Sarangi, R., Dan, D., Novatchkova, M., Rosenzweig, M., Gibson, D. G., Truong, D., Schramek, D., Zoranovic, T., Cronin, S. J. F., Angjeli, B., Brune, K., Dietzl, G., Maixner, W., Meixner, A., Thomas, W., Pospisilik, J. A., Alenius, M., Kress, M., Subramaniam, S., Garrity, P. A., Bellen, H. J., Woolf, C. J., and Penninger, J. M., 2010. A Genome-wide Drosophila screen for heat nociception identifies α2δ3 as an evolutionarily conserved pain gene. *Cell*, 143 , 628–638.

Neely, G. G., Kuba, K., Cammarato, A., Isobe, K., Amann, S., Zhang, L., Murata, M., Elmén, L., Gupta, V., Arora, S., Sarangi, R., Dan, D., Fujisawa, S., Usami, T., Xia, C. ping, Keene, A. C., Alayari, N. N., Yamakawa, H., Elling, U., Berger, C., Novatchkova, M., Koglgruber, R., Fukuda, K., Nishina, H., Isobe, M., Pospisilik, J. A., Imai, Y., Pfeufer, A., Hicks, A. A., Pramstaller, P. P., Subramaniam, S., Kimura, A., Ocorr, K., Bodmer, R., and Penninger, J. M., 2010. A Global In Vivo Drosophila RNAi Screen Identifies NOT3 as a Conserved Regulator of Heart Function. *Cell*, 141, 142–153.

Nehlig, A., Daval, J. L., and Debry, G., 1992. Caffeine and the central nervous system: mechanisms of action, biochemical, metabolic and psychostimulant effects. *Brain Research Reviews*, 17, 139–170.

Neilson, E., Bois, C., Clarke, T. K., Hall, L., Johnstone, E. C., Owens, D. G. C., Whalley, H. C., McIntosh, A. M., and Lawrie, S. M., 2018. Polygenic risk for SZ, transition and cortical gyrification: A high-risk study. *Psychological Medicine*, 48, 1532–1539.

Nesic, M. J., Stojkovic, B., and Maric, N. P., 2019. On the origin of SZ: Testing evolutionary theories in the post-genomic era. *Psychiatry and Clinical Neurosciences*, 723–730.

Nesvåg, R., Lawyer, G., Varnäs, K., Fjell, A. M., Walhovd, K. B., Frigessi, A., Jönsson, E. G., and Agartz, I., 2008. Regional thinning of the cerebral cortex in SZ: Effects of diagnosis, age and antipsychotic medication. *SZ Research*, 98 (1–3), 16–28.

Neumüller, R. A., Richter, C., Fischer, A., Novatchkova, M., Neumüller, K. G., and Knoblich, J. A., 2011. Genome-wide analysis of self-renewal in Drosophila neural stem cells by transgenic RNAi. *Cell Stem Cell*, 8, 580–593.

Neumüller, R. A., Richter, C., Fischer, A., Novatchkova, M., Neumüller, K. G., and Knoblich, J. A., 2011. Genome-wide analysis of self-renewal in Drosophila neural stem cells by transgenic RNAi. *Cell Stem Cell*, 8, 580–593.

Ni, J.D., Gurav, A.S., Liu, W., Ogunmowo, T.H., Hackbart, H., Elsheikh, A., Verdegaal, A.A., and Montell, C., 2019. Differential regulation of the Drosophila sleep homeostat by circadian and arousal inputs. *eLife*. 8, 2 - 27.

Nichols, C. D., Becnel, J., and Pandey, U. B., 2012. Methods to Assay Drosophila Behavior. *Journal of Visualized Experiments*, (March), 3–7.

Nicholson, A. W., 2014. Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdisciplinary Reviews: RNA*, 5, 31–48.

Nilsson, E. M., Laursen, K. B., Whitchurch, J., McWilliam, A., Ødum, N., Persson, J. L., Heery, D. M., Gudas, L. J., and Mongan, N. P., 2015. MiR137 is an androgen regulated repressor of an extended network of transcriptional coregulators. *Oncotarget*, 6 (34), 35710–25.

Nyegaard, M., Demontis, D., Foldager, L., Hedemand, A., Flint, T. J., Sørensen, K. M., Andersen, P. S., Nordentoft, M., Werge, T., Pedersen, C. B., Hougaard, D. M., Mortensen, P. B., Mors, O., and Børglum, A. D., 2010. CACNA1C (rs1006737) is associated with SZ. *Molecular Psychiatry*, 15, 119–121.

Ohi, K., Nishizawa, D., Shimada, T., Kataoka, Y., Hasegawa, J., Shioiri, T., Kawasaki, Y., Hashimoto, R., and Ikeda, K., 2020. Polygenetic Risk Scores for Major Psychiatric Disorders Among Schizophrenia Patients, Their First-Degree Relatives, and Healthy Participants. Neuropsychopharmacology. 23(3): 157–164.

Olde Loohuis, N. F. M., Ba, W., Stoerchel, P. H., Kos, A., Jager, A., Schratt, G., Martens, G. J. M., van Bokhoven, H., Nadif Kasri, N., and Aschrafi, A., 2015. MicroRNA-137 Controls AMPA-Receptor-Mediated Transmission and mGluR-Dependent LTD. *Cell Reports*, 11 (12), 1876–1884.

Olde Loohuis, N. F. M., Nadif Kasri, N., Glennon, J. C., van Bokhoven, H., H??bert, S. S., Kaplan, B. B., Martens, G. J. M., and Aschrafi, A., 2017. The SZ risk gene MIR137 acts as a hippocampal gene network node orchestrating the expression of genes relevant to nervous system development and function. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 73, 109–118.

Olswang-Kutz, Y., Gertel, Y., Benjamin, S., Sela, O., Pekar, O., Arama, E., Steller, H., Horowitz, M., and Segal, D., 2009. Erratum: Drosophila Past1 is involved in endocytosis and is required for germline development and survival of the adult fly (Journal of Cell Science vol. 122 (471-480)). *Journal of Cell Science*, 122 (11), 1929.

Ortner, N. J. and Striessnig, J., 2016. L-type calcium channels as drug targets in CNS disorders. *Channels*, 10, 7–13.

Osterwalder, T., Yoon, K. S., White, B. H., and Keshishian, H., 2001. A conditional tissue-specific transgene expression system using inducible GAL4. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (22), 12596–12601.

Oyama, S., Yamakawa, H., Sasagawa, N., Hosoi, Y., Futai, E., and Ishiura, S., 2009. Dysbindin-1, a SZ-related protein, functionally interacts with the DNA-dependent protein kinase complex in an isoform-dependent manner. *PLoS ONE*, 4.

Pacheco, A., Berger, R., Freedman, R., and Law, A. J., 2019. A VNTR Regulates miR-137 Expression Through Novel Alternative Splicing and Contributes to Risk for SZ. *Scientific Reports*, 9, 1–12.

Park, S., Sonn, J. Y., Oh, Y., Lim, C., and Choe, J., 2014. SIFamide and SIFamide receptor define a novel neuropeptide signaling to promote sleep in Drosophila. *Molecules and Cells*, 37, 295–301.

Pauls, D., Selcho, M., Gendre, N., Stocker, R. F., and Thum, A. S., 2010. Drosophila larvae establish appetitive olfactory memories via mushroom body neurons of embryonic origin. *Journal of Neuroscience*, 30 (32), 10655–10666.

Panoutsopoulou, K., Tachmazidou, I., and Zeggini, E., 2013. In search of low-frequency and rare variants affecting complex traits. *Human Molecular Genetics*. 22, 16 – 21.

Penadés, R., Bosia, M., Catalán, R., Spangaro, M., García-Rizo, C., Amoretti, S., Bioque, M., and Bernardo, M., 2019. The role of genetics in cognitive remediation in SZ: A systematic review. *SZ Research: Cognition*, 19.

Pendleton, R. G., Rasheed, A., Sardina, T., Tully, T., and Hillman, R., 2002. Effects of tyrosine hydroxylase mutants on locomotor activity in Drosophila: A study in functional genomics. *Behavior Genetics*, 32, 89–94.

Peschel, N. and Helfrich-Förster, C., 2011. Setting the clock - By nature: Circadian rhythm in the fruitfly Drosophila melanogaster. *FEBS Letters*, 585 (10), 1435–1442.

Pfeiffenberger, C., Lear, B. C., Keegan, K. P., and Allada, R., 2010. Locomotor activity level monitoring using the Drosophila activity monitoring (DAM) system. *Cold Spring Harbor Protocols*, 5 (11).

Pfeiffer, B. D., Ngo, T. T. B., Hibbard, K. L., Murphy, C., Jenett, A., Truman, J. W., and Rubin, G. M., 2010. Refinement of tools for targeted gene expression in Drosophila. *Genetics*, 186, 735–755.

Picao-Osario, J., Lago-Baldaia, I., Patraquim, P., and Alonso, C.R., 2017. Pervasive behavioural effects of microrna regulation in Drosophila. *Genetics*. 206, 1535 – 1548.

Piazza, N. and Wessells, R. J., 2011. *Drosophila models of cardiac disease*. Progress in Molecular Biology and Translational Science.

Piggins, H. D. and Guilding, C., 2011. The neural circadian system of mammals. *Essays Biochem*, 49, 1–17.

Pimentel, D., Donlea, J. M., Talbot, C. B., Song, S. M., Thurston, A. J. F., and Miesenböck, G., 2016. Operation of a homeostatic sleep switch. *Nature*, 536 (7616), 333–337.

Pinggera, A., Lieb, A., Benedetti, B., Lampert, M., Monteleone, S., Liedl, K. R., Tuluc, P., and Striessnig, J., 2015. CACNA1D de novo mutations in autism spectrum disorders activate cav1.3 l-type calcium channels. *Biological Psychiatry*, 77, 816–822.

Pocklington, A.J., Rees, E., O'Donovan, M.C., and Owen, M.J., 2015. Novel findings from CNVs implicate inhibitory and excitatory signalling complexes in Schizophrenia. *Neuron*. 86, 1203 – 1214.

Potdar, S., Daniel, D.K., Thomas,F.A., Lall, S., and Sheeba, V., 2018. Sleep deprivation negatively impacts reproductive output in Drosophila melanogaster. *Journal of experimental Biology.* 221.

Pollock, J. A., Ellisman, M. H., and Benzer, S., 1990. Subcellular localization of transcripts in Drosophila photoreceptor neurons: chaoptic mutants have an aberrant distribution. *Genes and Development*, 4, 806–821.

Pratt, A. J. and MacRae, I. J., 2009. The RNA-induced silencing complex: A versatile gene-silencing machine. *Journal of Biological Chemistry*, 284 (27), 17897–17901.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., and Sham, P. C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559–575.

Raabe, F. J., Slapakova, L., Rossner, M. J., Cantuti-Castelvetri, L., Simons, M., Falkai, P. G., and Schmitt, A., 2019. Oligodendrocytes as A New Therapeutic Target in SZ: From Histopathological Findings to Neuron-Oligodendrocyte Interaction. *Cells*, 8 (12), 1496.

Ramanathan, L., Gulyani, S., Nienhuis, R., and Siegel, J. M., 2002. Sleep deprivation decreases superoxide dismutase activity in rat hippocampus and brainstem. *NeuroReport*, 13 (11), 1387–1390.

Rana, T. M., 2007. Illuminating the silence: understanding the structure and function of small RNAs. *Nat.Rev.Mol.Cell Biol.*, 8 , 23–36.

Reiter, L. T., Potocki, L., Chien, S., Gribskov, M., and Bier, E., 2001. A Systematic Analysis of Human Disease-Associated Gene Sequences In. *Genome Research*, 1114–1125.

Renn, S. C. P., Park, J. H., Rosbash, M., Hall, J. C., and Taghert, P. H., 1999. A pdf neuropeptide gene mutation and ablation of PDF neurons each cause severe abnormalities of behavioral circadian rhythms in Drosophila. *Cell*, 99, 791–802.

Reyes, A., Melchionda, L., Nasca, A., Carrara, F., Lamantea, E., Zanolini, A., Lamperti, C., Fang, M., Zhang, J., Ronchi, D., Bonato, S., Fagiolari, G., Moggio, M., Ghezzi, D., and Zeviani, M., 2015. RNASEH1 Mutations Impair mtDNA Replication and Cause Adult-Onset Mitochondrial Encephalomyopathy. *American Journal of Human Genetics*, 97, 186–193.

Riemensperger, T. and Isabel, G., 2011. Behavioural consequences of dopamine deficiency in the Drosophila central nervous system. *PNAS.* 108 (2), 834-9.

Reis, A.S., Hermanns, T., Poeck, B., and Strauss, R., 2017. Serotonin modulates a depression-like state in Drosophila responsive to lithium treatment. *Nature communications.* 8 (15738).

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., Lin, D.-Y., Duan, J., Ophoff, R. A., Andreassen, O. A., Scolnick, E., Cichon, S., St. Clair, D., Corvin, A., Gurling, H., Werge, T., Rujescu, D., Blackwood, D. H. R., Pato, C. N., Malhotra, A. K., Purcell, S., Dudbridge, F., Neale, B. M., Rossin, L., Visscher, P. M., Posthuma, D., Ruderfer, D. M., Fanous, A., Stefansson, H., Steinberg, S., Mowry, B. J., Golimbet, V., De Hert, M., Jönsson, E. G., Bitter, I., Pietiläinen, O. P. H., Collier, D. A., Tosato, S., Agartz, I., Albus, M., Alexander, M., Amdur, R. L., Amin, F., Bass, N., Bergen, S. E., Black, D. W., Børglum, A. D., Brown, M. A., Bruggeman, R., Buccola, N. G., Byerley, W. F., Cahn, W., Cantor, R. M., Carr, V. J., Catts, S. V, Choudhury, K., Cloninger, C. R., Cormican, P., Craddock, N., Danoy, P. A., Datta, S., de Haan, L., Demontis, D., Dikeos, D., Djurovic, S., Donnelly, P., Donohoe, G., Duong, L., Dwyer, S., Fink-Jensen, A., Freedman, R., Freimer, N. B., Friedl, M., Georgieva, L., Giegling, I., Gill, M., Glenthøj, B., Godard, S., Hamshere, M., Hansen, M., Hansen, T., Hartmann, A. M., Henskens, F. A., Hougaard, D. M., Hultman, C. M., Ingason, A., Jablensky, A. V, Jakobsen, K. D., Jay, M., Jürgens, G., Kahn, R. S., Keller, M. C., Kenis, G., Kenny, E., Kim, Y., Kirov, G. K., Konnerth, H., Konte, B., Krabbendam, L., Krasucki, R., Lasseter, V. K., Laurent, C., Lawrence, J., Lencz, T., Lerer, F. B., Liang, K.-Y., Lichtenstein, P., Lieberman, J. A., Linszen, D. H., Lönnqvist, J., Loughland, C. M., Maclean, A. W., Maher, B. S., Maier, W., Mallet, J., Malloy, P., Mattheisen, M., Mattingsdal, M., McGhee, K. A., McGrath, J. J., McIntosh, A., McLean, D. E., McQuillin, A., Melle, I., Michie, P. T., Milanova, V., Morris, D. W., Mors, O., Mortensen, P. B., Moskvina, V., Muglia, P., Myin-Germeys, I., Nertney, D. A., Nestadt, G., Nielsen, J., Nikolov, I., Nordentoft, M., Norton, N., Nöthen, M. M., O'Dushlaine, C. T., Olincy, A., Olsen, L., O'Neill, F. A., Ørntoft, T. F., Owen, M. J., Pantelis, C., Papadimitriou, G., Pato, M. T., Peltonen, L., Petursson, H., Pickard, B., Pimm, J., Pulver, A. E., Puri, V., Quested, D., Quinn, E. M., Rasmussen, H. B., Réthelyi, J. M., Ribble, R., Rietschel, M., Riley, B. P., Ruggeri, M., Schall, U., Schulze, T. G., Schwab, S. G., Scott, R. J., Shi, J., Sigurdsson, E., Silverman, J. M., Spencer, C. C. A., Stefansson, K., Strange, A., Strengman, E., Stroup, T. S., Suvisaari, J., Terenius, L., Thirumalai, S., Thygesen, J. H., Timm, S., Toncheva, D., van den Oord, E., van Os, J., van Winkel, R., Veldink, J., Walsh, D., Wang, A. G., Wiersma, D., Wildenauer, D. B., Williams, H. J., Williams, N. M., Wormley, B., Zammit, S., Sullivan, P. F., O'Donovan, M. C., Daly, M. J., and Gejman, P. V, 2011. Genome-wide association study identifies five new SZ loci. *Nature Genetics*, 43 (10), 969–976.

Rizos, E., Siafakas, N., Skourti, E., Papageorgiou C, Tsoporis J, Parker TH, Christodoulou D.I., Spandidos, D.A., Katsantoni, E., and Zoumpourlis, V., 2016. miRNAs and their role in the correlation between SZ and cancer Review. *Mol Med Rep*. 14, 4942-4946.

Roberts, J., Scott, A. C., Howard, M. R., Breen, G., Bubb, V. J., Klenova, E., and Quinn, J. P., 2007. Differential regulation of the serotonin transporter gene by lithium is mediated by transcription factors, CCCTC binding protein and Y-box binding protein 1, through the polymorphic intron 2 variable number tandem repeat. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 27 (11), 2793–2801.

Robertson, C. W., 1936. The metamorphosis of Drosophila melanogaster, including an accurately timed account of the principal morphological changes. *Journal of Morphology*, 59, 351–399.

Rodriguez-Berriguete, G., Granata, G., Puliyadi, R., Tiwana, G., Prevo, R., Wilson, R. S., Yu, S., Buffa, F., Humphrey, T. C., McKenna, W. G., and Higgins, G. S., 2018. Nucleoporin 54 contributes to homologous recombination repair and post-replicative DNA integrity. *Nucleic Acids Research*, 46 (15), 7731–7746.

Rohde, M., Daugaard, M., Jensen, M. H., Helin, K., Nylandsted, J., and Jäättelä, M., 2005. Members of the heat-shock protein 70 family promote cancer cell growth by distinct mechanisms. *Genes and Development*, 19, 570–582.

Rose, K. M. and Lorenz, R., 2010. Sleep disturbances in dementia: What they are and what to do. *Journal of Gerontological Nursing*, 36, 9–14.

Roux, A. F., Nguyen, T. t. V., Squire, J. A., and Cox, D. W., 1994. A heat shock gene at 14q22: Mapping and expression. *Human Molecular Genetics*, 3 (10), 1819–1822.

Rubio-Gozalbo, M. E., Bakker, J. A., Waterham, H. R., and Wanders, R. J. A., 2004. Carnitine-acylcarnitine translocase deficiency, clinical, biochemical and genetic aspects. *Molecular Aspects of Medicine*, 25 (5–6), 521–532.

Ruby, J. G., Stark, A., Johnston, W. K., Kellis, M., Bartel, D. P., and Lai, E. C., 2007. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Research*, 17 (12), 1850–1864.

Rzezniczak, T. Z., Douglas, L. A., Watterson, J. H., and Merritt, T. J. S., 2011. Paraquat administration in Drosophila for use in metabolic studies of oxidative stress. *Analytical Biochemistry*, 419, 345–347.

Sato, F., Tsuchiya, S., Meltzer, S.J., Shimizu, K., 2011. MicroRNAs and epigenetics. 278(10), 1598-609.

Sample, P. A., Boden, C., Zhang, Z., Pascual, J., Lee, T., Zangwill, L. M., Weinreb, R. N., Crowston, J. G., Hoffmann, E. M., Medeiros, F. A., Sejnowski, T., and Goldbaum, M., 2007. *Methods*, 46 (10), 3684–3692.

Saraiva, C., Esteves, M., and Bernardino, L., 2017. MicroRNA: Basic concepts and implications for regeneration and repair of neurodegenerative diseases. *Biochemical Pharmacology*, 141, 118–131.

Sarnthein, J., Morel, A., Von Stein, A., and Jeanmonod, D., 2005. Thalamocortical theta coherence in neurological patients at rest and during a working memory task. *International Journal of Psychophysiology*, 57, 87–96.

Sarnthein, J. and Jeanmonod, D., 2007. High thalamocortical theta coherence in patients with Parkinson's disease. *Journal of Neuroscience*, 27, 124–131.

Savage, P. a, 2009. Recognition of a Ubiquitous Self. *Evolution*, 215, 215–220.

Savelyev, S. A., Rantamäki, T., Rytkönen, K. M., Castren, E., and Porkka-Heiskanen, T., 2012. Sleep homeostasis and depression: Studies with the rat clomipramine model of depression. *Neuroscience*, 212, 149–158.

Schnorrer, F., Schönbauer, C., Langer, C. C. H., Dietzl, G., Novatchkova, M., Schernhuber, K., Fellner, M., Azaryan, A., Radolf, M., Stark, A., Keleman, K., and Dickson, B. J., 2010. Systematic genetic analysis of muscle morphogenesis and function in Drosophila. *Nature*, 464 (7286), 287–291.

Scholl, U. I., Goh, G., Stölting, G., de Oliveira, R. C., Choi, M., Overton, J. D., Fonseca, A. L., Korah, R., Starker, L. F., Kunstman, J. W., Prasad, M. L., Hartung, E. a, Mauras, N., Benson, M. R., Brady, T., Shapiro, J. R., Loring, E., Nelson-Williams, C., Libutti, S. K., Mane, S., Hellman, P., Westin, G., Åkerström, G., Björklund, P., Carling, T., Fahlke, C., Hidalgo, P., and Lifton, R. P., 2013. Somatic and germline CACNA1D calcium channel mutations in aldosterone-producing adenomas and primary aldosteronism. *Nature genetics*, 45 , 1050–4.

Sehgal, A. and Mignot, E., 2011. Genetics of Sleep and Sleep Disorders. *Cell*, 146, 194–207.

Sekine, Y., Hatanaka, R., Watanabe, T., Sono, N., Iemura, S. ichiro, Natsume, T., Kuranaga, E., Miura, M., Takeda, K., and Ichijo, H., 2012. The Kelch Repeat Protein KLHDC10 Regulates Oxidative Stress-Induced ASK1 Activation by Suppressing PP5. *Molecular Cell*, 48 , 692–704.

Shah, V. O., Dorin, R. I., Sun, Y., Braun, M., and Zager, P. G., 1997. Aldose reductase gene expression is increased in diabetic nephropathy. *Journal of Clinical Endocrinology and Metabolism*, 82, 2294–2298.

Shaltiel-Karyo, R., Davidi, D., Menuchin, Y., Frenkel-Pinter, M., Marcus-Kalish, M., Ringo, J., Gazit, E., and Segal, D., 2012. A novel, sensitive assay for behavioral defects in Parkinsons disease model drosophila. *Parkinson's Disease*, 2012.

Shaw, P. J., Cirelli, C., Greenspan, R. J., Tononi, G., Campbell, S. S., Tobler, I., Zepelin, H., Rechtschaffen, A., Rechtschaffen, A., Tobler, I., ___., Stalder, J., Tobler, I., Neuner-Jehle, M., Kaiser, W., Steiner-Kaiser, J., Kaiser, W., Cirelli, C., Tononi, G., Hamblen, M., Mistlberger, R. E., Bergmann, B. M., Waldenar, W., Rechtschaffen, A., Tobler, I., Borbely, A. A., Groos, G., Edgar, D. M., Dement, W. C., Fuller, C. A., Stone, W. S., Dijk, D.-J., Duffy, J. F., Riel, E., Shanahan, T. L., Czeisler, C. A., Yanik, G., Glaum, S., Radulovacki, M., Aronstein, K., Auld, V., Ffrench-Constant, R., Yehuda, S., Dunkov, B. C., Rodriguez-Arnaiz, R., Pittendrigh, B., Ffrench-Constant, R. H., Feyereisen, R., Kuhl, D., Kennedy, T. E., Barzilai, A., Kandel, E., Rubin, D. M., Brodbeck, D., McGinty, D. J., Harper, R. M., Aston-Jones, G., Bloom, F. E., Siegel, J. M., Rogawksi, M. A., and Hendricks, J., 2000. Correlates of sleep and waking in Drosophila melanogaster. *Science (New York, N.Y.)*, 287 (5459), 1834–7.

Shaw, P. J., Tortoni, G., Greenspan, R. J., and Robinson, D. F., 2002. Stress response genes protect against lethal effects of sleep deprivation in Drosophila. *Nature*, 417 (6886), 287–291.

Shi, C., Zhang, K., Wang, X., Shen, Y., and Xu, Q., 2012. A study of the combined effects of the EHD3 and FREM3 genes in patients with major depressive disorder. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*, 159 B, 336–342.

Shu, Z., Huang, Y., Palmer, W.H., Tamori, Y., Xie, G., Wang, H., Liu, N., and Deng, W.M., 2017. Systematic analysis reveals tumor-enhancing and -suppressing microRNAs in Drosophila epithelial tumors. *Oncotarget*. 8 (65), 108825 – 108839.

Sidore, C., Busonero, F., Maschio, A., Porcu, E., Naitza, S., and 35 others, 2015. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nature genetics*. 47 (11), 1272 – 1281.

Singh, T., Kurki, M.I., Curtis, D., Purcell, S.M., Crooks, L., and 48 others, 2016. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nature neuroscience.* 19 (4), 571 – 577.

Singh, T., Walters, J.T.R., Johnstone, M., Curtis, D., Suvisaari, J., and 19 others, 2017. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nature genetics.* Online pre-publication.

Singh, K. and Singh, R. N., 1999. Metamorphosis of the central nervous system of Drosophila melanogaster Meigen (Diptera: Drosophilidae) during pupation. *Journal of Biosciences*, 24, 345–360.

Sitaraman, D., Aso, Y., Rubin, G. M., and Nitabach, M. N., 2015. Control of Sleep by Dopaminergic Inputs to the Drosophila Mushroom Body. *Frontiers in neural circuits*, 9 (November), 73.

Smith, D. J., Nicholl, B. I., Cullen, B., Martin, D., Ul-Haq, Z., Evans, J., Gill, J. M. R., Roberts, B., Gallacher, J., Mackay, D., Hotopf, M., Deary, I., Craddock, N., and Pell, J. P., 2013. Prevalence and characteristics of probable major depression and bipolar disorder within UK Biobank: Cross-sectional study of 172,751 participants. *PLoS ONE*, 8 (11), 1–7.

Smoller, J. W., Craddock, N., Kendler, K., Lee, P. H., and Neale, B. M., 2013. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*, 381 (9875), 1371–1379.

Smrt, R. D., Szulwach, K. E., Pfeiffer, R. L., Li, X., Guo, W., Pathania, M., Teng, Z., Luo, Y., Peng, J., and Bordey, A., 2011. NIH Public Access. *Mind*, 28, 1060–1070.

Sofola, O., Kerr, F., Rogers, I., Killick, R., Augustin, H., Gandy, C., Allen, M.J., Hardy, J., Lovestone, S., and Partridge, L., 2010. Inhibition of GSK-3 Ameliorates Ab Pathology in an Adult-Onset Drosophila Model of Alzheimer's Disease. *PLOS Genetics.* 6 (9).

Solovev, I., Dobrovolskaya, E., Shaposhnikov, M., Sheptyakov, M., and Moskalev, A., 2019. Neuron-specific overexpression of core clock genes improves stress-resistance and extends lifespan of Drosophila melanogaster. *Experimental Gerontology*, 117, 61–71.

Somers, K. J., Lennon, V. A., Rundell, J. R., Pittock, S. J., Drubach, D. A., Trenerry, M. R., Lachance, D. H., Klein, C. J., Aston, P. A., and McKeon, A., 2011. Psychiatric manifestations of voltage-gated potassium-channel complex autoimmunity. *Journal of Neuropsychiatry and Clinical Neurosciences*, 23, 425–433.

Spiegel, K., Knutson, K., Leproult, R., Tasali, E., and Cauter, E. Van, 2008. Sleep loss: a novel risk factor for insulin resistance and Type 2 diabetes. *J Appl Physiol*, 99 (54), 2008–2019.

Splawski, I., Timothy, K. W., Sharpe, L. M., Decher, N., Kumar, P., Bloise, R., Napolitano, C., Schwartz, P. J., Joseph, R. M., Condouris, K., Tager-Flusberg, H., Priori, S. G., Sanguinetti, M. C., and Keating, M. T., 2004. CaV1.2 calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell*, 119, 19–31.

Spring, A. M., Brusich, D. J., and Frank, C. A., 2016. C-terminal Src Kinase Gates Homeostatic Synaptic Plasticity and Regulates Fasciclin II Expression at the Drosophila Neuromuscular Junction. *PLoS Genetics*, 12, 1–31.

St Johnston, D., 2002. The art and design of genetic screens: Drosophila melanogaster. *Nature Reviews Genetics*, 3, 176–188.

Steeg, P. S., Kopper, L., Thorgeirsson, U. P., Talmadge, E., Liotta, L. A., Sobep, M. E., Boissan, M., Wendum, D., Arnaud-Dabernat, S., Munier, A., Debray, M., Lascu, I., Daniel, J.-Y., Lacombe, M.-L., Dorsam, R. T., Gutkind, J. S., O'Hayre, M., Vázquez-Prado, J., Kufareva, I., Stawiski, E. W., Handel, T. M., Seshagiri, S., Gutkind, J. S., Kramer, M. F., Domcke, S., Sinha, R., Levine, D. a, Sander, C., Schultz, N., Sheet, P. D., Promoter, C. M. V. I. E., Ie, C. M. V, Nhei, P., Chen, C., Ridzon, D. a, Broomer, A. J., Zhou, Z., Lee, D. H., Nguyen, J. T., Barbisin, M., Xu, N. L., Mahuvakar, V. R., Andersen, M. R., Lao, K. Q., Livak, K. J., Guegler, K. J., Biosystems, A., Kayali, S., Giraud, G., Morlé, F., Guyot, B., Czimmerer, Z., Hulvely, J., Simandi, Z., Varallyay, E., Havelda, Z., Szabo, E., Varga, A., Dezso, B., Balogh, M., Horvath, A., Domokos, B., Torok, Z., Nagy, L., Balint, B. L., Carman, C. V., Parent, J.-L., Day, P. W., Pronin, a. N., Sternweis, P. M., Wedegaertner, P. B., Gilman, a. G., Benovic, J. L., Kozasa, T., Galés, C., Van Durm, J. J. J., Schaak, S., Pontier, S., Percherancier, Y., Audet, M., Paris, H., Bouvier, M., Lyon, A. M., Dutta, S., Boguth, C. a, Skiniotis, G., Tesmer, J. J. G., Gaugler, L., Raamsdonk, C. D. Van, Bezrookove, V., Green, G., Brien, J. M. O., Simpson, E. M., Barsh, G. S., Bastian, B. C., Two, T., Hub, L. I., Sheets, F., Li, C., Eiymo Mwa Mpollo, M.-S., Gonsalves, C. S., Tahara, S. M., Malik, P., Kalra, V. K., Huang, J., Sun, Y., Zhang, J. J., Huang, X.-Y., Kramer, M. F., Rasheed, S. A. K., Teo, C. R., Beillard, E. J., Voorhoeve, P. M., Casey, P. J., Verma, P., Augustine, G. J., Tominaga, N., Kosaka, N., Ono, M., Katsuda, T., Yoshioka, Y., Tamura, K., Lötvall, J., Nakagama, H., Ochiya, T., Wilkie, T. M., Scherle, P. a, Strathmann, M. P., Slepak, V. Z., Simon, M. I., Park, D., Jhon, D. Y., Lee, C. W. H., Lee, K. H., Rhee, S. G., Pedone, K. H., Hepler, J. R., Park, D., Jhon, D. Y., Kriz, R., Knopf, J., Rhee, S. G., Wu, D., Katz, a, Simon, M. I., Cheung, T. H., Rando, T. a, Conrad, K. L., Tseng, K. Y., Uejima, J. L., Reimers, J. M., Heng, L.-J., Shaham, Y., Marinelli, M., Wolf, M. E., Lijnen, H. R., Christiaens, V., and Scroyen, L., 2013. Stem-loop RT-qPCR for miRNAs. *The Journal of biological chemistry*, 7, 355–62.

Stefana, M. I., Driscoll, P. C., Obata, F., Pengelly, A. R., Newell, C. L., MacRae, J. I., and Gould, A. P., 2017. Developmental diet regulates Drosophila lifespan via lipid autotoxins. *Nature Communications*, 8 .

Stessman, H. A. F., Willemsen, M. H., Fenckova, M., Penn, O., Hoischen, A., Xiong, B., Wang, T., Hoekzema, K., Vives, L., Vogel, I., Brunner, H. G., Van Der Burgt, I., Ockeloen, C. W., Schuurs-Hoeijmakers, J. H., Klein Wassink-Ruiter, J. S., Stumpel, C., Stevens, S. J. C., Vles, H. S., Marcelis, C. M., Van Bokhoven, H., Cantagrel, V., Colleaux, L., Nicouleau, M., Lyonnet, S., Bernier, R. A., Gerdts, J., Coe, B. P., Romano, C., Alberti, A., Grillo, L., Scuderi, C., Nordenskj??ld, M., Kvarnung, M., Guo, H., Xia, K., Piton, A., Gerard, B., Genevieve, D., Delobel, B., Lehalle, D., Perrin, L., Prieur, F., Thevenon, J., Gecz, J., Shaw, M., Pfundt, R., Keren, B., Jacquette, A., Schenck, A., Eichler, E. E., and Kleefstra, T., 2016. Disruption of POGZ Is Associated with Intellectual Disability and Autism Spectrum Disorders. *American Journal of Human Genetics*, 98, 541–552.

Strauss, R. and Heisenberg, M., 1993. A higher control center of locomotor behavior in the Drosophila brain. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 13, 1852–1861.

Striessnig, J., Pinggera, A., Kaur, G., Bock, G., and Tuluc, P., 2014. L-type Ca2+ channels in heart and brain. *Wiley Interdisciplinary Reviews: Membrane Transport and Signaling*, 3, 15–38.

Stuart, J. M., Segal, E., Koller, D., and Kim, S. K., 2003. A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 302 (5643), 249–255.

Sun, J., Folk, D., Bradley, T. J., and Tower, J., 2002. Induced overexpression of mitochondrial Mn-superoxide dismutase extends the life span of adult Drosophila melanogaster. *Genetics*, 161, 661–672.

Sun, Y. J., Yu, Y., Zhu, G. C., Sun, Z. H., Xu, J., Cao, J. H., and Ge, J. X., 2015. Association between single nucleotide polymorphisms in MiR219-1 and MiR137 and susceptibility to SZ in a Chinese population. *FEBS Open Bio*, 5, 774–778.

Szmeja, Z., 2008. Recenzja podręcznika pt. „CT Teaching Manual. A Systematic Approach to CT Reading". *Otolaryngologia Polska*, 62, 117.

Tamim, S., Vo, D.T., Uren, P.J., Qiao, M., Bindewald, E., Kasprzak WK, Shapiro BA, Nakaya, H.I., Burns, S.C., Araujo, P.R., Nakano I, Radek AJ, Kuersten S, Smith AD, and Penalva, L.O., 2019. Genomic analyses reveal broad impact of miR-137 on genes associated with malignant transformation and neuronal differentiation in glioblastoma cells. PLoS One. 22, 9.

Technau, G. and Heisenberg, M., 1982. Neural reorganization during metamorphosis of the corpora pedunculata in Drosophila melanogaster. *Nature*.

Thimgan, M. S., Seugnet, L., Turk, J., and Shaw, P. J., 2015. Identification of Genes Associated with Resilience/Vulnerability to Sleep Deprivation and Starvation in Drosophila. *Sleep*, 38, 801–814.

Thomas KT, Gross C, Bassell GJ., 2018. microRNAs Sculpt Neuronal Communication in a Tight Balance That Is Lost in Neurological Disease. *Front Mol Neurosci.* 12 (11), 455.

Thomson, D. W., Bracken, C. P., and Goodall, G. J., 2011. Experimental strategies for microRNA target identification. *Nucleic Acids Research*, 39 (16), 6845–6853.

Tomita, J., Ueno, T., Mitsuyoshi, M., Kume, S., and Kume, K., 2015. The NMDA Receptor Promotes Sleep in the Fruit Fly, Drosophila melanogaster. *PloS one*, 10, e0128101.

Tooney, P. A., 2016. Attention: SZ Risk Gene Product miR-137 Now Targeting EFNB2. *EBioMedicine*, 12, 10–11.

Toulopoulou, T., van Haren, N., Zhang X, Sham PC, Cherny SS, Campbell DD, Picchioni M, Murray, R., Boomsma, D.I., Hulshoff Pol, H.E., Brouwer, R., Schnack, H., Fañanás, L., Sauer, H., Nenadic, I., Weisbrod, M., Cannon, T.D., and Kahn RS., 2015. Reciprocal causation models of cognitive vs volumetric cerebral intermediate phenotypes for SZ in a pan-European twin cohort. *Mol Psychiatry*. 20(11), 1386-96.

Truman, J. W., 1990. Metamorphosis of the central nervous system of Drosophila. *Journal of Neurobiology*, 21, 1072–1084.

Tucholski, J., Simmons, M. S., Pinner, A. L., Haroutunian, V., McCullumsmith, R. E., and Meador-Woodruff, J. H., 2013. Abnormal N-linked glycosylation of cortical AMPA receptor subunits in SZ. *SZ Research*, 146 (1–3), 177–183.

Tully, T., Cambiazo, V., and Kruse, L., 1994. Memory through metamorphosis in normal and mutant Drosophila. *Journal of Neuroscience*, 14, 68–74.

Um, S. M., Ha, S., Lee, H., Kim, J., Kim, K., Shin, W., Cho, Y. S., Roh, J. D., Kang, J., Yoo, T., Noh, Y. W., Choi, Y., Bae, Y. C., and Kim, E., 2018. NGL-2 Deletion Leads to Autistic-like Behaviors Responsive to NMDAR Modulation. *Cell Reports*, 23 (13), 3839–3851.

Vassos, E., Di Forti, M., and Coleman, J., 2017. An examination of polygenic score risk prediction in individuals with first-episode psychosis. Biol Psychiatry. 81:470–477.

Vallès, A., Martens, G. J. M., De Weerd, P., Poelmans, G., and Aschrafi, A., 2014. MicroRNA-137 regulates a glucocorticoid receptor-dependent signalling network: Implications for the etiology of SZ. *Journal of Psychiatry and Neuroscience*, 39, 312–320.

Van Swinderen, B. and Andretic, R., 2011. Dopamine in Drosophila: setting arousal thresholds in a miniature brain. *Proceedings of the Royal Society B: Biological Sciences*, 278 (1707), 906–913.

Veverytsa, L. and Allan, D. W., 2013. Subtype-specific neuronal remodeling during drosophila metamorphosis. *Fly*, 7.

Vogel, B. O., Lett, T. A., Erk, S., Mohnke, S., Wackerhagen, C., Brandl, E. J., Romanczuk-Seiferth, N., Otto, K., Schweiger, J. I., Tost, H., Nöthen, M. M., Rietschel, M.,

Degenhardt, F., Witt, S. H., Meyer-Lindenberg, A., Heinz, A., and Walter, H., 2018. The influence of MIR137 on white matter fractional anisotropy and cortical surface area in individuals with familial risk for psychosis. *SZ Research*, 195, 190–196.

Wang, J. W. and Wu, C., 1996. In Vivo Functional Role of the Drosophila Hyperkinetic Gating and Inactivation of Shaker K + Channels Subunit in. *Biophysical Journal*, 71 , 3167–3176.

Wang, K., Li, M., and Hakonarson, H., 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38 (16), 1–7.

Wang, Y., Billon, C., Walker, J. K., and Burris, T. P., 2015. Therapeutic Effect of a Synthetic RORα/γ Agonist in an Animal Model of Autism. *ACS Chemical Neuroscience*, acschemneuro.5b00159.

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S., Montojo, J., Shao, Q., Wright, G., Bader, G. D., and Morris, Q., 2010. The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*, 38 (SUPPL. 2), 214–220.

Weinberger, D.R., 1987. Implications of normal brain development for the pathogenesis of SZ. *Arch Gen Psychiatry*. 44, 660-9.

Welch, J, W., 1991. The role of heat-shock proteins as molecular chaperones. *Current Opinion in Cell Biology*, 3, 1033–1038.

Whitmore, L. M. and Smith, T. C., 2019. Isolating the Association of Sleep, Depressive State, and Other Independent Indicators for Suicide Ideation in United States Teenagers. *Archives of Suicide Research*, 23 , 471–490.

Willemsen, M. H., Vallès, A., Kirkels, L. a M. H., Mastebroek, M., Olde Loohuis, N., Kos, A., Wissink-Lindhout, W. M., de Brouwer, A. P. M., Nillesen, W. M., Pfundt, R., Holder-Espinasse, M., Vallée, L., Andrieux, J., Coppens-Hofman, M. C., Rensen, H., Hamel, B. C. J., van Bokhoven, H., Aschrafi, A., and Kleefstra, T., 2011. Chromosome 1p21.3 microdeletions comprising DPYD and MIR137 are associated with intellectual disability. *Journal of medical genetics*, 48 (12), 810–8.

Wise, C.D., and Stein, L., 1973. Dopamine-beta-hydroxylase deficits in the brains of schizophrenic patients. Science. 181(4097), 344-7.

Woo, J., Kwon, S. K., and Kim, E., 2009. The NGL family of leucine-rich repeat-containing synaptic adhesion molecules. *Molecular and Cellular Neuroscience*, 42, 1–10.

Wright, C., Calhoun, V. D., Ehrlich, S., Wang, L., Turner, J. A., and Perrone-Bizzozero, N. I., 2015. Meta gene set enrichment analyses link miR-137-regulated pathways with SZ risk. *Frontiers in Genetics*, 6 (MAR), 1–10.

Wright, C., Turner, J. A., Calhoun, V. D., and Perrone-Bizzozero, N., 2013. Potential Impact of miR-137 and Its Targets in SZ. *Frontiers in Genetics*, 4 (April), 1–14.

Wrzosek, M., Łukaszkiewicz, J., Wrzosek, M., Serafin, P., Jakubczyk, A., Klimkiewicz, A., Matsumoto, H., Brower, K. J., and Wojnar, M., 2011. Association of polymorphisms in HTR2A, HTR1A and TPH2 genes with suicide attempts in alcohol dependence: A preliminary report. *Psychiatry Research*, 190, 149–151.

Wu, G., Feng, X., and Stein, L., 2010. A human functional protein interaction network and its application to cancer data analysis. *Genome Biology*, 11.

Wu, M. N., Ho, K., Crocker, A., Yue, Z., and Koh, K., 2010. Activity, but not the Adenosine Receptor. *J Neurosci*, 29 (35), 11029–11037.

Wu, M. N., Koh, K., Yue, Z., Joiner, W. J., and Sehgal, A., 2008. A genetic screen for sleep and circadian mutants reveals mechanisms underlying regulation of sleep in Drosophila. *Sleep*, 31, 465–472.

Xiao, C., Calado, D. P., Galler, G., Thai, T. H., Patterson, H. C., Wang, J., Rajewsky, N., Bender, T. P., and Rajewsky, K., 2007. MiR-150 Controls B Cell Differentiation by Targeting the Transcription Factor c-Myb. *Cell*, 131, 146–159.

Xu, P., Vernooy, S. Y., Gou, M. and Hay, B. A., 2003. Molecular evolution of enolase. *Current Biology* 13, 790–795.

Xue, Y., abd Zhang, Y., 2018. Emerging roles for microrna in the regulation of the Drosophila circadian clock. *Neuroscience*. 19 (1).

Yan, H.-M., Hu, H., Ahmed, A., Feng, B.-B., Liu, J., Jia, Z.-J., and Wang, H., 2017. Carnitine-acylcarnitine translocase deficiency with c.199-10 T>G and novel c.1A>G mutation: Two case reports and brief literature review. *Medicine*, 96 (45), e8549.

Yang, J., Lui, J. C., Vedantam, S., and Gustafsson, S., 2015. Using Predicted Gene Functions.

Yin, J., Lin, J., Luo, X., Chen, Y., Li, Z., Ma, G., and Li, K., 2014. MiR-137: A new player in SZ. *International Journal of Molecular Sciences*, 15, 3262–3271.

Yuan, Q., Joiner, W. J., and Sehgal, A., 2006. A Sleep-Promoting Role for the Drosophila Serotonin Receptor 1A. *Current Biology*, 16 (11), 1051–1062.

Zamore PD, Haley B. Ribo-gnome: the big world of small RNAs. Science. 2005 Sep 2;309(5740):1519-24.

Zeng, M., Zhu, L., Li, L., and Kang, C., 2017. miR-378 suppresses the proliferation, migration and invasion of colon cancer cells by inhibiting SDAD1. *Cellular and Molecular Biology Letters*, 22, 1–13.

Zhang, W., Reeves, G. R., and Tautz, D., 2019. Identification of a genetic network for an ecologically relevant behavioural phenotype in Drosophila melanogaster. *Molecular Ecology*, (December 2019), 1–17.

Zheng, W., Feng, G., and Ren, D., 1995. Cloning and characterization of a calcium channel alpha 1 subunit from Drosophila melanogaster with similarity to the rat brain type D isoform. *Journal of neuroscience.* 15(2), 1132-43.

Zheng F, Gallagher JP., 1995. Pharmacologically distinct, pertussis toxin-resistant inward currents evoked by metabotropic glutamate receptor (mGluR) agonists in dorsolateral septal nucleus (DLSN) neurons. *Journal of neuroscience*. 15, 504-10.

Zhong, Y., Budnik, V., and Wu, C. F., 1992. Synaptic plasticity in Drosophila memory and hyperexcitable mutants: role of cAMP cascade. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 12, 644–651.

Zhou, J., Xu, L., Duan, X., Liu, W., Zhao, X., Wang, X., Shang, W., Fang, X., Yang, H., Jia, L., Bai, J., Zhao, J., Wang, L., and Tong, C., 2019. Large-scale RNAi screen identified Dhpr as a regulator of mitochondrial morphology and tissue homeostasis. *Science Advances*, 5, 1–16.

Zhu, X., Need, A.C., Petrovski, S., and Goldstein, D.B., 2014. One gene, many neuropsychiatric disorders: lessons from mendelian diseases. *Nature neuroscience*. 17, 773 – 781.

Zhuo, C. J., Hou, W. H., Jiang, D. G., Tian, H. J., Wang, L. N., Jia, F., Zhou, C. H., and Zhu, J. J., 2020. Circular RNAs in early brain development and their influence and clinical significance in neuropsychiatric disorders. *Neural Regeneration Research*, 15, 817–823.

Zimmerman, J. E., Chan, M. T., Jackson, N., Maislin, G., and Pack, A. I., 2012. Genetic background has a major impact on differences in sleep resulting from environmental influences in Drosophila. *Sleep*, 35, 545–57.

12 Supplementary information

12.1 Targetscan miR-137 target comparison

Table: The predicted targets of human and *Drosophila melanogaster miR-137* were downloaded from Targetscan.org, and the human targets were converted into *Drosophila* orthologs using DIOPT. These orthologs were then checked against the fly miR-137 targets to create a list of 278 genes which were predicted to be directly regulated by *miR-137* in both species.

Human gene	fly	Human gene	fly gene	Human gene	fly gene	Human gene	fly gene
ABCC9	MRP	FLI1	Ets21C	PAQR3	CG33203	SLC22A23	CG3690
ABCG4	Atet	FNBP1L	nwk	PAQR3	AdipoR	SLC24A2	Nckx30C
ADRB1	Octbeta3R	GABBR2	GABA-B-R3	PDE10A	Pde6	SLC24A3	Nckx30C
AGO1	AGO1	GABBR2	GABA-B-R1	PDE1B	Pde1c	SLC25A5	sesB
AGO3	AGO1	GAN	kel	PDE4A	dnc	SLC46A3	CG8008
AGO4	AGO1	GBP2	atl	PDE4A	Pde6	SLC4A7	CG8177
AKT2	Akt1	GLRA3	Rdl	PDE4D	dnc	SMURF1	Su(dx)
AKT2	Pkc53E	GPR22	SIFaR	PDE4D	Pde6	SMURF2	Su(dx)
AKT2	S6k	HCN2	Ih	PDE7A	Pde6	SNX27	CG32758
ANKRD44	Tnks	HIPK2	Hipk	PHACTR4	CG32264	SOCS6	Socs16D
ANO4	CG10353	HLF	Pdp1	PHF20	upSET	SOX11	Sox100B
ANO8	CG10353	HNRNPDL	sqd	PITPNA	vib	SPATA13	RhoGEF3
ARF6	Arf51F	HNRNPDL	Rbp6	PKN2	Pkc53E	SRC	Csk
BCL11B	CG9650	HNRNPDL	msi	PLCB1	Plc21C	SRGAP3	nwk
BHLHE22	Oli	IGLON5	DIP-theta	PLEKHB2	CG11000	SSBP2	Ssdp
BICRAL	CG11873	IGLON5	CG13506	PLEKHG3	GEFmeso	SSBP3	Ssdp
BRD1	rno	IL1RAPL1	CG13506	PLPPR4	CG11426	STK3	hpo
CACFD1	fwe	IL1RAPL1	CG45263	PLXDC2	l(1)G0289	STRBP	Zn72D
CACNA1C	cac	JADE1	rno	PPP1CB	flw	STRN	Cka
CACNA1C	para	JADE2	rno	PPP1CB	Pp1alpha	STYX	CG7378
CACNA1C	NaCP60E	JDP2	kay	PPP1R16B	MYPT-75D	SUZ12	Su(z)12
CACNA1C	Ca-alpha1T	KCNAB3	Hk	PPP6R2	fmt	SV2A	CG3690
CACNA1D	cac	KCNC1	Shaw	PPP6R3	fmt	SYN3	Syn
CACNA1D	para	KCNC2	Shaw	PRKAR2B	Pka-R2	SYNCRIP	Syp
CACNA1D	NaCP60E	KCNC3	Shaw	PROX1	pros	SYNJ1	CG6805
CACNA1D	Ca-alpha1T	KCNH1	eag	PTPN2	Ptp61F	TBC1D12	CG5916
CACNA1G	Ca-alpha1T	KDM1A	Su(var)3-3	PTPRD	Ptp36E	TBC1D12	GAPsec
CACNA1G	cac	KLF12	CG42741	PTRH2	CG1307	TBC1D12	CG42795
CACNA1H	Ca-alpha1T	KLHL3	kel	RAB8A	Rab10	TBX15	bi
CACNA1H	cac	LCOR	Eip93F	RAD54L2	CG4049	TBX18	bi
CACNA1I	Ca-alpha1T	LIFR	Dscam1	RASIP1	cno	TBX3	bi
CACNA1I	cac	LMTK2	InR	RASIP1	didum	TCF12	da
CADM2	dpr6	LRP12	arr	RBM24	Rbp6	TCF3	da
CADM2	beat-VII	LRP6	arr	RBM24	msi	TCF4	da
CCNG2	CycE	LRRN3	Fili	RBM24	sqd	TCHP	CG17230
CCNI	CycE	LRRN3	caps	RGS7BP	CG34351	TEAD1	sd
CCNJ	CycE	LVRN	CG8773	RHOBTB1	RhoBTB	TEF	Pdp1
CD2AP	cindr	MAF	tj	RNF150	mura	TENT4A	Trf4-1
CD69	rgn	MAFK	tj	RNF152	mei-P26	TMED5	opm
CHD1	Chd1	MAN2A1	alpha-Man	RNF152	wech	TMEM132B	dtn
CHRM2	mAChR-A	MAP2K4	Mkk4	RNF38	mura	TMSB15B	cib
CHRM3	mAChR-A	MAX	Max	RNF4	elfless	TPCN1	para
CHST1	CG31637	MBD5	sba	RORA	Hr3	TPCN1	cac
CISD1	Cisd2	MBD6	sba	RORB	Hr3	TPCN1	Ca-alpha1T
CNOT6L	twin	MEGF11	CG7381	RREB1	peb	TPCN1	NaCP60E
CPLX1	cpx	MSI1	Rbp6	RYR2	RyR	TRIM3	wech
CREB5	kay	MSI1	sqd	RYR3	RyR	TRIM3	mei-P26

CUL3	Cul5	MSI1	msi	SASH3	SKIP	TRPS1	GATAd
CUL4A	Cul5	MSI2	Rbp6	SASH3	CG34377	TTC28	pins
DAG1	Dg	MSI2	msi	SCN1A	para	UBE2H	UbcE2H
DENND4B	Crag	MSI2	sqd	SCN1A	NaCP60E	WWP2	Su(dx)
DENND5B	Crag	MXD1	Max	SCN2A	para	YTHDF3	Ythdf
DFFB	Drep4	MXD1	Myc	SCN2A	NaCP60E	ZBTB18	CG12605
DNMT3A	sba	MYO1B	Myo95E	SCN8A	para	ZBTB21	erm
DSCAML1	Dscam1	MYO1C	Myo95E	SCN8A	NaCP60E	ZBTB34	ab
DUSP1	CG7378	MYO1D	Myo95E	SCN8A	cac	ZBTB37	ab
DUSP10	CG7378	MYO1F	Myo95E	SCN8A	Ca-alpha1T	ZBTB4	CG10274
DUSP16	CG7378	NACC2	erm	SCRT1	scrt	ZBTB4	Cf2
DUSP4	CG7378	NBEAL2	CG43367	SCRT1	CG12605	ZBTB7B	erm
DUSP5	CG7378	NEFH	Lam	SCRT2	scrt	ZCCHC2	CG10492
DUSP8	CG7378	NEGR1	DIP-theta	SCRT2	CG12605	ZNF148	CG12773
DYRK1A	Hipk	NEGR1	CG13506	SEMA4D	Sema2a	ZNF217	CG9650
EIF1	eIF1	NEUROD1	Oli	SEMA4D	Sema2b	ZNF449	CG32772
EIF4E3	eIF4EHP	NEUROD2	Oli	SEMA4D	Sema1a	ZNF496	CG32772
ELAVL3	fne	NEUROD4	Oli	SGPL1	Sply	ZNF516	CG9650
ELAVL3	Rbp9	NLGN4X	alpha-Est8	SLC16A10	CG13907	ZNF76	CG32772
ELFN2	Fili	OPCML	DIP-theta	SLC16A10	CG8034	ZNF770	CG4374
ERBIN	Lap1	OPCML	CG13506	SLC17A6	VGlut	ZSWIM4	CG34401
ERBIN	scrib	OSBPL8	CG42668	SLC22A18	Glut1	-	-
ERG	Ets21C	PAPLN	nolo	SLC22A23	CG44098	-	-

12.2   T-tests

12.2.1   Deficiency Crosses

24-hour	w1118	ko/+	w1118	mirko
Mean	816.7205	966.25	816.7205	1240.403
Variance	15247.25	10435.42	15247.25	30034.66
Observations	322	32	322	77
Hypothesized Mean Difference	0		0	
df	352		95	
t Stat	-6.62602		-20.2579	
P(T<=t) one-tail	6.47E-11		1.5E-36	
t Critical one-tail	1.649194		1.661052	
	sig		sig	

24-hour	w1118	w/df	w1118	ko/df
Mean	816.7205	996.3333	816.7205	1214.742
Variance	15247.25	5972	15247.25	2527.331
Observations	322	18	322	31
Hypothesized Mean Difference	0		0	
df	22		73	
t Stat	-9.22452		-35.0603	
P(T<=t) one-tail	2.57E-09		1.01E-47	
t Critical one-tail	1.717144		1.665996	
	sig		sig	

24-hour	mirko	ko/df	ko/+	w/df
Mean	1240.403	1214.742	966.25	996.3333
Variance	30034.66	2527.331	10435.42	5972
Observations	77	31	32	18
Hypothesized Mean Difference	0		0	
df	100		48	
t Stat	1.181644		-1.08509	
P(T<=t) one-tail	0.120075		0.14165	
t Critical one-tail	1.660234		1.677224	
	not sig		not sig	

Day	w1118	ko/+	w1118	mirko
Mean	316.3602	413.75	316.3602	602.9091
Variance	6061.988	5992.258	6061.988	11290.14
Observations	322	32	322	77
Hypothesized Mean Difference	0		0	
df	352		96	
t Stat	-6.75192		-22.2774	
P(T<=t) one-tail	3.02E-11		5.17E-40	
t Critical one-tail	1.649194		1.660881	
	sig		sig	

Day	w1118	w/df	w1118	ko/df
Mean	316.3602	482.6111	316.3602	595.3226
Variance	6061.988	3539.428	6061.988	1030.292
Observations	322	18	322	31
Hypothesized Mean Difference	0		0	
df	338		71	
t Stat	-8.90993		-38.6623	
P(T<=t) one-tail	1.61E-17		9.78E-50	
t Critical one-tail	1.649374		1.6666	
	sig		sig	

Day	mirko	ko/df	ko/+	w/df
Mean	602.9091	595.3226	413.75	482.6111
Variance	11290.14	1030.292	5992.258	3539.428
Observations	77	31	32	18
Hypothesized Mean Difference	0		0	
df	101		48	
t Stat	0.565685		-3.26524	
P(T<=t) one-tail	0.286431		0.001011	
t Critical one-tail	1.660081		1.677224	
	not sig		sig	

Night	w1118	ko/+	w1118	mirko
Mean	500.3602	552.5	500.3602	637.4935
Variance	5179.789	3527.806	5179.789	5911.622
Observations	322	32	322	77
Hypothesized Mean Difference	0		0	
df	352		397	
t Stat	-3.96461		-14.821	
P(T<=t) one-tail	4.45E-05		3.62E-40	
t Critical one-tail	1.649194		1.648701	
	sig		sig	

Night	w1118	w/df	w1118	ko/df
Mean	500.3602	513.7222	500.3602	619.4194
Variance	5179.789	1783.154	5179.789	1109.318
Observations	322	18	322	31
Hypothesized Mean Difference	0		0	
df	23		62	
t Stat	-1.2452		-16.5311	
P(T<=t) one-tail	0.112799		1.05E-24	
t Critical one-tail	1.713872		1.669804	
	not sig		sig	

Night	mirko	ko/df	ko/+	w/df
Mean	637.4935	619.4194	552.5	513.7222
Variance	5911.622	1109.318	3527.806	1783.154
Observations	77	31	32	18
Hypothesized Mean Difference	0		0	
df	105		48	
t Stat	1.703602		2.439886	
P(T<=t) one-tail	0.045707		0.009218	
t Critical one-tail	1.659495		1.677224	
	not sig		sig	

12.2.2   Gal4/UAS sponge knockdown

24-hour	w1118	elav/sponge	w1118	fb/sponge
Mean	816.7205	1151.533	816.7205	1037
Variance	15247.25	8476.436	15247.25	7224.667
Observations	322	45	322	16
Hypothesized Mean Difference	0		0	
df	68		18	
t Stat	-21.8075		-9.86212	
P(T<=t) one-tail	1.04E-32		5.53E-09	
t Critical one-tail	1.667572		1.734064	
	sig		sig	

24-hour	w1118	mb/sponge	w1118	mirsponge/+
Mean	816.7205	1035.067	816.7205	811.5625
Variance	15247.25	17133.35	15247.25	10372.8
Observations	322	15	322	16
Hypothesized Mean Difference	0		0	
df	335		336	
t Stat	-6.67711		0.164262	
P(T<=t) one-tail	5.07E-11		0.434812	
t Critical one-tail	1.649415		1.649401	
	sig		not sig	

24-hour	w1118	elav/+	w1118	fb/+
Mean	816.7205	762.0667	816.7205	811.4375
Variance	15247.25	9427.495	15247.25	8819.196
Observations	322	15	322	16
Hypothesized Mean Difference	0		0	
df	335		336	
t Stat	1.689177		0.168632	
P(T<=t) one-tail	0.046058		0.433094	
t Critical one-tail	1.649415		1.649401	
	sig		not sig	

24-hour	w1118	mb/+	mirko	elav/sponge
Mean	816.7205	725.0625	1240.403	1151.533
Variance	15247.25	13659.93	30034.66	8476.436
Observations	322	16	77	45
Hypothesized Mean Difference	0		0	
df	336		119	
t Stat	2.904796		3.695113	
P(T<=t) one-tail	0.001959		0.000167	
t Critical one-tail	1.649401		1.657759	
	sig		sig	

24-hour	mirko	fb/sponge	mirko	mb/sponge
Mean	1240.403	1037	1240.403	1035.067
Variance	30034.66	7224.667	30034.66	17133.35
Observations	77	16	77	15
Hypothesized Mean Difference	0		0	
df	45		90	
t Stat	7.011373		4.345778	
P(T<=t) one-tail	4.9E-09		1.82E-05	
t Critical one-tail	1.679427		1.661961	
	sig		sig	

Day	w1118	elav/sponge	w1118	fb/sponge
Mean	316.3602	516.8	316.3602	441.125
Variance	6061.988	5535.118	6061.988	2868.25
Observations	322	45	322	16
Hypothesized Mean Difference	0		0	
df	365		18	
t Stat	-16.2616		-8.86459	
P(T<=t) one-tail	2.07E-45		2.76E-08	
t Critical one-tail	1.649039		1.734064	
	sig		sig	

Day	w1118	mb/sponge	w1118	mirsponge/+
Mean	316.3602	470	316.3602	344.1875
Variance	6061.988	10007.29	6061.988	2978.829
Observations	322	15	322	16
Hypothesized Mean Difference	0		0	
df	335		336	
t Stat	-7.37102		-1.4115	
P(T<=t) one-tail	6.68E-13		0.079512	
t Critical one-tail	1.649415		1.649401	
	sig		not sig	

Day	w1118	elav/+	w1118	fb/+
Mean	316.3602	356.5333	316.3602	319.5
Variance	6061.988	1490.552	6061.988	2956
Observations	322	15	322	16
Hypothesized Mean Difference	0		0	
df	20		336	
t Stat	-3.69516		-0.15927	
P(T<=t) one-tail	0.000717		0.436775	
t Critical one-tail	1.724718		1.649401	
	sig		not sig	

Day	w1118	mb/+	mirko	elav/sponge
Mean	316.3602	286.0625	602.9091	516.8
Variance	6061.988	4890.863	11290.14	5535.118
Observations	322	16	77	45
Hypothesized Mean Difference	0		0	
df	336		116	
t Stat	1.525855		5.244048	
P(T<=t) one-tail	0.063993		3.58E-07	
t Critical one-tail	1.649401		1.658096	
	not sig		sig	

Day	mirko	fb/sponge	mirko	mb/sponge
Mean	602.9091	441.125	602.9091	470
Variance	11290.14	2868.25	11290.14	10007.29
Observations	77	16	77	15
Hypothesized Mean Difference	0		0	
df	44		90	
t Stat	8.961894		4.47172	
P(T<=t) one-tail	8.77E-12		1.13E-05	
t Critical one-tail	1.68023		1.661961	
	sig		sig	

Night	w1118	elav/sponge	w1118	fb/sponge
Mean	500.3602	634.7333	500.3602	595.875
Variance	5179.789	2745.109	5179.789	3093.717
Observations	322	45	322	16
Hypothesized Mean Difference	0		0	
df	365		336	
t Stat	-12.0788		-5.22858	
P(T<=t) one-tail	8.63E-29		1.5E-07	
t Critical one-tail	1.649039		1.649401	
	sig		sig	

Night	w1118	mb/sponge	w1118	mirsponge/+
Mean	500.3602	565.0667	500.3602	467.375
Variance	5179.789	2449.781	5179.789	4828.383
Observations	322	15	322	16
Hypothesized Mean Difference	0		0	
df	335		336	
t Stat	-3.44181		1.792058	
P(T<=t) one-tail	0.000326		0.037012	
t Critical one-tail	1.649415		1.649401	
	sig		sig	

Night	w1118	elav/+	w1118	fb/+
Mean	500.3602	405.5333	500.3602	491.9375
Variance	5179.789	6659.124	5179.789	3754.196
Observations	322	15	322	16
Hypothesized Mean Difference	0		0	
df	335		336	
t Stat	4.958586		0.45974	
P(T<=t) one-tail	5.65E-07		0.323	
t Critical one-tail	1.649415		1.649401	
	sig		not sig	

Night	w1118	mb/+	mirko	elav/sponge
Mean	500.3602	439	637.4935	634.7333
Variance	5179.789	3790.267	5911.622	2745.109
Observations	322	16	77	45
Hypothesized Mean Difference	0		0	
df	336		117	
t Stat	3.348705		0.235152	
P(T<=t) one-tail	0.000452		0.407251	
t Critical one-tail	1.649401		1.657982	
	sig		not sig	

Night	mirko	fb/sponge	mirko	mb/sponge
Mean	637.4935	491.9375	637.4935	565.0667
Variance	5911.622	3754.196	5911.622	2449.781
Observations	77	16	77	15
Hypothesized Mean Difference	0		0	
df	91		29	
t Stat	7.107434		4.674226	
P(T<=t) one-tail	1.29E-10		3.13E-05	
t Critical one-tail	1.661771		1.699127	
	sig		sig	

12.2.3 Caffeine treatment

24-hour	w1118	w1118 + caff	mirko	mirko + caff
Mean	784.125	593.25	1257.625	1286.25
Variance	1907.267857	51062.21429	2003.125	2774.5
Observations	8	8	8	8
Pooled Variance	26484.74107		2388.8125	
Hypothesized Mean Difference	0		0	
df	14		14	
t Stat	2.345747056		-1.171344006	
P(T<=t) one-tail	0.017119902		0.130505482	
t Critical one-tail	1.761310136		1.761310136	
	sig		not sig	

Day	w1118	w1118 + caff	mirko	mirko + caff
Mean	251	215.125	588	643.25
Variance	1828	12659.55	1912.857	1896.786
Observations	8	8	8	8
Pooled Variance	7243.777		1904.821	
Hypothesized Mean Difference	0		0	
df	14		14	
t Stat	0.843023		-2.53183	
P(T<=t) one-tail	0.206694		0.011973	
t Critical one-tail	1.76131		1.76131	
	not sig		sig	

Night	w1118	w1118 _caff	mirko	mirko + caff
Mean	533.125	378.125	669.625	643
Variance	1651.268	21982.13	750.8393	909.4286
Observations	8	8	8	8
Pooled Variance	11816.7		830.1339	
Hypothesized Mean Difference	0		0	
df	14		14	
t Stat	2.851764		1.848185	
P(T<=t) one-tail	0.006403		0.042904	
t Critical one-tail	1.76131		1.76131	
	sig		sig	

12.2.4   LiCl treatment

24-hour	w1118	w1118 _licl	mirko	mirko + licl
Mean	944.8571429	792.875	1314.375	1279.25
Variance	1493.142857	4454.125	2880.553571	2811.928571
Observations	7	8	8	8
Pooled Variance	3087.517857		2846.241071	
Hypothesized Mean Difference	0		0	
df	13		14	
t Stat	5.284893117		1.316771712	
P(T<=t) one-tail	7.38353E-05		0.104531897	
t Critical one-tail	1.770933396		1.761310136	
	sig		not sig	

Day	w1118	w1118 _licl	mirko	mirko + licl
Mean	365.8571	304	658	638.75
Variance	727.1429	859.4286	1027.714	1997.071
Observations	7	8	8	8
Pooled Variance	798.3736		1512.393	
Hypothesized Mean Difference	0		0	
df	13		14	
t Stat	4.229949		0.989985	
P(T<=t) one-tail	0.000492		0.169492	
t Critical one-tail	1.770933		1.76131	
	sig		not sig	

Night	w1118	w1118 _licl	mirko	mirko + licl
Mean	579	488.875	656.375	640.5
Variance	866.6667	2746.696	552.8393	2434
Observations	7	8	8	8
Pooled Variance	1878.99		1493.42	
Hypothesized Mean Difference	0		0	
df	13		14	
t Stat	4.017273		0.821586	
P(T<=t) one-tail	0.000732		0.212542	
t Critical one-tail	1.770933		1.76131	
	sig		not sig	

12.2.5   3IY treatment

24-hour	w1118	w1118 + 3IY	mirko	mirko+3IY
Mean	841.5	821.625	1239.75	1105
Variance	19555.14286	27309.69643	2769.642857	17436.5
Observations	8	8	8	5
Pooled Variance	23432.41964		8103.045455	
Hypothesized Mean Difference	0		0	
df	14		11	
t Stat	0.259674058		2.625807931	
P(T<=t) one-tail	0.399447454		0.011791491	
t Critical one-tail	1.761310136		1.795884819	
	not sig		sig	

Day	w1118	w1118+3IY	mirko	mirko+3IY
Mean	321.5	308.375	647	544.4
Variance	7011.143	8532.839	656	9671.8
Observations	8	8	8	5
Pooled Variance	7771.991		3934.473	
Hypothesized Mean Difference	0		0	
df	14		11	
t Stat	0.297758		2.86921	
P(T<=t) one-tail	0.38513		0.007632	
t Critical one-tail	1.76131		1.795885	
	not sig		sig	

Night	w1118	w1118 + 3IY	mirko	mirko+3IY
Mean	520	513.25	592.75	560.6
Variance	7424.857	6413.643	2347.357	1246.3
Observations	8	8	8	5
Pooled Variance	6919.25		1946.973	
Hypothesized Mean Difference	0		0	
df	14		11	
t Stat	0.162295		1.278084	
P(T<=t) one-tail	0.436696		0.113766	
t Critical one-tail	1.76131		1.795885	
	not sig		not sig	

12.3 qPCR results and standard curves

12.3.1 qPCR attempt 1

Samples of 20 fly brains were dissected for w[1118], miR-137[KO], elav-Gal4/UAS-miR-137-sponge, Oregon-R, and elav-Gal4/UAS-miR-137 genotypes and subjected to RNA extraction, reverse transcription, and comparative qPCR as described in chapter 2.5.
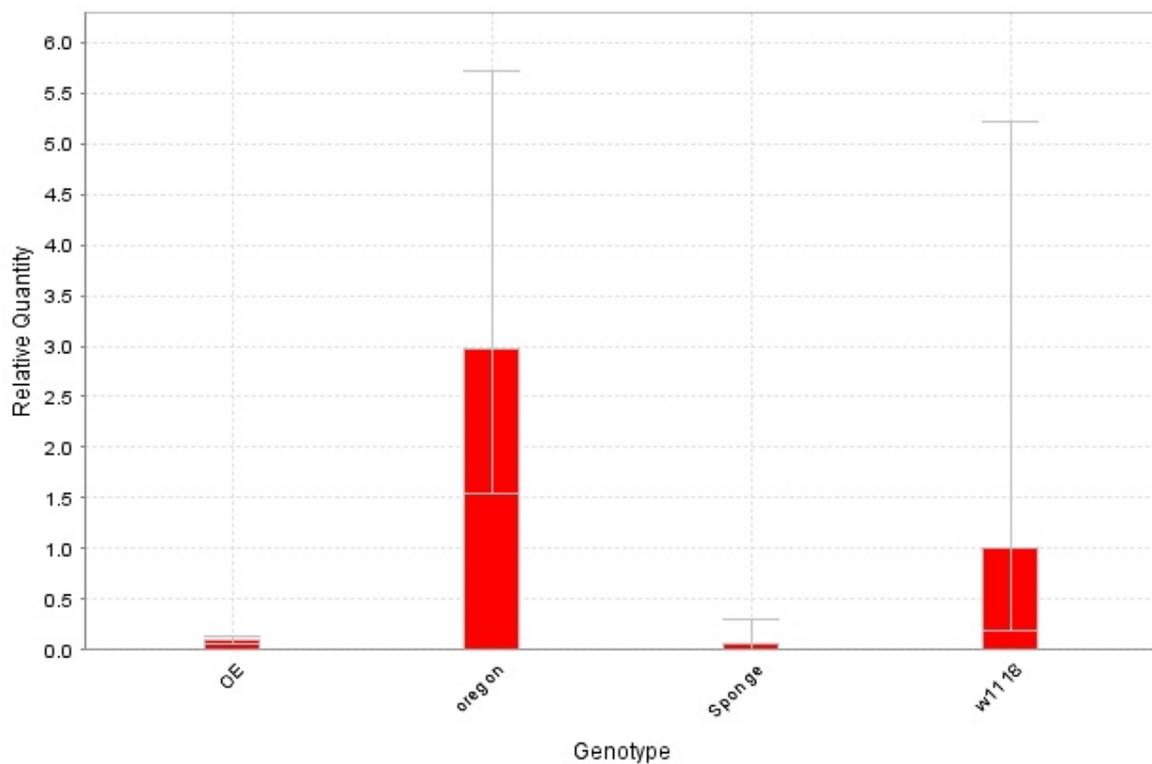
Figure: Comparative expression values of miR-137 in the over expression, Oregon-R, sponge knockdown, and w[1118] controls.

Values were calculated automatically by Applied biosystems Step-One plus software using standard curve dilution values of a housekeeping gene. No expression was found in the miR-137[KO] strain and as such is not displayed.
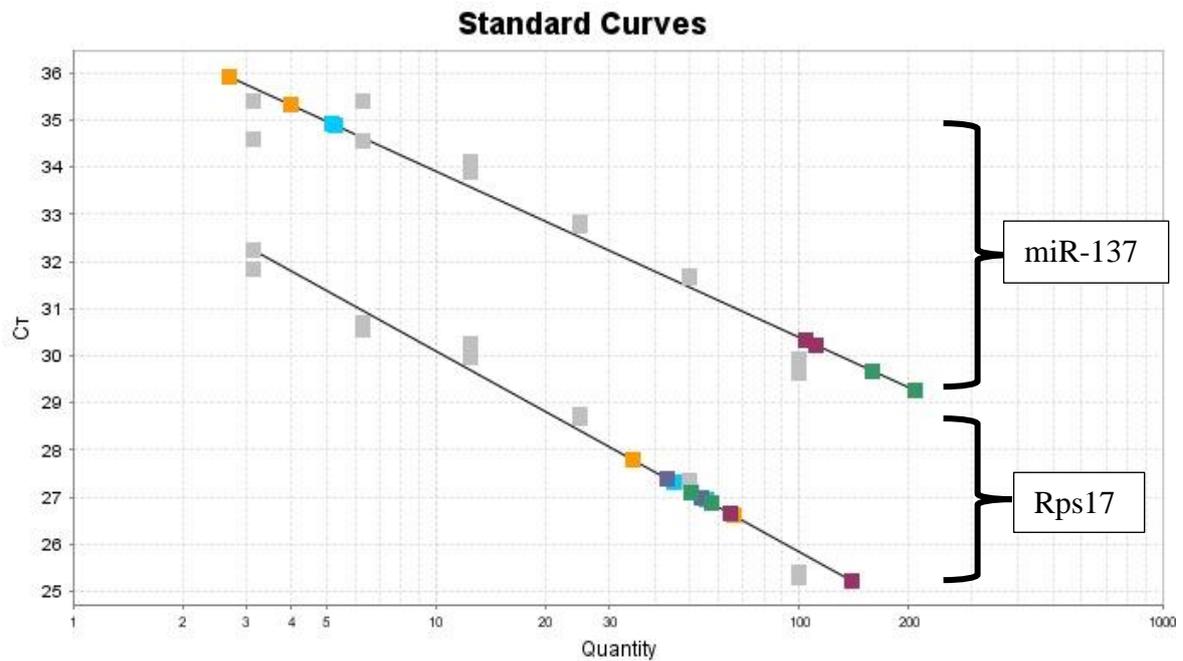
Figure: Standard curves from comparative *miR-137* qPCR

Oregon (red)*, w^{1118}* (dark blue)*, miR-137^{KO}* (green)*, elav-Gal4/UAS-miR-137-sponge* (light blue), and *elav-Gal4.UAS-miR-137* (yellow) *D. melanogaster* brains were dissected subjected to RNA extraction, reverse-transcription, and comparative qPCR to assess the change in expression levels. The cDNA samples were mixed together from each genotype and then dilutions in a ratio of 1:2 for 6 steps were used as a standard curve (grey). The upper line is automatically calculated from the dots for miR-137 expression, and the lower line is rps17 housekeeping gene control. The dots are the measured readings for the samples. CT is the "threshold cycle" value that is indicative of the number of cycles before fluorescence levels were above the threshold for reading, and Quantity is a comparative figure based on the standard curve dilution factors starting at "100" and decreasing in a ratio of 1:2.

12.3.2   qPCR attempt 2

Samples of 20 fly brains were dissected for w^{1118}, miR-137^{KO}, elav-Gal4/UAS-miR-137-sponge, Oregon-R, and elav-Gal4/UAS-miR-137 genotypes and subjected to RNA extraction, reverse transcription, and comparative qPCR as described in chapter 2.5.

Figure: Comparative expression values of miR-137 in the over expression, Oregon-R, sponge knockdown, and w¹¹¹⁸ controls.

Values were calculated automatically by Applied biosystems Step-One plus software using standard curve dilution values of a housekeeping gene. No expression was found in the miR-137^{KO} strain and as such is not displayed.
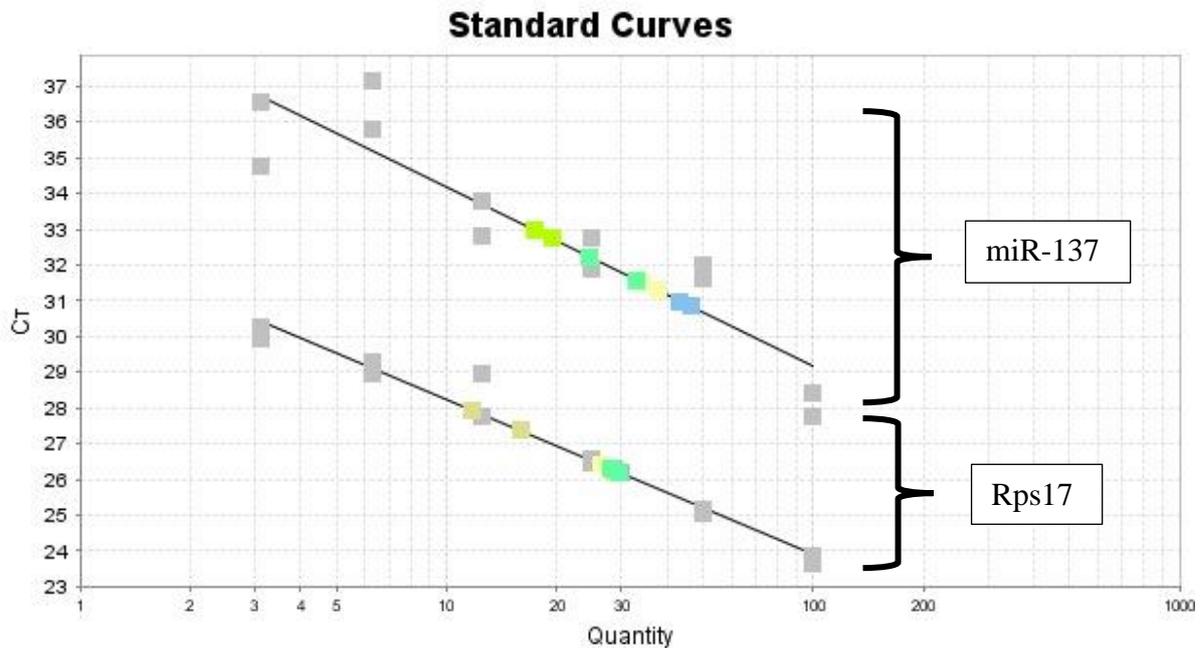
Standard Curves

Figure: Standard curves from comparative *miR-137* qPCR

Oregon (blue)*, w^1118^* (dark green)*, miR-137^KO^* (beige)*, elav-Gal4/UAS-miR-137-sponge* (light green)*, and elav-Gal4.UAS-miR-137* (yellow) *D. melanogaster* brains were dissected subjected to RNA extraction, reverse-transcription, and comparative qPCR to assess the change in expression levels. The cDNA samples were mixed together from each genotype and then dilutions in a ratio of 1:2 for 6 steps were used as a standard curve (grey). The upper line is automatically calculated from the dots for miR-137 expression, and the lower line is rps17 housekeeping gene control. The dots are the measured readings for the samples. CT is the "threshold cycle" value that is indicative of the number of cycles before fluorescence levels were above the threshold for reading, and Quantity is a comparative figure based on the standard curve dilution factors starting at "100" and decreasing in a ratio of 1:2.

12.3.3 qPCR attempt 3

Samples of 20 fly brains were dissected for w^1118^, miR-137^KO^, elav-Gal4/UAS-miR-137-sponge, presumed backcross KO, and heterozygous KO genotypes and subjected to RNA extraction, reverse transcription, and comparative qPCR as described in chapter 2.5.
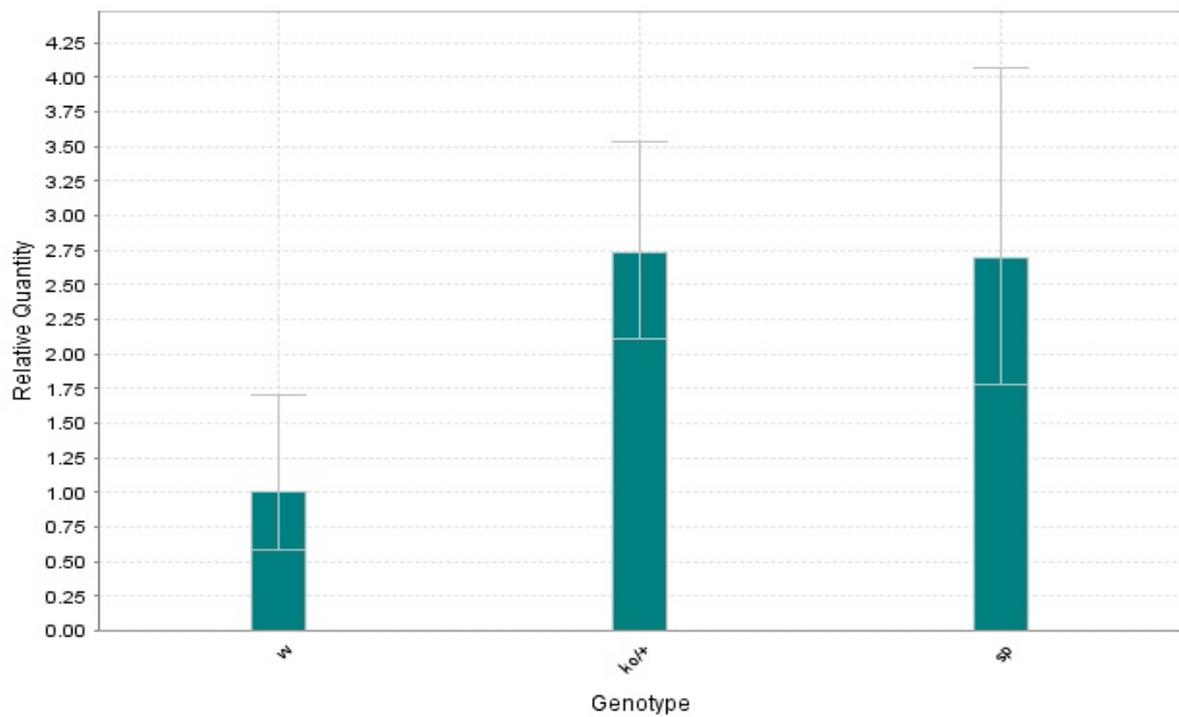
Figure: Comparative expression values of miR-137 in the presumed backcross KO, heterozygous KO, sponge knockdown, and w[1118] controls.

Values were calculated automatically by Applied biosystems Step-One plus software using standard curve dilution values of a housekeeping gene. No expression was found in the miR-137[KO] and backcross strains and as such are not displayed.
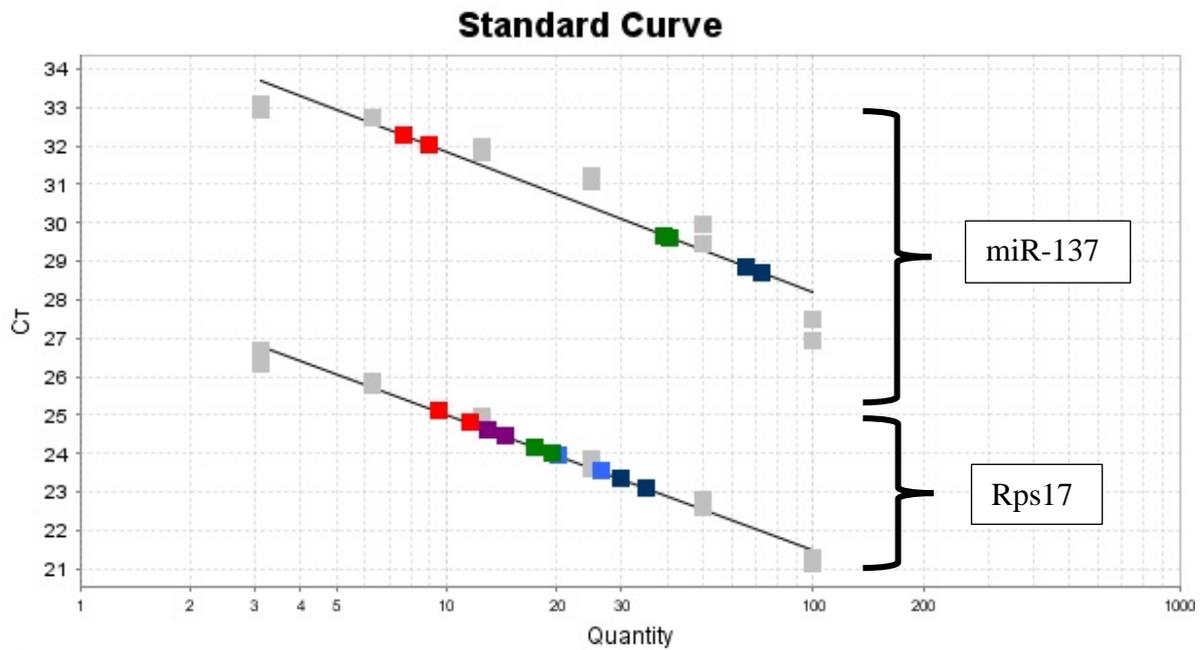
Figure: Standard curves from comparative *miR-137* qPCR

Presumed backcross KO (blue), *w^1118^* (red), *miR-137^KO^* (purple), *elav-Gal4/UAS-miR-137-sponge* (dark blue), and heterozygous KO (green) *D. melanogaster* brains were dissected subjected to RNA extraction, reverse-transcription, and comparative qPCR to assess the change in expression levels. The cDNA samples were mixed together from each genotype and then dilutions in a ratio of 1:2 for 6 steps were used as a standard curve (grey). The upper line is automatically calculated from the dots for miR-137 expression, and the lower line is rps17 housekeeping gene control. The dots are the measured readings for the samples. CT is the "threshold cycle" value that is indicative of the number of cycles before fluorescence levels were above the threshold for reading, and Quantity is a comparative figure based on the standard curve dilution factors starting at "100" and decreasing in a ratio of 1:2.

12.4   Manhatten Plots from SNP2GENE

20002_1262 Non-cancer related illness: Self-reported Parkinson's Disease

Parameter 20002_1262 refers to the initial UKbiobank assessment centre whereupon the participant was asked to disclose any history of medical disorders to a qualified nurse or doctor in a verbal interview. Participants who were unable to remember, or were uncertain of, the specific disorder were able to describe the symptoms, and this was later classified by the medical professional.

There were 974 people who reported a history of Parkinson's disease out of 383782 participants. This data was then cleaned and imputed by Neale *et al.* into 361194 samples, of which 652 cases against 360489 controls.
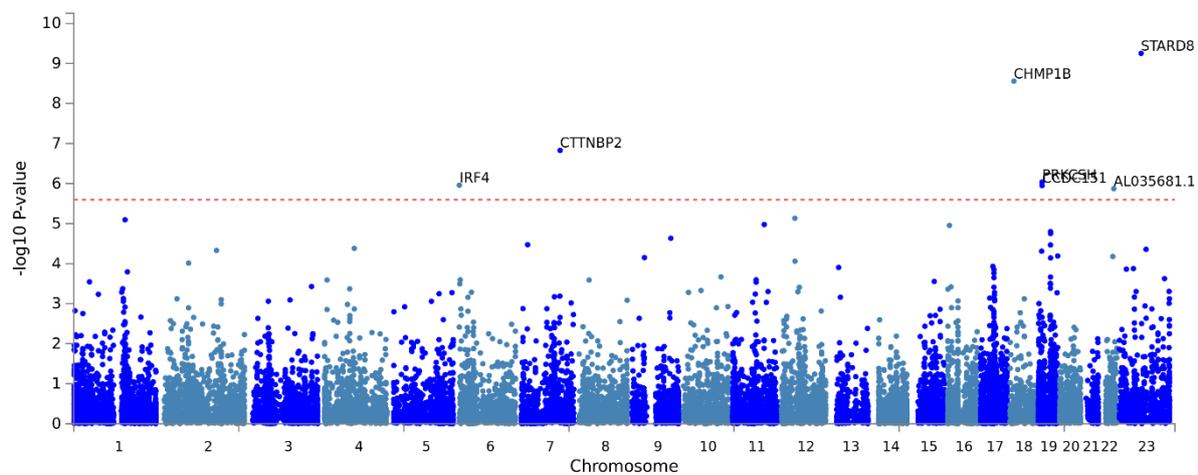
Figure: Manhattan plot for UKBiobank parameter 20002_1262 Non-cancer related illness self-reported Parkinson's Disease. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: STARD8, CHMP1B, CTTNBP2, IRF4, CCDC151, PRKCSH, and AL035681.1.

20002_1286 Non-cancer related illness: Self-reported Depression

Parameter 20002_1286 refers to the initial UKbiobank assessment centre whereupon the participant was asked to disclose any history of medical disorders to a qualified nurse or doctor in a verbal interview. Participants who were unable to remember, or were uncertain of, the specific disorder were able to describe the symptoms, and this was later classified by the medical professional.

There were 32965 people who reported a history of Depression out of 383782 participants. This data was then cleaned and imputed by Neale *et al.* into 361194 samples, of which 20648 cases against 340493 controls.
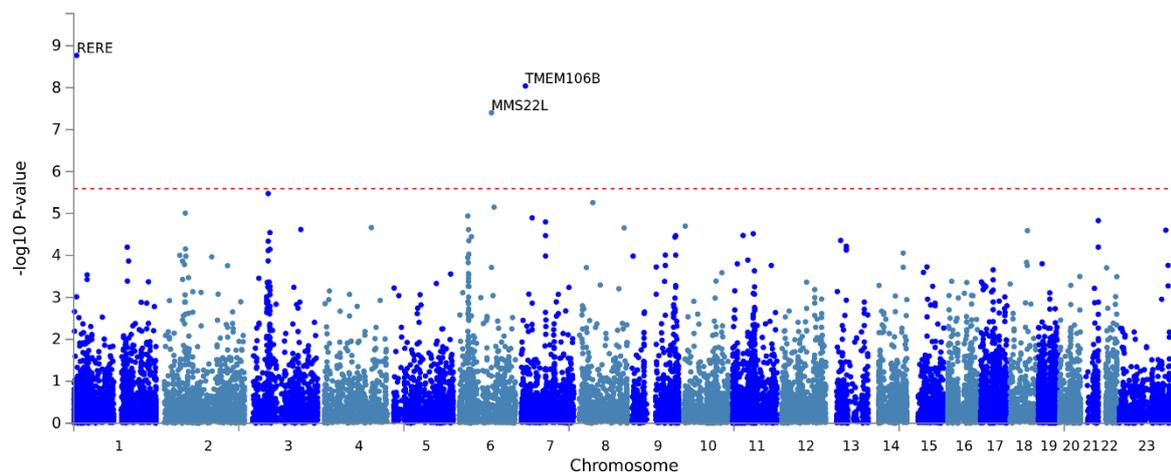
Figure: Manhattan plot for UKBiobank parameter 20002_1286 Non-cancer related illness self-reported Depression. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: RERE, TMEM106B, and MMS22L.

20002_1291 Non-cancer related illness: Self-reported Mania/Bipolar Disorder/Manic

Depression

Parameter 20002_1291 refers to the initial UKbiobank assessment centre whereupon the participant was asked to disclose any history of medical disorders to a qualified nurse or doctor in a verbal interview. Participants who were unable to remember, or were uncertain of, the specific disorder were able to describe the symptoms, and this was later classified by the medical professional.

There were 1530 people who reported a history of Mania/Bipolar Disorder/Manic Depression out of 383782 participants. This data was then cleaned and imputed by Neale *et al.* into 361194 samples, of which 1008 cases against 360133 controls.
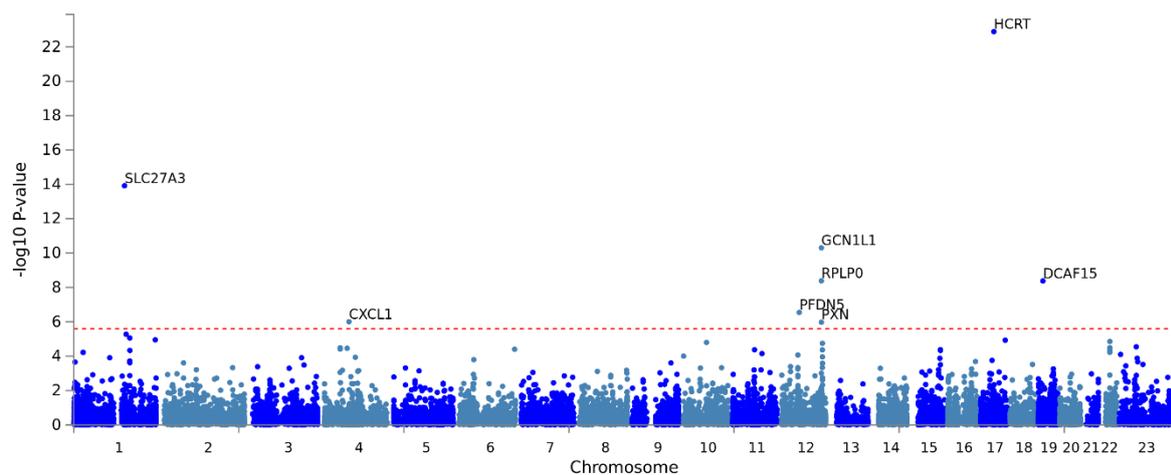
Figure: Manhattan plot for UKBiobank parameter 20002_1291 Non-cancer related illness self-reported Mania/Bipolar Disorder/Manic Depression. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: SLC27A3, CXCL1, PFDN5, GCN1L1, RPLP0, PXN, HCRT, and DCAF15.

20002_1482 Non-cancer related illness: Self-reported Chronic fatigue syndrome

Parameter 20002_1482 refers to the initial UKbiobank assessment centre whereupon the participant was asked to disclose any history of medical disorders to a qualified nurse or doctor in a verbal interview. Participants who were unable to remember, or were uncertain of, the specific disorder were able to describe the symptoms, and this was later classified by the medical professional.

There were 2397 people who reported a history of Chronic fatigue syndrome out of 383782 participants. This data was then cleaned and imputed by Neale *et al.* into 361194 samples, of which 1659 cases against 359482 controls.
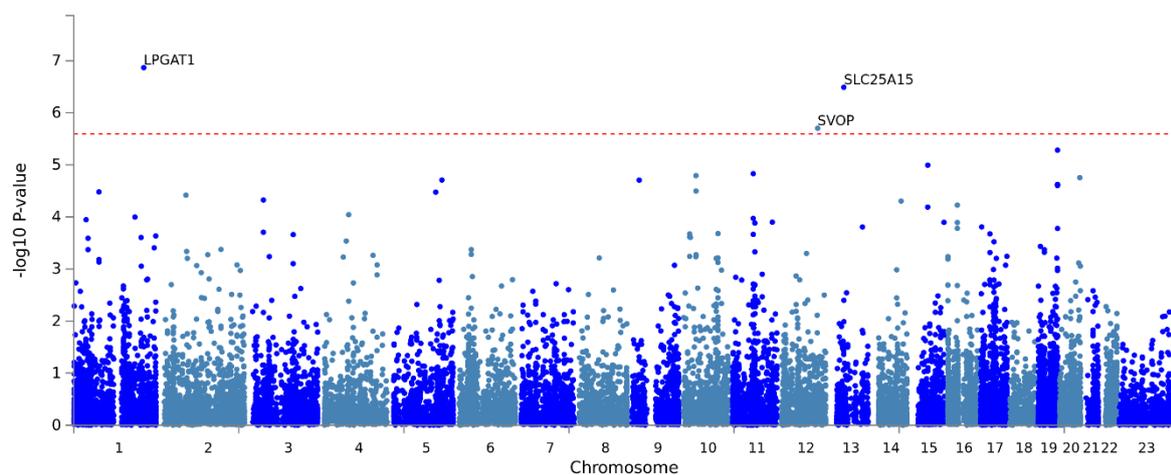
Figure: Manhattan plot for UKBiobank parameter 20002_1482 Non-cancer related illness self-reported Chronic fatigue syndrome. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: LPGAT1, SVOP, and SLC25A15.

20002_1614 Non-cancer related illness: Self-reported Stress

Parameter 20002_1614 refers to the initial UKbiobank assessment centre whereupon the participant was asked to disclose any history of medical disorders to a qualified nurse or

doctor in a verbal interview. Participants who were unable to remember, or were uncertain of, the specific disorder were able to describe the symptoms, and this was later classified by the medical professional.

There were 1083 people who reported a history of Stress out of 383782 participants. This data was then cleaned and imputed by Neale *et al.* into 361194 samples, of which 428 cases against 360713 controls.
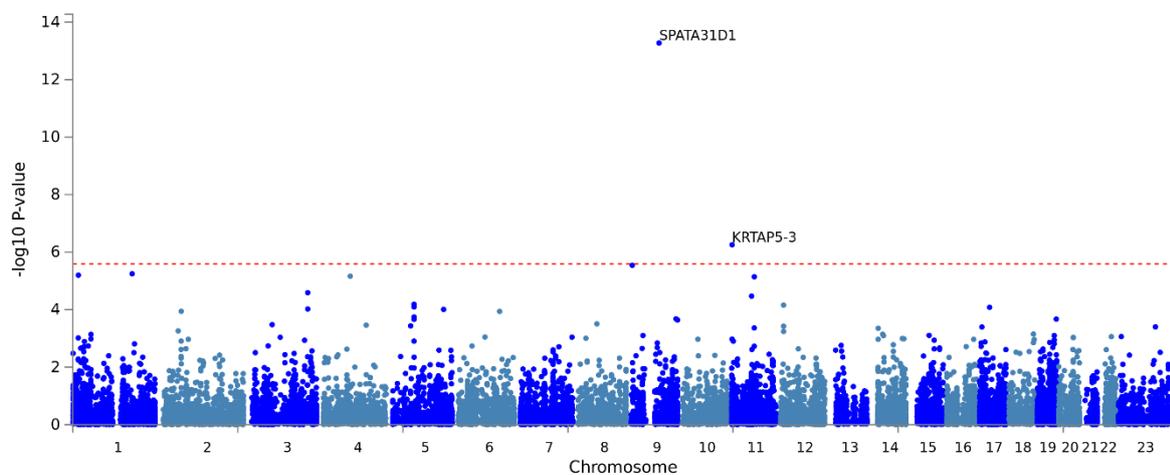
Figure: Manhattan plot for UKBiobank parameter 20002_1614 Non-cancer related illness self-reported Stress. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: SPATA31D1 and KRTAP5-3.

20003_1140867490 Treatment/Medication code: Lithium Product

Parameter 20003_1140867490 refers to the initial UKbiobank assessment centre whereupon the participant was asked to disclose any regular prescription medications to a qualified nurse in a verbal interview.

There were 520 people who reported regular use of a lithium-based medication out of 371433 participants. This data was then cleaned and imputed by Neale *et al.* into 361141 samples, of which 350 cases against 360791 controls.
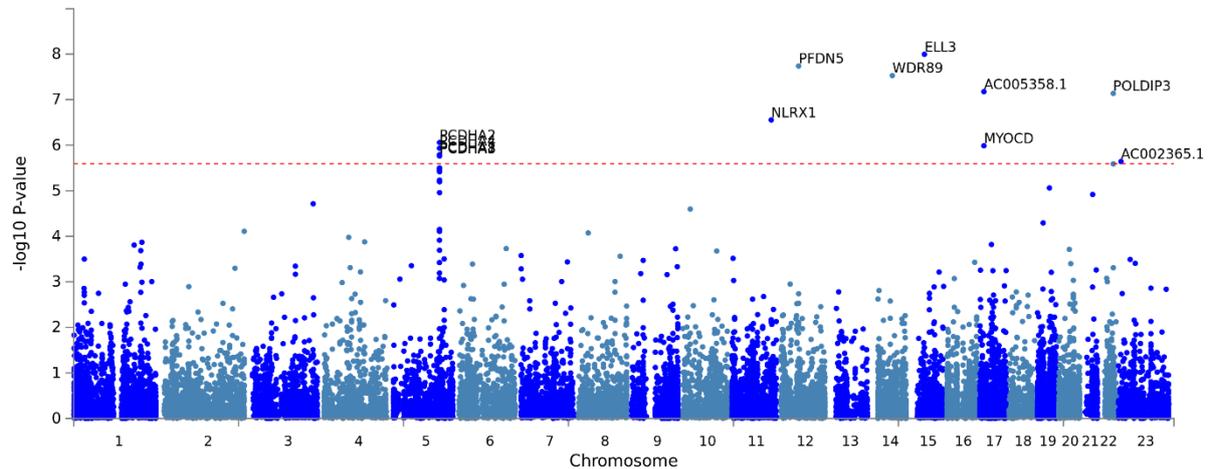
Figure: Manhattan plot for UKBiobank parameter 20003_1140867490 Treatment/Medication code: Lithium Product. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: PCDHA2, PCDHA3, PCDHA4, PCDHA8, NLRX1, PFDN5, WDR89, ELL3, MYOCD, AC005358.1, POLDIP3, and AC002365.1.

20126_1 Bipolar and major depression status: Bipolar I disorder

Parameter 20126_1 refers to a diagnosis by a mental health group headed by Professor Jill Pell from the Institute of Health & Wellbeing, University of Glasgow. The participants of the original UKBiobank survey were asked questions on a variety of general lifestyle factors, such as daily activity or average hours of sleep, and on previous feelings or moods, such as "Does your mood often go up and down?". The results were assessed and used to calculate probable diagnoses ranging from "single lifetime episode" to "probable disorder" as described in the paper by Smith *et al.* 2013.

Depressive and manic symptoms questionnaire results were collected for 172751 of the participants in the UKBiobank, of which 122982 were assessed by the mental health group

288

and found 808 people who were characterised as probable lifetime bipolar disorder I

sufferers. This data was then cleaned and imputed by Neale *et al.* into 86895 samples, of

which 556 cases against 86339 controls.
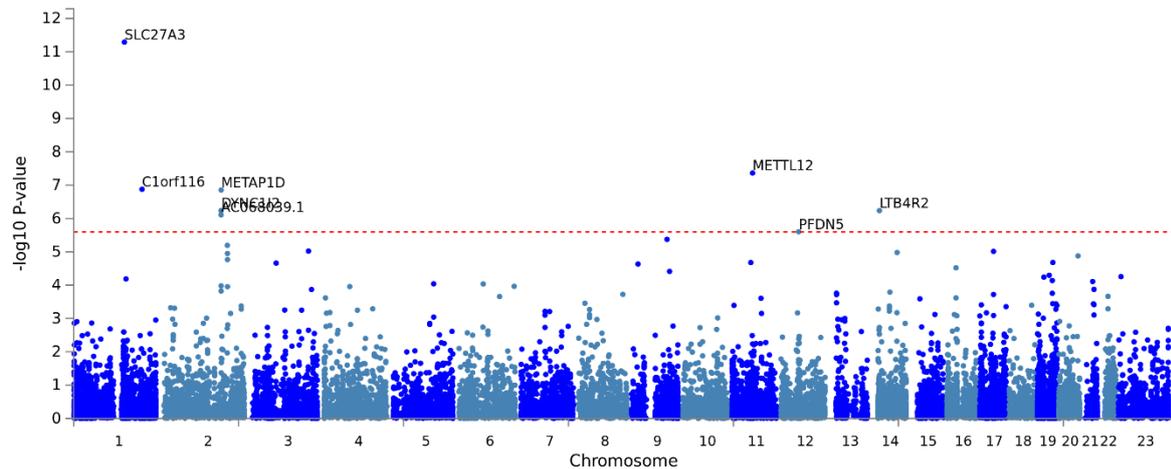
Figure: Manhattan plot for UKBiobank parameter 20216_1 Bipolar and major depression status: Bipolar I Disorder. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: SLC27A3, C1orf116, DYNC1I2, AC068039.1, METAP1D, METTL12, PFDN5, and LTB4R2.

20126_3 Bipolar and major depression status: Probable recurrent major depression

(severe)

Parameter 20126_3 refers to a diagnosis by a mental health group headed by Professor

Jill Pell from the Institute of Health & Wellbeing, University of Glasgow. The participants of

the original UKBiobank survey were asked questions on a variety of general lifestyle factors,

such as daily activity or average hours of sleep, and on previous feelings or moods, such as

"Does your mood often go up and down?". The results were assessed and used to calculate

probable diagnoses ranging from "single lifetime episode" to "probable disorder" as

described in the paper by Smith *et al.* 2013.

Depressive and manic symptoms questionnaire results were collected for 172751 of the participants in the UKBiobank, of which 122982 were assessed by the mental health group and found 8904 people who were characterised as probable lifetime recurrent severe major depression episode sufferers. This data was then cleaned and imputed by Neale *et al.* into 86895 samples, of which 6304 cases against 80591 controls.
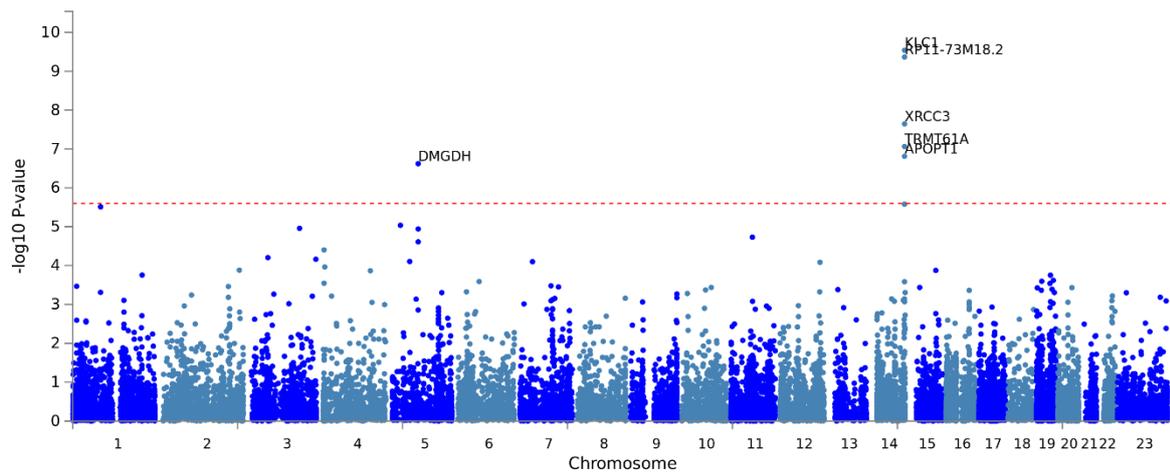
Figure: Manhattan plot for UKBiobank parameter 20216_3 Bipolar and major depression status: Bipolar I Disorder. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: DMGDH, TRMT61A, KLC1, APOPT1, RP11-73M18.2, and XRCC3.

20517 Trouble falling or staying asleep, or sleeping too much

Parameter 20517 refers to a self-reported parameter where the participants were emailed a survey. This parameter are the people who answered positively to the question: "Over the last 2 weeks, how often have you been bothered by any of the following problems?".

There were 157356 participants, with approximately 80000 refusing to answer or answering, "not at all". This data was cleaned and imputed by Neale *et al.* and a total set of 117822 was analysed.
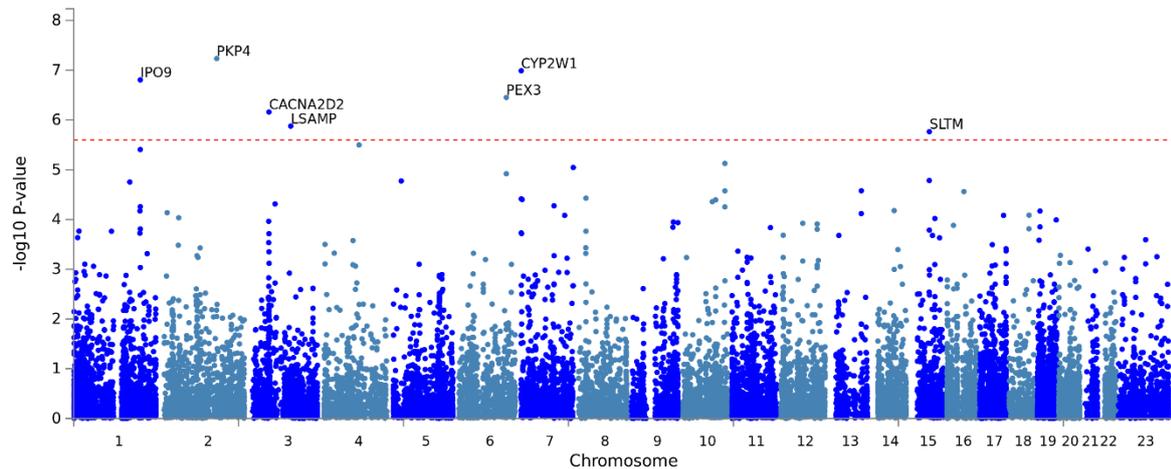
Figure: Manhattan plot for UKBiobank parameter 20517 Trouble falling or staying asleep or sleeping too much. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: IPO9, PKP4, CACNA2D2, LSAMP, PEX3, CYP2W1, and SLTM.

20534 Sleeping too much

Parameter 20534 refers to the responses of a chain of questions: "Have you had prolonged feelings of sadness or loss of interest in activities or hobbies?", then "Did your sleep change?" and ultimately "Was that sleeping too much?". The questions were asked as part of a mental health based online follow up questionnaire.

There were 60909 responses, with 12693 confirming too much sleep. This data was cleaned and imputed by Neale *et al.* and a total set of 45540 was analysed, with 9354 cases against 36186 controls.
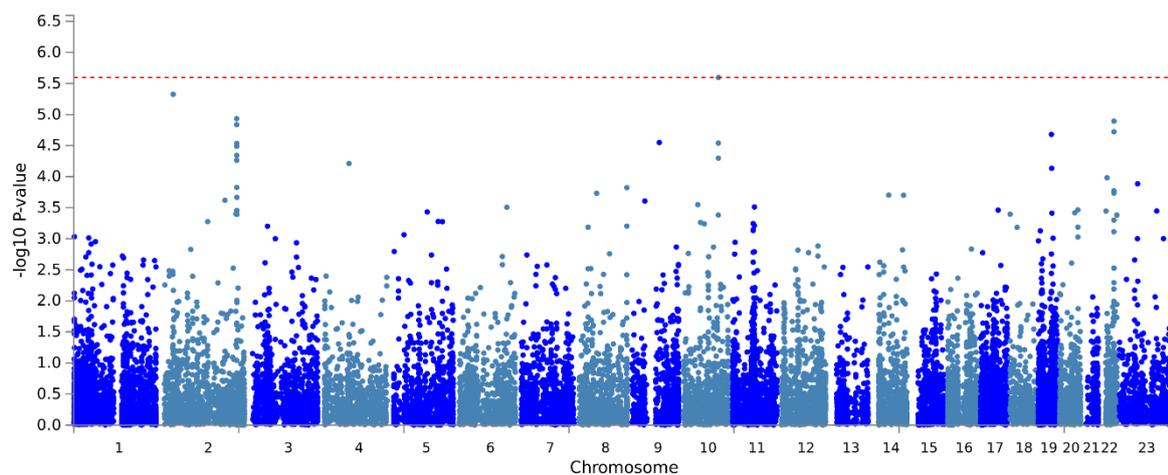
Figure: Manhattan plot for UKBiobank parameter 20534 sleeping too much. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. There were no genes that reached above the genome wide significance level.

20544_1 Mental health problems ever diagnosed by a professional: Social anxiety or social phobia

Parameter 20544_1 refers to participants disclosing if they had any current or historical diagnoses for social anxiety or phobias. The question was part of an online follow-up questionnaire revolving around mental health.

There were 50099 participants, of which 1962 claimed to have been diagnosed at some point in their life. This data was cleaned and imputed by Neale *et al.* and a total of 117716 was analysed, with 1474 cases against 116242 controls.
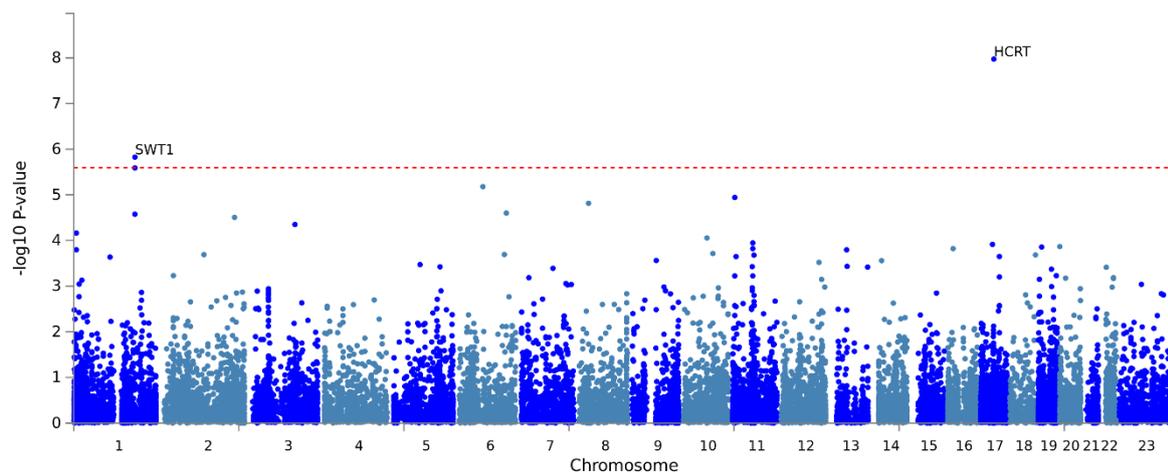
Figure: Manhattan plot for UKBiobank parameter 20544_1 Mental health problems ever diagnosed by a professional: Social anxiety or social phobia. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: SWT1 and HCRT.

20544_10 Mental health problems ever diagnosed by a professional: Mania, hypomania, bipolar or manic-depression.

Parameter 20544_10 refers to participants disclosing if they had any current or historical diagnoses for mania, hypomania, bipolar, or manic-depression. The question was part of an online follow-up questionnaire revolving around mental health.

There were 50099 participants, of which 837 claimed to have been diagnosed at some point in their life. This data was cleaned and imputed by Neale *et al.* and a total of 117708 was analysed, with 604 cases against 117104 controls.
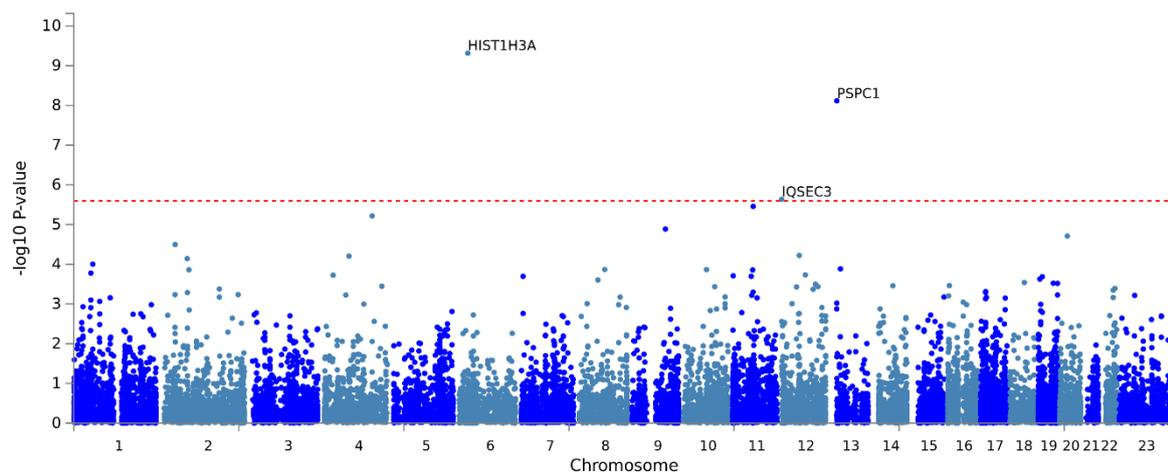
Figure: Manhattan plot for UKBiobank parameter 20544_10 Mental health problems ever diagnosed by a professional: mania, hypomania, bipolar, or manic-depression. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: HIST1H3A, IQSEC3, and PSPC1.

20544_14 Mental health problems ever diagnosed by a professional: Autism, Asperger's or autistic spectrum disorder.

Parameter 20544_14 refers to participants disclosing if they had any current or historical diagnoses for Autism, Asperger's or autistic spectrum disorder. The question was part of an online follow-up questionnaire revolving around mental health.

There were 50099 participants, of which 223 claimed to have been diagnosed at some point in their life. This data was cleaned and imputed by Neale *et al.* and a total of 117707 was analysed, with 166 cases against 117541 controls.
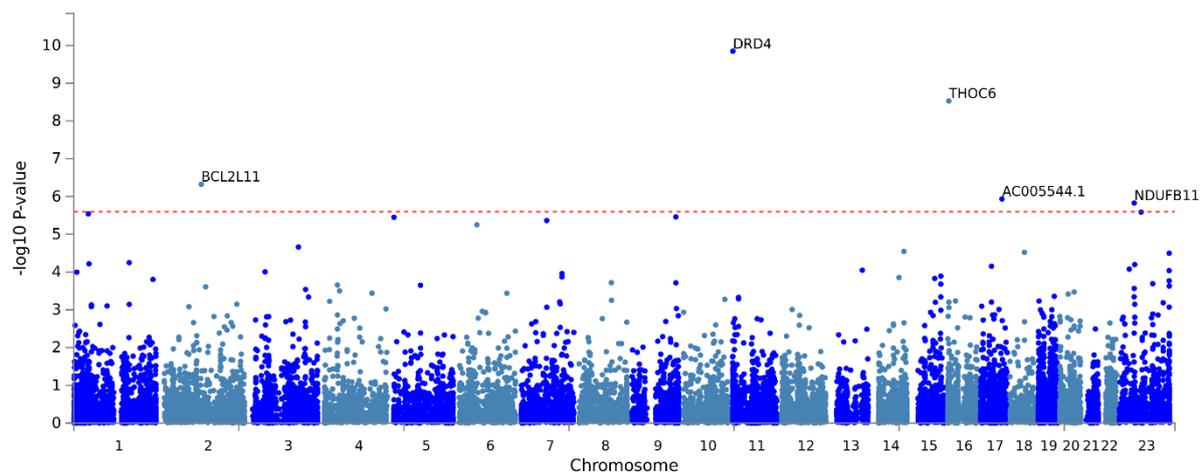
Figure: Manhattan plot for UKBiobank parameter 20544_14 Mental health problems ever diagnosed by a professional: Autism, Asperger's or autistic spectrum disorder. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: BCL2L11, DRD4, THOC6, AC005544.1, and NDUFB11.

20544_15 Mental health problems ever diagnosed by a professional: Anxiety, nerves, or generalized anxiety disorder.

Parameter 20544_15 refers to participants disclosing if they had any current or historical diagnoses for Anxiety, nerves, or generalized anxiety disorder. The question was part of an online follow-up questionnaire revolving around mental health.

There were 50099 participants, of which 22035 claimed to have been diagnosed at some point in their life. This data was cleaned and imputed by Neale *et al.* and a total of 117751 were analysed, with 16730 cases against 101021 controls.
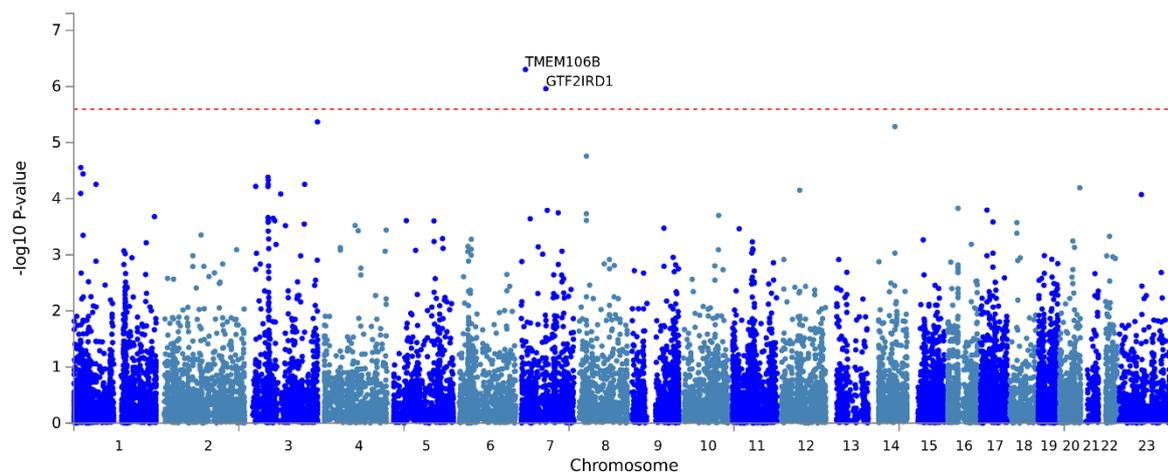
Figure: Manhattan plot for UKBiobank parameter 20544_15 Mental health problems ever diagnosed by a professional: Anxiety, nerves, or generalized anxiety disorder. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: TMEM106B and GTF2IRD1.

G20 ICD10 characterised Parkinson's disease

Parameter G20 is a summary of the specific diagnoses codes each participant had in their medical files from their episodes in hospital. The diagnoses codes are written in accordance with the World Health Organization International classification of disease version 10 (ICD10)

There were 410293 participants, of which 1935 had been diagnosed with G20 Parkinson's disease. This data was cleaned and imputed by Neale *et al.* and a total of 361194 were analysed, with 206 cases against 360988 controls.
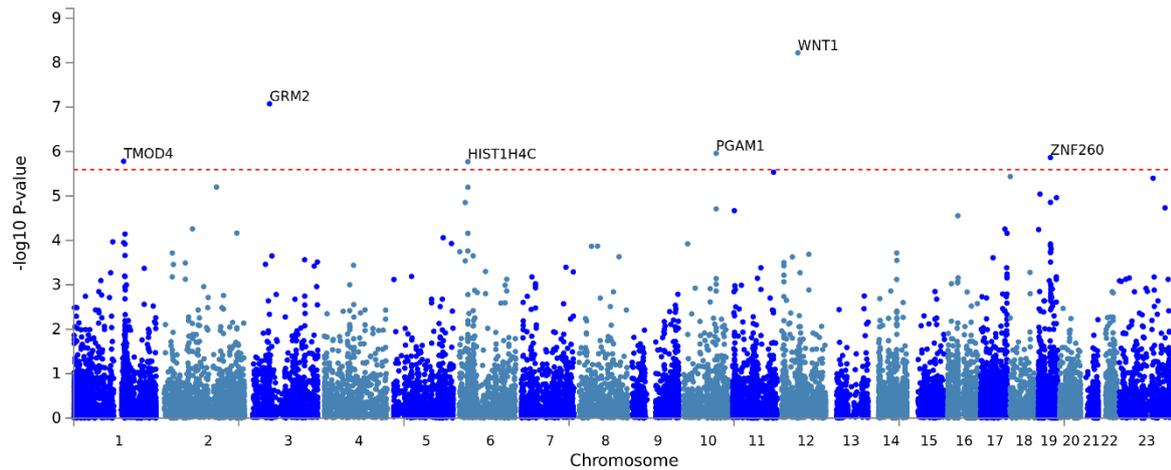
Figure: Manhattan plot for UKBiobank parameter G20 ICD10 characterised Parkinson's disease. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: TMOD4, GRM2, HIST1H4C, PGAM1, WNT1, and ZNF260.

G47 ICD10 characterised Sleep disorders

Parameter G47 is a summary of the specific diagnoses codes each participant had in their medical files from their episodes in hospital. The diagnoses codes are written in accordance with the World Health Organization International classification of disease version 10 (ICD10)

There were 410293 participants, of which 8708 had been diagnosed with G47 Sleep disorders. This group contains a collection of disorders such as sleep apnoea, insomnia, and catalepsy. This data was cleaned and imputed by Neale *et al.* and a total of 361194 were analysed, with 2723 cases against 358471 controls.
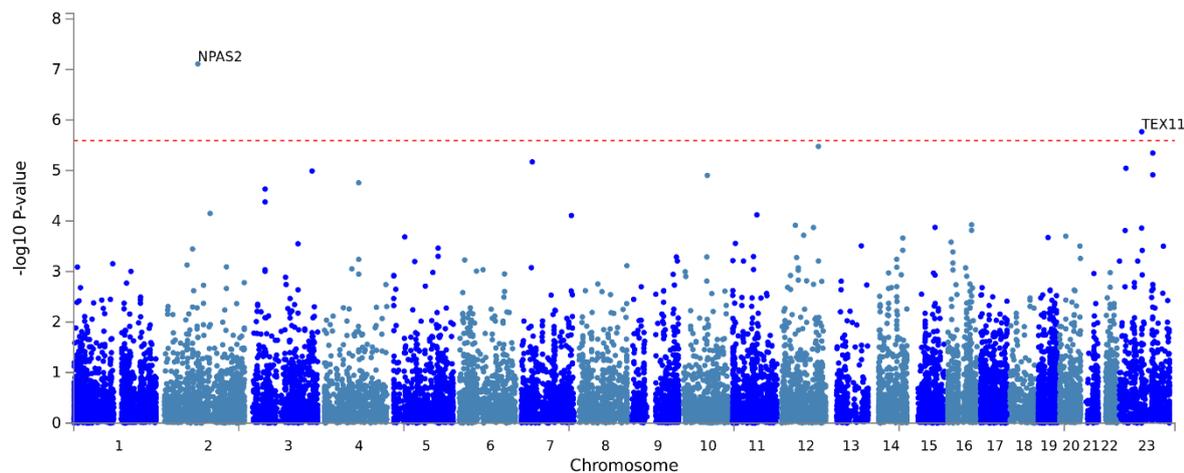
Figure: Manhattan plot for UKBiobank parameter G47 ICD10 characterised sleep disorders. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: NPAS2 and TEX11.

AD Alzheimer's Disease

This parameter is clinical diagnoses of Alzheimer's disease from the Finngen accumulative database. Neale lab cleaned and imputed the data creating a total of 361194 samples, with 119 cases against 361075 controls.
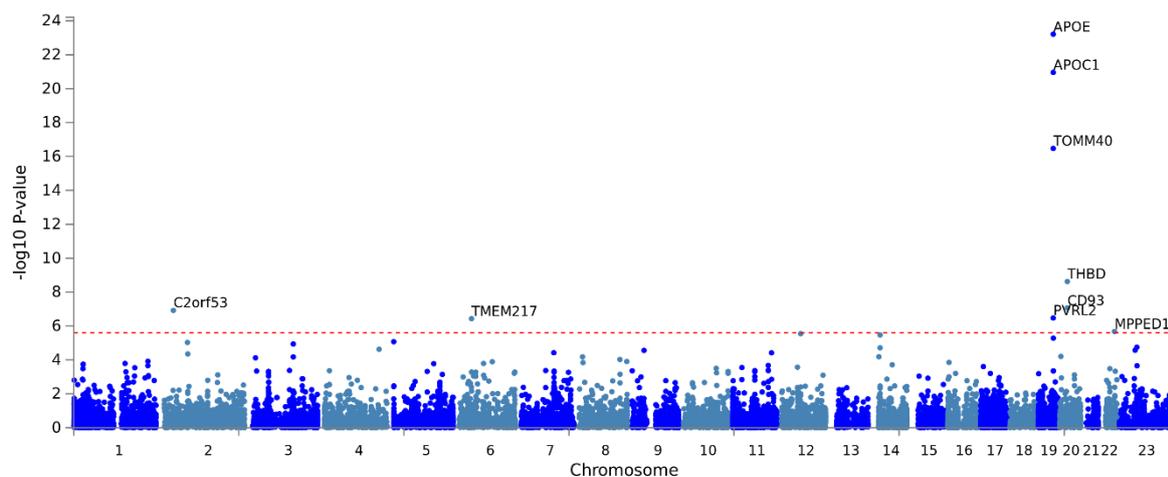
Figure: Manhattan plot for Finngen parameter AD Alzheimer's disease. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of

F5_Schizo: SZ, Schizotypal, and delusional disorders

This parameter is clinical diagnoses of SZ and delusion symptoms from the Finngen accumulative database. Neale lab cleaned and imputed the data creating a total of 361194 samples, with 461 cases against 360733 controls.
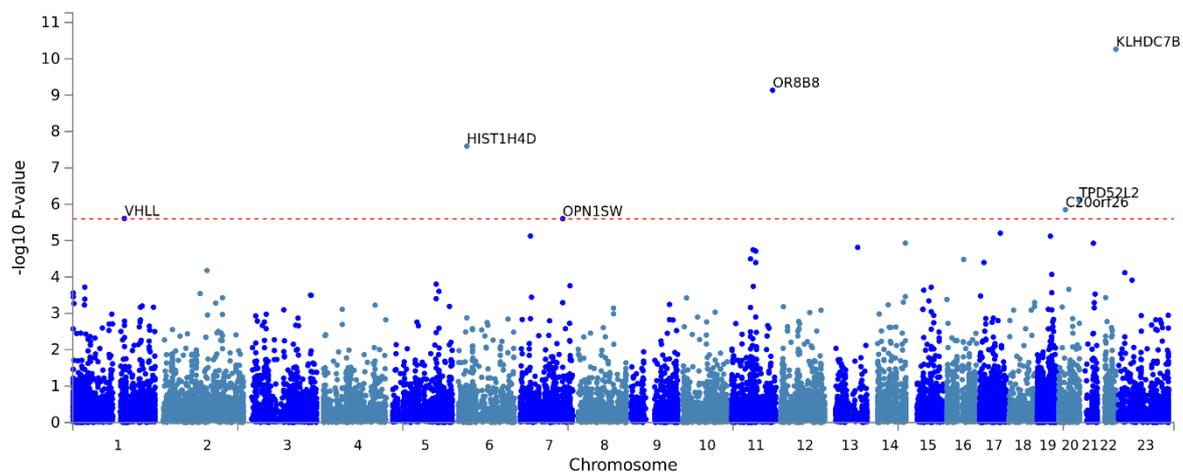
Figure: Manhattan plot for Finngen parameter f5_Schizo SZ, Schizotypal, and delusional disorders. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: VHLL, HIST1H4D, OPN1SW, OR8B8, C20orf26, TPD52L2, and KLHDC7B.

F5_Depressio: Depression

This parameter is clinical diagnoses of Depression from the Finngen accumulative database. Neale lab cleaned and imputed the data creating a total of 361194 samples, with 1145 cases against 360049 controls.
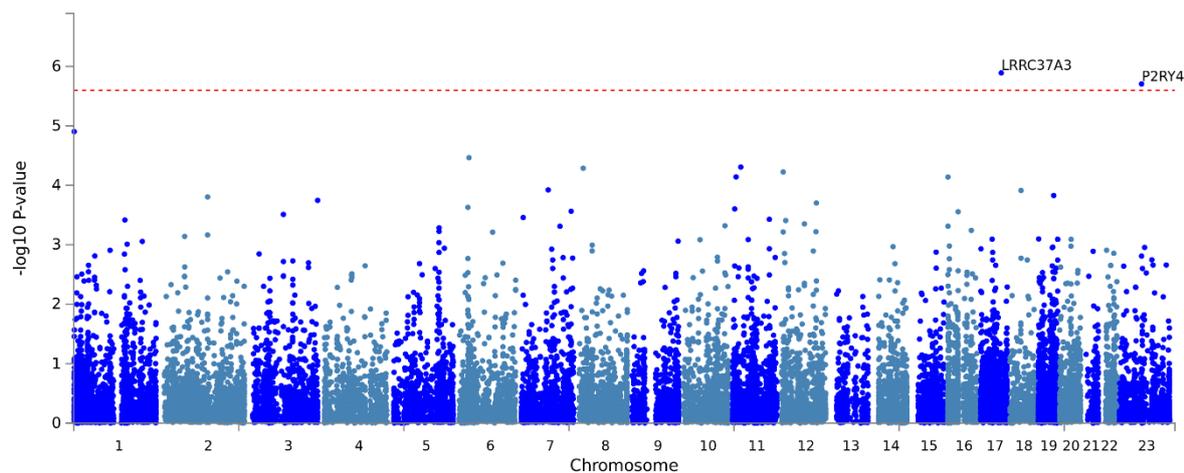
Figure: Manhattan plot for Finngen parameter f5_Depressio: Depression. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. Significant genes above the determined genome wide significant level were: LRRC37A3 and P2RY4.

F5_Dementia: Dementia

This parameter is clinical diagnoses of Depression from the Finngen accumulative database. Neale lab cleaned and imputed the data creating a total of 361194 samples, with 157 cases against 361037 controls.
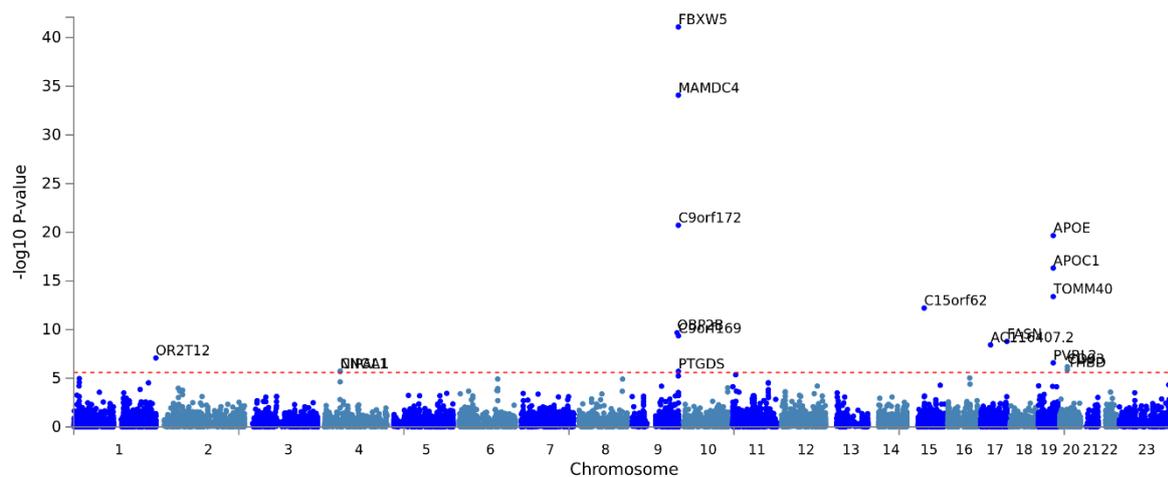
Figure: Manhattan plot for Finngen parameter f5_Dementia: Dementia. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA

F5_AllAnxious: All anxiety disorders

This parameter is and accumulation of clinical diagnoses of anxiety disorders from the Finngen accumulative database. Neale lab cleaned and imputed the data creating a total of 361194 samples, with 532 cases against 360662 controls.
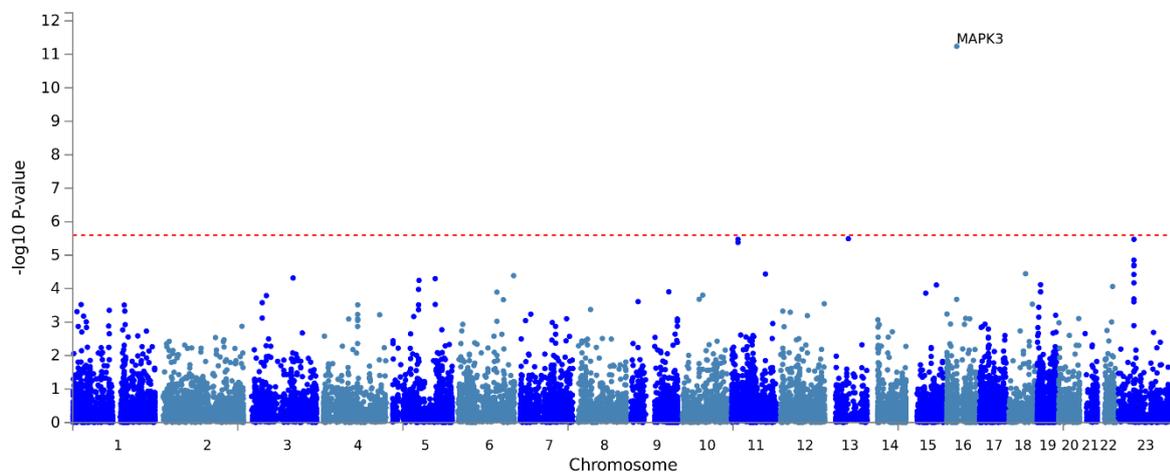
Figure: Manhattan plot for Finngen parameter f5_Allanxious: All anxiety disorders. The data was cleaned and imputed by the Neale lab then made available for public use. The Vrije Universiteit run FUMA (Function Annotation and Mapping of Genome Wide Association Studies) online programme was used analyse the data. The programme includes use of MAGMA to analyse SNP data to compute a gene-based Manhattan plot, highlighting the significant genes in this dataset. Input SNPs were mapped to 19712 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19712 = 2.537e-6. The significant gene above the determined genome wide significant level was MAPK3.

12.5 Richard Anney Cardiff University UKBiobank Parameters

UKBB Code	1160
UKBB Name	Sleep duration
Question	"About how many hours sleep do you get in every 24 hours? (please include naps)"
N in GWAS	335410
N (missing)	1789
Post Processing	drop if <1 drop if >23 prompt to confirm if <3 prompt to confirm if >12
Post Processing Categorisation	0 = <7 (n = 123280) 1 = [7,8] (n = 192438) 2 >=8 (n = 182697)

UKBB Code	1170
UKBB Name	Getting up in the morning
Question	"On an average day, how easy do you find getting up in the morning?"
N in GWAS	336510
N (missing)	698
Post Processing	If this varies a lot then answer this question in relation to last 4 weeks
Post Processing Categorisation	1 = not at all easy (n = 19763) 2 = not very easy (n = 69995) 3 = fairly easy (n = 246698) 4 = very easy (n = 160536)

UKBB Code	1180
UKBB Name	Morning / evening person (Chronotype)
Question	"Do you consider yourself to be?"
N in GWAS	301143
N (missing)	36056
Post Processing	If this varies a lot then answer this question in relation to last 4 weeks
Post Processing Categorisation	1 = definitely a morning person (n = 120401) 2 = more a morning than an evening person (n = 157414) 3 = more a evening than an morning person (n = 126359) 4 = definitely an evening person (n = 40118)

UKBB Code	1190
UKBB Name	Nap during day
Question	"Do you have a nap during the day?"
N in GWAS	335410
N (missing)	1789
Post Processing	If this varies a lot then answer this question in relation to last 4 weeks
Post Processing Categorisation	1 = Never/ rarely (n = 281145) 2 = Sometimes (n = 192704) 3 = Usually (n = 26888)

UKBB Code	1200
UKBB Name	Sleeplessness / insomnia
Question	"Do you have trouble falling asleep at night or do you wake up in the middle of the night?"
N in GWAS	336965
N (missing)	235
Post Processing	If this varies a lot then answer this question in relation to last 4 weeks
Post Processing Categorisation	1 = Never/ rarely (n = 120811) 2 = Sometimes (n = 238899) 3 = Usually (n = 141418)

UKBB Code	1220
UKBB Name	Daytime dozing / sleeping (narcolepsy)
Question	"How likely are you to doze off or fall asleep during the daytime when you don't mean to? (e.g. when working, reading or driving)"
N in GWAS	336965
N (missing)	234
Post Processing	If this varies a lot then answer this question in relation to last 4 weeks
Post Processing Categorisation	0 = Never/ rarely (n = 378813) 1 = Sometimes (n = 105988) 2 = Usually (n = 14053) 3 = All of the time (n = 43)