University of Kentucky
UKnowledge

Theses and Dissertations--Biology | Biology

2017

IDENTIFICATION OF NOVEL SLEEP RELATED GENES FROM LARGE SCALE PHENOTYPING EXPERIMENTS IN MICE

Shreyas Joshi
University of Kentucky, shreyas.joshi@live.com
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.159

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation

Joshi, Shreyas, "IDENTIFICATION OF NOVEL SLEEP RELATED GENES FROM LARGE SCALE PHENOTYPING EXPERIMENTS IN MICE" (2017). *Theses and Dissertations--Biology*. 42.
https://uknowledge.uky.edu/biology_etds/42

This Doctoral Dissertation is brought to you for free and open access by the Biology at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Biology by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Shreyas Joshi, Student

Dr. Bruce F. O'Hara, Major Professor

Dr. David Westneat, Director of Graduate Studies

IDENTIFICATION OF NOVEL SLEEP RELATED GENES FROM LARGE SCALE PHENOTYPING EXPERIMENTS IN MICE

DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By
Shreyas Joshi

Lexington, Kentucky

Director: Dr. Bruce F. O'Hara, Professor of Biology

Lexington, Kentucky

2017

Copyright © Shreyas Joshi 2017

ABSTRACT OF DISSERTATION

IDENTIFICATION OF NOVEL SLEEP RELATED GENES FROM LARGE SCALE PHENOTYPING EXPERIMENTS IN MICE

Humans spend a third of their lives sleeping but very little is known about the physiological and genetic mechanisms controlling sleep. Increased data from sleep phenotyping studies in mouse and other species, genetic crosses, and gene expression databases can all help improve our understanding of the process. Here, we present analysis of our own sleep data from the large-scale phenotyping program at The Jackson Laboratory (JAX), to identify the best gene candidates and phenotype predictors for influencing sleep traits.

The original knockout mouse project (KOMP) was a worldwide collaborative effort to produce embryonic stem (ES) cell lines with one of mouse's 21,000 protein coding genes knocked out. The objective of KOMP^2 is to phenotype as many as of these lines as feasible, with each mouse studied over a ten-week period (www.mousephenotype.org). The phenotyping for sleep behavior is done using our non-invasive Piezo system for mouse activity monitoring. Thus far, sleep behavior has been recorded in more than 6000 mice representing 343 knockout lines and nearly 2000 control mice. Control and KO mice have been compared using multivariate statistical approaches to identify genes that exhibit significant effects on sleep variables from Piezo data. Using these statistical approaches, significant genes affecting sleep have been identified. Genes affecting sleep in a specific sex and that specifically affect sleep during daytime and/or night have also been identified and reported.

The KOMP^2 consists of a broad-based phenotyping pipeline that consists of collection of physiological and biochemical parameters through a variety of assays. Mice enter the pipeline at 4 weeks of age and leave at 18 weeks. Currently, the IMPC (International Mouse Phenotyping Consortium) database consists of more than 33 million observations. Our final dataset prepared by extracting biological sample data for whom sleep recordings are

available consists of nearly 1.5 million observations from multitude of phenotyping assays. Through big data analytics and sophisticated machine learning approaches, we have been able to identify predictor phenotypes that affect sleep in mice. The phenotypes thus identified can play a key role in developing our understanding of mechanism of sleep regulation.

KEYWORDS: Sleep, Bioinformatics, Gene-Phenotype Association, KOMP2, Predictive Modeling

Shreyas Joshi

April 27, 2017

IDENTIFICATION OF NOVEL SLEEP RELATED GENES FROM LARGE SCALE PHENOTYPING EXPERIMENTS IN MICE

By

Shreyas Joshi

Dr. Bruce F. O'Hara

Director of Dissertation

Dr. Dave Westneat

Director of Graduate Studies

April 27, 2017

Date

Dedicated to my wife, Roopa and our parents whose love and support made this possible.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Bruce O'Hara for giving me the opportunity to pursue PhD under his guidance. His scientific wisdom and wealth of knowledge combined with his constant support and encouragement helped me immensely to learn new things and take on challenging projects. He has been a great mentor and has been vital in my growth as a researcher and in pursuit of my career goals. Special thanks to Dr. Jeramiah Smith who is my committee member and has been instrumental in helping me learn computational tools and techniques without which this project would not have been possible. I would also like to thank my other committee members, Dr. Robin Cooper, and Dr. Kevin Donohue for their guidance and support, and my outside examiner Dr. Melinda Wilson for taking the time and provide me with valuable feedback.

I would like to thank past members of O'Hara lab, Dr. Martin Striz and Dr. Mansi Sethi. Martin was the first person to work on sleep data from JAX and started a great collaboration. Mansi further contributed with data collation and preliminary analysis. I also want to thank Dr. Karen Svenson at The Jackson Laboratory for her constant support and collaboration in the project.

O'Hara lab has been a big part of my life and I feel blessed to have Chanung Wang and Lauren (Ren) Guerriero as labmates. They are family who have helped and supported me and it has been great to work alongside them. Special thanks to Abhisek Patel who despite his busy pre-med schedule, did great work in the lab and field on the meditation project.

I would also like to thank current and former Director of Graduate Studies, Dr. Dave Westneat and Dr. Brian Rymond for their dedication towards student success. Finally, I would like to thank all members of Biology department for being a part of this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Background

There is a consensus that all mammals, and perhaps all animals sleep (Campbell and Tobler 1984, Tobler 2005). And yet, the function of sleep is so far unknown and it remains one of the biggest mysteries of biological sciences. The general criteria to establish whether an organism sleeps or not is based on several parameters that include behavioral quiescence, increased arousal threshold, circadian control of sleep onset, homeostatic sleep rebound, postural change, physiological correlates, and impaired performance after sleep deprivation (Cirelli and Tononi 2008). There are other parameters as well that establish a strong connection of sleep with ontogeny (Kayser and Biron 2016). These include developmental control of sleep onset and amount, reduced sleep depth with developmental progression, physiological and brain abnormalities related to sleep loss, behavioral abnormalities caused by juvenile sleep loss, correlation of brain remodeling with early sleep, larval sleep, and embryonic sleep. Based on the above-mentioned criteria, it has been increasingly well established that in addition to mammals, model organisms such as *Danio rerio* (Zebrafish), *Drosophila melanogaster* (Fruitflies) and *Caenorhabditis elegans* (worms) also sleep (Campbell and Tobler 1984, McNamara, Capellini et al. 2008, Hartse 2011).

Two-Process Model

The history of sleep research, though relatively short, has been eventful, and our understanding of sleep has constantly grown through the years. Researchers have tried to understand sleep using a variety of neuroanatomical, physiological, and behavioral approaches. Examination of sleep behavior from a physiological perspective was first performed by Dr. Henry Pieron in 1913 (Piéron 1913). It was Dr. Nathaniel Kleitman who first studied regulation of sleep and wakefulness in 1920s and is considered as the "Father of American Sleep Research". His work led to the discovery of rapid eye movement (REM) during sleep (Aserinsky and Kleitman 1953). More recently, molecular, genetic, and pharmacological methods have been used as well. Although the basic functions of sleep are still unclear, sleep drive, sleep need and sleep propensity can be explained by the two-process model proposed by Alexander Borbely which states that sleep is regulated by a circadian process (Process C) and a homeostatic process (Process S) (Borbely 1982, Daan, Beersma et al. 1984, Borbely and Achermann 1992, Borbely 1998). Here, the sleep drive represents the desire to sleep, sleep need refers to the requirement and duration, and propensity refers to the inclination to sleep. The process C is governed by the endogenous circadian clock and is independent of prior sleep or wake. This endogenous circadian clock can be entrained by external stimuli like light and free runs under constant conditions with a period of approximately 24 hours i.e. after the entrainment, the clock maintains its periodicity even in the absence of stimulus. The process S, on the other hand, is governed by previous sleep or wake, with drive for sleep increasing as the wake prolongs. The Suprachiasmatic Nucleus (SCN) is the master

circadian pacemaker that is both necessary and sufficient for the generation and maintenance of most circadian rhythms (Ralph, Foster et al. 1990).

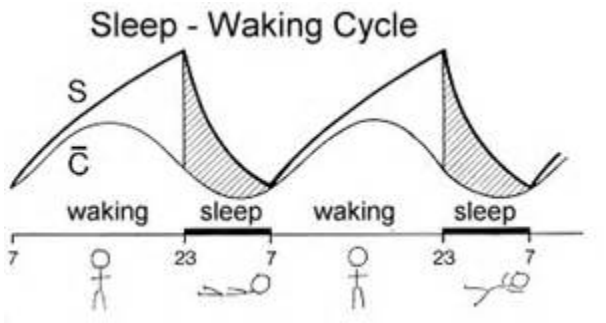


Figure 1. 1.: The two process model of sleep (Borbely 1982)

The two processes are independent but interact with one another. Two principal properties of Circadian rhythms are entrainment to environmental cues, and a self-sustained oscillation that maintains an approximately 24 h period and will free-run in the absence of any environmental cues. Entrainment or synchronization normally occurs via the 24 h solar cycle. The suprachiasmatic nucleus (SCN) is the dominant circadian pacemaker in mammals, and is located in the hypothalamus within the brain (Ralph, Foster et al. 1990, Provencio, Cooper et al. 1998). The SCN projections are sent to different parts of the brain, with a majority ending at the sub-paraventricular zone and the signals are relayed to dorsomedial hypothalamic nucleus. Projections from dorsomedial hypothalamic nucleus are sent to different hypothalamic regions that include inhibitory GABAergic projections to sleep promoting ventrolateral preoptic area (VLPO), and excitatory glutamatergic projections to wake promoting lateral hypothalamus (Saper, Scammell et al. 2005). The balance between wake and sleep is maintained via two

distinct groups of neurons from the tuberomammillary nucleus (TMN) and VLPO. The increase in firing rates in the VLPO neurons inhibits activity in wake promoting regions including TMN. Orexin/hypocretin is a wake promoting neurotransmitter and helps prevents rapid transition from wake to sleep. There are several hypotheses for the mechanisms underlying the homeostatic process, but they are not well understood (Dijk, Beersma et al. 1987). Adenosine has been proposed as a key molecule that plays a role in homeostatic sleep drive by acting on VLPO neurons through its accumulation in the basal forebrain during wakefulness (Porkka-Heiskanen and Kalinchuk 2011).

Molecular mechanisms of sleep

In terms of molecular mechanisms, the circadian regulation of sleep is much better understood than the homeostatic regulation. The molecular circadian pathway includes several genes. *CLOCK (Clk)* and *BMAL1* are two basic helix-loop-helix (BHLH)-PAS domain proteins that dimerize and bind to E-box element (a consensus cis-regulatory sequence) that drives transcription of Period genes (*Per1, Per2, Per3*) and Cryptochrome genes (*Cry1, Cry2*). Pers and Crys interact after translation and translocate to nucleus where they suppress Clk-Bmal1 activity, which in turn negatively regulates their own expression as well. This negative feedback loop appears to be at the heart of the circadian pacemaker. Pers and Crys degrade over time, and Clk-Bmal1 dimers are reactivated, leading to a re-start of the cycle once again after roughly 24 hours (Reppert and Weaver 2001, Lowrey and Takahashi 2004).

Circadian clock genes that influence both circadian timing and sleep homeostasis, have been identified using ENU/EMS (Ethyl methanesulfonate) mutagenesis techniques (Vitaterna, King et al. 1994, Kloss, Price et al. 1998, Kapfhamer, Valladares et al. 2002, O'Hara, Ding et al. 2007, Cirelli 2009). These gene include *Clock* and *Rab3a* in mice, and *Per* and *Dbt* in flies. Discovery of these clock genes led to identification of many others (*Bmal1/Cyc, Cry1,2, Per1,2,3,* etc.) that also underlie circadian and homeostatic aspects of sleep. However, many of these mutations produce only subtle phenotypes, which are difficult to detect and are affected by genetic background (Nadeau 2001). Approaches using traditional mouse strains, genetic crosses, and QTL strategies have also identified a modest number of genes that influence sleep including *Homer1a, Acads* (acyl-coenzyme A dehydrogenase), and *Rarb* (Retinoic acid receptor beta) (Tafti, Petit et al. 2003, Drager 2006, Maret, Dorsaz et al. 2007). A disadvantage of traditional QTL approaches is that they are subject to limited mapping resolution, and identifying the causal gene(s) is often difficult or not undertaken (Tabor, Risch et al. 2002, Churchill, Gatti et al. 2012), although this situation is improving with recent advances in mapping populations and related approaches (Jiang, Scarpa et al. 2015).

Sleep-wake monitoring in model organisms

Model organisms have played a key role in our understanding of sleep at both genetic and physiological levels. Much of this has been possible because of large-scale studies allowing for high-throughput screening for sleep related phenotypes (Joshi and O'Hara 2014, Sethi, Joshi et al. 2015, Funato, Miyoshi et al. 2016). Being able to perform

sleep-wake monitoring in a high-throughput manner is key for future developments in sleep research because that allows for studies to be done at a larger scale. Below is a short review that provides an overview of high-throughput sleep-wake monitoring techniques in the four most widely used model systems in sleep research: *Caenorhabditis elegans, Drosophila melanogaster, Danio rerio, and Mus musculus*.

Caenorhabditis elegans

C. elegans was introduced as a model organism for understanding development and the nervous system in 1960s (Brenner 1974). The life cycle of *C. elegans* is comprised of an embryonic stage, four larval stages, and adulthood. At the end of each larval stage, the organism undergoes molting wherein a new cuticle is synthesized and the older one is shed. Before each molt, there is a state called "*lethargus*" which is a quiescent behavioral state that is similar to sleep (Raizen, Zimmerman et al. 2008). It is a reversible state with reduced responsiveness to external stimuli and cessation of feeding and motion (Cassada and Russell 1975, Raizen, Zimmerman et al. 2008).

Initial studies for lethargus relied on single worm recordings on agar plates seeded with bacteria (Van Buskirk and Sternberg 2007, Raizen, Zimmerman et al. 2008). High-throughput approaches for activity monitoring in worms relies on microfluidics technology that allows for encapsulating worms in an array of droplets on a chip (Shi, Wen et al. 2010, Belfer, Chuang et al. 2013). In the method presented by Belfer et al, worms are housed in sub-microliter droplets with *E.coli* as food source, and is called the *Caenorhabditis*-in-drop (CiD) technique. Imaging can be performed by bright field

microscopy for several days throughout lethargus for at least 24 worms thus allowing for high-throughput analysis of complex behavior.

Drosophila melanogaster

Drosophila exhibits a sleep-like behavioral state called “Rest” (Hendricks, Finn et al. 2000). During rest, flies exhibited extended periods of immobility with reduced responsiveness to natural or mechanical stimulation, lasting up to 2.5 hours with uniformly characteristic resting locations and postures in isolated and groups housing conditions. Homeostatic regulation was also observed with significant rebound after sleep deprivation, supporting the idea that *Drosophila* “rest” is in fact a sleep or sleep-like state.

The current standard method for *Drosophila* sleep-wake monitoring involves signals recorded by infrared beam breaking, by flies that are housed within a tube. The system was first developed by Dr. Yoshiki Hotta who used an incandescent lamp with a filter transmitting wavelengths above 800 μm (Konopka and Benzer 1971). The system in its current high-throughput avatar is called Drosophila Activity Monitoring System (DAMS) (Trikinetics, Waltham, MA), and is widely used in Drosophila behavioral analysis. The DAM2 monitor can measure locomotor activity of 32 individual flies. The setup consists of 32 tubes with each tube containing food at one end, the fly, and a cotton plug or a plastic cap at the open end. Infrared beam breaking count is maintained for each channel throughout the duration of experiment and the data can be analyzed using the software produced with the system.

For sleep studies, a major limitation of DAMS is its low sensitivity towards fly movements. A small movement that can be an interruption in sleep might not be recorded unless the fly crosses the path of infrared beams. It is also not possible to assess the magnitude of movement, since a large movement from one end of the tube to another is counted as a single beam break. An alternative method with video analysis has been proposed that overcomes these limitations, and also provides more accurate data about sleep architecture, along with precise location of the fly within the tube.

Danio rerio

Zebrafish has emerged as a model for studying sleep and wake behavior in the last decade. Quiescent behavior that meets the criteria for a sleep like state exists in both larval and adult stages of Zebrafish (Yokogawa, Marin et al. 2007, Prober, Zimmerman et al. 2008, Zhdanova 2011). The quiescent behavior in Zebrafish is regulated by an endogenous rhythm with bouts of inactivity occurring mostly at night. These bouts of quiescence can last for several minutes or longer and are considered sleep if they last for at least 1 minute in larvae and for at least 6 seconds in adults. There is increased arousal threshold and decreased responsiveness during quiescence and increased sleep amount (rebound) after sleep deprivation thus fulfilling the criteria of homeostatic regulation of sleep.

Similar to *Drosophila,* infrared beam breaking is used extensively to study locomotor activity in Zebrafish. Though it is a very good technique to study circadian and gross locomotion activity, it suffers from the same limitations as discussed earlier for

sleep related studies in fruit flies. This makes automated digital video analysis of sleep-wake behavior an attractive alternative. There are several commercially available activity tracking systems for Zebrafish larvae and adults. The prominent of these are distributed by by Noldus IT (Wageningen, Netherlands) and Viewpoint Life Sciences (Montreal, Canada). Along with this, there is an option for researchers to develop their own setup with high-speed camera and custom algorithms developed for analyzing Zebrafish responses to outside stimuli.

Mus musculus

Among the mammalian systems, sleep has been extensively studied in mice. With a wide variety of inbred strains specific to sleep phenotypes, well-established behavioral assays, and a strong genetic toolbox, it is indeed an efficient model system for studying sleep. Although mice are nocturnal, the circadian system is well conserved across all mammals. Moreover, among the model organisms discussed thus far, the mouse best satisfies all sleep criteria, and as a mammal, its neuroanatomy, physiology, and genetics are much closer to humans thus making it the most attractive model system to study human sleep. Mouse sleep differs from humans in two aspects: mice sleep more during the day, and sleep in mice is highly fragmented. Similar to human sleep, the sleep measures in mice have traditionally been focused on neural surrogates in which the sleep is described in terms of Rapid Eye Movement (REM) and Non-Rapid Eye Movement (Non-REM or NREM) states.

Electroencephalogram (EEG) has been the gold standard for sleep studies in mice (as in humans) wherein sleep structure and the underlying neurophysiology can be assessed by analyzing EEG power spectra for different sleep-wake states (Davis, Davis et al. 1937, Lo, Chou et al. 2004). The EEG sleep studies in mice have played a great role in advancement of our understanding of sleep (Funato, Miyoshi et al. 2016). EEG recordings are generally straight-forward in humans, but in mice, the electrodes should be surgically implanted, making it an invasive, time-consuming and expensive process. These limitations make it unsuitable for use in high-throughput studies involving large numbers of animals. In addition, the invasive aspect of EEG with side effects of surgery and electrodes implanted for large periods of time can bring about a change in the sleep structure of an organisms, thus causing a confound in the study itself. When in-depth EEG studies are not required, ethical and cost restraints demand use of a non-invasive approach that can provide sleep variables like duration, bout length (duration of single sleep episode), number of bouts, and sleep latency, which in a high-throughput study can provide valuable information about sleep-wake behavior.

Two non-invasive techniques with a potential for high-throughput screening have gained prominence. Pack et al (McShane, Galante et al. 2012) used video monitoring with infrared beam breaking based object recognition algorithm to evaluate sleep. By comparing with EEG data, they identified that ≥40 s of inactivity was highly correlated with sleep in mice. Fisher et al have proposed a similar technique that also records additional variables that can be used for studying anxiety or behavioral inhibition phenotypes (Fisher, Godinho et al. 2012). Noldus IT (Wageningen, Netherlands) and

Pinnacle Technology (Lawrence, KS, USA) provide commercial systems that use video recording for sleep studies and proprietary software for data analysis.

Another approach is to monitor sleep-wake activity through piezoelectric sensors positioned on the floor of a mouse cage. Results from such sensors were first reported by Megens et al (Megens, Voeten et al. 1987) where they were used to study changes in activity in response to different drugs and dosages. Flores et al (Flores, Flores et al. 2007) implemented a novel pattern recognition algorithm that identified sleep when the animal assumed a specific posture and regular motions associated with breathing were recorded, and wake was classified in the form of large amplitude spikes. The limitations with this initial system included variable noise and a computationally expensive algorithm, where data from a single mouse took nearly 10 hours to process on a regular desktop PC.

Donohue et al (Donohue, Medonza et al. 2008) built on this initial attempt, and developed a piezoelectric sleep-wake monitor that overcomes these limitations with a simpler more robust algorithm and ergonomics, and became suitable for high-throughput sleep studies in mice. A single unit consists of four cages and detects motion through a Polyvinylidine Difluoride (PVDF) sensor on the cage floor. With respect to EEG, it has classification accuracy of more than 90% for sleep-wake. It can generate sleep statistics for large number of mice over long periods of time, and has been used in several large-scale sleep studies, with continual improvement in the software and hardware features (Philip, Sokoloff et al. 2011, Sethi, Joshi et al. 2015).

KOMP^2

International Knockout Mouse Consortium (IKMC) aimed to generate mutant embryonic stem cells (ES) for every coding gene in the mouse genome on the C57BL/6NJ (also called Black-6) background. This was the first generation of the project, also called as Knockout Mouse Project (KOMP). In the next stage, these mice are being phenotyped as part of Knockout Mouse Phenotyping Program (KOMP^2). As live mice are made from the ES cell lines, these single-gene knockouts undergo a core set of broad-based phenotyping screens at the KOMP^2 centers as part of the IMPC (Abbott 2010, Bradley, Anastassiadis et al. 2012, Brown and Moore 2012). At The Jackson Laboratory KOMP^2 Center (JAX-KOMP^2) sleep is part of this pipeline, and the results thus far are described in this report (Chapters 2 and 3). Such an undertaking has many advantages over individual lab efforts, as each mouse can be examined for many traits across many domains, quality control and bioinformatic approaches can more easily assess confounding variables, and the many different traits can be cross-correlated to reveal new relationships. This broad-based platform will expedite the functional annotation of genes, especially those that are currently most poorly understood, and promote continued utilization of these knockout mice for future studies (Bradley, Anastassiadis et al. 2012, Brown and Moore 2012).

In the first part of my thesis (Chapter 2), I have used data from the JAX Pipeline for sleep phenotyping from piezo system to identify gene knockouts in which sleep significantly differs from control mice. By analyzing data from more than 6000 mice belonging to 343 gene knockouts using multiple statistical approaches, I have identified 122 genes that affect sleep duration and quality. Sex-associated differences in genes

affecting sleep have been identified, along with genes that affect sleep only during a specific circadian phase. In the second part (Chapter 3), I have used sophisticated machine learning and data analysis techniques to identify predictor phenotypes for sleep behavior in KOMP^2. By performing partial data simulation and data modeling through multiple regression, decision trees, and random forest algorithms, I have identified key phenotypes in the KOMP^2 that can be used as a predictor for sleep behavior.

CHAPTER 2

Noninvasive sleep monitoring in large-scale screening of knock-out mice reveals novel sleep-related genes

Background

Sleep is a complex behavior common to all birds and mammals, and probably most or all other vertebrates and invertebrates with a nervous system (Tobler, Franken et al. 1992, Tobler 2005). Regulated by a multitude of neuronal processes and indirectly by gene networks, it is a process vital for an organism's health and survival. Sleep has been suggested to have a role in functions such as learning, memory consolidation, energy restoration, synaptic optimization and recently it has also been implicated in the clearance of metabolites, including Aβ (Tononi and Cirelli 2006, Tucker, Hirota et al. 2006, Marshall and Born 2007, Nishida, Pearsall et al. 2009, Diekelmann and Born 2010, Xie, Kang et al. 2013, Krueger, Frank et al. 2016). Aβ (Amyloid beta) are main components of the amyloid plaques found in the brains of patients suffering from Alzheimers disease.

Genetic manipulations have advanced our knowledge about some aspects of sleep, including influences on the sleep EEG, sleep disorders, brain areas regulating sleep processes, and molecular pathways underlying sleep and its regulation. However, relatively few gene mutations or gene knockouts in mice have been examined for effects on sleep, and there are still many unresolved questions regarding the biological need for sleep, functions of sleep, and the genetic and physiological basis of sleep homeostasis (Rechtschaffen 1998, Cirelli 2009, Vassalli and Dijk 2009) that could be addressed with insights from model organisms.

There have been numerous efforts to address these questions utilizing a variety of animal models, including mice. These efforts range from individual labs studying specific knockout mice, to large-scale QTL (quantitative trait loci) and genome-wide projects including phenotype-driven ENU (N-ethyl-N-nitrosourea) mutant screens involving many labs, and gene-driven knockout mouse phenotyping programs (Gondo, Fukumura et al. 2009).

A major bottleneck in large-scale genetic studies of sleep is the difficulty, expense, and time demands of traditional EEG/EMG studies. While knockout studies of selected target genes such as neurotransmitter receptors have found modest effects on at least one sleep parameter, relatively few genes have been examined (O'Hara, Jiang et al. 2017). Using a higher throughput, non-invasive approach allows for much larger numbers of mice to be examined (Flores, Flores et al. 2007, Pack, Galante et al. 2007, Donohue, Medonza et al. 2008, Philip, Sokoloff et al. 2011, Mang, Nicod et al. 2014). Our approach utilizes a sensitive piezoelectric film across the mouse cage floor, and is especially well suited to characterization of large-scale resources such as the International Knockout Mouse Consortium (IKMC) (Ringwald, Iyer et al. 2011).

Methods and Design

Generation of knockout mice

IKMC mouse mutants are generated on a C57BL/6NJ mouse background that have either null alleles, which have an entire locus removed or “knockout-first” alleles, which permits generation of conditional alleles by utilization of site-specific recombinase

as described previously (Skarnes, Rosen et al. 2011, Schofield, Hoehndorf et al. 2012). C57BL/6NJ is a substrain of the original C57BL/6 (commonly called B6 or Black-6) which is a well-characterized inbred strain and has been the most commonly used one in biomedical research. It also serves as a reference strain for the mouse genome, making it an ideal choice for this effort, although differences between the C57BL/6J and C57BL/6NJ substrains exist (Keane, Goodstadt et al. 2011, Simon, Greenaway et al. 2013, Mekada, Hirose et al. 2015). Here onwards, the C57BL/6NJ will be referred to as B6NJ for the remainder of chapter.

As part of the JAX-KOMP^2 phenotyping pipeline, each mouse is comprehensively phenotyped for over 200 measurements, from age 4-18 weeks, for a range of morphological, physiological and behavioral traits including many disease relevant parameters pertaining to neurobehavior, metabolism, immune, cardiovascular, sensory, and musculo-skeletal systems, followed by terminal collection of blood and histopathology (Morgan, Simon et al. 2012) (Fig 2.1). Additional tests such as light/dark and hole-board exploration tests, rotarod, and sleep, are unique to the JAX-KOMP^2 pipeline. Sleep is evaluated using a PiezoSleep System (*Signal Solutions, LLC, Lexington, KY*), a non-invasive, high throughput sleep-wake monitoring system (details provided in following sections). The primary traits analyzed are total sleep duration averaged across 24 h, 12 h light phase and 12 h dark phase, and average sleep bout lengths (across 24 h, 12 h light, and 12 h dark phase), and breath rate during sleep.

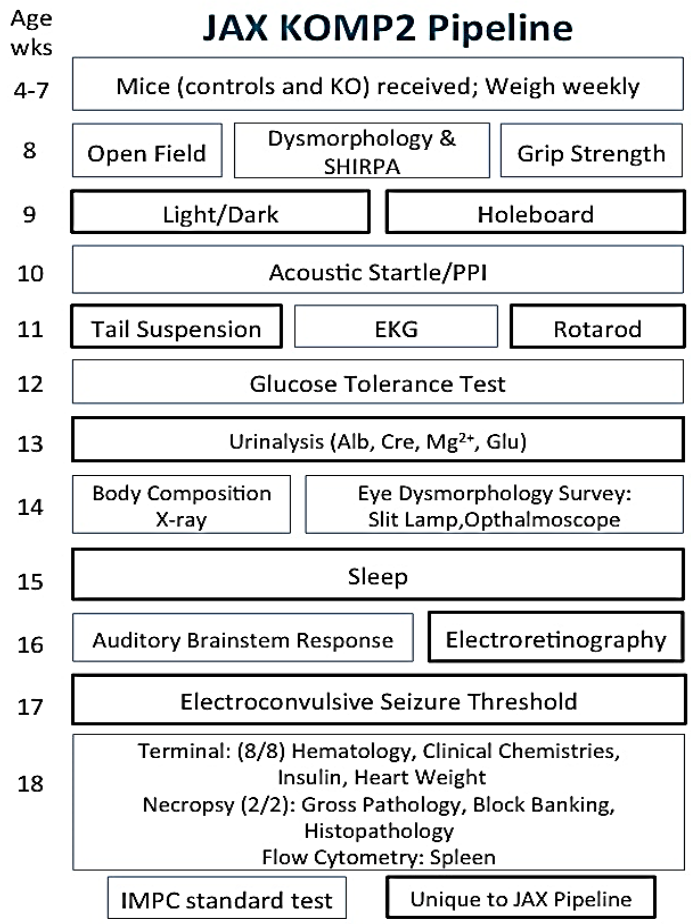


Figure 2. 1.: Phenotyping Pipeline at The Jackson Laboratory (Wt : Wildtype)

Sleep recording with piezoelectric system

Sleep and wake states were determined using the PiezoSleep System (Flores, Flores et al. 2007, Donohue, Medonza et al. 2008, Mang, Nicod et al. 2014). It is a quad-cage and piezoelectric sensor system that comprises plexiglass cages lined with piezoelectric films across the cage floor (*Signal Solutions Llc, Lexington, KY*). These piezoelectric films are highly sensitive to pressure changes caused by different movements and sleeping postures, and even pressure variations from breathing are also

detected. The software provided with the system records the changes in amplitude of electrical signals generated by force on the piezoelectric sensors. Sleep states are characterized by quasi-periodic signals with low variations in amplitude, whereas wakefulness and rest states are characterized by irregular transient and high amplitude pressure variations corresponding to voluntary body movements and weight shifting. The piezo system has been validated with simultaneous EEG and human scoring of mice demonstrating a classification accuracy of over 90% in mice (Donohue, Medonza et al. 2008, Mang, Nicod et al. 2014).

Animal housing and Phenotyping

This study utilizes control and KO mice generated on B6NJ background. From wean age, mice were housed at 3-5 mice per pen in pressurized, individually ventilated cages using pine shavings as bedding, with free access to acidified water and food (LabDiets 5K52, LabDiet, Scott Distributing, Hudson, NH). The housing facility was maintained on a 12:12 light/dark cycle starting at 0600. At 15 weeks of age, mice were removed from their home cages and placed into individual cages of the piezo system. The system used in the JAX-KOMP^2 pipeline is comprised of 17 4-cage units, allowing for simultaneous assessment of up to 72 mice per 5-day experiment window. During sleep testing, light cycle and food and water access were as that of standard housing conditions. A minimum amount of pine shavings was provided to each cage to allow sufficient detection of pressure signal. In each testing week, 10 control B6NJ animals (five females and five males) and 3-19 animals per KO line were tested. Females and males of all KO

strains were analyzed, with a pipeline throughput goal of testing eight animals of each sex for all screens.

Data Analysis

We began with a dataset with observations from 8849 mice (4446 females, and 4403 males). In the KOMP^2 pipeline, a rigorous quality check is performed before the data is approved for release. In the final dataset, observations only from mice that had been approved to have finished the KOMP phenotyping pipeline were included in the analysis. One of the outputs of piezo sleep-wake system is data confidence metric that ranges from 0 through 1 to assess the signal quality and/or outlying signal behavior. Any of the sleep recordings with a data confidence value below the threshold of 0.6 were excluded from the analysis. The control mice were also screened for outliers based on extreme high/low values for their sleep/wake parameters. Multivariate outlier analysis was carried out by computing Mahalanobis distance (md) (Mitchell and Krzanowski 1985). In brief, it calculates multidimensional distance of each of the observations from the centroid mean vector of all measured variable scores, (Bassett, Gogakos et al. 2012). Data analysis was carried with both outliers included and excluded, and in this dataset the results remained the same. The goal of removing outliers is to remove the observations that might distort the segmentation of the data which is key in identifying candidate genotypes in later steps. As more data for sleep behavior becomes available in KOMP^2, it can potentially help in identifying genes that might be missed because of the increased variance in the data caused by the outliers. MD assesses cutoff values for outliers based on quantiles in chi-square distribution. At default quantile threshold of 0.975, a cutoff

value of 4.0 was generated with 114 Control mice having md values above the cutoff, and were thus excluded from the analysis. The final dataset used for analysis contained 6350 mice belonging to 343 KO strains and control mice. This final dataset consisted of 1884 Control (918 Females; 894 Males), and 4466 KO (2250 Females; 2216 Males) mice. One-way analysis of variance (ANOVA) was performed for each sleep variable and breath rate, and posteriori multiple comparisons were done using Dunnett's test to identify genotypes showing significant difference/s with respect to the control group (Dunnett 1955). For multiple comparisons, p-values adjusted for family wise error rate (FWER) were computed. Similar analysis was done to identify genotypes with sex-specific effects. 289 of the knockout cohorts containing at least 3 females and 3 males each were included in this analysis. Significance values for simple and interaction effects between Genotype and Sex were computed through Factorial ANOVA and effect sizes were estimated. For all of the analyses of the sleep-wake parameters under consideration (listed above), an adjusted p-value of less than 0.05 was considered significant.

Principal component analysis (PCA) was performed on six of the sleep variables under consideration: sleep durations and bout lengths across 24 hours, and 12 hours during light and dark phases. PCA is a method for multivariate analysis through dimension reduction i.e. it is used to reduce the number of variables into principal components (PC) that account for most of the variance of the original variables. The first principal component that accounts for majority of variance in the data can be considered the summary variable representing the data. Genotypes significant for the first principal component were identified and were considered to be the ones that affect overall sleep. In the final step, mean values for each sleep variable from piezo system were computed

for all genotypes and multivariate Mahalanobis distance outlier analysis was performed for the whole dataset. The genes identified as outliers were considered to have maximum effect on sleep variables. The PCA and outlier analysis described above should be considered an extension of the significance analysis performed through ANOVA and Dunnett's test. These were conducted with the goal of identifying the genes that appear in the results of all three analyses and will be considered key candidate genes affecting sleep in mice. A final list of all significant results was prepared and these were presented as the sleep related candidate genes.

Gene network analysis was conducted using GeneMANIA, which is a tool for analyzing sets of genes (Warde-Farley, Donaldson et al. 2010). The Cytoscape software for network visualization contains a dedicated plugin for GeneMania and was used in this analysis. For phenotype associations, data was collected for final list of sleep related candidate genes from IMPC database (mousephenotype.org) through batch query. The complete schematic of data analysis procedure has been shown in Figure 2.2.

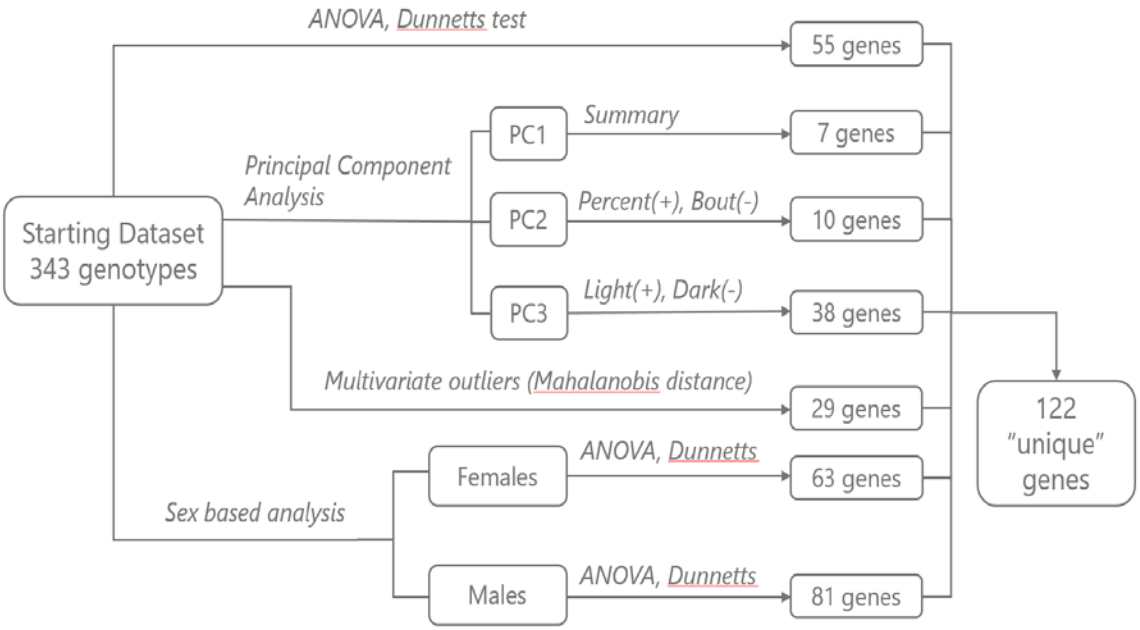


Figure 2. 2.: Schematic of data analysis procedure

Geneweaver software was used to perform geneset similarity analysis (Bubier, Langston et al. 2017). Data analysis was performed in R programming environment in the R Studio. The outlier analysis was done using Chemometrics package (Varmuza and Filzmoser 2016), and Dunnett's test was performed using DescTools package (Signorell 2015). Package ggplot2 was used to prepare plots (Wickham 2016).

Results

B6NJ/Control mice

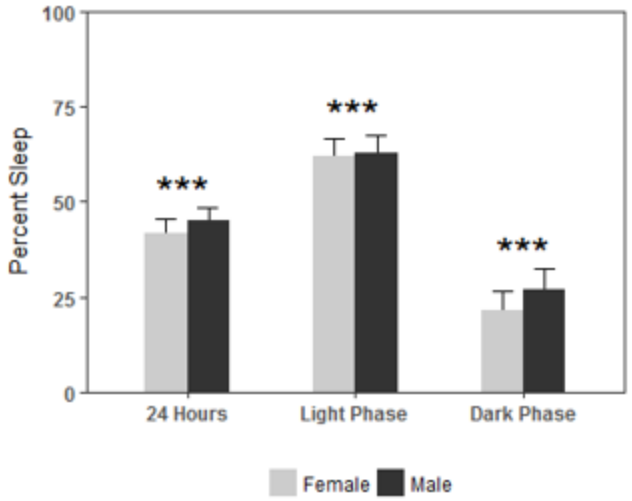
We found that female B6NJ have a significant reduction in total sleep time (41.83 ± 0.12%; Males: 45.11 ± 0.12%; *t-test, p<0.001*), as well as reduced sleep in both the

light (Females: 62.13 ± 0.16%; Males: 63.01 ± 0.15%; *t-test, p<0.001*) and dark periods (Females: 21.53 ± 0.17%; Males: 27.21 ± 0.18%; *t-test, p<0.001*) compared to B6NJ males, though it is less pronounced during the light phase as compared to dark phase (Fig 2.3A). Similar to sleep duration patterns, females also show shorter bout lengths measured across 24 h (Females: 369.68 ± 2.06 s; Males: 444.78 ± 2.55 s; *t-test, p<0.001*), and during both the dark (Females: 198.57 ± 1.44 s; Males: 275.85 ± 2.2 s; *t-test, p<0.001*) and light phases (Females: 539.29 ± 3.03 s; Males: 619.75 ± 3.55 s; *t-test, p<0.001*) (Fig 2.3B). Overall, B6NJ female mice have reduced sleep duration and shorter bout length than their male counterparts (Table 2.1).

Piezo Variable	Female	Male	p_value
Sleep Daily Percent	41.83 +- 0.12	45.11 +- 0.12	p<0.001
Sleep Dark Phase Percent	21.53 +- 0.17	27.21 +- 0.18	p<0.001
Sleep Light Phase Percent	62.13 +- 0.16	63.01 +- 0.15	p<0.001
Sleep Bout Lengths Mean	369.68 +- 2.06	444.78 +- 2.55	p<0.001
Dark Sleep Bout Lengths Mean	198.57 +- 1.44	275.85 +- 2.2	p<0.001
Light Sleep Bout Lengths Mean	539.29 +- 3.03	619.75 +- 3.55	p<0.001
Breath Rate During Sleep Mean	2.58 +- 0.01	2.82 +- 0.01	p<0.001

Table 2. 1.: Sleep comparison between male and female control mice. Comparison of mean and standard deviation values in males and females for piezo variables in control mice. P-values were calculated by comparing means via *t-test.*

A.



B.

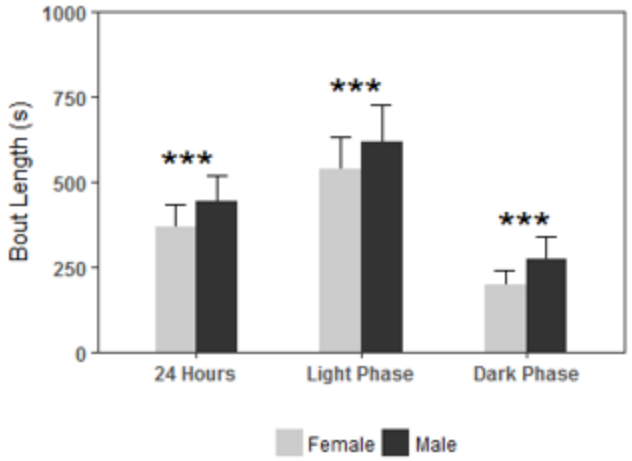
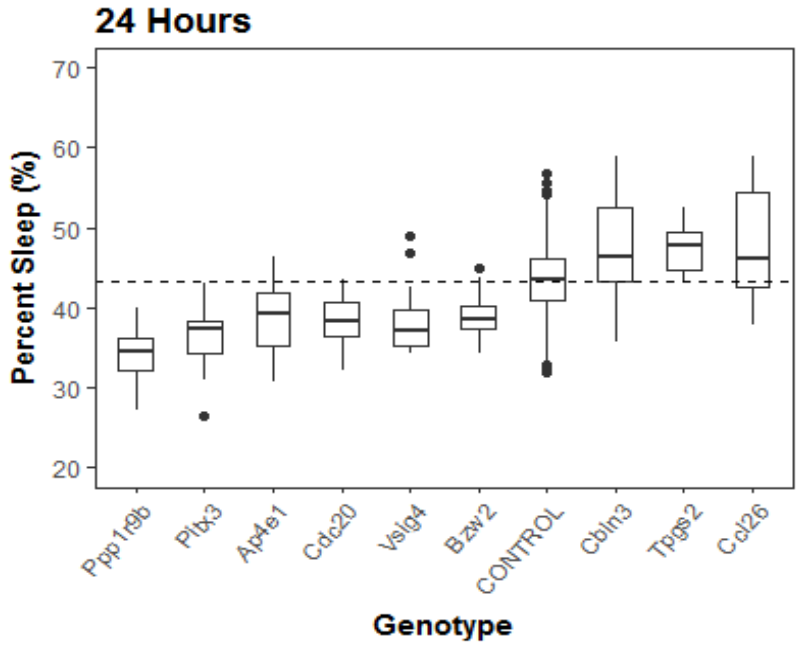


Figure 2. 3.: Sleep–wake patterns in control mice under baseline conditions. Average percent sleep across three consecutive days analyzed over (A) 24 h, dark phase, and light phase. Female mice show reduction in sleep duration across 24 h and during the light and dark phase. (B) depicts average bout length in seconds (s) over 24 h, dark phase, and light phase. Females had shorter average bout lengths across all phases. Error bars represent Standard Deviation. ***P < 0.001 in *t-test*.

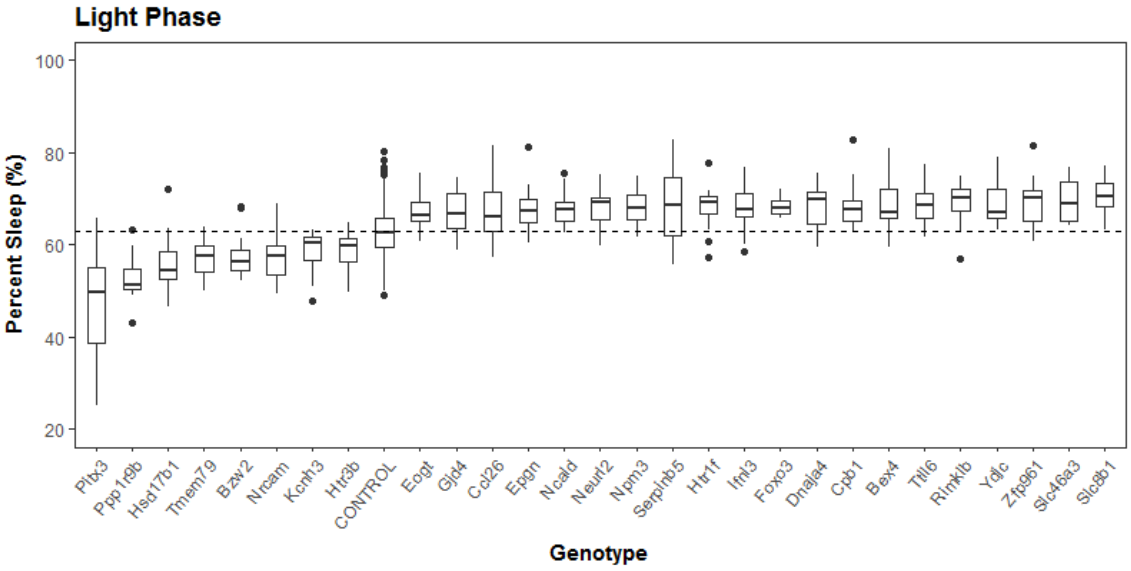
Knockout mice:

This report presents the analysis of piezo system sleep recordings from 343 KO strains. Of these, 55 KO strains showed significance at $p<0.05$ (adjusted for FWER) in Dunnett's post hoc analysis for one or more of the sleep variables recorded (Figure 2.4, Supplementary Table A.1). Across the 24-hour period, reduction in sleep percent (total sleep) was observed in *Ppp1r9b, Pitx3, Ap4e1, Cdc20, Vsig4, and Bzw2* KO strains, while increased total sleep was seen in *Cbln3, Tpgs2, and Ccl26* compared to control mice. During light phase, the reduced sleep percent was recorded in *Pitx3, Ppp1r9b, Hsd17b1, Tmem79, Bzw2, Nrcam, Kcnh3,* and *Htr3b* and longer sleep duration in *Eogt, Gjd4, Ccl26, epgn, Ncald, Neurl2, Npm3, Serpinb5, Htr1f, Ifnl3, Foxo3, Dnaja4, Cpb1, Bex4, Ttll6, Rimklb, Ydjc, Zfp961, Slc46a3,* and *Slc8b1*. KO mice. During dark phase, sleep durations was reduced for *Ppp1r9b, Vsig4, Rimklb,* and *Mylip* knockouts, and increased for *Cbln3, Macrod2,* and *Postn*. Additionally, as compared to controls, mean bout length was significantly reduced across 24h in *Pitx3, Hsd17b1, Myh1, Rnf10, Myo3b, Ap4e1,* and *Ppp1r9b*, and increased in *Tmem136* mutant mice. In light phase, bout lengths were significantly shorter in *Pitx3, Hsd17b1, Ptpru, Myh1, Ppp1r9b, Ap4e1,* and *Nrcam* and longer in *Arrb2, Adck2, Slc8b1, Zfp961, Htr1d, Zbtb4, Tmem136, Emp1,* and *Ipp* relative to control mice. During dark phase, significantly shorter mean bout lengths were seen in *Stx16, Rimklb, Rnf10, Myo3b, Bex4, Ppp1r9b, Tmem151b, Zzef1, Ap4e1, Rab27b, Tmod2, Nes* and *Mylip,* and longer bout lengths were found in case of *Ghrhr* (Figure 2.3, Supplementary Figure A.1).

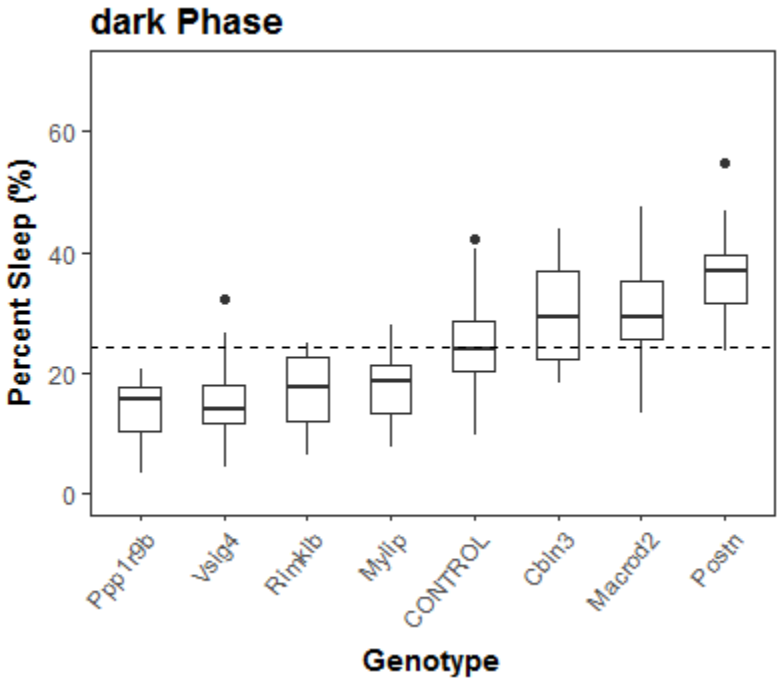
A.


24 Hours
Percent Sleep (%)
70
60
50
40
30
20
Ppp1r9b
Pitx3
Apbb1
Cdc20
Vsig4
Bzw2
CONTROL
Cbln3
Tpgs2
Ccl26
Genotype

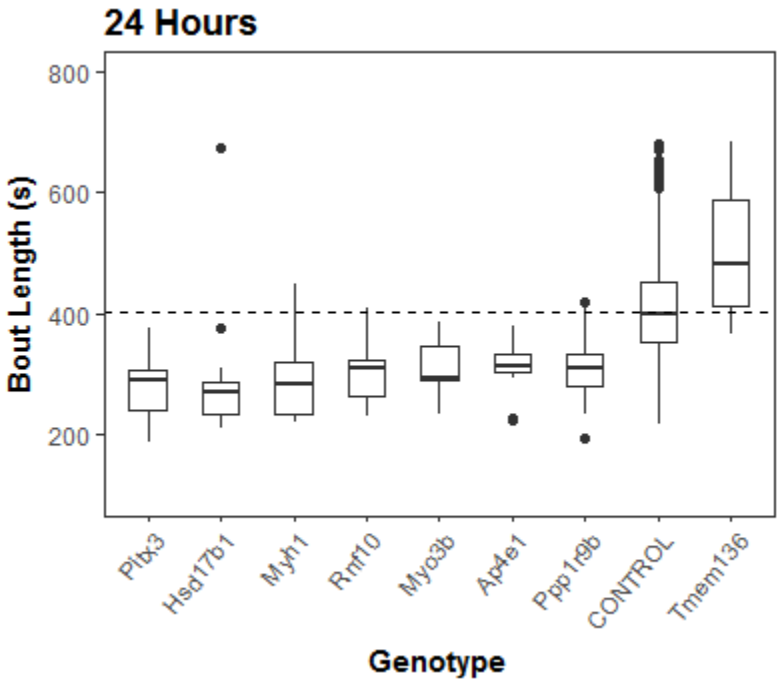
B.


Light Phase
Percent Sleep (%)
100
80
60
40
20
Pitx3
Ppp1r9b
Hsd17b1
Tmem79
Bzw2
Nrcam
Kcnh3
Htr3b
CONTROL
Eogt
Gjd4
Ccl26
Epgn
Ncald
Neurl2
Npm3
Serpinb5
Htr1f
Ifnl3
Foxo3
Dnaja4
Cpb1
Bex4
Tll6
Rlnklb
Ydjc
Zfp961
Slc46a3
Slc8b1
Genotype

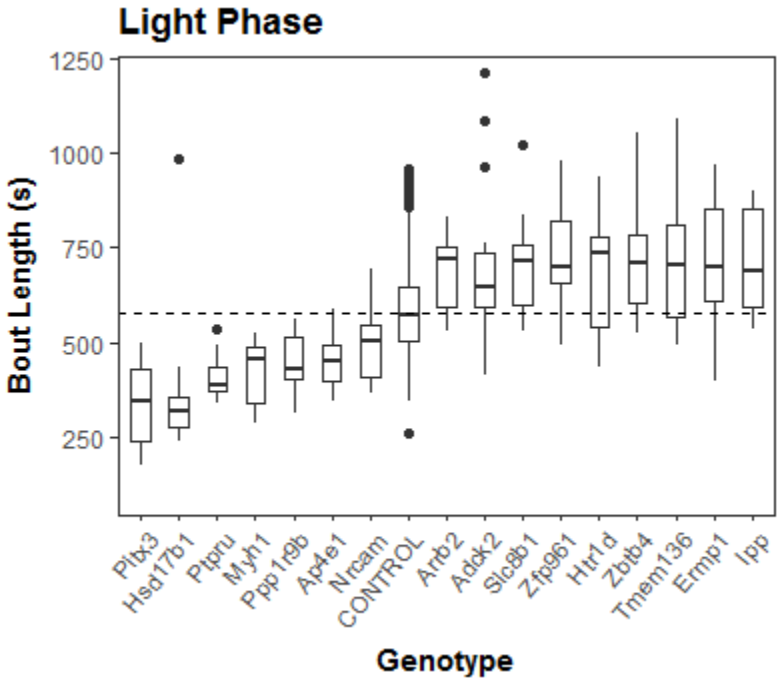
C.


dark Phase
Percent Sleep (%)
0
20
40
60
Ppp1r9b
Vsig4
Rimklb
Mylip
CONTROL
Cbln3
Macrod2
Postn
Genotype

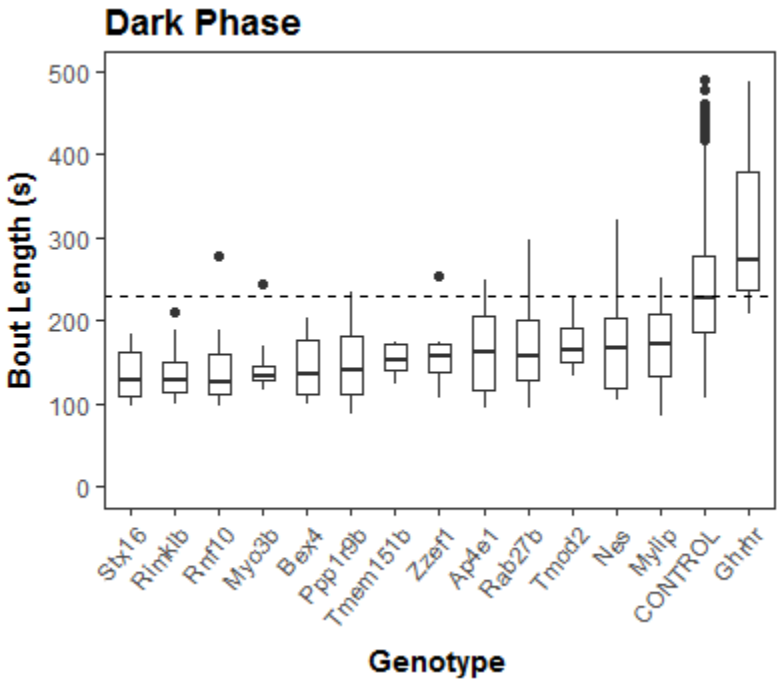
D.


24 Hours
Bout Length (s)
200
400
600
800
Pitx3
Hsd17b1
Myh1
Rnf10
Myo3b
Ap4e1
Ppp1r9b
CONTROL
Tmem136
Genotype

E.


Light Phase
Bout Length (s)
1250
1000
750
500
250
Pitx3
Hsd17b1
Ptpru
Myh1
Ppp1r9b
Ap4e1
Nrcam
CONTROL
Arrb2
Adck2
Slc8b1
Zfp961
Htr1d
Zbtb4
Tmem136
Ermp1
Ipp
Genotype

F.


Dark Phase
Bout Length (s)
500
400
300
200
100
0
Stx16
Rimklb
Rnf10
Myo3b
Bex4
Ppp1r9b
Tmem151b
Zzef1
Ap4e1
Rab27b
Tmod2
Nes
Mylip
CONTROL
Ghrhr
Genotype

G.

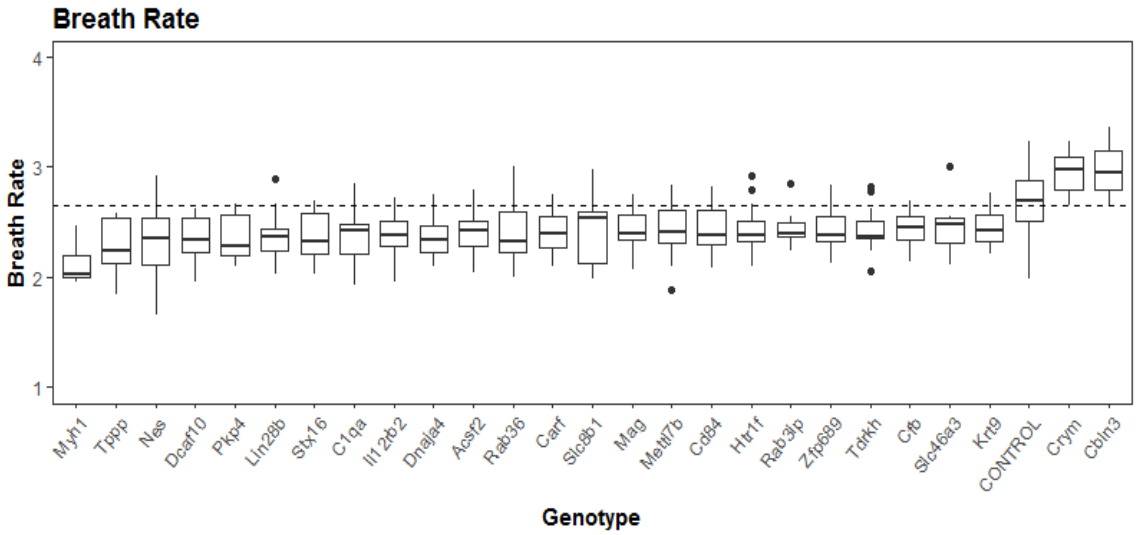


Figure 2. 4. Gene KO that differed significantly from control mice in percent sleep and mean bout lengths. Percent Sleep: 24 h (A), in light phase (B) and in dark phase (C); Bout Lengths: 24 h (D), in light phase (E) and in dark phase (F). (G) KO strains that differed significantly compared to control mice in their breathing rates.

Genes affecting specific sex

Similar analysis (as described above) was performed for 289 KO strains, each of which had at least 3 females and 3 males to evaluate sex-specific differences in sleep parameters measured. Along with significant simple effects of genotype and sex, a significant interaction effect was also observed for sleep percent and mean bout lengths over 24 hours, light phase and dark phase. Effect size was also calculated for each piezo parameter, which represents the strength of relationship between variables and tell about how much of an effect sex, genotype, or interaction had on a sleep parameter. Effect sizes have been reported as eta squared which represents the variance in piezo variable explained by either genotype, sex, or interaction between them. According to Cohen's

guidelines, effect sizes can be defined as small (0.01), medium (0.06), and large (0.13) (Cohen 1992). Based on this criterion, the interaction between genotype and sex, though significant had an intermediate effect with values nearly 0.05 for all piezo variables (Table 2.2).

Piezo Variable	Effect Size ($\eta^{2)}$	p-value
Daily Sleep Percent (24 h)	0.061	p<0.05
Light Phase Sleep Percent	0.053	p<0.05
Dark Phase Sleep Percent	0.052	p<0.05
Bout Length mean (24 h)	0.047	p<0.05
Light Phase Bout Length	0.044	p<0.05
Dark Phase Bout Length	0.051	p<0.05
Breath Rate during Sleep	0.046	p<0.05

Table 2. 2.: Effect size and p-values for sex-genotype interactions in piezo variables

For the overall sleep percent over 24 hours, *Pitx3* and *Ppp1r9b* had significantly reduced sleep in both males and females. Among males, significant reduction in sleep was also observed in *Cdc20, Chn1, Mettl7b,* and *Ptpn5*, and in females, it was reduced in *Ap4e1, Col18a1,* and *Tmem79,* and was significantly increased in *Cbln3, Ccl26, Rab24,* and *Ydjc*. During light phase, sleep percent is reduced in *Pitx3*, and *Ppp1r9b,* and increased in *Slc8b1* in both males and females. Among males, the genotypes *Cdc20, Bzw2, Hsd17b1, Parp16,* and *Prom2* had reduced sleep durations, and *Dnaja4, Epgn, Epha10, Htr1f,* and *Zfp961* had significantly increased sleep percent. Among the genotypes specific to females, *Col18a1, Kcnh3,* and *Nrcam* had reduced sleep percent, and *Ccl26, Ydjc, Cpb1, Npm3, Serpinb5, Slc46a3,* and *Zbtb4* had increased sleep percent during light phase. In dark phase, sleep durations were significantly reduced in *Ppp1r9b,* and increased in *Macrod2,* and *Postn* in both males and females. *Mettl7b, Ptpn5, Mylip*

and *Tmod* had lower values, and *Parp8* and *Postn* had significantly higher values as compared to control in males. Among the females, *Arrb2, Vsig4,* and *Zfyve26* had reduced sleep percent, and *Myh1*, and *Cbln3* had increased sleep percent values as compared to controls during dark phase. Mean bout lengths over 24 hours had significantly reduced values for *Pitx3* and *Hsd17b1* in both males and females. In males, Mean bout lengths were significantly lower for *Mettl7b, Ppp1r9b, Ap4e1, Rimklb,* and *Rnf10*, and higher for *Nfatc4*. Among females, these were lower for *Myh1*, and *Nes*, and higher for *Ghrhr, Prss56,* and *Tmem136*. During light phase, mean bout lengths were found to be significantly lower in *Pitx3* and *Hsd17b1* in both males and females. They were also reduced in *Mettl7b* in males and *Ptpru* in females, and increased in *Nfatc4* and *Nat1* genotypes in males, and *Slc8b1, Adck2, Zbtb4, Ipp,* and *Nrn1l* in females. For mean bout lengths during dark phase, there were no genotypes that affected both males and females. Among the genotypes specific to males, the values were significantly lower for *Ppp1r9b, Zfp961, Tmod2, Ap4e1, Rimklb, Rnf10, Stx16, Tmem151b,* and *Zzef1*, and none of the genotypes had significantly higher values. In females, lower values as compared to control were seen for *Pitx3, Hsd17b1*, and *Ptpru*, and higher values were seen for *Slc8b1*, *Adck2, Zbtb4*, *Ipp,* and *Nrn1l* genotypes (Supplementary Table A.2).

Principal Component Analysis

We performed additional analyses to identify sleep related genes based on multiple measures using Principal Component Analysis (PCA). We included standardized values for sleep percent and bout lengths for 24 hours, light phase and dark phase in our analysis. For our dataset, the first three principal components explained for more than

95% variability in the data, with PC1 accounting for 46.75% of the variability (Figure 2.5B). Correlation between the principal component and the original variables is described in terms of loadings. There is a point of inflection in the eigenvalues after 4 components (Figure 2.5A). However, fourth factor has eigenvalue less than 1, indicating that 3 components may best represent the data.

The first PC1 had high loading values for each variable (mean loading = 0.4; Table 2.3). ANOVA with multiple comparison through Dunnett's post hoc test of PC1 resulted in 7 genotypes that showed significance. These genotypes are *Ap4e1, Cdc20, Hsd17b1, Myo3b, Pitx3, Ppp1r9b,* and *Rnf10*. Interestingly, the PC2 had positive loadings for sleep percent and negative loadings for bout lengths, and *Arrb2, Cbln3, Dcaf10, Macrod2, Myh1, Parp8, Ppp1r9b, Rnf25, Tmem79,* and *Tpgs2* were significant. Similarly, PC3 had positive loadings for light phase and negative loadings for dark phase. The genes significant for PC3 include *Adck2, Ajap1, Arf2, Bex4, Cldn13, Cpb1, Dnaja4, Dnajc14, Eogt, Epgn, Epha10, Gipc3, H1fx, Hsd17b1, Htr1f, Ifnl3, Ipp, Kcnh3, Macrod2, Mylip, Nfatc4, Npm3, Nrcam, Parp16, Pitx3, Postn, Rab3, Rimklb, Serpinb5, Slc1a1, Slc46a3, Slc8b1, Stx16, Tdrkh, Tmod2, Ttll6, Zfp219,* and *Zfp961*. Based on the results from ANOVA of individual variables and PC3, we propose these genes as candidate genes that affect sleep in a specific circadian phase.

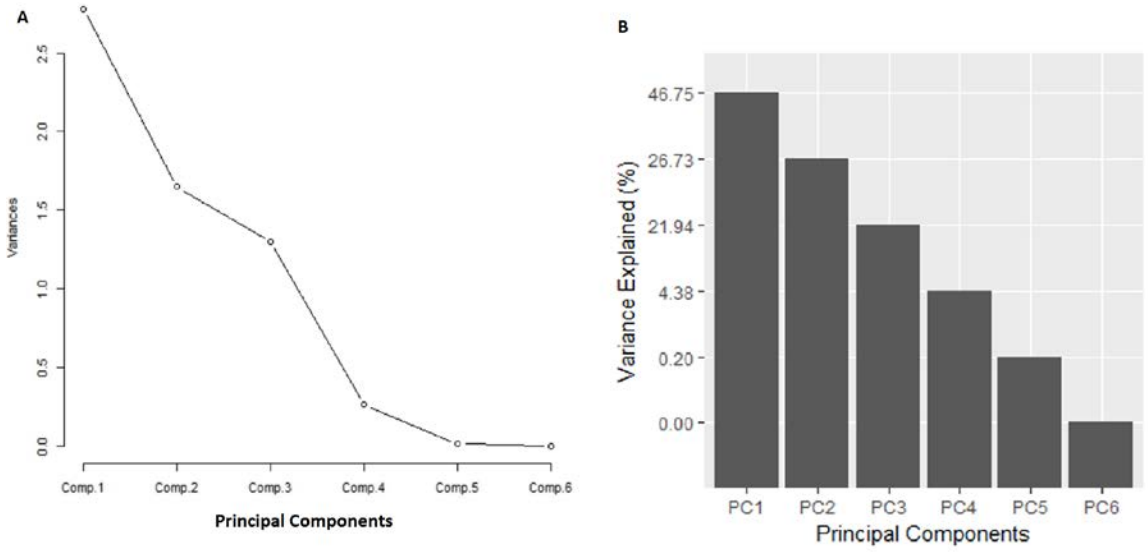


Figure 2. 5.: Principal Component Analysis Plots

A. Eigenvalue plot for principal components. B. Variance explained by individual principal components.

Piezo Variables	PC1	PC2	PC3	PC4
Sleep Daily Percent	-0.433	-0.536	-0.002	-0.015
Sleep Light Phase Percent	-0.347	-0.434	0.504	0.337
Sleep Dark Phase Percent	-0.287	-0.35	-0.632	-0.446
Sleep Bout Length Mean	-0.467	-0.482	-0.034	-0.187
Light Phase Bout Length Mean	-0.446	-0.291	0.439	-0.434
Dark Phase Bout Length Mean	-0.438	-0.291	-0.392	0.681

Table 2. 3.: Loading values for PCA

Outlier Analysis

The analysis through ANOVA and PCA is based on the differences between individual genotypes and the controls and it identifies genes that are deviating from the normal. As an additional strategy to identify genes affecting sleep, MD for multivariate outliers were calculated for means of sleep percent and bout length variables of all KO

strains. Candidate genes were identified with MD above cutoff value of 3.80 (at default quantile threshold of 0.975). The advantage with this analysis is that it helps in identifying genotypes that have extreme values that might have been missed earlier while comparing with controls. And if genes that were significant earlier appear again this analysis, then it acts as additional validation of results. Based on outlier analysis, the top candidate genes for sleep are *Akr1d1, Cacna2d3, Enox1, Fndc4, Galnt12, Gpr156, Hsd17b1, Ipp, Masp1, Mos, Myh1, Myo3a, Nek2, Ovch2, Pfdn6, Pitx3, Postn, Ppp1r9b, Ptpru, Rab3c, Rab3gap2, Rbm4b, Serpinb5, Srcin1, Tas2r138, Tmem79, Tnfsf18, Tubb4a,* and *Vsig4.*

Breath rate during sleep

Breath rate values during sleep were analyzed using the same methods and analytical approach as described above, genes affecting breath rate during sleep, and genes affecting breath rate in specific sex were identified. Significant difference was found in breath rate values between genotypes ($p<0.05$, *ANOVA*). In Dunnett's post-hoc analysis, 26 genotypes significantly differed from control with reduced values in *Myh1, Tppp, Nes, Dcaf10, Pkp4, Lin28b, Stx16, C1qa, Il12rb2, Dnaja4, Acsf2, Rab36, Carf, Slc8b1, Mag, Mettl7b, Cd84, Htr1f, Rab3ip, Zfp689, Tdrkh, Cfb, Slc46a3,* and *Krt9,* and increased valued for *Crym,* and *Cbln3* (Figure 2.4G). Among these, *Acsf2, Lin28b, Myh1, Nes, Rab36,* and *Tppp* were significant for both females and males. Genes specific to males include *Mettl7b, Ppp1r9b, Epgn, Postn, Stx16, Tmem151b, Zzef1, Adck2, Ajap1, BC030499, C1qa, Carf, Cers5, Ces4a, Cfb, Dcaf10, Il12rb2, Loxl1, Mag, Pkp4, Rab3ip, Rxfp4, Slc1a1, Tdrkh, Thsd1,* and *Zfp689* in which breath rate during sleep was

significantly reduced and Prokr1 that had values significantly higher than control mice. In female mice, breath rate was significantly higher than control mice in *Cbln3, Crym*, and *Igsf11*.

Key Candidate Genes for Sleep

Thus far gene lists were prepared for ANOVA, PCA, and MD analysis. Here PCA and MD can be considered more of validation steps rather than independent analysis. Genes that constantly appear in 2 or more of these lists can be considered as strong sleep related candidates. The gene lists prepared for ANOVA, PCA and MD analysis were analyzed to identify genotypes that appear in all of them. We consider them to be key candidate genes affecting sleep percent and bout lengths over 24 hours and light and dark phases. These genes are *Hsd17b1, Ipp, Myh1, Pitx3, Postn, Ppp1r9b, Serpinb5,* and *Tmem79* (Figure 2.6). A final list of 122 unique sleep related candidate genes was created by combining results from all our analysis.

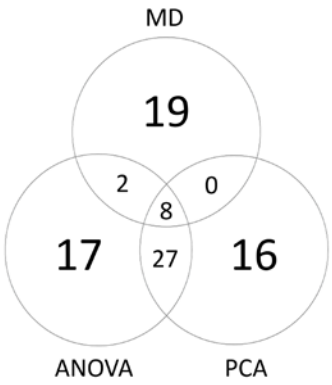


Figure 2. 6.: Venn diagram representing number of genes from ANOVA, PCA, and Outlier analyses

Gene Network Analysis

With the goal of investigating relationships between results from our analysis and some of the known circadian and sleep-influencing genes, network analysis was performed to find associations between genes through large set of data that include protein and genetic interactions, pathways, co-expression, co-localization, protein domain similarity, and predicted networks by OPHID (Online Predicted Human Interaction Database). The genes selected for this analysis are known circadian clock genes showing differential expression during sleep and wake (*Clock, Bmal, Per1,* and *Per2*), and *Homer1* which is associated with homeostatic regulation of sleep. Two recently reported genes *Sik3* and *Nalcn* that regulate non-rapid eye movement (NREM) and rapid eye movement (REM) sleep were also included in the network analysis. Gene networks for each of aforementioned genes in combination of our candidate genes was created and the nearest neighbors were selected. Each of our candidate genes was assessed for the number of previously known genes it interacted with (Supplementary Table A.3). *Kcnh3* was found to interact with all four circadian genes. Cacna2d3 also had four interactions with *Per1, Clock, Bmal1,* and *Nalcn*. Two interactions were observed for *Prom1* (with *Per2* and *Clock*), *Tpgs2* (with *Clock* and *Bmal1/Arntl*), and *Myo3a* (with *Bmal1/Arntl* and *Nalcn*) (Figure 6).

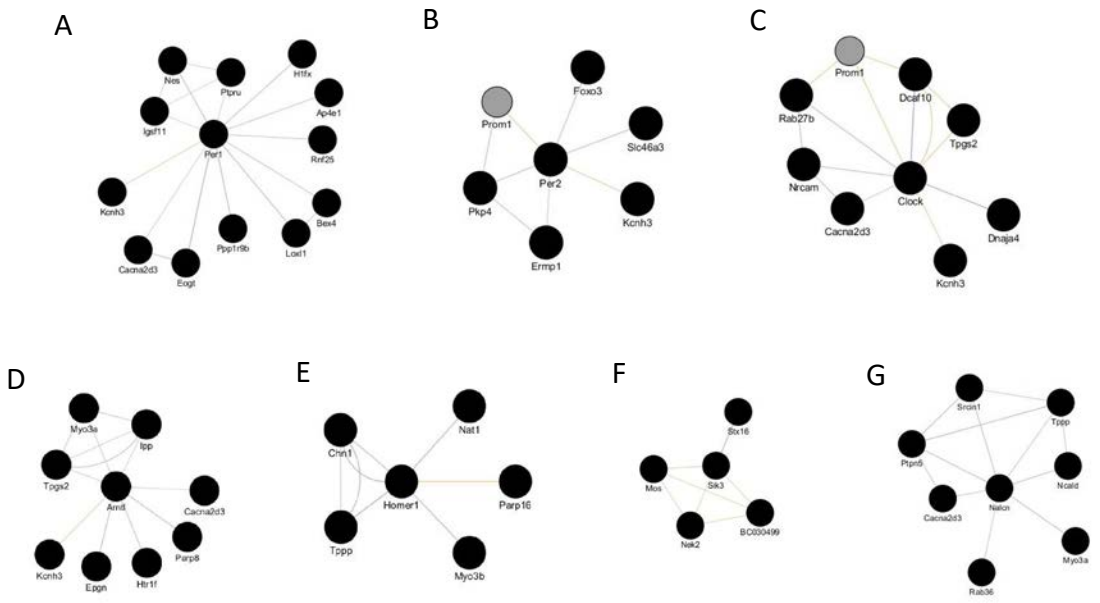


Figure 2. 7.: Gene networks predicted by GeneMANIA

Gene networks were created for previously known circadian and sleep related genes : A. Per1, B. Per2, C. Clock, D. Bmal1/Arntl, E. Homer1, F. Sik3, G. Nalcn. These networks are based on genetic interactions with each node representing single gene, and the edges represent physical associations, shared protein domains, co-expression, co-localization and their participation in the shared pathways for the candidate genes of interest.

Phenotype associations with IMPC database

Given that sleep is a complex trait and almost a third of the genes being studied in this analysis affect sleep, pleiotropy needs to be considered. Abundant pleiotropy has been reported for complex traits (Sivakumaran, Agakov et al. 2011). Some of the potential sleep regulating genes identified in our study can be broadly categorized into transcription factors (*Pitx3, Foxo3*), immune system related (*Ccl26, Vsig4*), membrane transporters (*Slc8b1, Slc46a3)* ion channels (*Kcnh3, Cacna2d3*), neurotransmitters receptors (*Htr3b, Htr1f, Htr1d*), signal transduction (*Ppp1r9b, Rab27b, Ap4e1*), and

metabolism (*Ghrhr*). A wide variety of genes belonging to these functional categories have been previously implicated in influencing sleep parameters. Furthermore, these categories are likely not exhaustive. Several of these genes have brain-associated functions such as myelin formation, neuronal differentiation, synaptic transmission, neuronal signal transduction and yet for many others such as *Adck2* and *Zfp961,* there is little information available in the literature. None of the target genes, except for *Ghrhr and Pitx3,* have been previously implicated in sleep regulation to the best of our knowledge. With a multitude of phenotypes and rich data in the KOMP^2 pipeline, cross trait analysis is a logical approach to understanding the effects of pleiotropy in this study. With a goal of studying what other phenotypes are observed for these genotypes in addition to sleep, phenotype data for all 122 candidate genes was collected and analyzed. In IMPC gene-phenotype associations, the phenotypes are reported as mammalian phenotype (MP) ontology terms. These are the phenotypes that the genes were found to be associated with based on the observations from assays and statistical analysis. Our candidate genotypes had 550 associations with 146 unique MP terms. Among the MP terms that showed high associations with genes, “abnormal sleep behavior” had 37 followed by “abnormal behavioral response to light” with 23, “decreased circulating glucose level” with 22, “decreased total body fat amount” with 20, “abnormal behavior” with 16, and “hyperactivity”, “abnormal bone structure”, and “decreased bone mineral content” with 14 genes each. When considered in terms of phenotype ontologies, more than half of the 550 associations observed belonged to either behavioral/neurological phenotype (167 associations) or homeostasis/metabolism phenotype (117 associations). These were followed by hematopoietic system phenotype with 65, growth/size/body

region phenotype with 45, skeleton phenotype with 40, cardiovascular system phenotype with 30, adipose tissue phenotype with 20, and vision/eye phenotype with 16 associations (Supplementary Table A.5, A.6).

Functional annotation of candidate genes

With the goal comparing our candidate genes with pre-existing functional genomic data, Geneweaver software system (Bubier, Langston et al. 2017) was used. Geneweaver contains a database and a suite of tools that enable cross species and cross platform comparison of data from multiple genomic experiments. The utility of the database for us lies in the fact that it provides data in the form of gene lists and computational tools to perform analysis that help in identifying genes that have been implicated previously in the phenotype of user's interest. By entering the keyword "sleep" as the search term, we were able to identify 26 genesets of interest from mouse and human. Using Jaccard Similarity Tool for pairwise comparisons, 10 genesets were identified that contained one or more of candidate genes we identified through our analysis. Six datasets that are "Sleep Disorders", "Sleep Stages", "Sleep Deprivation", "Abnormal Sleep Pattern", "REM Sleep", and "Abnormal frequency of paradoxical sleep" contained only one gene that was the Growth Hormone Releasing Hormone Receptor (*Ghrhr*). The other 4 genesets were QTL data for Dps (Delta power in slow wave sleep), and Rapid Eye Movement (REM) Sleep. *Tmem151b, Tubb4a, Cfb,* and *Pfdn6* were in the QTL region for Dps3 on chromosome 16, and *Tppp* in Dps1 QTL on chromosome 13. *Galnt12* and

Dcaf10 were in RemSlp1 QTL on chromosome 4, and *Htr1f*, *Igsf11*, and *Gpr156* RemSlp3 QTL on chromosome 16 (Supplementary Figure A.2).

Discussion

We have reported here the results from a large-scale phenotyping study that systematically assesses sleep in knockout mice. With sleep wake recordings from more than 6000 mice, consisting of more than 1800 controls and 343 gene knockouts, the KOMP^2 pipeline at The Jackson Laboratory has generated a wealth of information-rich gene centric data that is unprecedented in sleep research. Through our analysis, we have been able to identify sleep related genes that have been previously unknown to influence sleep. An experimental design using both sexes, and piezo monitoring system that consistently reports multiple sleep variables, has allowed for an in-depth analysis to identify genes that affect sleep in one or both sexes and during specific times of day. We found a considerably high hit rate with nearly a third (35%) of gene knockouts tested identified as candidate genes related to sleep behavior. This is perhaps not surprising given the dramatic changes in brain physiology and function that occur with sleep and wake transitions, and further suggests a potential to uncover many as yet unappreciated pathways affecting sleep. Although the purpose and functions of sleep are still unclear, we know that sleep is an essential physiological process and that even modest reductions in sleep have substantial effects on health and cognitive functions (Shaw and Franken 2003, Durmer and Dinges 2005, Cohen, Doyle et al. 2009, Cespedes, Bhupathiraju et al. 2016).

We identified several genes that had significant effect on sleep only during a specific circadian phase. Pitx3 (Paired like homeodomain transcription factor 3), Bzw2 (Basic Leucine Zipper W2 Domains 2), and Ccl26 (Chemokine (C-C motif) ligand 26) affect sleep durations only during the light phase, and Vsig4 (V-set and immunoglobulin domain containing 4), and Cbln3 (Cerebellin 3) during dark phase. Similarly, Pitx3, Myh1 (Myosin, heavy polypeptide 1), and Tmem136 (Transmembrane protein 136) affect bout lengths only during light phase, and Myo3b (Myosin IIIB), and Rnf10 (Ring finger protein 10) during dark phase. Most notable among these is the Pitx3 gene that affects both sleep percent and bout length only during light phase. Although little is known about the role of Pitx3 and sleep, Derwinska et al have reported that a hemizygous deletion on chromosome 10 involving Pitx3 resulted in sleep disturbances beginning early childhood in a Caucasian boy (Derwinska, Mierzewska et al. 2012). In addition, Pitx3 is well known for its role in regulating lens development and is therefore associated with ocular abnormalities as seen in a range of animals including xenopus, zebrafish, humans as well as mice where its deficiency is reflected as a form of aphakia (Khosrowshahian, Wolanski et al. 2005, Shi, Bosenko et al. 2005, Huang and He 2010). In the KOMP^2 pipeline, in addition to sleep, *Pitx3* is associated with a multitude of phenotypes including vision/eye, neurological/behavior, growth/size, homeostasis/metabolism, cardiovascular and skeleton. The *Pitx3* KO mice have anophthalmia or absence of eyes. In mice, *Pitx3* is also thought to be essential in development of dopaminergic neurons in Substantia Nigra (SN). Besides SN, Pitx3 is also expressed in Ventral Tegmental area (VTA) Reviewed by (Li, Dani et al. 2009). Not only these regions are associated with reward, addiction and movement, they also play an

important role in sleep and alertness reviewed by (Nishino 2013). *Pitx3* is the genotype that has the most significant reduction in sleep percent and bout lengths. These low numbers cannot be attributed to blindness since there is no evidence of effects on sleep duration caused by blindness. There might be sleep difficulties caused by circadian disorders like non-24 in unsighted individuals, but that does not necessarily translate to lower sleep durations (Weber, Cary et al. 1980). Another interesting aspect is that 16 sleep related candidate genotypes, including Pitx3, Cbln3, Hsd17b1, and Rnf10, have a vision/eye phenotype. These genotypes do not suffer from anopthalmia like Pitx3 mice, but have been associated with several other eye phenotypes. These include 'abnormal retina morphology', 'abnormal retinal pigmentation', 'persistence of hyaloid vascular system', 'abnormal eyelid aperture', 'impaired pupillary reflex', 'mydriasis', 'abnormal lens morphology', 'decreased cornea thickness', and 'fused cornea and lens' (Supplementary Table A.5). That makes it important to further investigate the role of eye morphology in maintenance of sleep architecture.

Along with *Pitx3*, *Ppp1r9b* and *Ap4e1* are the candidate genes that were found to be significant for both sleep percent and bout length. *Ppp1r9b* affects both sleep percent and bout lengths over 24 hours, light phase and dark phase. *Ppp1r9b* is the Protein Phosphatase 1 Regulatory Subunit 9B (or Neurabin II or Spinophilin), and as the name indicates is a regulatory subunit of protein phosphatase (PP1). It is a protein highly enriched in dendritic spines (Feng, Yan et al. 2000). Sleep is reported to promote formation of dendritic spines for memory consolidation (Yang, Lai et al. 2014). In addition, PP1 regulates AMPA channels that are believed to play a role in synaptic plasticity, and learning and memory (reviewed in (Prince and Abel 2013). Further,

Ppp1r9b is one of the substrates of *GSK3B* (Glycogen Synthase Kinase 3 Beta), which is a crucial circadian clock regulator (Kaasik, Kivimae et al. 2013). Casein Kinase I enzymes have been shown to play a critical role in regulating clock genes such as *Per2*, and *Ppp1r9b* may dephosphorylate some of these same sites and work in opposition (Fukuyama 2003 and Padiath 2005). There is increasing evidence that clock genes not only influence circadian aspects of sleep and wake, but are fundamentally tied to sleep homeostasis as well, which appears to be altered in the *Ppp1r9b* knockout mice (Franken, Thomason et al. 2007, Franken 2013). *Ap4e1* was found to be statistically significant for 24 hours, light phase and dark phase for bout length, and sleep percent over 24 hours. It codes for the Epsilon subunit 1 of Adaptor Protein (AP) 4 complex that is involved in vesicle trafficking. *Ap4e1* has been associated with Cerebral Palsy, and mutations in humans have been known to cause intellectual disabilities, of which abnormal sleep behavior is one of the symptoms (Moreno-De-Luca, Helmers et al. 2011). Like *Pitx3*, *Ap4e1* is associated with multiple additional aberrant phenotypes observed in the KOMP^2 pipeline (Supplementary Table A.6) , and is also found to be within the QTL Cplaq15 (Circadian Period of Locomotor Activity 15) (Hofstetter, Trofatter et al. 2003).

Another KO that showed significant changes in sleep-wake traits is *Kcnh3 (Kv12.2)* a subunit of potassium channels that regulate neuronal excitability. *Kcnh3* KO mice have shorter sleep duration in the light phase, and in gene network analysis, was found to interact with *Per1, Per2, Clock* and *Bmal1* through co-expression. A similar but less pronounced reduction was also seen in *Kcnh3* heterozygous mice (data not included). Its overexpression has been associated with deficits in learning, and its ablation with enhanced cognitive functions (spatial and working memory), hippocampal

hyperexcitability and spontaneous seizures (Miyake, Takahashi et al. 2009, Zhang, Bertaso et al. 2010). Many other Kv channels are known to modulate sleep-wake. A well-known example is Shaker in drosophila (Cirelli, Bushey et al. 2005). Flies mutant for the *Shaker* gene have reduced sleep and are short sleepers. In *Kcnc1, Kcnc3, Kcnc1/3 and Kv1.2* KO mice less NREM sleep has been observed, with a similar magnitude to sleep phenotypes observed for Kcnh3 reported here (Rechtschaffen 1998, Cirelli 2009). Hence, our study identifies and adds another novel potassium channel associated gene that affects sleep. Lack of certain K channels may reduce the resting membrane potential of neurons, leading to increased firing, and reduced sleep. While this would occur in both inhibitory and excitatory circuits, the net effect is presumably increased excitation. Recently, Ding et al (Ding, O'Donnell et al. 2016) has shown that reduction seen in brain extracellular potassium ion levels is associated with sleep and anesthetized mice supporting this hypothesis.

Several genes were found to affect breath rate. Majority of these genes affected breath rate only as is the case with *Tppp* (p25 alpha/p24) mice, which had shorter breath rate as compared to control mice. It has been suggested that *Tppp*, a tubulin-binding protein plays an important role in oligodendrocyte differentiation (Lehotzky, Lau et al. 2010). In addition, *Tppp* also functions as a glycogen synthase kinase 3 inhibitor (Martin, Vazquez et al. 2002). We don't know yet if breath rate regulating mechanisms are in any way associated with other brain functions or sleep-related pathways, but the ability to assess breathing should provide additional variables of interest to the larger efforts to understand the multiple roles of all protein coding genes.

Sex differences were seen for many of the KO lines as well as in B6NJ controls throughout our study. Previous studies in mice and humans also report similar sex differences in sleep and circadian rhythms, although with relatively small sample sizes. Sex differences in sleep have also been observed in BL6J mice for baseline sleep parameters such as REM and NREM sleep, with lower values observed in female mice (Koehl, Battle et al. 2006, Paul, Dugovic et al. 2006). Sex differences observed in human studies have generally been small, but include higher EEG power density in women (Carrier, Land et al. 2001). EEG profiles in primary insomnia patients also show sex differences (Buysse, Germain et al. 2008). Sex hormones are thought be one of the contributing factors for these differences (Collop, Adkins et al. 2004, Krishnan and Collop 2006, Pavlova and Sheikh 2011). In addition to small but significant sex differences in our control mice, several genotypes had highly significant sleep trait differences for a specific sex. KO strains in which sex differences were found might be helpful in identifying causes for the intriguing sexual dimorphism seen in sleep behavior, such as those in which female mice sleep less than males. In our data, the majority of significant findings are sex specific, suggesting at least some biochemical differences in female vs. male sleep regulation.

There are some limitations to this study. The current analysis software does not distinguish REM sleep from NREM sleep, however, an algorithm to distinguish these sleep states from the piezoelectric recordings is in development and will be utilized in future studies once it has had sufficient validation. The premise for the development of this algorithm is based on differences in breathing rhythm during REM and NREM sleep. Based on our analysis of gene set similarity with REM sleep QTLs and interactions

within networks for *Sik3* and *Nalcn*, we are in a position to propose 16 putative candidates for REM/NREM sleep. *Cacna2d3* (a calcium channel protein) is an interesting candidate since it interacts with *Nalcn* along with circadian Per1, Clock and Bmal1 genes. Five of these genes including *Stx16, Tppp, Rab36, Dcaf10,* and *Htr1f* also significantly affect breath rate during sleep and should be considered first in line for future studies. Based on similar evidence from Homer1 gene network and Dps QTLs, we have also been able to propose 9 novel candidate genes for homeostatic regulation of sleep. Here again, *Tppp* is a key candidate gene since it interacts with *Homer1* in co-expression network, and is also a part of Dps1 QTL.

Although we were able to detect many novel genes, the extensive filtering of sleep signals used in our study most likely excluded genes affecting sleep in subtle ways. Genes involved in sleep that have extensive redundancy or compensatory mechanisms in place to mask the effect of a gene ablation would also be missed. Finally, the knockout method is in general limited by the fact that a gene is ablated in all tissues, so KO of essential genes that may in fact be involved in sleep result in a non-viable animal, preventing sleep phenotyping. However, utilization of the conditional allele obtained from these KOs in the future may overcome this limitation. It is also unclear to what extent any of the sleep alterations from gene ablation are due to direct or indirect effects of the gene in question. These questions may be partially addressed by examining the multiple phenotypes for each knockout, and can be pursued in other ways for the most interesting cases. There is no reason to believe that there are any genes whose sole functions are related to sleep, as even the so-called core circadian clock genes are pleiotropic (Kyriacou and Hastings 2010, Rosenwasser 2010).

Our study demonstrates the utility of rapid-non-invasive sleep phenotyping in high throughput mouse screens. This initial set of approximately 350 genes was not selected to have sleep phenotypes and yet a high percentage were found to have altered sleep phenotypes, and of a magnitude as large as any that have been selected specifically for sleep studies over the past 25 years (Cirelli 2009). This supports the utility of an unbiased selection and phenotyping for mouse knockouts, especially given that a majority of genes are not well understood. Unlike most individual studies of KO mice that examine genes predicted to influence a trait of interest, the IMPC/KOMP^2 is a comprehensive, unbiased approach, having examined more than 2000 genes to date, and thus holds potential for detecting and identifying unexpected and pleiotropic effects of the knocked out genes (Brown and Moore 2012).With fewer than 350 KOs analyzed to date, we demonstrate the potential of this large scale effort to find novel sleep phenotypes for a significant percentage of coding genes.

CHAPTER 3

Ensemble approach to identify phenotype predictors for sleep behavior in KOMP2

Background

Identifying key candidate genes for a trait is important since they can help us identify and understand underlying pathways and biochemical mechanisms. The reverse genetics approach employed in KOMP^2 is a useful resource to understand as to which genes affect a specific phenotype. But it is also a great resource to study the data from a phenotypic standpoint. In the previous analysis to perform gene-phenotype associations between sleep related piezo variables and other phenotypes for genes that were significant, we performed a natural word processing to determine the frequency at which the mammalian phenotype (MP) ontology terms appeared for those genes. The analysis provided key insights including the fact that that a majority of genes significant for sleep related variables were also associated with behavioral/neurological phenotype ontology terms. Additionally, ontology terms for homeostasis/metabolism phenotypes also appeared with high frequency. We previously mentioned that sleep phenotyping is unique to the KOMP^2 pipeline at The Jackson Laboratory (also called JAX pipeline), and it is consistently done at the age of 15 weeks for each mouse. It means that there are data available for a multitude of phenotypes from the previous 14 weeks before a mouse enters the sleep studies and 3 weeks after that. In this part of the project, the driving question is how these phenotypes are affecting sleep. In other words, can we predict an abnormal sleep behavior based on the data available from other phenotypical assays

performed in the KOMP^2 pipeline, some of which may be useful throughout the larger IMPC.

The International Mouse Phenotyping Consortium (IMPC) website (www.mousephenotype.org) is a repository of all data generated from KOMP^2 and related phenotyping pipelines from centers across the world. There is a user interface that allows individual gene or phenotype level searches. Batch query is an option, and the website also provides several tools and visualizations for statistical analysis including all the documentation. Simple and batch querying is effective for basic data searches, but not suitable for complex queries where the output can be large, with several thousand or millions of data points. IMPC also provides APIs for data dissemination. API stands for Application Programming Interface which is a set of routines and protocols for developing software applications, and allows for two software programs to communicate with each other. IMPC APIs allow for remote computers to interact with their servers and gather data through customized queries. IMPC uses RESTful API based on representational state transfer (REST) technology that allows for gathering data through HTTP (Hypertext Transfer Protocol). IMPC REST APIs can be queried through Apache Solr, a Java based open source enterprise search platform that has a variety of search features and runs as a standalone full text search server. The Solr queries can be run standalone, or embedded inside another programming environment like R. The data is gathered in JSON format that can be transferred through the web and is language independent. In simple terms, this allows for any user with an internet access to connect to IMPC servers and gather data according to their requirements. Currently IMPC

provides four APIs that are for genotype-phenotype, statistical results, observation, and images.

Data science is an emerging interdisciplinary field formed from a combination of statistics, data mining, predictive analytics and machine learning with the goal of extracting knowledge and information from varied sources and forms of data. With the advances in computing, there has been an explosion of data that has resulted in datasets that are large in size and complex in formats and structure. These data-sets and the multiple of approaches to utilize and extract useful information is often referred to as Big Data, and is another emerging field. There has been a growing interest in applications of big data analytic approaches for sleep studies(Dean, Goldberger et al. 2016, Bianchi, Russo et al. 2017, Li, Cui et al. 2017). The IMPC currently houses phenotype data from more than a hundred thousand mice with millions of data points, and that is just the number of observations from phenotype assays. The observation and results from each phenotype has its own structure and format depending on the technique being used to gather the data. This complexity is often difficult for traditional processing applications, but newer Data science approaches, geared towards big data analytics, are equipped to deal with data of this size and complexity.

Several statistical approaches are available to perform multivariate analysis on high dimensional data. Multivariate multiple regression modeling helps to relate the dependent variable to several independent variables. By establishing a pattern of relationships between independent and dependent variables, the response can be estimated and the accuracy of the model is assessed. Classification and Regression Trees (CART), also called Decision Trees is another multivariate approach for high

dimensional data that produces a result in the form of a flowchart (Rokach and Maimon 2008). The data is split into branch-like segments with the root node at the top and branches below it. Modeling approaches depending on results from a single run of an algorithm are prone to confounding results because of overshadowing of one strong predictor over all others. Ensemble methods are general terms describing approaches that involve the use of multiple algorithms to gain better predictive performance (Opitz and Maclin 1999, Polikar 2006). Random Forest is one such approach that involves an ensemble of decision trees developed from randomly selected subsets of observations (Zhang, Ma et al. 2012). This helps in examining the contribution of predictors that might not have been significant in the presence of other strong predictor variables. Another advantage is that the random forests provide values for conditional permutation accuracy that helps in assessing conditional variables importance. Bootstrap aggregation (Bagging) takes place while running the algorithm. This means that a random subset of the data is used as a training dataset to develop a tree while the rest of the data remains out of the bag (oob) and acts as a test dataset. The accuracy of the model is assessed at each step on this test data and measured through out of bag (oob) error which is the mean prediction error on each training sample. The random forests have been shown to improve accuracy as compared to a single decision tree on its own.

Methods

Data Collection

Each mouse that enters the phenotyping pipeline in KOMP^2 is assigned a unique mouse id that makes it possible to track the data being collected throughout its life cycle in the pipeline. Our goal was to collect for all phenotypes from mice who underwent sleep phenotyping at The Jackson Laboratory, also referred to as the JAX pipeline. We used IMPC Observation API to query the database for mice from the JAX KOMP^2 pipeline. The total datapoints in IMPC are 33.07 million, with 2.49 million from JAX pipeline. We then queried the database for mouse ids of mice that underwent "Sleep-Wake" procedure. We obtained 65,702 observations for sleep related variables from 4,406 mice (1.43 million observations). We then gathered all the phenotype observation data available for these 4,406 mice in the JAX pipeline (Figure 3.1).

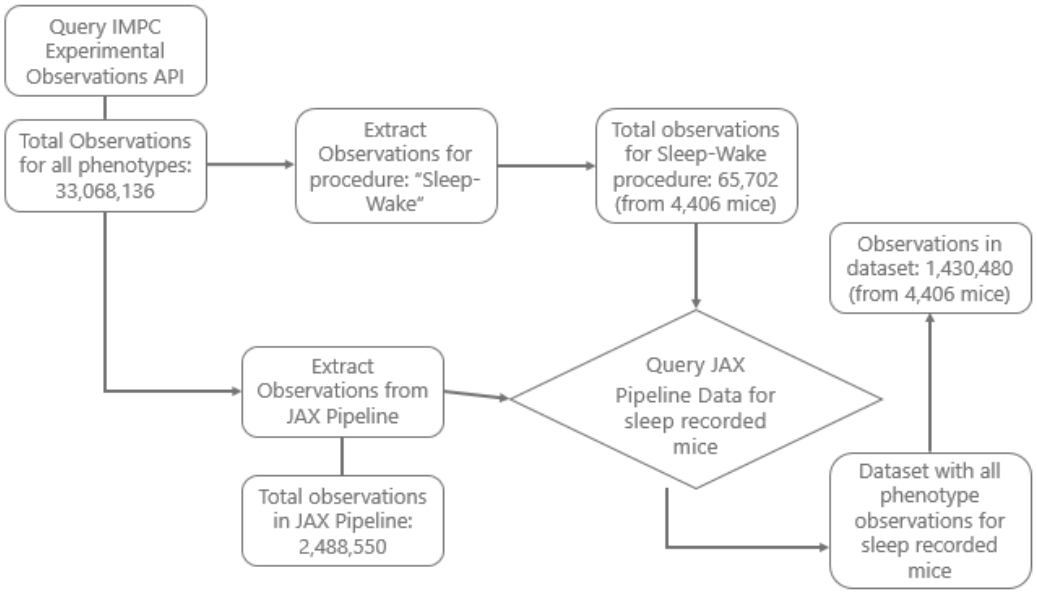


Figure3. 1.: Schematic of data collection procedure through IMPC API

Data Scrubbing and Imputation

IMPC data releases have already undergone substantial quality checks and validation so there was no need for parsing. Data contained 207 phenotype variables from 4,406 mice. However, the data we acquired had a substantial number of missing values. 135 variables with more than 50% missing data were dropped from the analysis, leaving 72 variables that could be included in further analysis (Table 3.1). Multiple imputation was carried out through Bayesian bootstrapping predictive mean matching.

Parameter	Procedure	Week
Locomotor activity	Combined SHIRPA and Dysmorphology	8
Forelimb and hindlimb grip strength normalized against body weight	Grip Strength	8
Forelimb grip strength normalized against body weight	Grip Strength	8
Center average speed	Open Field	8
Center distance travelled	Open Field	8
Center permanence time	Open Field	8
Distance traveled	Open Field	8
Total distance traveled	Open Field	8
Percentage center time	Open Field	8
Periphery average speed	Open Field	8
Periphery distance travelled	Open Field	8
Periphery permanence time	Open Field	8
Whole arena average speed	Open Field	8
Whole arena resting time	Open Field	8
Total holepokes	Hole-board Exploration	9
Dark side time spent	Light-Dark Test	9
Latency to first transition into dark	Light-Dark Test	9
Percent time in dark	Light-Dark Test	9
Percent time in light	Light-Dark Test	9
Side changes	Light-Dark Test	9
Time mobile dark side	Light-Dark Test	9

Time mobile light side	Light-Dark Test	9
% Pre Pulse Inhibition PP1	Acoustic Startle and Prepulse Inhibition (PPI)	10
% Pre Pulse Inhibition PP2	Acoustic Startle and Prepulse Inhibition (PPI)	10
% Pre Pulse Inhibition PP3	Acoustic Startle and Prepulse Inhibition (PPI)	10
Response amplitude BN	Acoustic Startle and Prepulse Inhibition (PPI)	10
Response amplitude PP1	Acoustic Startle and Prepulse Inhibition (PPI)	10
Response amplitude PP1_S	Acoustic Startle and Prepulse Inhibition (PPI)	10
Response amplitude PP2	Acoustic Startle and Prepulse Inhibition (PPI)	10
Response amplitude PP2_S	Acoustic Startle and Prepulse Inhibition (PPI)	10
Response amplitude PP3	Acoustic Startle and Prepulse Inhibition (PPI)	10
Response amplitude PP3_S	Acoustic Startle and Prepulse Inhibition (PPI)	10
Response amplitude S	Acoustic Startle and Prepulse Inhibition (PPI)	10
CV	Electrocardiogram (ECG)	11
Mean R amplitude	Electrocardiogram (ECG)	11
Mean SR amplitude	Electrocardiogram (ECG)	11
Number of Signals	Electrocardiogram (ECG)	11
pNN5(6>ms)	Electrocardiogram (ECG)	11
PQ	Electrocardiogram (ECG)	11
PR	Electrocardiogram (ECG)	11
QRS	Electrocardiogram (ECG)	11
QTc	Electrocardiogram (ECG)	11
QTc dispersion	Electrocardiogram (ECG)	11
rmSSD	Electrocardiogram (ECG)	11
RR	Electrocardiogram (ECG)	11
ST	Electrocardiogram (ECG)	11
HR (Heart Rate)	Electrocardiogram (ECG)	12
HRV (Heart Rate Variability)	Electrocardiogram (ECG)	12
Area under glucose response curve	Intraperitoneal glucose tolerance test (IPGTT)	12
Blood glucose concentration	Intraperitoneal glucose tolerance test (IPGTT)	12
Fasted blood glucose concentration	Intraperitoneal glucose tolerance test (IPGTT)	12
Initial response to glucose challenge	Intraperitoneal glucose tolerance test (IPGTT)	12
BMC/body weight	Body Composition (DEXA lean/fat)	14
Body length	Body Composition (DEXA lean/fat)	14
Body weight	Body Composition (DEXA lean/fat)	14
Body weight curve	Body Composition (DEXA lean/fat)	14
Bone Area (BMC/BMD)	Body Composition (DEXA lean/fat)	14
Bone mineral content excluding skull	Body Composition (DEXA lean/fat)	14
Bone mineral density excluding skull	Body Composition (DEXA lean/fat)	14
Fat mass	Body Composition (DEXA lean/fat)	14
Fat/body weight	Body Composition (DEXA lean/fat)	14
Lean mass	Body Composition (DEXA lean/fat)	14

Lean/body weight	Body Composition (DEXA lean/fat)	14
Sleep bout lengths mean	Sleep-Wake	15
Sleep daily percent	Sleep-Wake	15
6 kHz evoked ABR Threshold	Auditory Brain Stem Response (ABR)	16
12 kHz evoked ABR Threshold	Auditory Brain Stem Response (ABR)	16
18 kHz evoked ABR Threshold	Auditory Brain Stem Response (ABR)	16
24 kHz evoked ABR Threshold	Auditory Brain Stem Response (ABR)	16
30 kHz evoked ABR Threshold	Auditory Brain Stem Response (ABR)	16
Heart weight	Heart weight	18

Table 3. 1.: Variables included in the phenotype predictor analysis.

These variables had less than 50% missing values in the JAX pipeline and were included in the further analysis.

Data modeling

Data was divided into training and test datasets in 80:20 ratio. Multivariate simple linear regression was carried out to identify the phenotypes variables that significantly affect sleep. A combination of both forward and backward stepwise regression was carried out, followed by backward elimination until all variables in the model were significant. This additional backward elimination step was carried out because of multicollinearity in the data (Figure 3.2). Relative importance of variables was calculated after bootstrapping for 1000 iterations. Regression Trees were built for Daily Sleep Percent and similar analysis was carried out for Mean Bout Length variable. In the final step, Random Forest regression was performed by building 5000 trees and important variables were identified based on percent increase in mean squared error by randomly permuting a variable. In addition to that, we also made a baseline model comparing original values and mean values for daily sleep percent and mean bout length. Model accuracy was assessed for each model by calculating Mean Squared Error (MSE) from

original and predicted values. The same training and test datasets were used for all models.

Data analysis was performed in R programming environment. The phenotype observation data was collected from IMPC Observation API using package solr (Chamberlain 2015). Regression trees were built using rpart package (Therneau, Atkinson et al. 2010), and package party was used for plotting the results (Hothorn, Hornik et al. 2015). Package randomForest was used to perform Random Forest analysis (Liaw and Wiener 2002).

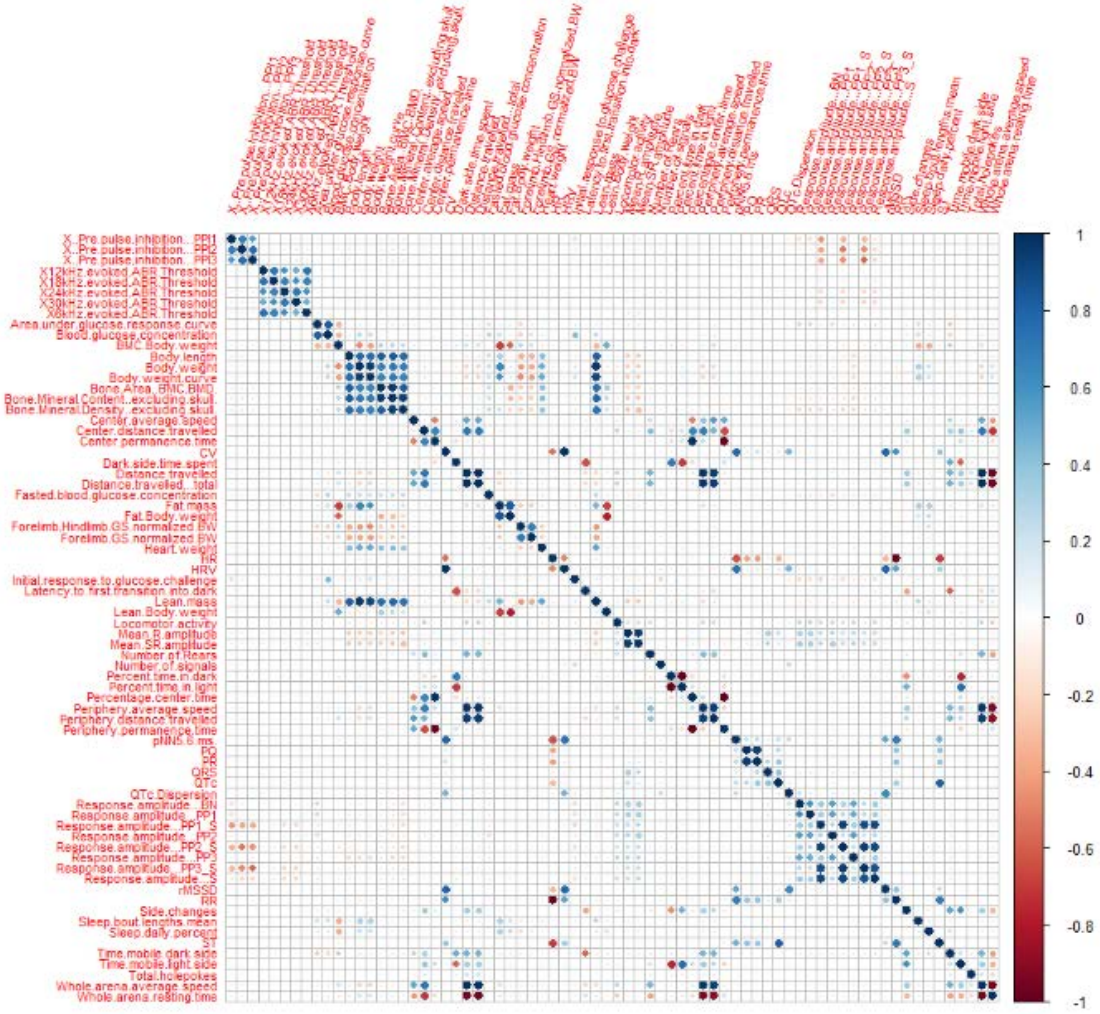


Figure 3. 2.: Correlation plot of KOMP2 variables. Blue color represents positive correlations and Red represents negative correlations Sizes of the dots are in proportion to their values.

Results

Sleep Daily Percent

Based on the results from three different data analysis approaches, the ratio of bone mineral content to body weight (BMC/BW) is the most significant predictor for sleep percent over 24 hours. BMC/BW is followed by fat mass and fat body weight

variables in their importance in the regression model (Figure 3.3). In addition to these, the random forest model also includes body weight, and lean mass (Figure 3.4). Aforementioned variables are from body composition phenotype assessed at 14 weeks of age in KOMP^2. Body weight curve is also a variable with high importance in random forest result. It records the weekly pattern of weight changes from 4th to 16th week in the pipeline. Time mobile dark side is a part of Light-Dark Test carried out in 9th week, and Mean SR amplitude variable from Electrocardiogram (ECG) in 11th week. The light-dark test results are indices of bright space anxiety in mice where more time spent in the dark side as compared to control mice is a sign of anxiety. Both of these variables were found to be important in both regression and random forest analysis. In the regression model, additional variables from Open Field (Total distance traveled, and Periphery distance traveled) and Grip Strength measurement (Forelimb and Hindlimb grip strength normalized against body weight) were also found to be significant. Decision tree model identified three key variables among the ones mentioned above affecting duration of sleep with BMC/BW was the root node. Data with SR amplitude values less than 0.76 are more likely to have lower sleep durations. It is a variable from electrocardiogram procedure, and represents the amplitude between S and R points in the PQRST waveforms seen in ECG. Fat mass values higher than 4.24 are more likely to have higher sleep percent values (Figure 3.7).

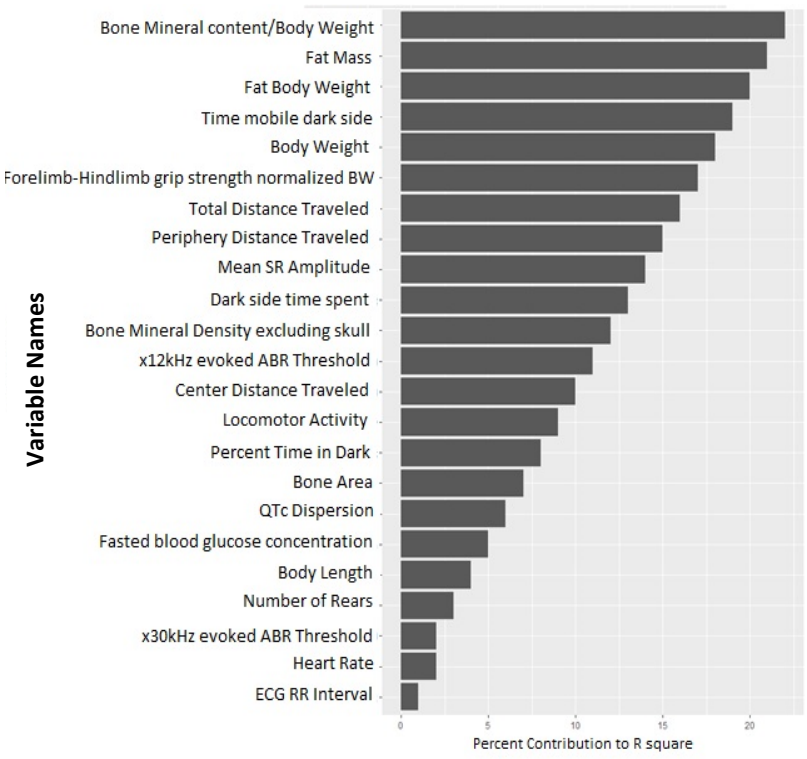


Figure3. 3.: Variable importance in multiple regression model for Daily Sleep Percent

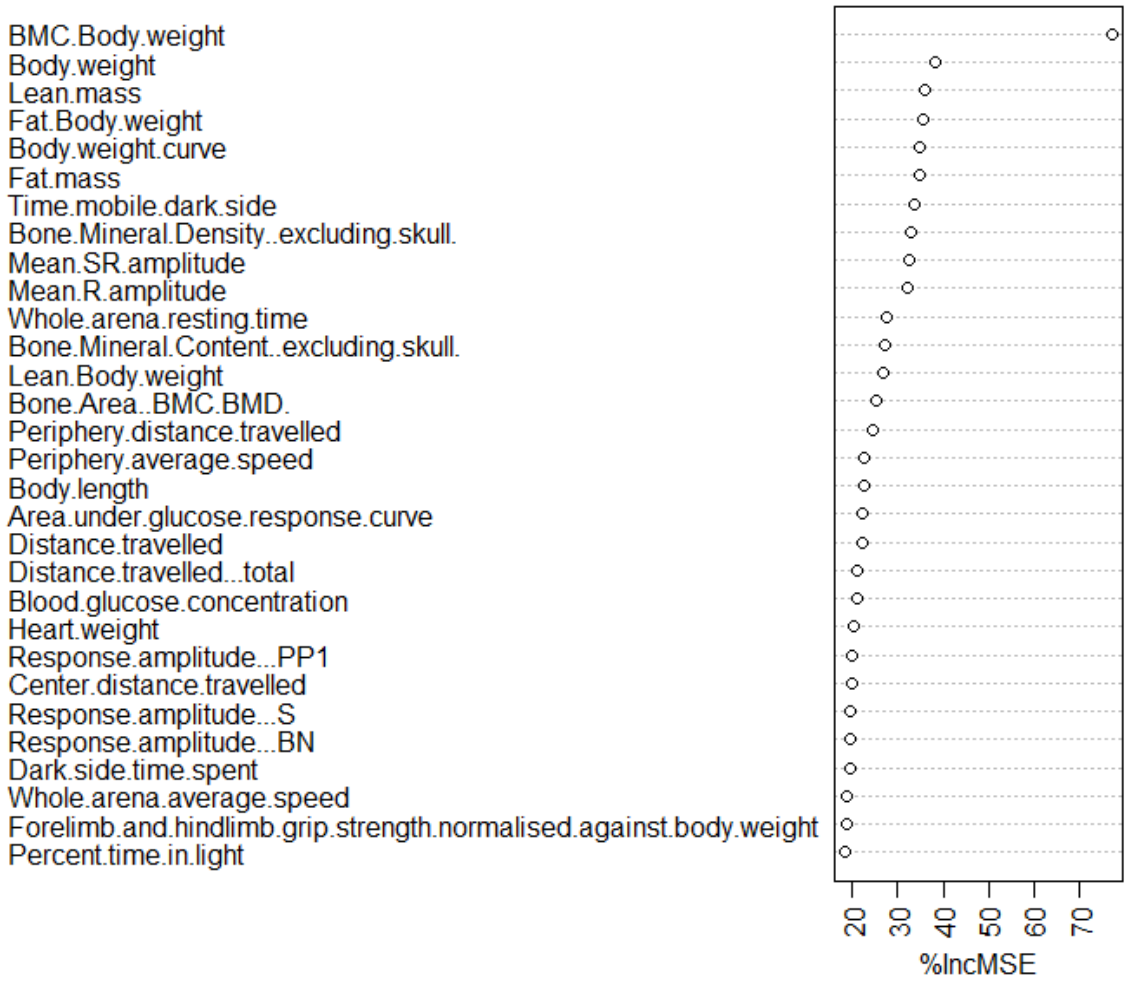


Figure3. 4.: Variable importance in Random Forest model for Daily Sleep Percent. %IncMSE represents the percent increase in mean square error by permuting the variable.

Mean Bout Length

For mean bout length values, the importance of variables differed considerably between multiple regression and random forest (Figure 3.5, 3.6). In multiple regression model, Whole arena average speed was the variable with highest relative importance, followed by Time mobile dark side (Light-Dark test, Week 9), Response Amplitude-PP1

S (Acoustic Startle and Prepulse Inhibition, Week 10), and Mean SR Amplitude, and Mean R Amplitude (ECG, Week 11). Body Composition variables like fat and lean body mass and weight do appear but with numbers much lower than what was seen for sleep daily percent. The variable importance in random forest model was much closer to sleep percent with BMC/BW being the variable with highest importance followed by Body weight, Body weight curve, Lean mass and Fat mass. Area under the glucose response curve (Intraperitoneal Glucose Tolerance Test, Week 12) also appeared with high importance. The decision tree model for mean bout lengths had Body Weight at the root node, with splits at BMD/BW, and Area under glucose response curve. Based on this model, mice with body weight lower than 23.37 grams, and BMC/BW more than 0.018 are more likely to have shorter bout lengths, and mice with weights higher than 23.27 and BMC/BW less than 0.018 more likely to have longer mean bout lengths (Figure 3.8).

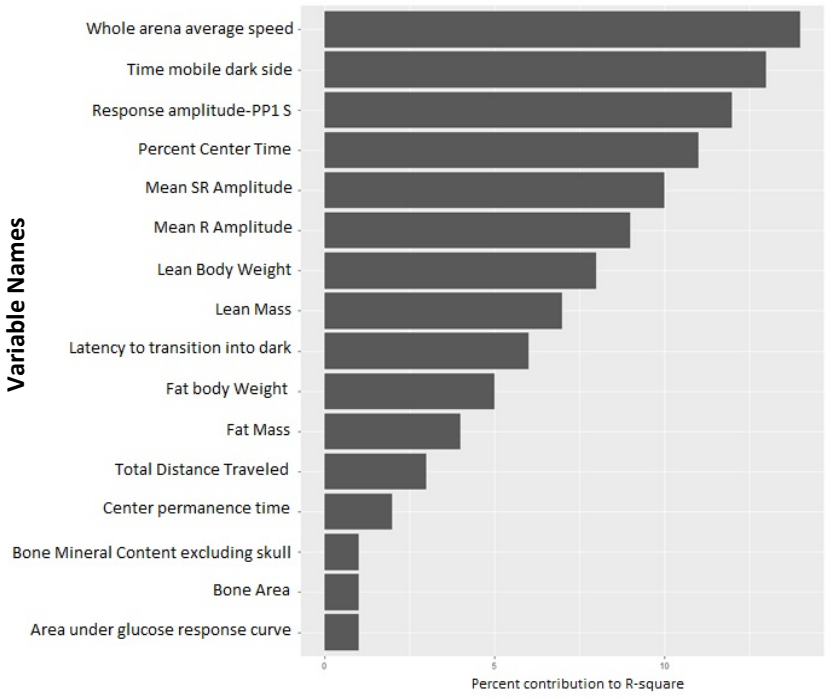


Figure3. 5.: Variable importance in multiple regression model for Mean Bout Lengths

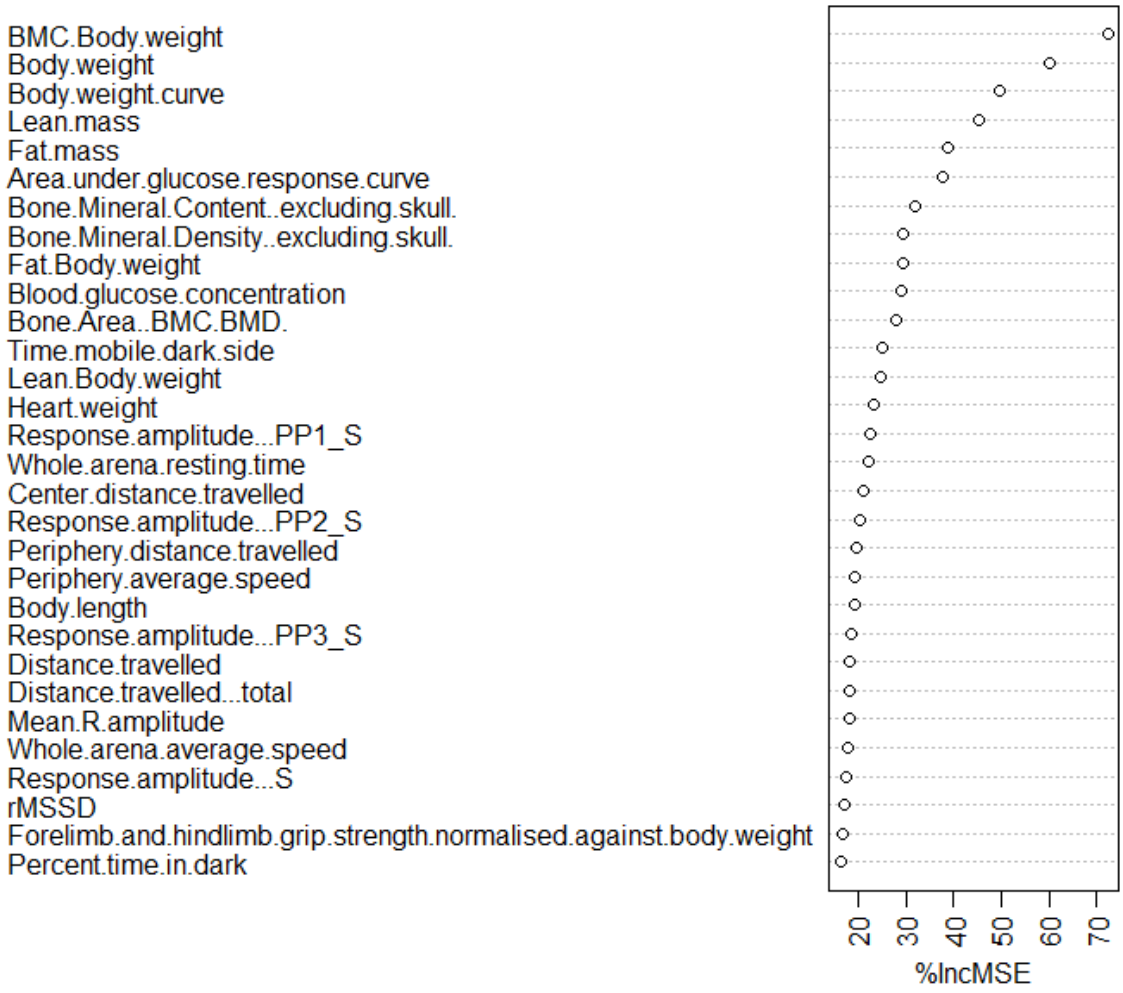


Figure 3. 6.: Variable importance in Random Forest model for Mean Bout Lengths. %IncMSE represents the percent increase in mean square error by permuting the variable.

Model Comparison

For both daily sleep percent and mean bout length variables, random forest model was found to be most accurate with minimum MSE values. For daily sleep percent variable, the baseline model had MSE of 20.56. The multiple regression model was much improved with MSE value 17.19. The Decision tree model had MSE of 18.04, though

better than baseline but less accurate than the regression model. MSE for random forest model was 16.84 which is the lowest among all. Similar trend was seen for mean bout length models where all three performed better than the baseline. The baseline model had MSE of 7964.65, followed by decision tree (MSE = 6636.99), multiple regression (MSE = 6849.84), and random forest (MSE = 6552.71).

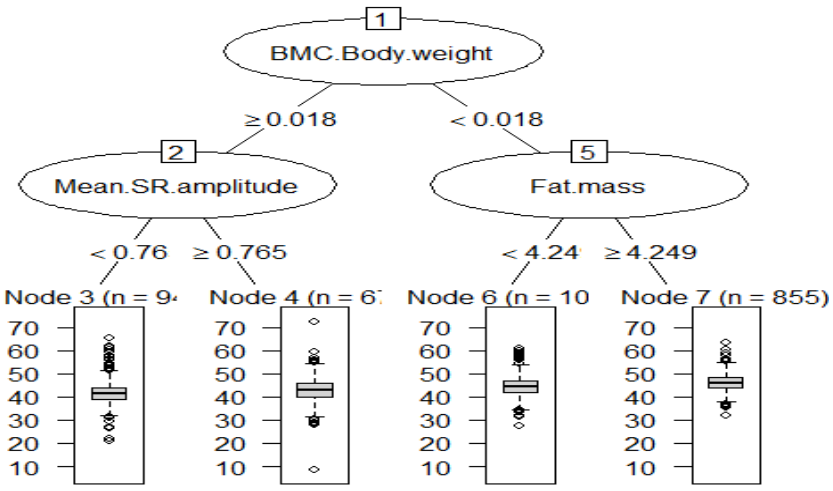


Figure 3. 7.: Decision Tree for Sleep Daily Percent

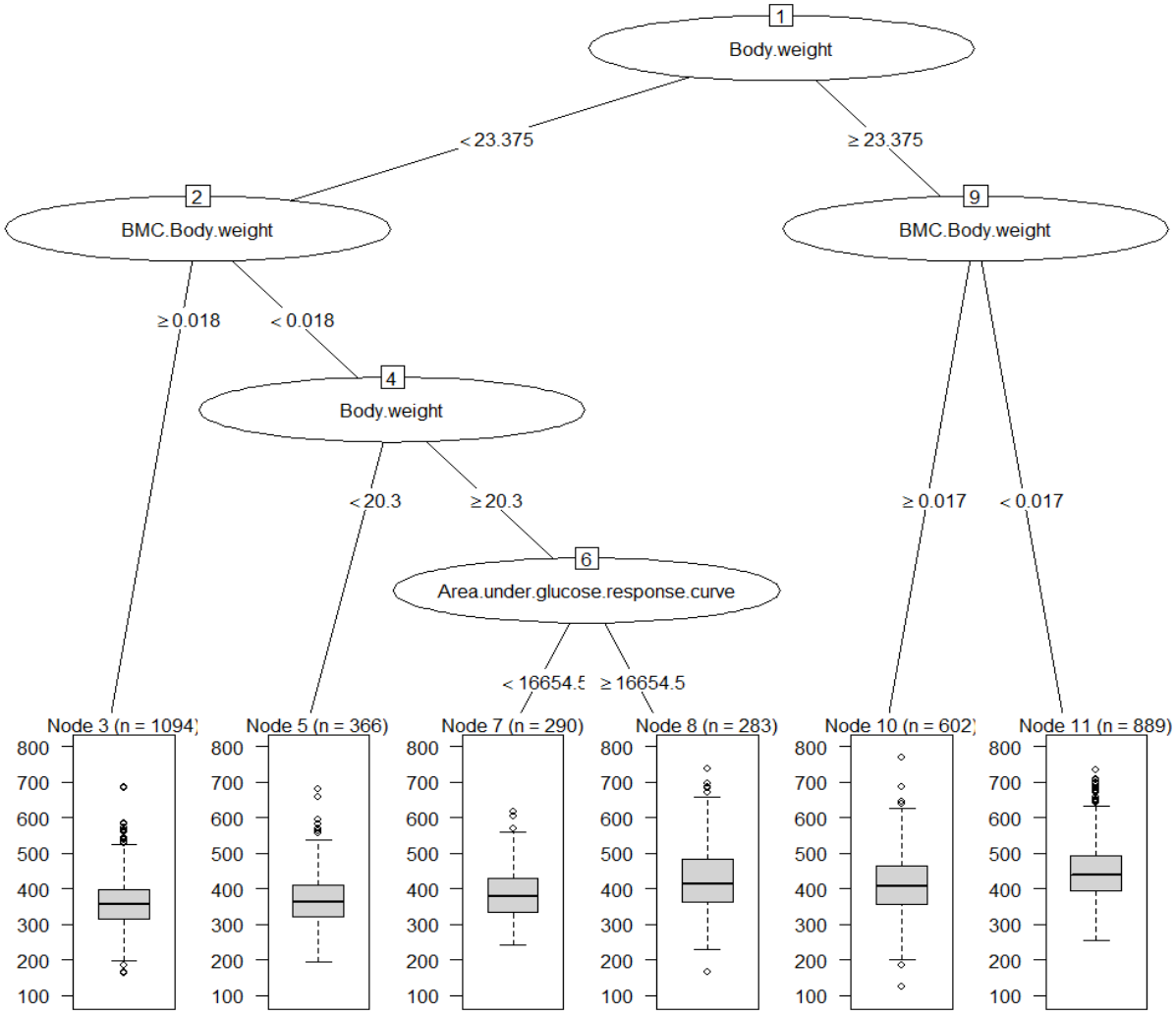


Figure 3. 8.: Decision Tree for Mean Bout Length

Discussion

Predictive modeling was carried out through multiple data models and key predictor phenotypes were identified for sleep behavior. Body composition parameters were found to be main predictors for both sleep percent and bout length. Data modeling is nuanced by several factors that can affect the accuracy and predictive ability of the

model. Multicollinearity, distribution and variance are all such factors. We observed similar MSE values for multiple regression and random forest models. In an approach like ours where the data is divided into training and test subsets, a major limitation with linear modeling approaches (i.e. multiple regression) is that a change in the training subset might bring about a change in the composition of variables in the model. Ensemble approaches, wherein the final model is based on the results from thousands of tree models are more robust to aforementioned changes in data. Though there were no dramatic changes in MSE values to suggest improvements in the model, but the goal of identifying phenotype predictors for sleep behavior was largely accomplished. The results from decision tree provide further evidence of the validity of results along with insights into how the data for predictors is affecting dependent variable.

The main predictors for sleep behavior in our analysis: BMC/BW, Body weight, Fat mass and Lean mass, are all part of Body Composition phenotype in KOMP^2. Traditionally, bone mineral content and body composition have treated as separate phenotype in studies (Manzoni, Brambilla et al. 1996, Henderson and Madsen 1999), but in KOMP^2, they are part of the same procedure. Limited literature is available for the role of bone mineral content with respect to sleep. A pertinent hypothesis would be the possible role of clock genes. Clock genes that maintain rhythmicity of several physiological systems by regulating gene expression are present in virtually all cells including bone cells (Rosen 2008). Osteoclasts that break down old bone contain *Cry1, Per1, Per2,* and *Bmal1*, and osteoblasts that make new bone tissue contain *Per1, Per2, Cry1, Clock,* and *Bmal1* (Fu, Patel et al. 2005, Fujihara, Kondo et al. 2014). Circadian rhythmicity has been seen in the function of osteoblasts and osteoclasts which is

identified by expression levels of Bone Turnover Marker genes. The process of bone resorption and some bone formation increases through the night and peaks early in the morning, with lowest levels in the afternoon (Luchavova, Zikan et al. 2011, Michalska, Luchavova et al. 2012). Abnormal sleep behavior has the potential to disrupt this rhythmicity and may cause decrease in rates of bone formation (Specker, Binkley et al. 2007, Rao, Blackwell et al. 2009). Another possibility is that the change in bone mineral content might be an indirect effect of body weight. Excessive reduction or gain in body weight has been known to affect bone mineral content (Simon, Holmes et al. 1985). Moreover, sleep deprivation or abnormal sleep behavior has been associated with obesity (Chaput, Despres et al. 2008, Mavanji, Billington et al. 2012).

Another possible indirect effect on sleep can be postmenopausal bone mineral loss caused by cessation of ovarian function. Estrogen replacement therapy (ERT) prevents this bone loss and resultant osteoporosis (Lindsay, Hart et al. 1984, Sheldon 1984). Sleep difficulties also increase in frequency during menopause. These include insomnia (Ameratunga, Goldin et al. 2012), breathing irregularities (Saaresranta, Polo-Kantola et al. 2001), difficulty in maintaining sleep, and hot flashes (Woodward and Freedman 1994) . Hot flashes are sudden feeling of intense warmth on face, neck and chest sometimes with profuse sweating. Although ERT ablates some of the aforementioned sleep difficulties especially hot flashes (Sarti, Chiantera et al. 2005), it was not found to affect overall sleep quality (Becerra, Becerra et al. 2016).

The amount of data generated in KOMP^2 is increasing constantly and provides unprecedented opportunities to investigate and understand phenotypes. In all phenotyping pipelines across the world, 3,328 out of nearly 20,000 mouse genes phenotyped

(according to IMPC Data Release 5.0). The data we analyzed represented only 209 knockout lines that have been phenotyped for sleep behavior in the JAX pipeline. The model that we have created and the predictor phenotypes that we have identified can help in identifying candidates for sleep behavior based on the data from rest of the centers as well. So even if the gene is not a part of JAX pipeline, individual researchers can still acquire those gene knockout mice and conduct their own sleep studies. Our results also open doors for studying genes previously reported in literature for the phenotypes mentioned in this analysis. On a big data scale, it can be conducted by using natural language processing algorithms to collate information about genes and phenotypes.

CHAPTER 4

Conclusions

This study utilizes data generated from the KOMP^2 project at The Jackson Laboratory, a large-scale project intended to phenotype knockout mice in alignment with the IMPC. The scale of data used in this report for studying sleep behavior in mice is unprecedented in the field. The KOMP^2 deals with the dark matter of the genes (i.e. the genes that have not been extensively studied). Even in this scenario, genes significant for sleep related variables appear at an incredibly high hit rate. This study reports on more than 6000 mice representing 343 different gene knockouts, along with over 1800 BL6NJ control mice all assessed for sleep and wake as a part of the unique behavioral phenotyping pipeline at JAX. Our findings showed altered sleep and wake in many of the knockout lines compared to the controls. Many of these strains also exhibit sex differences in sleep traits. In all, large number of genes were identified in which targeted deletion resulted in high to modest effects on one or more of the observed sleep-wake traits.

A total of 122 genes were identified that significantly affected sleep in KOMP^2 data. Out of these, 55 were found to be significant for individual piezo variables, 81 genes identified in sex based analysis, 51 in principal component analysis, and 29 in outlier analysis. Key genes have been identified that affect overall sleep like *Ppp1r9b*, and genes like *Pitx3* that affect sleep only during specific circadian phase. The main limitation with a large-scale study like KOMP^2 is that the congenital absence of gene function cannot provide any information for genes with embryonic lethal phenotype. In

such cases, conditional knockout of genes at different stages of development can provide better insights into their mechanism and function. Another limitation is the use of a single genetic background. Though C57BL/6 has been a robust model system, recent studies have shown that strong gene-phenotype interactions can be seen in different mouse strains that are in diametrically opposite direction (Sittig, Carbonetto et al. 2016). Since the goal of our report is to generalize these findings to humans, it would therefore be ideal to examine the effects of the candidate genes on sleep behavior in mice belonging to several genetic backgrounds. However, this is impractical in most cases. In cases where a gene is of sufficient interest, diverse genetic backgrounds could take advantage of the Collaborative Cross (CC) inbred strains that are ideal for studying complex behavioral traits like sleep (Philip, Sokoloff et al. 2011, Chesler 2014). These populations have been developed from an eight-way cross of common laboratory and wild-derived strains which translates to a high level of precision and allelic diversity. A combination of conditional gene knockouts with CC resources would help in identifying and tracing variations in sleep behavior for candidate genes and the underlying neurobiological phenomena. The CC mice, and the associated DO mice (from the same parents) may also display "naturally" occurring allelic variations including both gene "knockout" and more subtle variations, which may also shed light on the influence of each gene on sleep traits. The DO mice have the disadvantage of each mouse having a unique genotype, that cannot be studied in replicate in the present or the future, but also have the advantage of being a healthier animal with more normal levels of heterozygosity.

Breath rate is a supplemental variable from piezo whose role is underplayed in the analysis since it is a relatively new measure. It is separate variable in itself that is

different from the measures of sleep duration and consolidation. The breath rate can help in identifying phenotypes related to obstructive sleep apnea, and REM/NREM sleep. The breathing pattern during REM sleep is ataxic or irregular, as compared to the rhythmic pattern seen during NREM. Rather than measuring the mean breath rate during whole duration of sleep, gathering data in smaller time intervals during subjective night can provide insights into the breathing patterns during those periods, thus allowing for better differentiation between REM/NREM sleep. The algorithm to carry out such differentiation is in the works, the application of which will open avenues to identify genes affecting different stages of sleep. Obstructive sleep apnea (OSA) is a syndrome that affects 10 – 17% of US population (Young, Finn et al. 2008, Punjabi, Caffo et al. 2009). The symptoms are oxygen desaturation, loud snoring, frequent arousals, and disruption of sleep, and has been associated with cardiovascular disease (Shahar, Whitney et al. 2001, Marin, Carrizo et al. 2005), hypertension (Nieto, Young et al. 2000), cognitive impairment, and metabolic abnormalities (Young, Peppard et al. 2002). The piezo sleep variables along with breathing rate in combination with results from multitude of physiological phenotyping assays make KOMP^2 a great resource to further our understanding of this disorder.

Functional annotation of genes identified in our analysis was carried out by cross-referencing them with the mammalian phenotype terms that they were associated with in IMPC. In the field of ontology, sleep has been traditionally treated as a monolithic entity that is wither “normal” or “abnormal”. The current mammalian phenotype ontology for sleep behavior consists of “NREM sleep pattern”, “paradoxical sleep pattern”, “fragmentation of sleep-wake states”, and “narcolepsy”. Only 37 out of 122 genes being

reported in current analysis are captured by these ontologies. Though this is much better than treating sleep as a single entity, it is still far from covering the complexities of sleep behavior. Now that we have a large amount of sleep data from mice, it is possible to include more ontological terms for sleep behavior. To begin with, these would include "short sleepers", "long sleepers", "abnormal sleep during light phase", and "abnormal sleep during dark phase".

The follow up study to identify predictor phenotypes in Chapter 3 showed "Bone mineral content divided by body weight" as the most important predictor for sleep percent and mean bout lengths over 24 h. In this analysis, I have completely focused on phenotypes while taking the genotypes out of the equation. There are a couple of advantages with this approach. First, it helps in identifying the predictors for the phenotypes of interest. Second, on the more abstract level, it provides a peek into the fine tapestry of relationships that exist between phenotypes, the magnitude of which would be difficult to study if the focus is solely on the genes.

The key to any data analysis exercise is to ask prudent questions and then prepare and analyze the data to answer those questions. The results reported in this thesis answers several questions regarding a number of genes which appear to influence sleep, that in turn may help in gaining a better understanding of the underlying processes governing sleep. The data available through KOMP^2 continue to increase and the analysis presented in this report sets the groundwork for future identification of novel sleep related genes and predictor phenotypes that will empower the sleep research community. It is a small step forward in understanding of the phenomenon called sleep that has been a mystery, and something that a majority of the human race can't get enough of.

APPENDIX

List of Abbreviations

ANOVA	Analysis of Variance
API	Application Programming Interface
BMC/BW	Bone Mineral Content divided by Body Weight
CART	Classification and Regression Trees
ECG	Electrocardiogram
EEG/EKG	Electroencephalogram
ERT	Estrogen Replacement Therapy
HTTP	Hypertext Transfer Protocol
IKMC	Internation Knockout Mouse Consortium
IMPC	Internation Mouse Phenotyping Consortium
JAX	The Jackson Laboratory
KOMP	Knockout Mouse Project
$KOMP^2$	Knockout Mouse Phenotyping Program
MD	Mahalanobis Distance
MSE	Mean Squared Error
PCA	Principal Component Analaysis
PVDF	Polyvinylidine Difluoride
QTL	Quantitative Trait Loci
REST	Representational State Transfer
SCN	Suprachiasmatic Nucleus
TMN	Tuberomammillary Nucleus
VLPO	Ventrolateral Preoptic Area

Supplementary Tables and Figures

Gene name	Sleep Daily Percent	Sleep Percent during Light Phase	Sleep Percent during Dark Phase	Bout Lengths Mean	Light phase bout lengths mean	Dark phase bout lengths mean	Breath Rate
1700016K19Rik	1	1	1	1	1	1	1
4921509C19Rik	1	0.195	1	1	1	1	1
A1cf	1	1	1	1	1	1	1
Abca16	1	1	1	1	1	1	1
Abca7	1	1	1	1	1	1	1
Abcg2	1	1	1	1	1	1	1
Acap1	1	1	1	1	1	1	1
Acsf2	1	1	1	0.882	1	1	0
Acsm2	1	1	1	1	1	1	1
Actrt3	1	0.995	1	1	1	1	1
Adad2	1	1	1	1	1	1	1
Adck2	1	0.279	1	1	0.012	1	0.053
Adck5	1	1	1	1	1	1	1
Adgrb2	1	0.468	1	1	1	1	1
Adora2b	1	1	1	1	1	1	1
Ahrr	1	1	1	1	1	1	1
AI464131	1	1	1	1	1	1	1
Ajap1	1	0.479	1	1	1	0.499	0.166
Akap11	1	0.984	1	1	1	1	1
Akip1	1	1	1	1	1	1	1
Akr1b8	1	0.991	1	1	1	1	1
Akr1d1	1	1	1	1	1	1	1
Ap4e1	0.002	0.082	0.416	0.001	0.009	0.009	0.999
Arf2	1	0.981	1	1	0.063	1	1
Arhgef10	1	1	1	1	1	1	1
Arpc5l	1	1	1	1	1	1	1
Arrb1	1	1	0.7	1	1	1	0.831
Arrb2	0.305	1	0.11	1	0.026	1	1
Arrdc1	1	1	1	1	1	1	0.996
Arsk	1	1	1	1	1	1	1
Asb10	1	1	1	1	1	1	1
Bbox1	1	1	1	1	1	1	1
BC030499	0.891	0.716	1	1	1	1	0.073
BC100451	1	0.517	1	1	1	1	0.994

Bex4	1	0.042	1	0.796	1	0.025	0.855
Bhlhe40	1	1	1	1	1	1	1
Bmp2k	1	1	1	1	1	0.995	1
Btg2	1	1	1	1	1	1	1
Bzw2	0.013	0.003	1	1	0.838	1	0.124
C1qa	1	1	1	1	0.997	1	0
C1qb	1	1	1	1	1	1	0.243
C1qtnf5	1	1	1	1	1	1	0.166
C3	1	1	1	1	1	1	0.994
C9	1	1	1	1	1	1	1
Cacna2d3	1	1	1	1	1	1	1
Car12	1	1	1	1	1	1	1
Carf	1	0.603	1	1	1	0.985	0.001
Cast	1	1	1	0.836	0.907	1	1
Cbln3	0.001	1	0.001	1	1	1	0
Ccdc120	1	1	1	1	1	1	1
Ccl26	0.003	0.009	0.92	1	0.906	1	1
Cd33	1	1	1	1	1	1	1
Cd84	0.823	0.81	1	1	1	1	0.015
Cdc20	0.008	0.941	0.125	1	1	0.314	1
Cdh4	1	1	1	1	1	1	1
Cdk15	0.081	0.075	1	1	1	1	0.686
Cdk19	1	0.997	1	1	1	1	1
Ceacam16	0.82	1	1	1	1	1	1
Cers5	0.839	0.293	1	1	1	1	0.054
Ces4a	1	1	1	1	1	1	0.154
Cfb	1	0.981	1	1	1	1	0.012
Chek2	1	1	1	1	1	1	1
Chn1	0.119	0.464	0.999	1	1	1	1
Cited4	1	1	1	1	1	1	1
Cldn13	1	0.808	1	1	0.137	1	0.998
Cldn19	1	1	1	1	1	1	1
Clvs1	1	1	1	1	1	1	0.989
Cml2	1	1	1	1	1	1	1
Col18a1	1	0.808	1	1	1	1	1
Cp	1	0.991	1	1	1	1	1
Cpb1	1	0.002	1	1	0.997	0.922	0.977
Crym	1	1	1	1	1	1	0.017
Cyb5d1	1	0.989	1	1	1	1	1
Cyb5d2	1	1	1	1	1	1	1
Dcaf10	1	1	1	0.165	1	0.069	0
Dennd2d	1	1	1	1	1	1	0.954
Dixdc1	1	1	1	1	1	1	1

Dnaja4	1	0.001	1	1	0.628	1	0
Dnajb3	1	1	1	1	1	1	1
Dnajb7	1	1	1	1	1	1	0.89
Dnajc14	1	0.842	0.44	1	1	1	0.999
Dnajc28	1	1	1	1	1	1	1
Dnajc5g	1	1	1	1	0.999	1	0.746
Dnajc7	1	1	1	0.753	1	0.924	0.071
Dnase1l2	1	1	1	1	1	1	1
Dntt	1	1	1	1	1	1	1
Dpf1	1	1	1	0.996	1	1	1
Efna4	1	1	1	1	1	1	1
Elk1	1	1	1	1	1	1	1
Enox1	1	0.975	1	1	1	1	1
Eogt	1	0.034	1	1	0.806	1	1
Epb4.1l4a	1	1	1	1	1	1	1
Epgn	1	0.008	1	1	1	1	0.426
Epha10	1	0.069	1	1	1	1	0.971
Ermp1	1	1	1	1	0.023	1	1
Espnl	1	1	1	1	1	1	1
Esrra	1	1	1	1	1	1	1
F2rl1	1	1	1	1	1	1	1
Fam161a	1	1	1	1	1	1	0.307
Fam186b	0.997	1	1	0.247	0.08	0.997	1
Fam217b	1	1	1	1	1	1	1
Far2	0.824	0.911	1	1	1	1	0.983
Fastkd5	1	0.999	1	1	1	1	1
Fdxacb1	1	0.993	1	1	1	1	0.239
Fndc4	1	1	1	1	1	1	1
Foxi2	1	0.162	1	1	1	1	0.105
Foxo3	0.998	0.015	1	1	0.994	1	1
Foxred2	1	1	1	1	1	1	1
Galnt12	1	1	1	1	1	1	1
Gfod1	1	1	1	1	1	1	1
Ghrhr	1	1	0.879	0.986	1	0.019	1
Ghsr	1	0.993	1	1	1	1	1
Gimap6	0.989	0.178	1	1	1	0.905	0.996
Gimap8	1	1	1	1	1	1	1
Gipc3	1	0.204	1	1	0.215	1	1
Gjd4	0.993	0.047	1	1	1	0.999	0.915
Glycam1	1	1	1	1	1	0.931	1
Gpnmb	1	1	1	1	1	1	1
Gpr142	1	1	1	0.997	1	0.598	1
Gpr156	0.996	1	1	1	1	1	1

Gpr183	1	0.978	1	1	1	1	1
Gpr19	1	1	1	1	1	1	1
H1fx	1	1	1	1	0.999	0.393	1
Hc	1	1	1	1	1	1	1
Hdac10	1	1	1	1	1	1	1
Hemgn	1	1	1	0.97	1	0.11	1
Heyl	1	1	0.999	1	1	1	1
Hfe2	1	1	1	1	1	1	1
Hsd17b1	1	0	0.998	0	0	1	1
Hsd17b11	1	0.679	1	1	1	1	0.601
Hsf2	1	1	1	1	1	1	1
Hsf4	1	1	1	1	1	1	1
Hspb1	1	1	1	1	1	0.977	1
Hspb2	1	1	1	1	1	1	1
Hspb3	1	1	1	0.991	1	1	1
Htr1a	1	0.919	1	1	1	1	1
Htr1d	1	1	1	1	0.001	1	0.996
Htr1f	0.965	0.002	1	1	1	1	0.004
Htr3b	1	0.025	1	1	1	1	1
Htr7	1	1	1	1	1	1	1
Hyal3	1	1	1	1	1	1	1
Ifnk	1	0.999	1	1	1	1	1
Ifnl3	0.99	0.004	1	1	1	1	0.204
Igsf11	1	1	0.67	1	1	1	0.233
Il12rb2	1	0.78	1	1	0.765	1	0.009
Il24	0.997	1	0.434	0.998	0.761	1	1
Ipp	1	0.553	1	1	0	1	0.614
Iqcj	1	1	1	1	1	1	1
Iqgap2	1	1	1	1	1	1	1
Irf8	1	1	1	1	1	1	1
Jam2	1	0.492	1	1	1	1	0.092
Jmjd8	1	1	1	1	1	1	1
Kcnh3	1	0.028	1	1	1	0.747	1
Klk14	1	1	1	1	1	1	1
Krt77	1	1	1	1	1	1	1
Krt9	1	0.883	1	1	1	1	0.033
Lcn2	1	1	1	1	1	1	1
Lima1	1	1	1	1	1	1	1
Limch1	1	1	1	1	1	1	1
Lin28b	1	1	1	1	1	1	0
Lipn	1	1	1	1	1	1	1
Il24	1	1	1	1	1	1	1
Lman1l	1	1	1	1	0.995	1	1

Loxl1	1	1	1	1	1	0.823	0.906
Lpar6	1	1	1	1	1	1	1
Lrch1	1	1	1	1	1	1	1
Lrrc15	1	1	1	1	1	1	1
Ltbp2	1	1	0.986	1	1	1	1
Macrod2	0.986	0.941	0	0.965	0.163	1	1
Mag	1	1	0.492	1	1	1	0.001
Masp1	1	0.999	1	1	1	1	1
Mdk	1	1	1	1	1	1	1
Mdp1	1	0.681	1	1	1	1	1
Mepce	1	1	1	1	1	1	0.92
Mettl21c	1	0.894	1	1	1	1	0.077
Mettl7b	1	1	0.93	0.308	0.563	0.129	0.002
Mfsd10	1	0.998	1	1	1	1	1
Mlx	1	1	1	1	1	1	1
Mmp8	1	1	1	1	1	1	1
Mos	1	1	1	1	1	1	1
Moxd1	1	0.651	1	1	1	1	0.999
Mpdz	1	1	1	1	1	1	1
Mrgpre	1	1	1	1	1	1	1
Mtmr3	1	1	1	1	1	1	1
Myh1	1	1	0.233	0.001	0.001	0.999	0
Mylip	1	0.589	0.007	1	1	0.044	1
Myo3a	1	1	1	1	1	1	1
Myo3b	1	1	0.999	0.017	0.997	0.003	1
Myo7b	1	1	1	1	1	1	0.995
Nat1	1	0.993	1	1	0.143	1	1
Ncald	0.573	0.003	1	1	1	1	0.096
Nefh	1	1	1	0.635	1	0.054	0.833
Nek11	1	1	1	1	1	1	1
Nek2	1	1	1	1	1	1	1
Nes	1	1	1	0.463	1	0.04	0
Neurl2	0.99	0.005	1	1	1	1	0.987
Nfatc4	1	0.633	1	0.272	0.052	1	1
Nmb	1	1	1	1	1	1	1
Nmrk2	1	1	1	1	1	1	0.268
Npm3	1	0.008	1	1	0.298	1	0.615
Nrcam	1	0	0.999	0.658	0.011	1	1
Nrn1l	1	0.998	1	1	0.375	1	0.96
Nrsn1	1	1	1	1	1	1	1
Nsun7	1	1	1	1	0.955	1	1
Nt5c	1	0.997	1	1	1	1	1
Nt5c1b	1	1	1	1	1	1	1

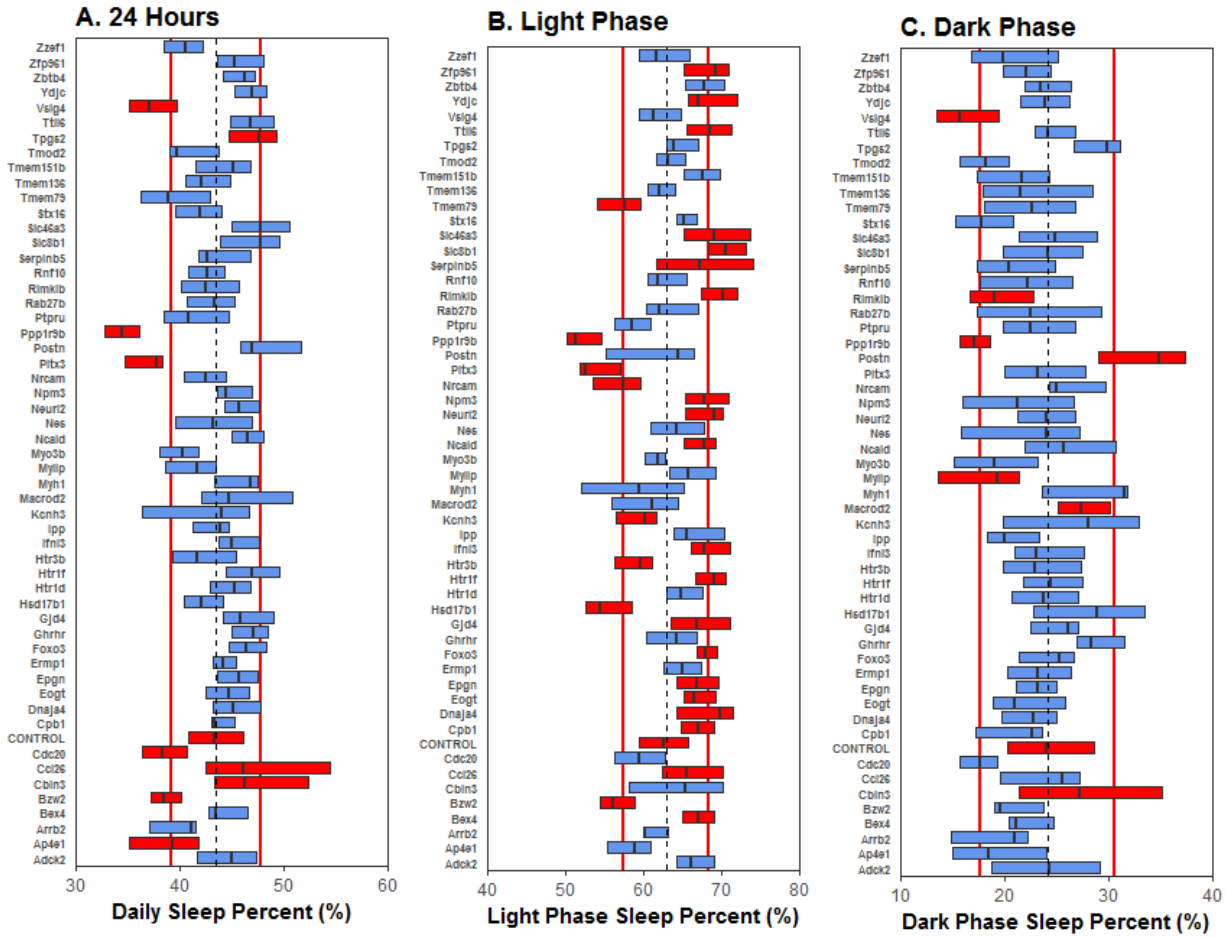
Oard1	1	1	1	1	1	1	1
Ocstamp	1	1	1	1	1	1	0.698
Ogn	1	1	1	1	1	1	0.987
Osm	1	0.653	1	1	1	1	0.992
Ovch2	1	1	1	1	1	1	1
Pacsin2	1	1	1	1	1	1	1
Paqr7	1	1	1	1	1	1	1
Pard6a	1	1	1	0.916	0.911	1	1
Parp16	1	0.146	0.985	1	1	0.732	1
Parp8	0.382	1	0.636	1	0.995	1	0.993
Pcdh12	1	1	1	1	1	1	1
Pcsk4	1	1	1	1	1	1	1
Pfdn6	1	1	0.983	1	1	1	1
Pitx3	0	0	1	0	0	1	0.928
Pkn3	1	1	1	1	1	1	1
Pkp4	1	1	1	0.821	1	0.7	0
Pla2g2d	1	1	1	1	1	1	1
Plekha3	1	1	1	1	1	0.14	1
Plk5	1	1	1	1	1	1	1
Pnmt	1	1	1	1	0.905	0.999	0.99
Postn	0.125	0.94	0	1	1	0.8	0.065
Ppp1r26	1	1	1	1	1	1	1
Ppp1r9b	0	0	0	0	0	0	0.824
Prkab1	1	1	1	0.999	1	0.588	1
Prokr1	1	1	1	1	1	1	0.665
Prom2	1	0.257	1	1	1	0.62	1
Prss40	0.999	1	1	1	1	1	1
Prss56	1	1	1	1	1	1	1
Ptpn20	1	1	1	1	1	1	1
Ptpn5	0.18	1	0.326	1	1	0.999	1
Ptpru	1	0.965	1	0.611	0	1	1
Pycr1	1	1	0.998	1	1	1	0.949
R3hcc1l	1	1	1	1	0.999	1	1
Rab11fip5	1	1	1	1	1	1	1
Rab20	1	1	1	1	1	1	1
Rab24	0.704	0.093	1	1	1	1	1
Rab27b	1	1	1	0.178	1	0.024	0.956
Rab36	1	0.964	1	1	1	1	0
Rab39	1	1	1	1	1	1	1
Rab3c	1	1	1	1	1	1	1
Rab3gap2	1	1	1	1	1	1	1
Rab3ip	0.999	0.066	1	1	0.887	0.824	0.033
Rab43	1	1	1	1	1	1	1

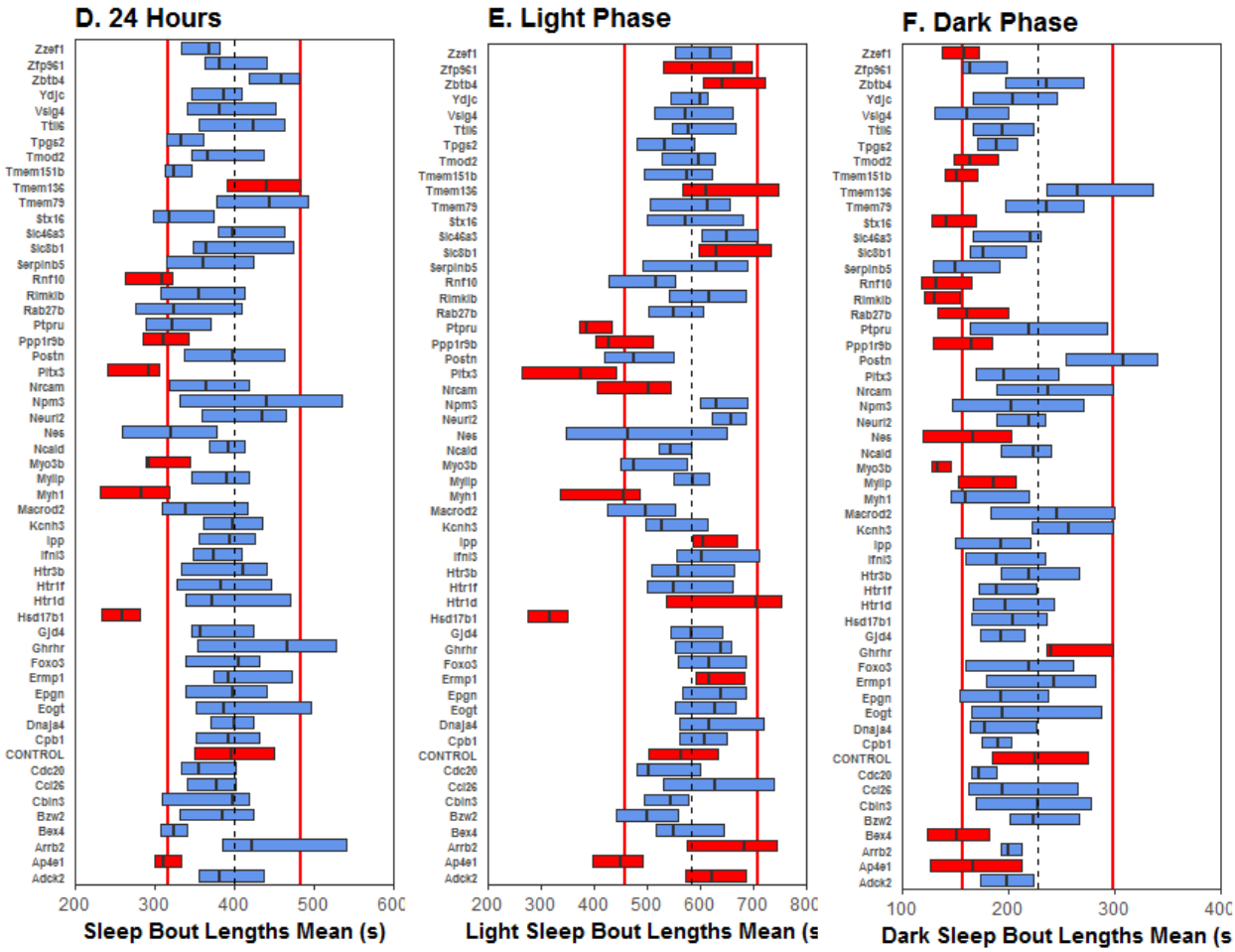
Rab5a	1	0.998	1	1	1	0.998	0.172
Rabac1	1	1	1	1	1	1	1
Rad21l	1	0.999	1	1	1	1	1
Rap2b	1	1	1	0.967	1	0.969	1
Rbm4b	1	1	1	1	1	1	1
Resp18	1	1	0.978	1	1	1	1
Rexo4	1	1	1	1	1	1	1
Rhbdl2	1	1	1	1	1	1	0.999
Rilpl2	1	1	1	1	1	1	1
Rimklb	1	0.001	0.009	1	1	0	0.614
Rin3	1	1	1	1	0.994	1	1
Rnf10	1	1	1	0	0.874	0	1
Rnf112	1	1	1	1	1	1	1
Rnf13	1	1	1	1	1	1	0.052
Rnf133	1	1	1	1	1	1	1
Rnf25	1	0.158	1	0.234	1	0.056	1
Rps6kl1	1	1	1	1	1	1	1
Rrad	1	1	1	1	0.987	1	1
Rxfp4	1	0.389	1	1	0.934	1	0.077
Scg2	1	1	1	1	1	1	1
Serpina1f	1	1	1	1	1	1	1
Serpina7	1	1	1	1	1	1	1
Serpinb5	1	0.007	1	1	0.242	0.176	0.955
Setd6	1	0.991	1	1	0.984	1	1
Sfxn4	1	1	1	1	1	1	1
Sgta	1	0.999	1	1	1	1	1
Sh3tc2	0.851	0.824	1	1	1	1	1
Slc1a1	1	1	0.082	1	1	0.299	0.841
Slc24a5	1	1	1	0.748	1	0.925	1
Slc25a35	1	1	1	1	1	1	1
Slc46a3	0.076	0	1	1	0.708	1	0.035
Slc8a3	1	1	1	1	1	1	1
Slc8b1	0.366	0	1	1	0.027	0.994	0.011
Smoc2	1	0.07	1	1	0.989	1	1
Snx15	1	0.197	1	1	1	1	0.835
Sorbs2	1	1	1	1	0.97	1	1
Sox18	1	1	1	1	0.194	1	0.821
Sp5	1	1	1	1	1	0.993	1
Spag4	1	1	1	1	1	1	1
Spp1	1	1	1	1	1	1	1
Sprr1b	1	1	1	1	1	0.986	1
Sprr3	1	1	1	1	1	1	1
Sptssb	1	1	1	1	1	1	1

Srcin1	1	1	1	1	1	1	1
Stag3	1	1	1	1	1	0.375	1
Stat5b	1	1	1	1	1	1	1
Stk16	1	1	1	1	1	1	1
Stx16	1	0.81	0.734	0.596	1	0.001	0.006
Stx19	1	1	1	1	1	1	0.084
Syce1	1	1	1	1	1	1	1
Syce1l	1	1	1	1	1	1	1
Sycp3	1	1	1	1	1	1	1
Syn3	1	1	1	1	1	1	0.179
Tas2r138	1	1	1	1	1	1	1
Tbc1d4	1	1	1	1	1	1	1
Tbx22	1	1	1	1	1	1	1
Tdrkh	1	0.072	1	1	1	1	0.012
Tex101	1	1	1	1	1	1	1
Tex29	1	1	1	1	1	1	1
Thsd1	1	0.989	1	1	1	1	0.307
Timp3	1	1	1	1	1	0.996	1
Tmem136	1	1	1	0.003	0.001	0.999	1
Tmem151b	1	0.868	1	0.64	1	0.038	1
Tmem181a	1	1	1	1	1	1	1
Tmem79	0.112	0.003	1	0.851	0.053	1	1
Tmod2	1	1	0.144	1	1	0.045	0.07
Tnfsf18	0.468	0.992	1	0.856	0.991	0.952	0.667
Tpcn1	1	1	1	1	1	1	1
Tpgs2	0.046	1	0.247	0.394	1	0.997	0.99
Tppp	1	1	1	0.581	1	0.691	0
Tprn	0.939	1	0.85	1	1	1	1
Trip13	1	1	1	1	1	1	1
Try4	1	1	1	1	1	1	1
Tspan18	1	1	1	0.926	1	0.977	1
Tssk5	1	1	1	1	1	1	1
Ttll10	1	1	1	1	1	1	1
Ttll6	0.695	0.002	1	1	0.112	1	1
Tubb4a	1	1	1	1	1	1	1
Vcpkmt	1	0.999	1	1	1	1	1
Vegfb	0.809	0.26	1	1	1	1	1
Vsig4	0.032	1	0	1	1	0.074	0.957
Vsig8	1	1	1	1	1	1	1
Wee2	1	1	1	1	0.205	1	1
Ydjc	0.453	0	1	1	1	1	1
Ypel1	1	1	1	1	1	1	1
Zbtb32	1	0.998	1	1	1	1	1

Zbtb4	1	0.073	1	0.999	0.016	1	1
Zdhhc11	1	1	1	1	1	1	1
Zfp14	1	1	1	1	0.798	1	1
Zfp219	1	1	0.743	1	1	1	1
Zfp689	1	1	1	1	1	1	0.01
Zfp961	1	0	1	1	0.01	0.068	0.974
Zfyve26	1	1	0.076	1	1	0.961	1
Zzef1	1	1	0.792	0.998	1	0.009	0.384

Supplementary Table A. 1.: p-values of individual genotypes for piezo variables from Dunnett's Post Hoc analysis.





Supplementary Figure A. 1.: Visual representation of the differences in values for sleep variables in 55 KO strains that were found to be significant for one of more of these variables. The genotypes are arranged alphabetically for sleep percent (A. 24h, B. dark phase, C. light phase), and bout length (D. 24h, E. dark phase, F. light phase) measured in seconds as depicted on y axis. Each bar represents values between first and third quantile and the median (dot) for each KO cohort. Boxes marked in red represent the genes significant for a specific piezo variable.

A.

	Daily sleep Percent		Light Phase Percent		Dark Phase Percent	
Genotype	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Cdc20	p<0.05 ↓	1	p<0.05 ↓	1	0.09	1
Chn1	p<0.05 ↓	1	0.11	1	1	1
Mettl7b	p<0.05 ↓	1	1	0.82	p<0.05 ↓	1
Pitx3	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓	1	0.33
Ppp1r9b	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓
Ptpn5	p<0.05 ↓	1	1	1	p<0.05 ↓	1
Bzw2	0.25	0.41	p<0.05 ↓	0.94	1	1
Dnaja4	1	1	p<0.05 ↑	0.96	1	1
Epgn	1	1	p<0.05 ↑	1	1	1
Epha10	1	1	p<0.05 ↑	1	1	0.98
Hsd17b1	1	1	p<0.05 ↓	1	0.36	1
Htr1f	1	1	p<0.05 ↑	0.81	1	1
Parp16	1	1	p<0.05 ↓	1	0.94	1
Prom2	1	1	p<0.05 ↓	1	1	1
Slc8b1	1	0.11	p<0.05 ↑	p<0.05 ↑	1	1
Zfp961	1	1	p<0.05 ↑	0.4	1	1
Macrod2	1	1	1	1	p<0.05 ↑	p<0.05 ↑
Mylip	1	1	0.9	1	p<0.05 ↓	0.51
Parp8	0.1	1	1	1	p<0.05 ↑	1
Postn	1	0.93	0.7	1	p<0.05 ↑	p<0.05 ↑
Tmod2	0.33	1	1	1	p<0.05 ↓	1
Ap4e1	0.99	p<0.05 ↓	1	0.67	1	0.31
Nfatc4	1	1	1	1	1	1
Rimklb	1	1	0.2	0.21	0.31	0.08
Rnf10	1	1	1	1	1	1
Nat1	1	1	0.29	1	1	1
Stx16	1	1	1	1	0.61	1
Tmem151b	1	1	1	1	1	1
Zzef1	0.91	1	1	1	0.66	1
Adck2	1	1	0.77	1	1	1
Myh1	1	0.33	0.42	1	1	p<0.05 ↑
Nes	1	1	0.98	1	1	1
Cbln3	1	p<0.05 ↑	1	0.06	0.44	p<0.05 ↑
Ccl26	1	p<0.05 ↑	1	p<0.05 ↑	1	0.32
Col18a1	1	p<0.05 ↓	1	p<0.05 ↓	1	1
Rab24	1	p<0.05 ↑	1	0.51	1	0.98
Tmem79	1	p<0.05 ↓	0.92	0.06	1	1
Ydjc	1	p<0.05 ↑	1	p<0.05 ↑	1	1
Cpb1	1	0.69	1	p<0.05 ↑	0.29	1
Kcnh3	1	0.39	1	p<0.05 ↓	0.89	1
Npm3	1	1	1	p<0.05 ↑	1	1
Nrcam	1	1	0.83	p<0.05 ↓	1	0.97
Serpinb5	1	0.91	1	p<0.05 ↑	1	1
Slc46a3	1	0.11	0.13	p<0.05 ↑	1	1
Zbtb4	1	0.35	1	p<0.05 ↑	1	1
Arrb2	0.27	0.8	1	1	0.39	p<0.05 ↓

Vsig4	1	0.06	1	1	0.94	p<0.05 ↓
Zfyve26	1	0.94	1	1	0.87	p<0.05 ↓
Ghrhr	1	1	1	1	1	0.91
Prss56	1	1	1	1	1	0.88
Tmem136	1	1	1	1	1	1
Ipp	1	1	1	0.09	1	1
Nrn1l	1	1	1	1	1	1
Ptpru	1	1	1	1	1	1

B

	Sleep Bout Length		Light Phase Bout Length		Dark Phase Bout Length	
Genotype	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Cdc20	1	1	1	1	0.18	1
Chn1	1	1	1	1	1	1
Mettl7b	p<0.05 ↓	1	p<0.05 ↓	1	0.12	1
Pitx3	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓	0.89	p<0.05 ↓
Ppp1r9b	p<0.05 ↓	0.67	0.1	0.14	p<0.05 ↓	0.14
Ptpn5	1	1	1	1	0.13	1
Bzw2	1	1	0.45	1	1	1
Dnaja4	1	1	1	1	0.94	1
Epgn	1	1	1	1	0.66	1
Epha10	1	1	1	0.98	1	0.98
Hsd17b1	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓	p<0.05 ↓	1	p<0.05 ↓
Htr1f	1	1	1	1	1	1
Parp16	1	1	1	1	0.57	1
Prom2	1	1	1	1	0.58	1
Slc8b1	1	1	0.94	p<0.05 ↑	1	p<0.05 ↑
Zfp961	1	1	p<0.05 ↑	1	p<0.05 ↓	1
Macrod2	1	0.78	1	0.09	1	0.09
Mylip	1	1	1	1	0.15	1
Parp8	1	1	0.8	1	1	1
Postn	1	1	1	1	1	1
Tmod2	1	1	1	1	p<0.05 ↓	1
Ap4e1	p<0.05 ↓	0.68	0.39	0.5	p<0.05 ↓	0.5
Nfatc4	p<0.05 ↑	1	p<0.05 ↑	1	1	1
Rimklb	p<0.05 ↓	1	1	0.92	p<0.05 ↓	0.92
Rnf10	p<0.05 ↓	0.99	0.53	1	p<0.05 ↓	1
Nat1	0.99	1	p<0.05 ↑	1	1	1
Stx16	0.18	1	1	1	p<0.05 ↓	1
Tmem151b	0.07	1	1	1	p<0.05 ↓	1
Zzef1	0.96	1	1	1	p<0.05 ↓	1
Adck2	1	1	1	p<0.05 ↑	0.95	p<0.05 ↑
Myh1	0.1	p<0.05 ↓	0.08	0.2	1	0.2
Nes	1	p<0.05 ↓	1	0.95	0.23	0.95

Cbln3	1	0.99	1	1	1	1
Ccl26	1	1	1	0.96	1	0.96
Col18a1	1	1	1	1	1	1
Rab24	1	0.93	1	1	1	1
Tmem79	0.91	1	0.17	1	1	1
Ydjc	1	1	1	0.93	1	0.93
Cpb1	1	1	1	0.78	0.71	0.78
Kcnh3	1	0.64	0.99	1	1	1
Npm3	1	1	1	0.76	1	0.76
Nrcam	1	0.98	0.94	0.4	1	0.4
Serpinb5	1	1	1	0.17	0.94	0.17
Slc46a3	1	1	1	0.51	1	0.51
Zbtb4	1	0.19	1	p<0.05 ↑	1	p<0.05 ↑
Arrb2	1	0.9	1	0.16	1	0.16
Vsig4	1	1	1	1	0.99	1
Zfyve26	1	1	1	1	1	1
Ghrhr	1	p<0.05 ↑	1	1	1	1
Prss56	1	p<0.05 ↑	1	0.91	1	0.91
Tmem136	0.76	p<0.05 ↑	0.16	0.05	1	0.05
Ipp	1	1	1	p<0.05 ↑	1	p<0.05 ↑
Nrn1l	1	0.12	1	p<0.05 ↑	1	p<0.05 ↑
Ptpru	1	0.6	0.43	p<0.05 ↓	1	p<0.05 ↓

Supplementary Table A. 2.: Sex differences in significant genotypes for (A) sleep percent and (B) bout lengths with direction of change. ↑ represents values higher than Control animals, and ↓ represent values lower than Controls. P-values were calculating through Dunnett's Post Hoc analysis.

Genotype	Per1	Per2	Clock	Bmal1/Arntl	Homer1	Sik3	Nalcn
Ap4e1	1	0	0	0	0	0	0
BC030499	0	0	0	0	0	1	0
Bex4	1	0	0	0	0	0	0
Cacna2d3	1	0	1	1	0	0	1
Chn1	0	0	0	0	1	0	0
Dcaf10	0	0	1	0	0	0	0
Dnaja4	0	0	1	0	0	0	0
Eogt	1	0	0	0	0	0	0
Epgn	0	0	0	1	0	0	0
Ermp1	0	1	0	0	0	0	0
Foxo3	0	1	0	0	0	0	0

H1fx	1	0	0	0	0	0	0
Htr1f	0	0	0	1	0	0	0
Igsf11	1	0	0	0	0	0	0
Ipp	0	0	0	1	0	0	0
Kcnh3	1	1	1	1	0	0	0
Loxl1	1	0	0	0	0	0	0
Mos	0	0	0	0	0	1	0
Myo3a	0	0	0	1	0	0	1
Myo3b	0	0	0	0	1	0	0
Nat1	0	0	0	0	1	0	0
Ncald	0	0	0	0	0	0	1
Nek2	0	0	0	0	0	1	0
Nes	1	0	0	0	0	0	0
Nrcam	0	0	1	0	0	0	0
Parp16	0	0	0	0	1	0	0
Parp8	0	0	0	1	0	0	0
Pkp4	0	1	0	0	0	0	0
Ppp1r9b	1	0	0	0	0	0	0
Prom1	0	1	1	0	0	0	0
Ptpn5	0	0	0	0	0	0	1
Ptpru	1	0	0	0	0	0	0
Rab27b	0	0	1	0	0	0	0
Rab36	0	0	0	0	0	0	1
Rnf25	1	0	0	0	0	0	0
Slc46a3	0	1	0	0	0	0	0
Srcin1	0	0	0	0	0	0	1
Stx16	0	0	0	0	0	1	0
Tpgs2	0	0	1	1	0	0	0
Tppp	0	0	0	0	1	0	1

Supplementary Table A. 3.: Binary table representing connections in gene network analysis. 1 indicates that gene is part of the network.

Genotype	Count
Ap4e1	20
Pitx3	20
Rnf10	20
Myh1	20
Ghrhr	20

Chn1	20
Ncald	18
Tppp	15
Foxo3	14
Macrod2	14
Vsig4	12
Hsd17b1	12
Postn	12
Mag	12
Ydjc	11
Bzw2	10
Rimklb	9
Stx16	9
Pkp4	9
Ppp1r9b	8
Tmem79	8
Zzef1	8
Cdc20	7
Tpgs2	7
Arrb2	7
Prom2	7
Kcnh3	6
Ajap1	6
H1fx	6
Ptpn5	6
Loxl1	6
Rab24	6
Prss56	6
Tmem151b	5
Dcaf10	5
Parp8	5
Cldn13	5
Rab3ip	5
Prokr1	5
Rab36	5
Cbln3	4
Htr3b	4
Htr1f	4
Serpinb5	4
Epgn	4
Mylip	4

Adck2	4
Tmod2	4
Rnf25	4
Mettl7b	4
Cers5	4
Thsd1	4
Ccl26	3
Neurl2	3
Eogt	3
Slc8b1	3
Zbtb4	3
Ptpru	3
Gipc3	3
Slc1a1	3
Acsf2	3
Cfb	3
Dnaja4	2
Slc46a3	2
Npm3	2
Tmem136	2
Myo3b	2
Arf2	2
Zfp219	2
Sox18	2
Carf	2
Lin28b	2
Rxfp4	2
Zfp689	2
Col18a1	2
Crym	2
Nrcam	1
Cpb1	1
Gjd4	1
Ifnl3	1
Ttll6	1
Zfp961	1
Bex4	1
Ermp1	1
Htr1d	1
Ipp	1
Rab27b	1

Nes	1
Dnajc14	1
Epha10	1
Nfatc4	1
Parp16	1
Tdrkh	1
Nat1	1
BC030499	1
C1qa	1
Ces4a	1
Il12rb2	1
Zfyve26	1
Nrn1l	1
Igsf11	NA
Akr1d1	NA
Cacna2d3	NA
Enox1	NA
Fndc4	NA
Galnt12	NA
Gpr156	NA
Masp1	NA
Mos	NA
Myo3a	NA
Nek2	NA
Ovch2	NA
Pfdn6	NA
Rab3c	NA
Rab3gap2	NA
Rbm4b	NA
Srcin1	NA
Tas2r138	NA
Tnfsf18	NA
Tubb4a	NA
Cd84	NA
Krt9	NA

Supplementary Table A. 4.: Count of gene-MP Term associations. Column "Count" represents the number of mammalian phenotype (MP) terms a single gene was associated with.

MP Term	Count
abnormal sleep behavior	37
abnormal behavioral response to light	23
decreased circulating glucose level	22
Info not available	22
decreased total body fat amount	20
abnormal behavior	16
hyperactivity	14
abnormal bone structure	14
decreased bone mineral content	14
decreased lean body mass	12
increased circulating sodium level	11
decreased bone mineral density	11
increased circulating alkaline phosphatase level	10
decreased circulating insulin level	10
decreased grip strength	10
increased lean body mass	9
decreased body length	8
increased hematocrit	8
increased circulating potassium level	8
impaired righting response	8
increased fasted circulating glucose level	7
hypoactivity	7
convulsive seizures	7
abnormal retinal pigmentation	7
increased bone mineral content	6
decreased leukocyte cell number	6
increased mean corpuscular volume	6
increased mean corpuscular hemoglobin	6
increased hemoglobin content	6
decreased circulating triglyceride level	6
abnormal retina morphology	6
abnormal QRS complex	5
increased erythrocyte cell number	5
improved glucose tolerance	5
decreased circulating free fatty acid level	5
abnormal coat appearance	5
shortened RR interval	4
increased body length	4
increased bone mineral density	4
increased or absent threshold for auditory brainstem response	4

preweaning lethality, incomplete penetrance	4
increased exploration in new environment	4
decreased vertical activity	4
increased grip strength	4
decreased exploration in new environment	4
increased vertical activity	4
decreased startle reflex	4
decreased anxiety-related response	4
persistence of hyaloid vascular system	4
decreased heart weight	3
cardiovascular system phenotype	3
decreased mean corpuscular hemoglobin concentration	3
thrombocytopenia	3
decreased threshold for auditory brainstem response	3
increased circulating HDL cholesterol level	3
decreased circulating HDL cholesterol level	3
decreased circulating cholesterol level	3
decreased circulating chloride level	3
decreased prepulse inhibition	3
increased total body fat amount	2
increased heart rate	2
shortened PQ interval	2
increased heart weight	2
increased mean corpuscular hemoglobin concentration	2
decreased mean corpuscular hemoglobin	2
increased blood urea nitrogen level	2
increased circulating glucose level	2
decreased fasted circulating glucose level	2
increased circulating chloride level	2
decreased blood urea nitrogen level	2
enlarged lymph nodes	2
abnormal coat/hair pigmentation	2
increased prepulse inhibition	2
increased coping response	2
straub tail	2
limb grasping	2
abnormal motor coordination/ balance	2
abnormal kidney morphology	2
male infertility	2
decreased heart rate variability	1
abnormal heart morphology	1

shortened ST segment	1
abnormal retinal blood vessel morphology	1
prolonged QRS complex duration	1
thin ventricular wall	1
shortened PR interval	1
prolonged PQ interval	1
prolonged ST segment	1
absent teeth	1
increased body weight	1
abnormal head morphology	1
enlarged spleen	1
increased gamma-delta T cell number	1
decreased alpha-beta T cell number	1
increased CD4-positive, CD25-positive, alpha-beta regulatory T cell number	1
decreased erythrocyte cell number	1
increased memory CD4-positive, CD25-positive, alpha-beta regulatory T cell number	1
decreased Ly6C-positive NK T cell number	1
increased marginal zone B cell number	1
decreased hemoglobin content	1
increased leukocyte cell number	1
decreased mean corpuscular volume	1
decreased memory-marker CD4-negative NK T cell number	1
decreased effector memory T-helper cell number	1
increased CD8-positive, naive alpha-beta T cell number	1
decreased macrophage cell number	1
increased CD8-positive, alpha-beta T cell number	1
decreased basophil cell number	1
decreased eosinophil cell number	1
decreased circulating amylase level	1
increased circulating alanine transaminase level	1
increased circulating bilirubin level	1
decreased circulating serum albumin level	1
decreased circulating alkaline phosphatase level	1
decreased urine creatinine level	1
increased circulating phosphate level	1
increased circulating iron level	1
decreased circulating iron level	1
decreased circulating phosphate level	1
decreased circulating alanine transaminase level	1
immune system phenotype	1
abnormal nail morphology	1

short tail	1
abnormal digit morphology	1
preweaning lethality, complete penetrance	1
embryonic lethality prior to tooth bud stage	1
embryonic lethality prior to organogenesis	1
prenatal lethality prior to heart atrial septation	1
abnormal social/conspecific interaction	1
impaired pupillary reflex	1
jumpy	1
absent startle reflex	1
abnormal vocalization	1
abnormal motor learning	1
decreased coping response	1
increased startle reflex	1
tremors	1
abnormal whole-body plethysmography	1
increased tidal volume	1
decreased pulmonary respiratory rate	1
increased sacral vertebrae number	1
anophthalmia	1
abnormal eyelid aperture	1
mydriasis	1
abnormal lens morphology	1
decreased cornea thickness	1
fused cornea and lens	1

Supplementary Table A. 5.: Frequency of MP Terms. Column "Count" represents the number of times a mammalian phenotype (MP) term was associated with a gene.

marker_symbol	mp_term
Ppp1r9b	hyperactivity|increased exploration in new environment|abnormal behavior|abnormal sleep behavior|abnormal bone structure|decreased lean body mass|decreased body length|increased coping response
Cbln3	abnormal retina morphology|increased blood urea nitrogen level|shortened RR interval|straub tail

Ap4e1	limb grasping|enlarged spleen|increased circulating alkaline phosphatase level|increased mean corpuscular hemoglobin concentration|decreased mean corpuscular hemoglobin|hyperactivity|decreased vertical activity|decreased heart weight|decreased grip strength|enlarged lymph nodes|increased gamma-delta T cell number|decreased heart rate variability|increased sacral vertebrae number|decreased alpha-beta T cell number|abnormal heart morphology|cardiovascular system phenotype|abnormal kidney morphology|decreased circulating glucose level|increased lean body mass|abnormal bone structure
Cdc20	preweaning lethality, complete penetrance|embryonic lethality prior to tooth bud stage|increased grip strength|improved glucose tolerance|decreased circulating glucose level|increased circulating glucose level|embryonic lethality prior to organogenesis
Bzw2	increased body length|increased bone mineral content|decreased circulating glucose level|decreased total body fat amount|increased lean body mass|abnormal sleep behavior|abnormal bone structure|shortened RR interval|increased bone mineral density|increased circulating glucose level
Tpgs2	decreased bone mineral content|decreased exploration in new environment|abnormal sleep behavior|abnormal bone structure|male infertility|decreased body length|decreased circulating free fatty acid level
Pitx3	hypoactivity|increased circulating alkaline phosphatase level|increased heart rate|decreased vertical activity|decreased bone mineral content|anophthalmia|decreased exploration in new environment|decreased lean body mass|abnormal bone structure|abnormal social/conspecific interaction|shortened ST segment|abnormal behavioral response to light|increased grip strength|increased exploration in new environment|decreased circulating triglyceride level|decreased circulating amylase level|decreased total body fat amount|abnormal eyelid aperture|abnormal sleep behavior|increased body weight
Ccl26	decreased circulating insulin level|abnormal behavioral response to light|improved glucose tolerance
Vsig4	decreased leukocyte cell number|hyperactivity|increased circulating alanine transaminase level|improved glucose tolerance|decreased circulating glucose level|increased vertical activity|abnormal behavior|decreased grip strength|increased circulating potassium level|abnormal behavioral response to light|shortened PQ interval|abnormal sleep behavior
Nrcam	increased fasted circulating glucose level
Htr3b	increased body length|decreased bone mineral density|increased bone mineral density|decreased grip strength
Kcnh3	cardiovascular system phenotype|increased circulating HDL cholesterol level|abnormal retinal pigmentation|increased fasted circulating glucose level|abnormal sleep behavior|increased CD4-positive, CD25-positive, alpha-beta regulatory T cell number

Ncald	increased circulating alkaline phosphatase level|decreased circulating HDL cholesterol level|hyperactivity|decreased bone mineral content|decreased circulating glucose level|increased circulating sodium level|abnormal behavior|impaired pupillary reflex|decreased bone mineral density|decreased lean body mass|abnormal bone structure|mydriasis|decreased mean corpuscular hemoglobin concentration|decreased total body fat amount|abnormal sleep behavior|increased total body fat amount|decreased body length|abnormal retinal blood vessel morphology
Foxo3	hypoactivity|increased mean corpuscular volume|thrombocytopenia|increased mean corpuscular hemoglobin|decreased prepulse inhibition|abnormal retina morphology|decreased bone mineral density|prolonged QRS complex duration|abnormal behavioral response to light|abnormal retinal pigmentation|decreased startle reflex|decreased total body fat amount|decreased erythrocyte cell number|abnormal sleep behavior
Rimklb	abnormal behavioral response to light|decreased leukocyte cell number|hyperactivity|increased bone mineral content|abnormal behavior|decreased circulating triglyceride level|decreased total body fat amount|increased lean body mass|abnormal sleep behavior
Htr1f	increased mean corpuscular volume|cardiovascular system phenotype|increased circulating sodium level|short tail
Dnaja4	decreased circulating glucose level|increased circulating bilirubin level
Ydjc	increased memory CD4-positive, CD25-positive, alpha-beta regulatory T cell number|increased circulating alkaline phosphatase level|abnormal behavioral response to light|increased hemoglobin content|decreased circulating glucose level|increased circulating sodium level|increased hematocrit|abnormal retinal pigmentation|abnormal retina morphology|increased lean body mass|increased erythrocyte cell number
Cpb1	Info not available
Gjd4	Info not available
Ifnl3	Info not available
Serpinb5	increased mean corpuscular volume|decreased total body fat amount|decreased grip strength|decreased body length
Slc46a3	abnormal coat appearance|abnormal behavioral response to light
Neurl2	abnormal coat appearance|decreased circulating glucose level|abnormal QRS complex
Eogt	hypoactivity|decreased circulating insulin level|abnormal behavior
Epgn	decreased circulating insulin level|increased hematocrit|decreased circulating glucose level|impaired righting response
Npm3	abnormal behavioral response to light|decreased circulating glucose level
Slc8b1	abnormal coat appearance|decreased circulating glucose level|impaired righting response
Ttll6	decreased circulating glucose level

Hsd17b1	increased circulating alkaline phosphatase level|decreased circulating glucose level|abnormal behavior|abnormal retina morphology|decreased exploration in new environment|impaired righting response|increased circulating potassium level|abnormal behavioral response to light|abnormal retinal pigmentation|abnormal lens morphology|abnormal sleep behavior|abnormal motor coordination/ balance
Zfp961	decreased threshold for auditory brainstem response
Bex4	Info not available
Tmem79	abnormal coat appearance|abnormal behavioral response to light|abnormal retinal pigmentation|decreased fasted circulating glucose level|decreased startle reflex|decreased total body fat amount|increased lean body mass|abnormal sleep behavior
Macrod2	hyperactivity|improved glucose tolerance|convulsive seizures|increased vertical activity|abnormal behavior|abnormal retina morphology|decreased lean body mass|decreased circulating insulin level|abnormal behavioral response to light|increased bone mineral content|increased fasted circulating glucose level|increased blood urea nitrogen level|abnormal sleep behavior|decreased body length
Mylip	hyperactivity|increased exploration in new environment|abnormal behavior|abnormal sleep behavior
Postn	increased circulating alkaline phosphatase level|decreased bone mineral content|decreased heart weight|decreased bone mineral density|abnormal bone structure|decreased lean body mass|impaired righting response|increased circulating potassium level|abnormal behavioral response to light|decreased total body fat amount|abnormal sleep behavior|decreased body length
Tmem136	decreased bone mineral content|abnormal sleep behavior
Rnf10	limb grasping|hypoactivity|hyperactivity|improved glucose tolerance|jumpy|abnormal digit morphology|decreased cornea thickness|decreased grip strength|preweaning lethality, incomplete penetrance|straub tail|decreased circulating serum albumin level|prenatal lethality prior to heart atrial septation|decreased Ly6C-positive NK T cell number|thin ventricular wall|increased marginal zone B cell number|decreased circulating alkaline phosphatase level|decreased body length|decreased circulating cholesterol level|increased circulating HDL cholesterol level|increased circulating chloride level
Myo3b	convulsive seizures|abnormal sleep behavior
Myh1	absent startle reflex|decreased leukocyte cell number|increased hemoglobin content|decreased vertical activity|decreased bone mineral content|convulsive seizures|decreased exploration in new environment|decreased bone mineral density|decreased lean body mass|abnormal bone structure|decreased grip strength|abnormal behavioral response to light|increased hematocrit|abnormal vocalization|decreased startle reflex|decreased total body fat amount|decreased anxiety-related response|increased erythrocyte cell number|decreased urine creatinine level|abnormal sleep behavior

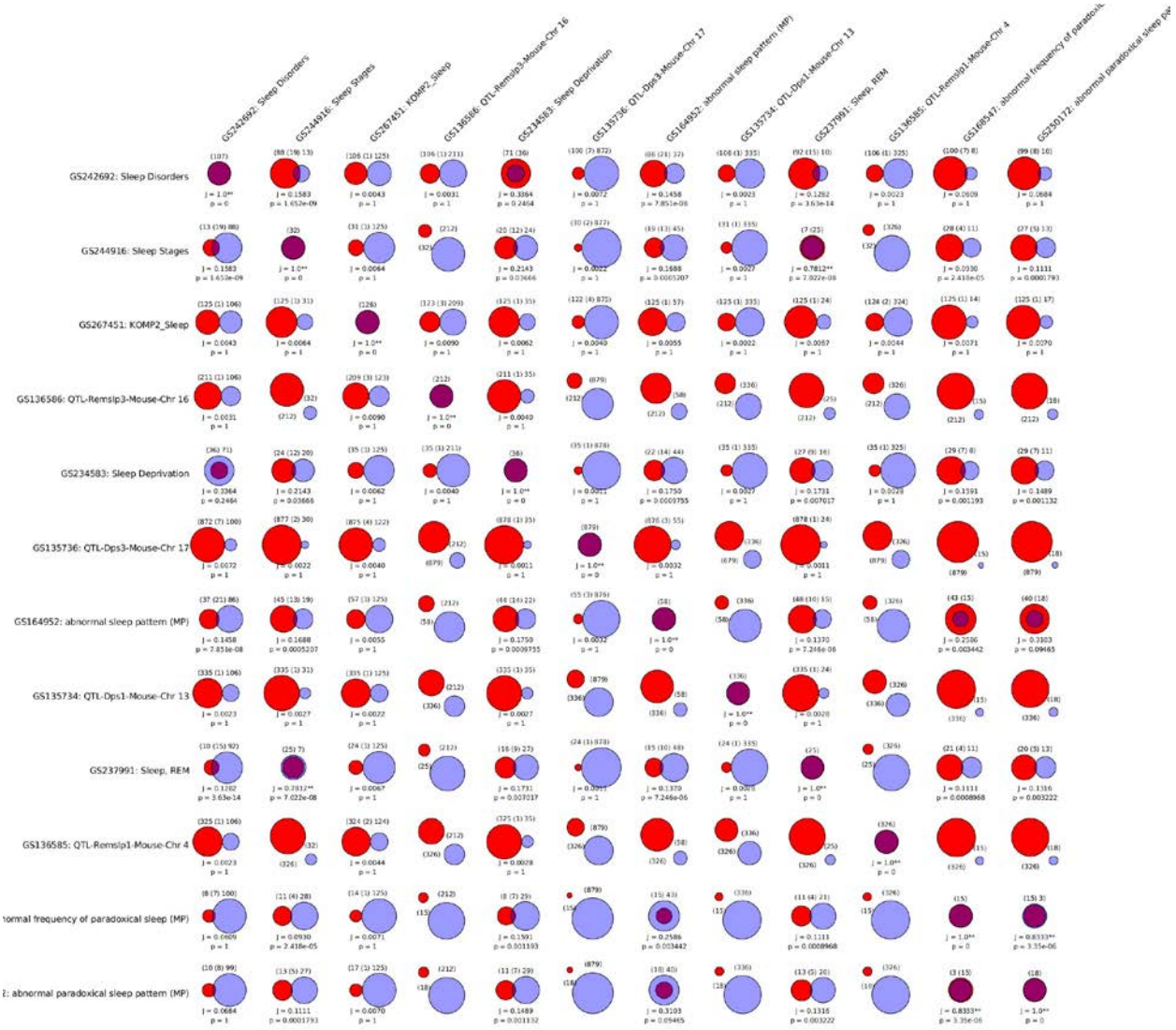
Arrb2	increased circulating phosphate level|increased circulating alkaline phosphatase level|increased circulating HDL cholesterol level|thrombocytopenia|increased grip strength|decreased total body fat amount|increased circulating potassium level
Ermp1	decreased prepulse inhibition
Zbtb4	hyperactivity|fused cornea and lens|persistence of hyaloid vascular system
Htr1d	Info not available
Adck2	abnormal motor learning|increased circulating sodium level|abnormal behavior|increased circulating potassium level
Ipp	decreased leukocyte cell number
Ptpru	increased vertical activity|abnormal sleep behavior|persistence of hyaloid vascular system
Ghrhr	decreased circulating cholesterol level|decreased circulating HDL cholesterol level|decreased mean corpuscular hemoglobin|decreased vertical activity|thrombocytopenia|decreased bone mineral content|decreased hemoglobin content|decreased circulating glucose level|decreased heart weight|decreased bone mineral density|decreased lean body mass|abnormal bone structure|decreased grip strength|shortened PR interval|abnormal behavioral response to light|increased grip strength|decreased mean corpuscular hemoglobin concentration|shortened PQ interval|abnormal sleep behavior|increased leukocyte cell number
Rab27b	Info not available
Zzef1	increased mean corpuscular hemoglobin|decreased total body fat amount|absent teeth|abnormal sleep behavior|decreased grip strength|decreased circulating free fatty acid level|increased coping response|increased circulating potassium level
Tmem151b	abnormal behavioral response to light|hyperactivity|increased circulating sodium level|abnormal behavior|abnormal sleep behavior
Stx16	decreased circulating insulin level|increased exploration in new environment|increased bone mineral content|abnormal behavior|decreased circulating triglyceride level|decreased total body fat amount|increased lean body mass|abnormal sleep behavior|abnormal QRS complex
Nes	abnormal sleep behavior
Tmod2	abnormal behavioral response to light|hyperactivity|abnormal behavior|abnormal sleep behavior
Dcaf10	increased mean corpuscular volume|abnormal behavioral response to light|abnormal behavior|increased mean corpuscular hemoglobin|decreased anxiety-related response
Parp8	hypoactivity|decreased coping response|abnormal behavioral response to light|abnormal retina morphology|abnormal sleep behavior
Rnf25	decreased circulating chloride level|hypoactivity|abnormal behavioral response to light|abnormal behavior

Ajap1	decreased bone mineral content|increased startle reflex|increased circulating sodium level|abnormal sleep behavior|abnormal bone structure|increased circulating potassium level
Arf2	abnormal coat/hair pigmentation|increased hematocrit
Cldn13	decreased mean corpuscular volume|decreased fasted circulating glucose level|increased erythrocyte cell number|decreased blood urea nitrogen level|increased circulating potassium level
Dnajc14	increased or absent threshold for auditory brainstem response
Epha10	impaired righting response
Gipc3	decreased startle reflex|decreased prepulse inhibition|increased or absent threshold for auditory brainstem response
H1fx	increased circulating sodium level|increased fasted circulating glucose level|decreased total body fat amount|increased lean body mass|abnormal sleep behavior|shortened RR interval
Nfatc4	Info not available
Parp16	Info not available
Rab3ip	decreased circulating insulin level|increased circulating alkaline phosphatase level|decreased bone mineral content|decreased bone mineral density|increased prepulse inhibition
Slc1a1	increased circulating iron level|decreased total body fat amount|persistence of hyaloid vascular system
Tdrkh	Info not available
Zfp219	convulsive seizures|decreased circulating glucose level
Chn1	hypoactivity|increased circulating alkaline phosphatase level|decreased circulating HDL cholesterol level|increased mean corpuscular volume|hyperactivity|decreased circulating iron level|increased hemoglobin content|decreased bone mineral content|decreased circulating glucose level|increased circulating sodium level|abnormal behavior|increased mean corpuscular hemoglobin|increased lean body mass|decreased bone mineral density|impaired righting response|preweaning lethality, incomplete penetrance|decreased circulating insulin level|abnormal behavioral response to light|decreased mean corpuscular hemoglobin concentration|increased hematocrit
Mettl7b	decreased circulating chloride level|abnormal coat/hair pigmentation|decreased circulating glucose level|abnormal sleep behavior
Ptpn5	increased heart weight|decreased leukocyte cell number|increased bone mineral content|abnormal sleep behavior|abnormal bone structure|increased prepulse inhibition
Prom2	decreased circulating insulin level|hyperactivity|abnormal retinal pigmentation|abnormal nail morphology|decreased memory-marker CD4-negative NK T cell number|decreased grip strength|abnormal QRS complex
Nat1	abnormal sleep behavior
Sox18	abnormal behavior|decreased blood urea nitrogen level

Acsf2	abnormal behavioral response to light|increased mean corpuscular hemoglobin|decreased total body fat amount
BC030499	Info not available
C1qa	decreased circulating glucose level
Carf	prolonged PQ interval|prolonged ST segment
Cers5	decreased bone mineral content|increased circulating sodium level|decreased lean body mass|abnormal motor coordination/ balance
Ces4a	Info not available
Cfb	increased hemoglobin content|increased hematocrit|increased erythrocyte cell number
Il12rb2	decreased circulating free fatty acid level
Lin28b	increased heart rate|shortened RR interval
Loxl1	increased body length|convulsive seizures|increased fasted circulating glucose level|decreased threshold for auditory brainstem response|abnormal sleep behavior|abnormal QRS complex
Mag	abnormal whole-body plethysmography|convulsive seizures|decreased circulating glucose level|decreased circulating phosphate level|decreased circulating insulin level|increased tidal volume|tremors|decreased circulating triglyceride level|decreased total body fat amount|abnormal sleep behavior|increased or absent threshold for auditory brainstem response|decreased pulmonary respiratory rate
Pkp4	decreased bone mineral content|increased bone mineral content|increased hematocrit|decreased total body fat amount|increased lean body mass|decreased bone mineral density|abnormal sleep behavior|abnormal bone structure|impaired righting response
Prokr1	increased body length|abnormal retinal pigmentation|decreased circulating glucose level|abnormal sleep behavior|increased bone mineral density
Rab36	decreased circulating cholesterol level|increased mean corpuscular hemoglobin concentration|increased hemoglobin content|abnormal head morphology|decreased circulating triglyceride level
Rxfp4	decreased circulating chloride level|increased fasted circulating glucose level
Thsd1	increased mean corpuscular volume|increased mean corpuscular hemoglobin|decreased total body fat amount|decreased circulating free fatty acid level
Tppp	increased circulating alkaline phosphatase level|hyperactivity|increased hemoglobin content|decreased bone mineral content|increased circulating chloride level|convulsive seizures|decreased bone mineral density|decreased lean body mass|abnormal bone structure|abnormal behavioral response to light|increased hematocrit|decreased anxiety-related response|increased erythrocyte cell number|abnormal sleep behavior|decreased body length
Zfp689	abnormal sleep behavior|preweaning lethality, incomplete penetrance
Col18a1	increased fasted circulating glucose level|increased vertical activity

Rab24	abnormal coat appearance|decreased bone mineral content|decreased anxiety-related response|decreased bone mineral density|abnormal bone structure|decreased lean body mass
Zfyve26	increased or absent threshold for auditory brainstem response
Prss56	decreased effector memory T-helper cell number|increased CD8-positive, naive alpha-beta T cell number|decreased leukocyte cell number|decreased macrophage cell number|immune system phenotype|increased CD8-positive, alpha-beta T cell number
Nrn1l	decreased circulating glucose level
Crym	abnormal sleep behavior|decreased grip strength
Igsf11	male infertility
Akr1d1	persistence of hyaloid vascular system
Cacna2d3	decreased circulating glucose level|increased circulating sodium level|decreased circulating free fatty acid level
Enox1	decreased threshold for auditory brainstem response|decreased lean body mass|increased bone mineral density|increased total body fat amount
Fndc4	Info not available
Galnt12	Info not available
Gpr156	Info not available
Masp1	Info not available
Mos	Info not available
Myo3a	abnormal behavioral response to light
Nek2	decreased basophil cell number|abnormal kidney morphology|decreased eosinophil cell number|enlarged lymph nodes
Ovch2	Info not available
Pfdn6	impaired righting response|preweaning lethality, incomplete penetrance
Rab3c	Info not available
Rab3gap2	Info not available
Rbm4b	Info not available
Srcin1	increased heart weight|decreased total body fat amount|abnormal QRS complex
Tas2r138	Info not available
Tnfsf18	decreased lean body mass
Tubb4a	Info not available
Cd84	increased circulating sodium level|decreased circulating triglyceride level|decreased total body fat amount|decreased circulating alanine transaminase level
Krt9	decreased circulating insulin level

Supplementary Table A. 6.: Mammalian Phenotype (MP) terms associated with sleep related candidate genes from KOMP[2]



Supplementary Figure A. 2.: Jaccard Analysis for geneset similarities from 12 shortlisted genelists from Geneweaver

REFERENCES

Abbott, A. (2010). "Mouse project to find each gene's role." Nature **465**(7297): 410.

Ameratunga, D., J. Goldin and M. Hickey (2012). "Sleep disturbance in menopause." Intern Med J **42**(7): 742-747.

Aserinsky, E. and N. Kleitman (1953). "Regularly occurring periods of eye motility, and concomitant phenomena, during sleep." Science **118**(3062): 273-274.

Bassett, J. H., A. Gogakos, J. K. White, H. Evans, R. M. Jacques, A. H. van der Spek, P. Sanger Mouse Genetics, R. Ramirez-Solis, E. Ryder, D. Sunter, A. Boyde, M. J. Campbell, P. I. Croucher and G. R. Williams (2012). "Rapid-throughput skeletal phenotyping of 100 knockout mice identifies 9 new genes that determine bone strength." PLoS Genet **8**(8): e1002858.

Becerra, M. B., B. J. Becerra and M. Teodorescu (2016). "Healthcare burden of obstructive sleep apnea and obesity among asthma hospitalizations: Results from the U.S.-based Nationwide Inpatient Sample." Respir Med **117**: 230-236.

Belfer, S. J., H. S. Chuang, B. L. Freedman, J. Yuan, M. Norton, H. H. Bau and D. M. Raizen (2013). "Caenorhabditis-in-drop array for monitoring C. elegans quiescent behavior." Sleep **36**(5): 689-698G.

Bianchi, M. T., K. Russo, H. Gabbidon, T. Smith, B. Goparaju and M. B. Westover (2017). "Big data in sleep medicine: prospects and pitfalls in phenotyping." Nat Sci Sleep **9**: 11-29.

Borbely, A. A. (1982). "A two process model of sleep regulation." Hum Neurobiol **1**(3): 195-204.

Borbely, A. A. (1998). "Processes underlying sleep regulation." Horm Res **49**(3-4): 114-117.

Borbely, A. A. and P. Achermann (1992). "Concepts and models of sleep regulation: an overview." J Sleep Res **1**(2): 63-79.

Bradley, A., K. Anastassiadis, A. Ayadi, J. F. Battey, C. Bell, M. C. Birling, J. Bottomley, S. D. Brown, A. Burger, C. J. Bult, W. Bushell, F. S. Collins, C. Desaintes, B. Doe, A. Economides, J. T. Eppig, R. H. Finnell, C. Fletcher, M. Fray, D. Frendewey, R. H. Friedel, F. G. Grosveld, J. Hansen, Y. Herault, G. Hicks, A. Horlein, R. Houghton, M. Hrabe de Angelis, D. Huylebroeck, V. Iyer, P. J. de Jong, J. A. Kadin, C. Kaloff, K. Kennedy, M. Koutsourakis, K. C. Lloyd, S. Marschall, J. Mason, C. McKerlie, M. P. McLeod, H. von Melchner, M. Moore, A. O. Mujica, A. Nagy, M. Nefedov, L. M. Nutter, G. Pavlovic, J. L. Peterson, J. Pollock, R. Ramirez-Solis, D. E. Rancourt, M. Raspa, J. E. Remacle, M. Ringwald, B. Rosen, N. Rosenthal, J. Rossant, P. Ruiz Noppinger, E. Ryder, J. Z. Schick, F. Schnutgen, P. Schofield,

C. Seisenberger, M. Selloum, E. M. Simpson, W. C. Skarnes, D. Smedley, W. L. Stanford, A. F. Stewart, K. Stone, K. Swan, H. Tadepally, L. Teboul, G. P. Tocchini-Valentini, D. Valenzuela, A. P. West, K. Yamamura, Y. Yoshinaga and W. Wurst (2012). "The mammalian gene function resource: the International Knockout Mouse Consortium." Mamm Genome **23**(9-10): 580-586.

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

Brown, S. D. and M. W. Moore (2012). "The International Mouse Phenotyping Consortium: past and future perspectives on mouse phenotyping." Mamm Genome **23**(9-10): 632-640.

Brown, S. D. and M. W. Moore (2012). "Towards an encyclopaedia of mammalian gene function: the International Mouse Phenotyping Consortium." Dis Model Mech **5**(3): 289-292.

Bubier, J. A., M. A. Langston, E. J. Baker and E. J. Chesler (2017). "Integrative Functional Genomics for Systems Genetics in GeneWeaver.org." Methods Mol Biol **1488**: 131-152.

Buysse, D. J., A. Germain, M. L. Hall, D. E. Moul, E. A. Nofzinger, A. Begley, C. L. Ehlers, W. Thompson and D. J. Kupfer (2008). "EEG spectral analysis in primary insomnia: NREM period effects and sex differences." Sleep **31**(12): 1673-1682.

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

Carrier, J., S. Land, D. J. Buysse, D. J. Kupfer and T. H. Monk (2001). "The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20-60 years old)." Psychophysiology **38**(2): 232-242.

Cassada, R. C. and R. L. Russell (1975). "The dauerlarva, a post-embryonic developmental variant of the nematode Caenorhabditis elegans." Dev Biol **46**(2): 326-342.

Cespedes, E. M., S. N. Bhupathiraju, Y. Li, B. Rosner, S. Redline and F. B. Hu (2016). "Long-term changes in sleep duration, energy balance and risk of type 2 diabetes." Diabetologia **59**(1): 101-109.

Chamberlain, S. (2015). "solr: General Purpose R Interface to Solr. R package version 0.1.6.".

Chaput, J. P., J. P. Despres, C. Bouchard and A. Tremblay (2008). "The association between sleep duration and weight gain in adults: a 6-year prospective study from the Quebec Family Study." Sleep **31**(4): 517-523.

Chesler, E. J. (2014). "Out of the bottleneck: the Diversity Outcross and Collaborative Cross mouse populations in behavioral genetics research." Mamm Genome **25**(1-2): 3-11.

Churchill, G. A., D. M. Gatti, S. C. Munger and K. L. Svenson (2012). "The Diversity Outbred mouse population." Mamm Genome **23**(9-10): 713-718.

Cirelli, C. (2009). "The genetic and molecular regulation of sleep: from fruit flies to humans." Nat Rev Neurosci **10**(8): 549-560.

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

Cirelli, C. and G. Tononi (2008). "Is sleep essential?" PLoS Biol **6**(8): e216.

Cohen, J. (1992). "A power primer." Psychol Bull **112**(1): 155-159.

Cohen, S., W. J. Doyle, C. M. Alper, D. Janicki-Deverts and R. B. Turner (2009). "Sleep habits and susceptibility to the common cold." Arch Intern Med **169**(1): 62-67.

Collop, N. A., D. Adkins and B. A. Phillips (2004). "Gender differences in sleep and sleep-disordered breathing." Clin Chest Med **25**(2): 257-268.

Daan, S., D. G. Beersma and A. A. Borbely (1984). "Timing of human sleep: recovery process gated by a circadian pacemaker." Am J Physiol **246**(2 Pt 2): R161-183.

Davis, H., P. A. Davis, A. L. Loomis, E. N. Harvey and G. Hobart (1937). "Changes in Human Brain Potentials during the Onset of Sleep." Science **86**(2237): 448-450.

Dean, D. A., 2nd, A. L. Goldberger, R. Mueller, M. Kim, M. Rueschman, D. Mobley, S. S. Sahoo, C. P. Jayapandian, L. Cui, M. G. Morrical, S. Surovec, G. Q. Zhang and S. Redline (2016). "Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource." Sleep **39**(5): 1151-1164.

Derwinska, K., H. Mierzewska, A. Goszczanska, E. Szczepanik, Z. L. Xia, K. Kusmierska, J. Tryfon, A. Kutkowska-Kazmierczak, E. Bocian, T. Mazurczak, E. Obersztyn and P. Stankiewicz (2012). "Clinical improvement of the aggressive neurobehavioral phenotype in a patient with a deletion of PITX3 and the absence of L-DOPA in the cerebrospinal fluid." American Journal of Medical Genetics Part B-Neuropsychiatric Genetics **159B**(2): 236-242.

Diekelmann, S. and J. Born (2010). "The memory function of sleep." Nat Rev Neurosci **11**(2): 114-126.

Dijk, D. J., D. G. Beersma and S. Daan (1987). "EEG power density during nap sleep: reflection of an hourglass measuring the duration of prior wakefulness." J Biol Rhythms **2**(3): 207-219.

Ding, F., J. O'Donnell, Q. Xu, N. Kang, N. Goldman and M. Nedergaard (2016). "Changes in the composition of brain interstitial ions control the sleep-wake cycle." Science **352**(6285): 550-555.

Donohue, K. D., D. C. Medonza, E. R. Crane and B. F. O'Hara (2008). "Assessment of a non-invasive high-throughput classifier for behaviours associated with sleep and wake in mice." Biomedical Engineering Online **7**: 14.

Donohue, K. D., D. C. Medonza, E. R. Crane and B. F. O'Hara (2008). "Assessment of a non-invasive high-throughput classifier for behaviours associated with sleep and wake in mice." Biomed Eng Online **7**: 14.

Drager, U. C. (2006). "Retinoic acid signaling in the functioning brain." Sci STKE **2006**(324): pe10.

Dunnett, C. W. (1955). "A multiple comparison procedure for comparing several treatments with a control." Journal of the American Statistical Association **50**(272): 1096-1121.

Durmer, J. S. and D. F. Dinges (2005). "Neurocognitive consequences of sleep deprivation." Semin Neurol **25**(1): 117-129.

Feng, J., Z. Yan, A. Ferreira, K. Tomizawa, J. A. Liauw, M. Zhuo, P. B. Allen, C. C. Ouimet and P. Greengard (2000). "Spinophilin regulates the formation and function of dendritic spines." Proc Natl Acad Sci U S A **97**(16): 9287-9292.

Fisher, S. P., S. I. Godinho, C. A. Pothecary, M. W. Hankins, R. G. Foster and S. N. Peirson (2012). "Rapid assessment of sleep-wake behavior in mice." J Biol Rhythms **27**(1): 48-58.

Flores, A. E., J. E. Flores, H. Deshpande, J. A. Picazo, X. Xie, P. Franken, H. C. Heller, D. A. Grahn and B. F. O'Hara (2007). "Pattern recognition of sleep in rodents using piezoelectric signals generated by gross body movements." IEEE transactions on biomedical engineering **54**(2): 225-233.

Flores, A. E., J. E. Flores, H. Deshpande, J. A. Picazo, X. M. S. Xie, P. Franken, H. C. Heller, D. A. Grahn and B. F. O'Hara (2007). "Pattern recognition of sleep in rodents using piezoelectric signals generated by gross body movements." Ieee Transactions on Biomedical Engineering **54**(2): 225-233.

Franken, P. (2013). "A role for clock genes in sleep homeostasis." Curr Opin Neurobiol **23**(5): 864-872.

Franken, P., R. Thomason, H. C. Heller and B. F. O'Hara (2007). "A non-circadian role for clock-genes in sleep homeostasis: a strain comparison." BMC Neurosci **8**: 87.

Fu, L., M. S. Patel, A. Bradley, E. F. Wagner and G. Karsenty (2005). "The molecular clock mediates leptin-regulated bone formation." Cell **122**(5): 803-815.

Fujihara, Y., H. Kondo, T. Noguchi and A. Togari (2014). "Glucocorticoids mediate circadian timing in peripheral osteoclasts resulting in the circadian expression rhythm of osteoclast-related genes." Bone **61**: 1-9.

Funato, H., C. Miyoshi, T. Fujiyama, T. Kanda, M. Sato, Z. Wang, J. Ma, S. Nakane, J. Tomita, A. Ikkyu, M. Kakizaki, N. Hotta-Hirashima, S. Kanno, H. Komiya, F. Asano, T. Honda, S. J. Kim, K. Harano, H. Muramoto, T. Yonezawa, S. Mizuno, S. Miyazaki, L. Connor, V. Kumar, I. Miura, T. Suzuki, A. Watanabe, M. Abe, F. Sugiyama, S. Takahashi, K. Sakimura, Y. Hayashi, Q. Liu, K. Kume, S. Wakana, J. S. Takahashi and M. Yanagisawa (2016). "Forward-genetics analysis of sleep in randomly mutagenized mice." Nature **539**(7629): 378-383.

Gondo, Y., R. Fukumura, T. Murata and S. Makino (2009). "Next-generation gene targeting in the mouse for functional genomics." BMB Rep **42**(6): 315-323.

Hartse, K. M. (2011). "The phylogeny of sleep." Handb Clin Neurol **98**: 97-109.

Henderson, R. C. and C. D. Madsen (1999). "Bone mineral content and body composition in children and young adults with cystic fibrosis." Pediatr Pulmonol **27**(2): 80-84.

Hendricks, J. C., S. M. Finn, K. A. Panckeri, J. Chavkin, J. A. Williams, A. Sehgal and A. I. Pack (2000). "Rest in Drosophila is a sleep-like state." Neuron **25**(1): 129-138.

Hofstetter, J. R., J. A. Trofatter, K. L. Kernek, J. I. Nurnberger and A. R. Mayeda (2003). "New quantitative trait loci for the genetic variance in circadian period of locomotor activity between inbred strains of mice." J Biol Rhythms **18**(6): 450-462.

Hothorn, T., K. Hornik, C. Strobl, A. Zeileis and M. T. Hothorn (2015). "Package 'party'." Package Reference Manual for Party Version 0.9-998 **16**: 37.

Huang, B. and W. He (2010). "Molecular characteristics of inherited congenital cataracts." Eur J Med Genet **53**(6): 347-357.

Jiang, P., J. R. Scarpa, K. Fitzpatrick, B. Losic, V. D. Gao, K. Hao, K. C. Summa, H. S. Yang, B. Zhang, R. Allada, M. H. Vitaterna, F. W. Turek and A. Kasarskis (2015). "A systems approach identifies networks and genes linking sleep and stress: implications for neuropsychiatric disorders." Cell Rep **11**(5): 835-848.

Joshi, S. and B. F. O'Hara (2014). "Identifying candidate genes for variation in sleep-related quantitative traits."

Kaasik, K., S. Kivimae, J. J. Allen, R. J. Chalkley, Y. Huang, K. Baer, H. Kissel, A. L. Burlingame, K. M. Shokat, L. J. Ptacek and Y. H. Fu (2013). "Glucose sensor O-GlcNAcylation coordinates with phosphorylation to regulate circadian clock." Cell Metab **17**(2): 291-302.

Kapfhamer, D., O. Valladares, Y. Sun, P. M. Nolan, J. J. Rux, S. E. Arnold, S. C. Veasey and M. Bucan (2002). "Mutations in Rab3a alter circadian period and homeostatic response to sleep loss in the mouse." Nat Genet **32**(2): 290-295.

Kayser, M. S. and D. Biron (2016). "Sleep and Development in Genetically Tractable Model Organisms." Genetics **203**(1): 21-33.

Keane, T. M., L. Goodstadt, P. Danecek, M. A. White, K. Wong, B. Yalcin, A. Heger, A. Agam, G. Slater, M. Goodson, N. A. Furlotte, E. Eskin, C. Nellaker, H. Whitley, J. Cleak, D. Janowitz, P. Hernandez-Pliego, A. Edwards, T. G. Belgard, P. L. Oliver, R. E. McIntyre, A. Bhomra, J. Nicod, X. Gan, W. Yuan, L. van der Weyden, C. A. Steward, S. Bala, J. Stalker, R. Mott, R. Durbin, I. J. Jackson, A. Czechanski, J. A. Guerra-Assuncao, L. R. Donahue, L. G. Reinholdt, B. A. Payseur, C. P. Ponting, E. Birney, J. Flint and D. J. Adams (2011). "Mouse genomic variation and its effect on phenotypes and gene regulation." Nature **477**(7364): 289-294.

Khosrowshahian, F., M. Wolanski, W. Y. Chang, K. Fujiki, L. Jacobs and M. J. Crawford (2005). "Lens and retina formation require expression of Pitx3 in Xenopus pre-lens ectoderm." Dev Dyn **234**(3): 577-589.

Kloss, B., J. L. Price, L. Saez, J. Blau, A. Rothenfluh, C. S. Wesley and M. W. Young (1998). "The Drosophila clock gene double-time encodes a protein closely related to human casein kinase Iepsilon." Cell **94**(1): 97-107.

Koehl, M., S. Battle and P. Meerlo (2006). "Sex differences in sleep: the response to sleep deprivation and restraint stress in mice." Sleep **29**(9): 1224-1231.

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

Krishnan, V. and N. A. Collop (2006). "Gender differences in sleep disorders." Curr Opin Pulm Med **12**(6): 383-389.

Krueger, J. M., M. G. Frank, J. P. Wisor and S. Roy (2016). "Sleep function: Toward elucidating an enigma." Sleep Med Rev **28**: 46-54.

Kyriacou, C. P. and M. H. Hastings (2010). "Circadian clocks: genes, sleep, and cognition." Trends Cogn Sci **14**(6): 259-267.

Lehotzky, A., P. Lau, N. Tokesi, N. Muja, L. D. Hudson and J. Ovadi (2010). "Tubulin polymerization-promoting protein (TPPP/p25) is critical for oligodendrocyte differentiation." Glia **58**(2): 157-168.

Li, J., J. A. Dani and W. Le (2009). "The role of transcription factor Pitx3 in dopamine neuron development and Parkinson's disease." Curr Top Med Chem **9**(10): 855-859.

Li, X., L. Cui, S. Tao, J. Chen, X. Zhang and G. Q. Zhang (2017). "HyCLASSS: A Hybrid Classifier for Automatic Sleep Stage Scoring." IEEE J Biomed Health Inform.

Liaw, A. and M. Wiener (2002). "Classification and regression by randomForest." R news **2**(3): 18-22.

Lindsay, R., D. M. Hart and D. M. Clark (1984). "The minimum effective dose of estrogen for prevention of postmenopausal bone loss." Obstet Gynecol **63**(6): 759-763.

Lo, C. C., T. Chou, T. Penzel, T. E. Scammell, R. E. Strecker, H. E. Stanley and P. Ivanov (2004). "Common scale-invariant patterns of sleep-wake transitions across mammalian species." Proc Natl Acad Sci U S A **101**(50): 17545-17548.

Lowrey, P. L. and J. S. Takahashi (2004). "Mammalian circadian biology: elucidating genome-wide levels of temporal organization." Annu Rev Genomics Hum Genet **5**: 407-441.

Luchavova, M., V. Zikan, D. Michalska, I. Raska, Jr., A. A. Kubena and J. J. Stepan (2011). "The effect of timing of teriparatide treatment on the circadian rhythm of bone turnover in postmenopausal osteoporosis." Eur J Endocrinol **164**(4): 643-648.

Mang, G. M., J. Nicod, Y. Emmenegger, K. D. Donohue, B. F. O'Hara and P. Franken (2014). "Evaluation of a piezoelectric system as an alternative to electroencephalogram/ electromyogram recordings in mouse sleep studies." Sleep **37**(8): 1383-1392.

Manzoni, P., P. Brambilla, A. Pietrobelli, L. Beccaria, A. Bianchessi, S. Mora and G. Chiumello (1996). "Influence of body composition on bone mineral content in children and adolescents." Am J Clin Nutr **64**(4): 603-607.

Maret, S., S. Dorsaz, L. Gurcel, S. Pradervand, B. Petit, C. Pfister, O. Hagenbuchle, B. F. O'Hara, P. Franken and M. Tafti (2007). "Homer1a is a core brain molecular correlate of sleep loss." Proceedings of the National Academy of Sciences of the United States of America **104**(50): 20090-20095.

Marin, J. M., S. J. Carrizo, E. Vicente and A. G. Agusti (2005). "Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study." Lancet **365**(9464): 1046-1053.

Marshall, L. and J. Born (2007). "The contribution of sleep to hippocampus-dependent memory consolidation." Trends Cogn Sci **11**(10): 442-450.

Martin, C. P., J. Vazquez, J. Avila and F. J. Moreno (2002). "P24, a glycogen synthase kinase 3 (GSK 3) inhibitor." Biochim Biophys Acta **1586**(1): 113-122.

Mavanji, V., C. J. Billington, C. M. Kotz and J. A. Teske (2012). "Sleep and obesity: a focus on animal models." Neurosci Biobehav Rev **36**(3): 1015-1029.

McNamara, P., I. Capellini, E. Harris, C. L. Nunn, R. A. Barton and B. Preston (2008). "The Phylogeny of Sleep Database: A New Resource for Sleep Scientists." Open Sleep J **1**: 11-14.

McShane, B. B., R. J. Galante, M. Biber, S. T. Jensen, A. J. Wyner and A. I. Pack (2012). "Assessing REM sleep in mice using video data." Sleep **35**(3): 433-442.

Megens, A., J. Voeten, J. Rombouts, T. F. Meert and C. Niemegeers (1987). "Behavioural activity of rats measured by a new method based on the piezo-electric principle." Psychopharmacology **93**(3): 382-388.

Mekada, K., M. Hirose, A. Murakami and A. Yoshiki (2015). "Development of SNP markers for C57BL/6N-derived mouse inbred strains." Exp Anim **64**(1): 91-100.

Michalska, D., M. Luchavova, V. Zikan, I. Raska, Jr., A. A. Kubena and J. J. Stepan (2012). "Effects of morning vs. evening teriparatide injection on bone mineral density and bone turnover markers in postmenopausal osteoporosis." Osteoporos Int **23**(12): 2885-2891.

Mitchell, A. F. S. and W. J. Krzanowski (1985). "The Mahalanobis Distance and Elliptic Distributions." Biometrika **72**(2): 464-467.

Miyake, A., S. Takahashi, Y. Nakamura, K. Inamura, S. Matsumoto, S. Mochizuki and M. Katou (2009). "Disruption of the ether-a-go-go K+ channel gene BEC1/KCNH3 enhances cognitive function." J Neurosci **29**(46): 14637-14645.

Moreno-De-Luca, A., S. L. Helmers, H. Mao, T. G. Burns, A. M. Melton, K. R. Schmidt, P. M. Fernhoff, D. H. Ledbetter and C. L. Martin (2011). "Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability." J Med Genet **48**(2): 141-144.

Morgan, H., M. Simon and A. M. Mallon (2012). "Accessing and mining data from large-scale mouse phenotyping projects." Int Rev Neurobiol **104**: 47-70.

Nadeau, J. H. (2001). "Modifier genes in mice and humans." Nat Rev Genet **2**(3): 165-174.

Nieto, F. J., T. B. Young, B. K. Lind, E. Shahar, J. M. Samet, S. Redline, R. B. D'Agostino, A. B. Newman, M. D. Lebowitz and T. G. Pickering (2000). "Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study." JAMA **283**(14): 1829-1836.

Nishida, M., J. Pearsall, R. L. Buckner and M. P. Walker (2009). "REM sleep, prefrontal theta, and the consolidation of human emotional memory." Cereb Cortex **19**(5): 1158-1166.

Nishino, S. (2013). Neurotransmitters and Neuropharmacology of Sleep/Wake Regulations A2 - Kushida, Clete A. Encyclopedia of Sleep. Waltham, Academic Press**:** 395-406.

O'Hara, B. F., J. Ding, R. L. Bernat and P. Franken (2007). "Genomic and proteomic approaches towards an understanding of sleep." CNS Neurol Disord Drug Targets **6**(1): 71-81.

O'Hara, B. F., P. Jiang, F. W. Turek and P. Franken (2017). Chapter 29 - Genetics and Genomic Basis of Sleep in Rodents A2 - Kryger, Meir. Principles and Practice of Sleep Medicine (Sixth Edition). T. Roth and W. C. Dement, Elsevier**:** 296-309.e295.

Opitz, D. and R. Maclin (1999). "Popular ensemble methods: An empirical study." Journal of Artificial Intelligence Research **11**: 169-198.

Pack, A. I., R. J. Galante, G. Maislin, J. Cater, D. Metaxas, S. Lu, L. Zhang, R. Von Smith, T. Kay, J. Lian, K. Svenson and L. L. Peters (2007). "Novel method for high-throughput phenotyping of sleep in mice." Physiol Genomics **28**(2): 232-238.

Paul, K. N., C. Dugovic, F. W. Turek and A. D. Laposky (2006). "Diurnal sex differences in the sleep-wake cycle of mice are dependent on gonadal function." Sleep **29**(9): 1211-1223.

Pavlova, M. and L. S. Sheikh (2011). "Sleep in women." Semin Neurol **31**(4): 397-403.

Philip, V. M., G. Sokoloff, C. L. Ackert-Bicknell, M. Striz, L. Branstetter, M. A. Beckmann, J. S. Spence, B. L. Jackson, L. D. Galloway, P. Barker, A. M. Wymore, P. R. Hunsicker, D. C. Durtschi, G. S. Shaw, S. Shinpock, K. F. Manly, D. R. Miller, K. D. Donohue, C. T. Culiat, G. A. Churchill, W. R. Lariviere, A. A. Palmer, B. F. O'Hara, B. H. Voy and E. J. Chesler (2011). "Genetic analysis in the Collaborative Cross breeding population." Genome Res **21**(8): 1223-1238.

Philip, V. M., G. Sokoloff, C. L. Ackert-Bicknell, M. Striz, L. Branstetter, M. A. Beckmann, J. S. Spence, B. L. Jackson, L. D. Galloway, P. Barker, A. M. Wymore, P. R. Hunsicker, D. C. Durtschi, G. S. Shaw, S. Shinpock, K. F. Manly, D. R. Miller, K. D. Donohue, C. T. Culiat, G. A. Churchill, W. R. Lariviere, A. A. Palmer, B. F. O'Hara, B. H. Voy and E. J. Chesler (2011). "Genetic analysis in the Collaborative Cross breeding population." Genome Research **21**(8): 1223-1238.

Piéron, H. (1913). Le problème physiologique du sommeil, Masson.

Polikar, R. (2006). "Ensemble based systems in decision making." IEEE Circuits and systems magazine **6**(3): 21-45.

Porkka-Heiskanen, T. and A. V. Kalinchuk (2011). "Adenosine, energy metabolism and sleep homeostasis." Sleep Med Rev **15**(2): 123-135.

Prince, T. M. and T. Abel (2013). "The impact of sleep loss on hippocampal function." Learn Mem **20**(10): 558-569.

Prober, D. A., S. Zimmerman, B. R. Myers, B. M. McDermott, Jr., S. H. Kim, S. Caron, J. Rihel, L. Solnica-Krezel, D. Julius, A. J. Hudspeth and A. F. Schier (2008). "Zebrafish TRPA1 channels are required for chemosensation but not for thermosensation or mechanosensory hair cell function." J Neurosci **28**(40): 10102-10110.

Provencio, I., H. M. Cooper and R. G. Foster (1998). "Retinal projections in mice with inherited retinal degeneration: implications for circadian photoentrainment." J Comp Neurol **395**(4): 417-439.

Punjabi, N. M., B. S. Caffo, J. L. Goodwin, D. J. Gottlieb, A. B. Newman, G. T. O'Connor, D. M. Rapoport, S. Redline, H. E. Resnick, J. A. Robbins, E. Shahar, M. L. Unruh and J. M. Samet (2009). "Sleep-disordered breathing and mortality: a prospective cohort study." PLoS Med **6**(8): e1000132.

Raizen, D. M., J. E. Zimmerman, M. H. Maycock, U. D. Ta, Y. J. You, M. V. Sundaram and A. I. Pack (2008). "Lethargus is a Caenorhabditis elegans sleep-like state." Nature **451**(7178): 569-572.

Ralph, M. R., R. G. Foster, F. C. Davis and M. Menaker (1990). "Transplanted suprachiasmatic nucleus determines circadian period." Science **247**(4945): 975-978.

Rao, M. N., T. Blackwell, S. Redline, M. L. Stefanick, S. Ancoli-Israel, K. L. Stone and G. Osteoporotic Fractures in Men Study (2009). "Association between sleep architecture and measures of body composition." Sleep **32**(4): 483-490.

Rechtschaffen, A. (1998). "Current perspectives on the function of sleep." Perspect Biol Med **41**(3): 359-390.

Reppert, S. M. and D. R. Weaver (2001). "Molecular analysis of mammalian circadian rhythms." Annu Rev Physiol **63**: 647-676.

Ringwald, M., V. Iyer, J. C. Mason, K. R. Stone, H. D. Tadepally, J. A. Kadin, C. J. Bult, J. T. Eppig, D. J. Oakley, S. Briois, E. Stupka, V. Maselli, D. Smedley, S. Liu, J. Hansen, R. Baldock, G. G. Hicks and W. C. Skarnes (2011). "The IKMC web portal: a central point of entry to data and resources from the International Knockout Mouse Consortium." Nucleic Acids Res **39**(Database issue): D849-855.

Rokach, L. and O. Maimon (2008). Data mining with decision trees : theory and applications. New Jersey ; London, World Scientific.

Rosen, C. J. (2008). "Bone remodeling, energy metabolism, and the molecular clock." Cell Metab **7**(1): 7-10.

Rosenwasser, A. M. (2010). "Circadian clock genes: non-circadian roles in sleep, addiction, and psychiatric disorders?" Neurosci Biobehav Rev **34**(8): 1249-1255.

Saaresranta, T., P. Polo-Kantola, E. Rauhala and O. Polo (2001). "Medroxyprogesterone in postmenopausal females with partial upper airway obstruction during sleep." Eur Respir J **18**(6): 989-995.

Saper, C. B., T. E. Scammell and J. Lu (2005). "Hypothalamic regulation of sleep and circadian rhythms." Nature **437**(7063): 1257-1263.

Sarti, C. D., A. Chiantera, A. Graziottin, F. Ognisanti, C. Sidoli, M. Mincigrucci, F. Parazzini and A. Gruppo di Studio Iper (2005). "Hormone therapy and sleep quality in women around menopause." Menopause **12**(5): 545-551.

Schofield, P. N., R. Hoehndorf and G. V. Gkoutos (2012). "Mouse genetic and phenotypic resources for human genetics." Hum Mutat **33**(5): 826-836.

Sethi, M., S. S. Joshi, M. Striz, N. Cole, J. Ryan, M. E. Lhamon, A. Agarwal, S. J. S. Rizzo, J. M. Denegre and R. E. Braun (2015). "Analysis of sleep traits in knockout mice from the large-scale KOMP2 population using a non-invasive, high-throughput piezoelectric system." BMC Bioinformatics **16**(Suppl 15): P15.

Shahar, E., C. W. Whitney, S. Redline, E. T. Lee, A. B. Newman, F. J. Nieto, G. T. O'Connor, L. L. Boland, J. E. Schwartz and J. M. Samet (2001). "Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the Sleep Heart Health Study." Am J Respir Crit Care Med **163**(1): 19-25.

Shaw, P. J. and P. Franken (2003). "Perchance to dream: solving the mystery of sleep through genetic analysis." J Neurobiol **54**(1): 179-202.

Sheldon, A. E. (1984). "Effect of estrogen dose in postmenopausal bone loss." N Engl J Med **311**(9): 604-605.

Shi, W., H. Wen, Y. Lu, Y. Shi, B. Lin and J. Qin (2010). "Droplet microfluidics for characterizing the neurotoxin-induced responses in individual Caenorhabditis elegans." Lab Chip **10**(21): 2855-2863.

Shi, X., D. V. Bosenko, N. S. Zinkevich, S. Foley, D. R. Hyde, E. V. Semina and T. S. Vihtelic (2005). "Zebrafish pitx3 is necessary for normal lens and retinal development." Mech Dev **122**(4): 513-527.

Signorell, A. (2015). "DescTools: Tools for descriptive statistics." R package version 0.99 **15**.

Simon, M. M., S. Greenaway, J. K. White, H. Fuchs, V. Gailus-Durner, S. Wells, T. Sorg, K. Wong, E. Bedu, E. J. Cartwright, R. Dacquin, S. Djebali, J. Estabel, J. Graw, N. J. Ingham, I.

J. Jackson, A. Lengeling, S. Mandillo, J. Marvel, H. Meziane, F. Preitner, O. Puk, M. Roux, D. J. Adams, S. Atkins, A. Ayadi, L. Becker, A. Blake, D. Brooker, H. Cater, M. F. Champy, R. Combe, P. Danecek, A. di Fenza, H. Gates, A. K. Gerdin, E. Golini, J. M. Hancock, W. Hans, S. M. Holter, T. Hough, P. Jurdic, T. M. Keane, H. Morgan, W. Muller, F. Neff, G. Nicholson, B. Pasche, L. A. Roberson, J. Rozman, M. Sanderson, L. Santos, M. Selloum, C. Shannon, A. Southwell, G. P. Tocchini-Valentini, V. E. Vancollie, H. Westerberg, W. Wurst, M. Zi, B. Yalcin, R. Ramirez-Solis, K. P. Steel, A. M. Mallon, M. H. de Angelis, Y. Herault and S. D. Brown (2013). "A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains." Genome Biol **14**(7): R82.

Simon, M. R., K. R. Holmes and A. M. Olsen (1985). "Bone mineral content of limb bones of male weanling rats subjected to 30 and 60 days of simulated increases in body weight." Acta Anat (Basel) **121**(1): 7-11.

Sittig, L. J., P. Carbonetto, K. A. Engel, K. S. Krauss, C. M. Barrios-Camacho and A. A. Palmer (2016). "Genetic Background Limits Generalizability of Genotype-Phenotype Relationships." Neuron **91**(6): 1253-1259.

Sivakumaran, S., F. Agakov, E. Theodoratou, J. G. Prendergast, L. Zgaga, T. Manolio, I. Rudan, P. McKeigue, J. F. Wilson and H. Campbell (2011). "Abundant pleiotropy in human complex diseases and traits." Am J Hum Genet **89**(5): 607-618.

Skarnes, W. C., B. Rosen, A. P. West, M. Koutsourakis, W. Bushell, V. Iyer, A. O. Mujica, M. Thomas, J. Harrow, T. Cox, D. Jackson, J. Severin, P. Biggs, J. Fu, M. Nefedov, P. J. de Jong, A. F. Stewart and A. Bradley (2011). "A conditional knockout resource for the genome-wide study of mouse gene function." Nature **474**(7351): 337-342.

Specker, B. L., T. Binkley, M. Vukovich and T. Beare (2007). "Volumetric bone mineral density and bone size in sleep-deprived individuals." Osteoporos Int **18**(1): 93-99.

Tabor, H. K., N. J. Risch and R. M. Myers (2002). "Candidate-gene approaches for studying complex genetic traits: practical considerations." Nat Rev Genet **3**(5): 391-397.

Tafti, M., B. Petit, D. Chollet, E. Neidhart, F. de Bilbao, J. Z. Kiss, P. A. Wood and P. Franken (2003). "Deficiency in short-chain fatty acid beta-oxidation affects theta oscillations during sleep." Nature Genetics **34**(3): 320-325.

Therneau, T. M., B. Atkinson and M. B. Ripley (2010). The rpart package.

Tobler, I. (2005). "Phylogeny of sleep regulation." Principles and practice of sleep medicine **4**: 77-90.

Tobler, I. I., P. Franken, L. Trachsel and A. A. Borbely (1992). "Models of sleep regulation in mammals." J Sleep Res **1**(2): 125-127.

Tononi, G. and C. Cirelli (2006). "Sleep function and synaptic homeostasis." Sleep Med Rev **10**(1): 49-62.

Tucker, M. A., Y. Hirota, E. J. Wamsley, H. Lau, A. Chaklader and W. Fishbein (2006). "A daytime nap containing solely non-REM sleep enhances declarative but not procedural memory." Neurobiol Learn Mem **86**(2): 241-247.

Van Buskirk, C. and P. W. Sternberg (2007). "Epidermal growth factor signaling induces behavioral quiescence in Caenorhabditis elegans." Nat Neurosci **10**(10): 1300-1307.

Varmuza, K. and P. Filzmoser (2016). Introduction to multivariate statistical analysis in chemometrics, CRC press.

Vassalli, A. and D. J. Dijk (2009). "Sleep function: current questions and new approaches." Eur J Neurosci **29**(9): 1830-1841.

Vitaterna, M. H., D. P. King, A. M. Chang, J. M. Kornhauser, P. L. Lowrey, J. D. McDonald, W. F. Dove, L. H. Pinto, F. W. Turek and J. S. Takahashi (1994). "Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior." Science **264**(5159): 719-725.

Warde-Farley, D., S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G. D. Bader and Q. Morris (2010). "The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function." Nucleic Acids Res **38**(Web Server issue): W214-220.

Weber, A. L., M. S. Cary, N. Connor and P. Keyes (1980). "Human non-24-hour sleep-wake cycles in an everyday environment." Sleep **2**(3): 347-354.

Wickham, H. (2016). ggplot2: elegant graphics for data analysis, Springer.

Woodward, S. and R. R. Freedman (1994). "The thermoregulatory effects of menopausal hot flashes on sleep." Sleep **17**(6): 497-501.

Xie, L., H. Kang, Q. Xu, M. J. Chen, Y. Liao, M. Thiyagarajan, J. O'Donnell, D. J. Christensen, C. Nicholson, J. J. Iliff, T. Takano, R. Deane and M. Nedergaard (2013). "Sleep drives metabolite clearance from the adult brain." Science **342**(6156): 373-377.

Yang, G., C. S. Lai, J. Cichon, L. Ma, W. Li and W. B. Gan (2014). "Sleep promotes branch-specific formation of dendritic spines after learning." Science **344**(6188): 1173-1178.

Yokogawa, T., W. Marin, J. Faraco, G. Pezeron, L. Appelbaum, J. Zhang, F. Rosa, P. Mourrain and E. Mignot (2007). "Characterization of sleep in zebrafish and insomnia in hypocretin receptor mutants." PLoS Biol **5**(10): e277.

Young, T., L. Finn, P. E. Peppard, M. Szklo-Coxe, D. Austin, F. J. Nieto, R. Stubbs and K. M. Hla (2008). "Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort." Sleep **31**(8): 1071-1078.

Young, T., P. E. Peppard and D. J. Gottlieb (2002). "Epidemiology of obstructive sleep apnea: a population health perspective." Am J Respir Crit Care Med **165**(9): 1217-1239.

Zhang, C., Y. Ma and SpringerLink (Online service) (2012). Ensemble machine learning methods and applications. New York, Springer.

Zhang, X., F. Bertaso, J. W. Yoo, K. Baumgartel, S. M. Clancy, V. Lee, C. Cienfuegos, C. Wilmot, J. Avis, T. Hunyh, C. Daguia, C. Schmedt, J. Noebels and T. Jegla (2010). "Deletion of the potassium channel Kv12.2 causes hippocampal hyperexcitability and epilepsy." Nat Neurosci **13**(9): 1056-1058.

Zhdanova, I. V. (2011). "Sleep and its regulation in zebrafish." Rev Neurosci **22**(1): 27-36.

VITA

Author's Name: Shreyas Joshi

Education:

2008 MSc Bioinformatics, Nottingham Trent University, Nottingham, UK
2006 BSc Microbiology, MVM College, Bhopal, India

Publications:

1. Joshi, S. and B. O'Hara (2014). "Identifying candidate genes for variation in sleep-related quantitative traits." BMC Bioinformatics 15(Suppl 10): P13-P13.

2. Joshi, S. S., T. J. Lesser, J. W. Olsen and B. F. O'Hara (2016). "The importance of temperature and thermoregulation for optimal human sleep." Energy and Buildings 131: 153-157.

3. Sawant, S., R. Gokulan, H. Dongre, M. Vaidya, D. Chaukar, K. Prabhash, A. Ingle, S. Joshi, P. Dange and S. Joshi (2016). "Prognostic role of Oct4, CD44 and c-Myc in radio–chemo-resistant oral cancer patients and their tumourigenic potential in immunodeficient mice." Clinical oral investigations 20(1): 43-56.

4. Sawant, S. S., D. A. Chaukar, S. S. Joshi, P. P. Dange, S. Kannan, S. Kane, A. K. D'Cruz and M. M. Vaidya (2011). "Prognostic value of tissue polypeptide antigen in oral squamous cell carcinoma." Oral oncology 47(2): 114-120.

5. Sethi, M., S. S. Joshi, M. Striz, N. Cole, J. Ryan, M. E. Lhamon, A. Agarwal, S. J. S. Rizzo, J. M. Denegre and R. E. Braun (2015). "Analysis of sleep traits in knockout mice from the large-scale KOMP2 population using a non-invasive, high-throughput piezoelectric system." BMC Bioinformatics 16(Suppl 15): P15.

6. Sethi, M., S. S. Joshi, R. L. Webb, T. L. Beckett, K. D. Donohue, M. P. Murphy, B. F. O'Hara and M. J. Duncan (2015). "Increased fragmentation of sleep–wake cycles in the 5XFAD mouse model of Alzheimer's disease." Neuroscience 290: 80-89.