



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Gerstner, JR;Koberstein, JN;Watson, AJ;Zapero, N;Risso, D;Speed, TP;Frank, MG;Peixoto, L

Title:

Removal of unwanted variation reveals novel patterns of gene expression linked to sleep homeostasis in murine cortex

Date:

2016-10-25

Citation:

Gerstner, J. R., Koberstein, J. N., Watson, A. J., Zapero, N., Risso, D., Speed, T. P., Frank, M. G. & Peixoto, L. (2016). Removal of unwanted variation reveals novel patterns of gene expression linked to sleep homeostasis in murine cortex. *BMC Genomics*, 17 (Suppl 8), <https://doi.org/10.1186/s12864-016-3065-8>.

Persistent Link:

<https://hdl.handle.net/11343/259469>

License:

[CC BY](#)

RESEARCH

Open Access



# Removal of unwanted variation reveals novel patterns of gene expression linked to sleep homeostasis in murine cortex

Jason R. Gerstner<sup>1</sup>, John N. Koberstein<sup>1</sup>, Adam J. Watson<sup>2</sup>, Nikolai Zaperov<sup>1</sup>, Davide Risso<sup>3</sup>, Terence P. Speed<sup>4,5,6</sup>, Marcos G. Frank<sup>1\*</sup> and Lucia Peixoto<sup>1\*</sup>

From 6th SolBio International Conference 2016 (SolBio-IC&W-2016)  
Riviera Maya, Mexico. 22-26 April 2016

## Abstract

**Background:** Why we sleep is still one of the most perplexing mysteries in biology. Strong evidence indicates that sleep is necessary for normal brain function and that sleep need is a tightly regulated process. Surprisingly, molecular mechanisms that determine sleep need are incompletely described. Moreover, very little is known about transcriptional changes that specifically accompany the accumulation and discharge of sleep need. Several studies have characterized differential gene expression changes following sleep deprivation. Much less is known, however, about changes in gene expression during the compensatory response to sleep deprivation (i.e. recovery sleep).

**Results:** In this study we present a comprehensive analysis of the effects of sleep deprivation and subsequent recovery sleep on gene expression in the mouse cortex. We used a non-traditional analytical method for normalization of genome-wide gene expression data, Removal of Unwanted Variation (RUV). RUV improves detection of differential gene expression following sleep deprivation. We also show that RUV normalization is crucial to the discovery of differentially expressed genes associated with recovery sleep. Our analysis indicates that the majority of transcripts upregulated by sleep deprivation require 6 h of recovery sleep to return to baseline levels, while the majority of downregulated transcripts return to baseline levels within 1–3 h. We also find that transcripts that change rapidly during recovery (i.e. within 3 h) do so on average with a time constant that is similar to the time constant for the discharge of sleep need.

**Conclusions:** We demonstrate that proper data normalization is essential to identify changes in gene expression that are specifically linked to sleep deprivation and recovery sleep. Our results provide the first evidence that recovery sleep is comprised of two waves of transcriptional regulation that occur at different times and affect functionally distinct classes of genes.

**Keywords:** Sleep, Sleep deprivation, Circadian, Microarray, Gene, mRNA, Transcriptomics

\* Correspondence: marcos.frank@wsu.edu; lucia.peixoto@wsu.edu

<sup>1</sup>Washington State University, Elson S. Floyd College of Medicine, Spokane, WA 99202, USA

Full list of author information is available at the end of the article



## Background

Sleep is thought to be controlled by two processes, 1) a homeostatic process that determines sleep need (or pressure), and 2) a circadian process that determines the timing of sleep and wakefulness. A robust index for sleep need is known as delta power, which refers to “delta” (1–4 Hz) oscillations in the electroencephalogram (EEG) of non-rapid eye movement (NREM) sleep. Delta power increases with increased sleep pressure, and declines following sleep. Therefore, sleep deprivation increases delta power, which then naturally decreases during recovery sleep. Previous studies have shown that EEG delta power is under genetic control [1], suggesting that specific genes contribute to sleep homeostasis. Nevertheless, the molecular mechanisms that regulate sleep need remain incompletely described.

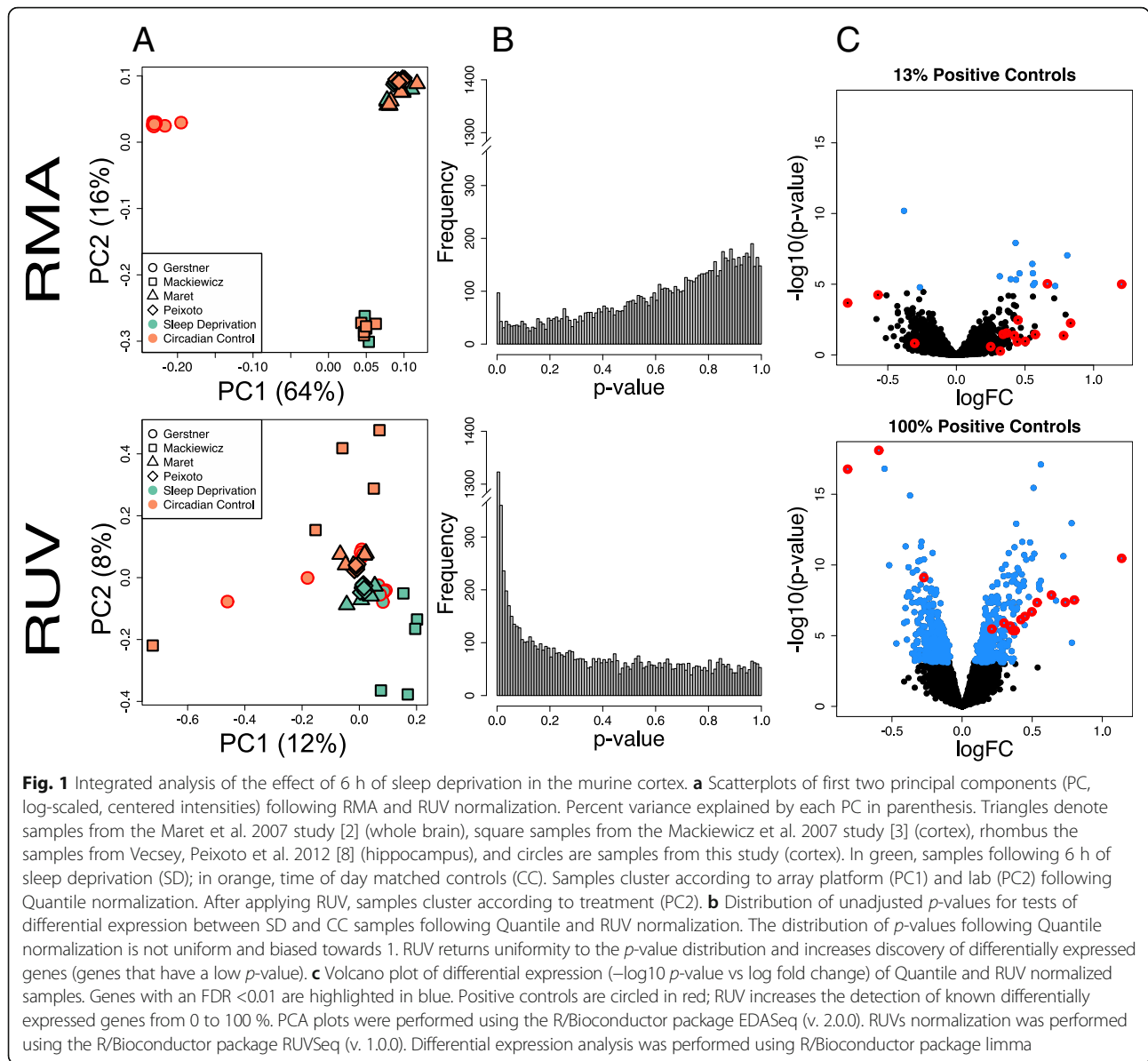
Genome-wide technologies have been used to interrogate gene expression changes that follow sleep deprivation in the mouse brain [2, 3], but there is little agreement between studies. There are also no genome-wide studies that characterize transcriptional changes that occur during subsequent recovery sleep. A major challenge when analyzing genome-wide data in the brain in response to behavior is isolating the signal of interest from other factors (batch effects) that simultaneously influence gene expression. We have previously shown that batch effects are widespread in genome-wide studies of gene expression in experimental neuroscience [4]. Several data normalization strategies are available to correct these batch effects, including Removal of Unwanted Variation (RUV). RUV adjusts for batch effects by performing factor analysis on control genes or replicate samples [5–7]. We have also shown that RUV normalization leads to increased power and reproducibility of results [4]. We now employ RUV to generate an integrated cross-laboratory analysis of differential gene expression following sleep deprivation in the mouse brain. We also provide the first comprehensive genome-wide assessment of transcripts from mouse cortex during recovery sleep. Our analysis improves detection of differentially expressed genes following sleep deprivation and shows that recovery sleep reverses the transcriptional changes it causes. This latter process occurs in waves that happen at different times during recovery sleep and affect functionally distinct classes of genes.

## Results

We first wanted to obtain a reliable estimate of the differential gene expression changes caused by sleep deprivation. To do this, we subjected mice to 5–6 h of sleep deprivation and different amounts of recovery sleep. Control mice (left undisturbed in home cages) were sacrificed at matching circadian times and cortical mRNA was isolated from experimental and control groups for microarray analysis (see methods for more

details). We then integrated our data with that from three other studies performed in mouse cortex, whole brain and hippocampus [2, 3, 8]. We began with a traditional method of normalization, Robust multi-array average (RMA) plus quantile normalization [9], to integrate the data. The resulting principal component (PC) analysis showed that 64 % of the variability was due to platform (PC 1) while 16 % was due to different laboratories (PC 2; Fig. 1a). These results show that traditional normalization methods are not well suited to integrate data across laboratories or array platforms. Following RUV normalization, however, samples group according to treatment (sleep deprivation versus controls) and not laboratories, platform (Fig. 1a) or tissue type. We then evaluated the impact of normalization method in differential expression analysis. RUV restored the *p*-value histogram of differential expression to its expected distribution (Fig. 1b), increased detection of differentially expressed genes, and recovered 100 % of positive control genes known to respond to sleep deprivation (Fig. 1c). Positive control genes used in our study (see Methods) are listed in Additional file 1. The list of differentially expressed genes in response to sleep deprivation regardless of tissue for the integrated study is available in Additional file 2. Our results show that a method as effective as RUV normalization is required for the proper characterization of differentially expressed genes across labs and platforms in meta-analysis studies, as we have shown for RNA-seq data [4].

We were also interested in determining the effects of recovery sleep on sleep deprivation-induced gene expression, as this has been minimally explored. Separate groups of mice were sleep-deprived and allowed to recover for 1, 2, 3, and 6 h (RS1, RS2, RS3, and RS6, respectively) to identify changes in genome-wide expression during recovery sleep. We again compared the effect of using traditional normalization (RMA) versus RUV on the characterization of differential gene expression after sleep deprivation or varying lengths of recovery sleep. The PC scatterplots revealed a lack of grouping of biological replicates with RMA, indicative of the presence of confounding factors or “batch effects” in the data (Fig. 2a). Unlike our previous analysis (Fig. 1), no obvious grouping attributable to technical factors was found. Following RUV normalization, the PCA plots showed the expected segregation of groups according to treatment (Fig. 2a). The effect of data normalization is also evident in the analysis of differential expression. RUV analysis improved the detection of differentially expressed genes following one hour of recovery sleep (RS1, Fig. 2b), and recovered 86 % of the positive controls (Fig. 2). Improvement was also seen at all other time-points (RS2, RS3 and RS6, Additional file 3). These results show that RUV normalization, or similar methodology geared to address batch effects, is necessary for

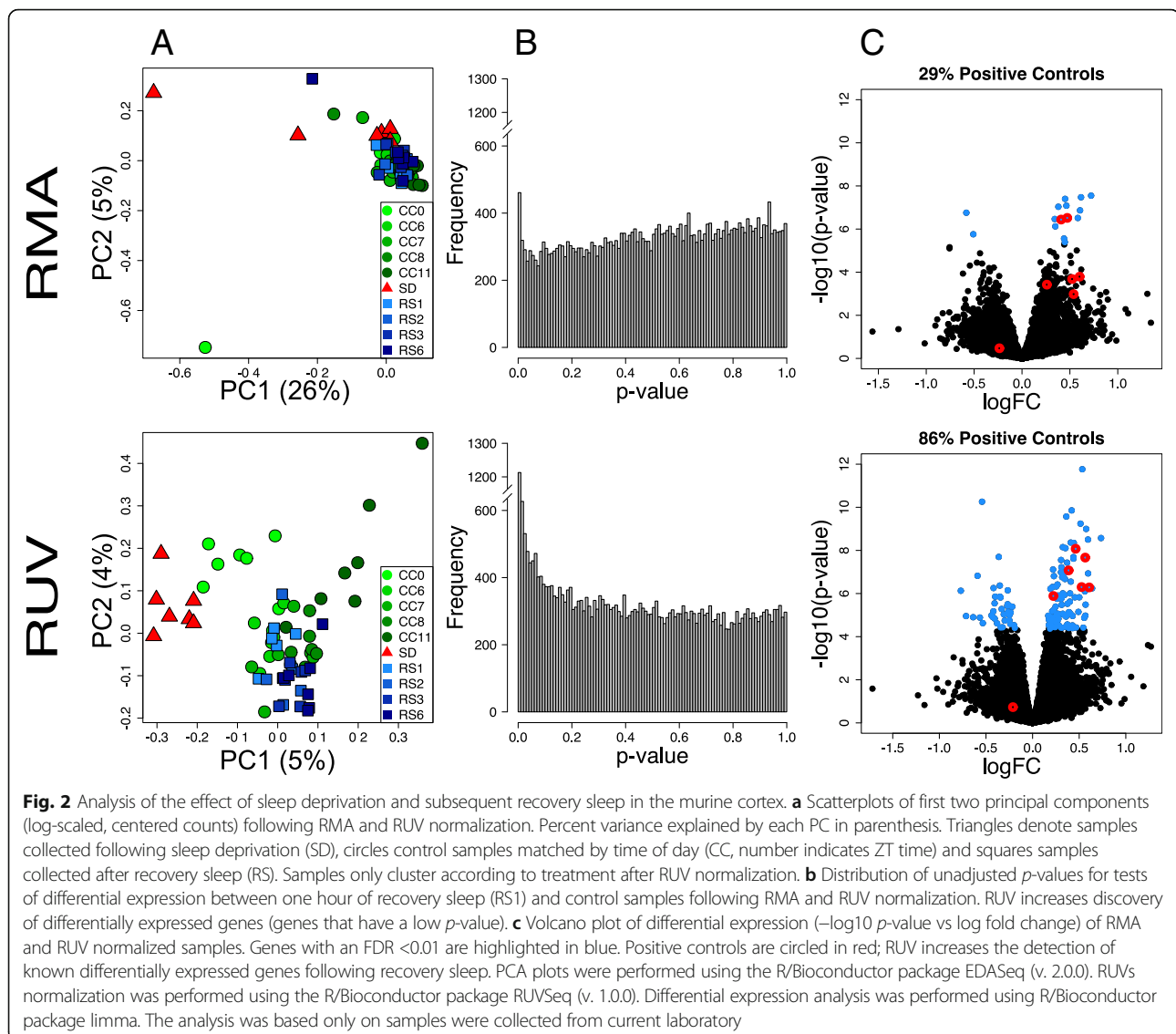


accurate detection of differential expression not only across but also within studies. Genes differentially expressed at all time points in this study relative to controls can be found in Additional file 4. It is also known that time of day by itself can profoundly impact brain gene expression in the brain. To address the possible confound of circadian time in the fold changes observed following RS, we compared three controls samples to each other, CC0, CC6 and CC11, which correspond to circadian time ZT = 0, ZT = 6 and ZT = 11 respectively. The results are available in Additional file 5 and were taken into account on the interpretation of gene expression changes brought on by RS.

The discharge of sleep need in mammals is a gradual process, including events that normalize within the first 3 h (e.g. NREM EEG delta power) and others that may

take 6–12 h to appear (e.g. changes in sleep time) [1, 10]. We therefore plotted the number of significant probesets altered by sleep deprivation and recovery sleep relative to their respective circadian time-point controls following both RMA and RUV normalization (Fig. 3a). Compared to RMA, RUV analysis greatly improved the detection of significant genes for all groups examined. Its impact is more pronounced on the detection of differential expression during recovery sleep. Overall, the number of genes up-regulated outnumbered those that are downregulated at all time-points.

We then generated a heatmap of differentially expressed genes detected using RUV normalization relative to circadian time-matched controls (Fig. 3b). The heatmap and subsequent clustering of gene expression across time-

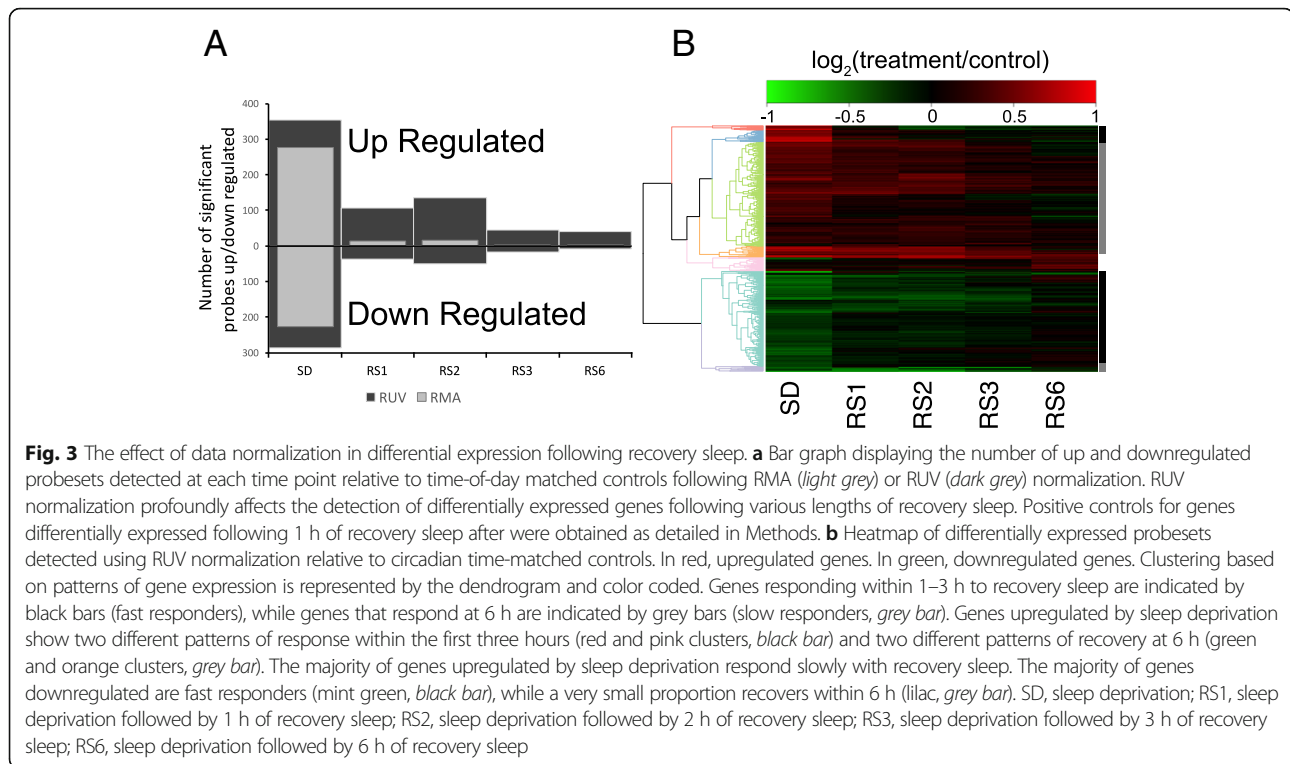


points revealed different patterns of gene expression relative to the amount of recovery sleep. We identified seven clusters of genes based on their expression pattern across recovery sleep (Additional file 6). These seven clusters can be grouped into three categories: genes that recover within 1–3 h (“fast responders”), genes that recover by 6 h of recovery sleep (“slow responders”), and genes unaffected by sleep deprivation, but upregulated by the 6<sup>th</sup> hour of recovery. Examples of genes belonging to each category can be found in Fig. 4 and Additional file 6.

The majority of genes upregulated by sleep deprivation were slow responders, while the majority of genes downregulated by sleep deprivation were fast responders (the exception is a small cluster of 16 genes that includes *Cirbp* and *Dbp*), supporting the pattern observed in Fig. 3a. Interestingly, a more fine-grained analysis of these results showed that transcripts known to be upregulated by sleep

deprivation (such as *Arc* and *Homer1*) [2, 3] include both ‘slow’ and ‘fast’ responders. Transcripts that recover by 1–3 h included *Arc*, *Per1*, *Per2*, *Egr1*, and *Egr2*, while transcripts that recovered by 6 h included *Homer1*, *BDNF*, *Fosb*, *Hspa5*, and *Npas2* (Fig. 4). Many of these transcripts also followed a normal circadian expression pattern, which are shown independently (Additional file 5). The smaller set of 46 transcripts upregulated by the 6<sup>th</sup> hour of recovery sleep (pink cluster, Fig. 3b, Additional file 6) included transcripts with less clearly defined functions, including micro RNA and non-coding RNA and transcripts involved in RNA-binding protein sequestration (*Neat1*) [11].

The identification of functional properties of differentially expressed genes associated with changes in sleep need is crucial for understanding mechanisms that regulate sleep and sleep function. In order to identify functional classes and pathways that respond to different



amounts of recovery sleep, we performed functional annotation analysis of genes upregulated or downregulated by sleep deprivation within “fast” and “slow” responder groups. Figure 5 represents clusters of enriched functional terms among the different classes of genes (see Additional file 7 for details on clustered terms). Genes downregulated by sleep deprivation were mostly “fast” responders ( $N = 310$ ) and were enriched in the following functional clusters: cell adhesion, protein catabolic process, Ras GTPase binding, GTP signaling and cell cycle. In addition the following terms were enriched, although they did not cluster with other similar terms: transcriptional corepression, alternative splicing, and neurogenesis (Additional file 7). For genes upregulated by sleep deprivation, “fast” responders ( $N = 52$ ) were enriched in the following functional classes: MAPK signaling and regulation, GTPase signaling, and transcriptional regulation. The “slow” responders ( $N = 375$ ) represented nine different enriched clusters, generally with lower enrichment scores than the upregulated fast responders despite the fact that there were substantially more genes in this group.

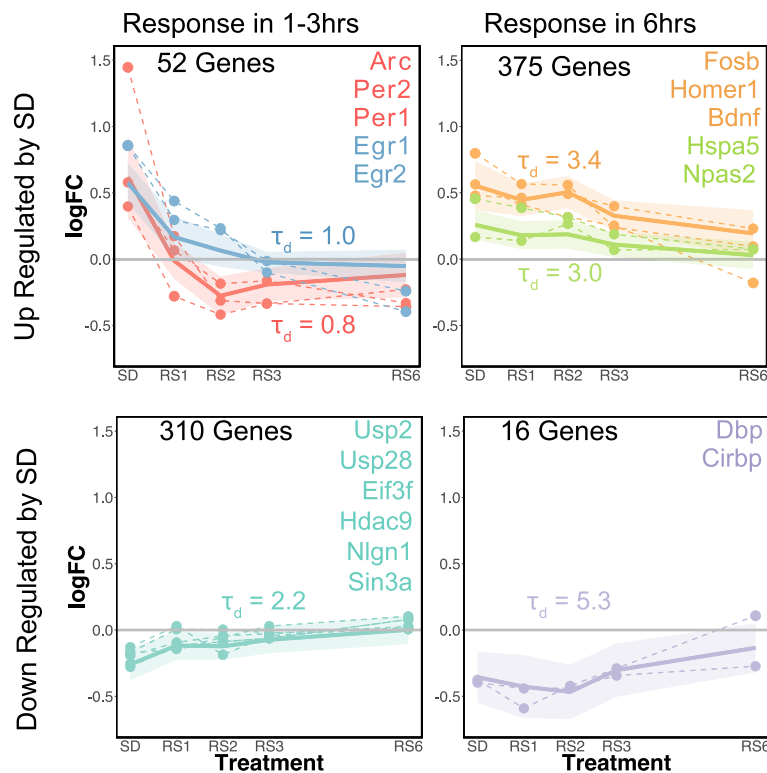
We also computed the decay time constant ( $\tau_d$ ) for the changes in slow and fast responding transcripts to see if they approximated the  $\tau_d$  for the discharge of sleep need [10]. As shown in Fig. 4, these varied across functional classes. Interestingly, when we averaged  $\tau_d$  across the different clusters of fast-responder genes the value

was identical to that reported for the discharge of sleep need in this mouse strain (average  $\tau_d$  fast responder genes: 1.3;  $\tau_d$  for discharge of sleep need in c57/bl6: 1.3 [10]).

## Discussion

The molecular determinants of sleep homeostasis are not well known. Furthermore, transcriptional changes that track both the accumulation and discharge of sleep need have not been well characterized. Here, we present a fully integrated meta-analysis of the effects of sleep deprivation on mouse brain gene expression, by combining our data with three other previous studies [2, 3, 8]. We also provide additional evidence for genome-wide changes in cortical gene expression following various lengths of subsequent recovery sleep. We report that discovery of gene expression changes linked to either sleep deprivation or recovery sleep, and not batch effects, requires a non-traditional method of data normalization (RUV). Therefore, our study more accurately reflects true biological variance due to changes in sleep need, and vastly improves both single laboratory and meta-analysis studies of gene expression previously conducted in the absence of RUV [12].

Our results represent the first genome-wide examination of differentially expressed cortical genes that includes the response to sleep deprivation and subsequent changes across the recovery sleep period. The majority

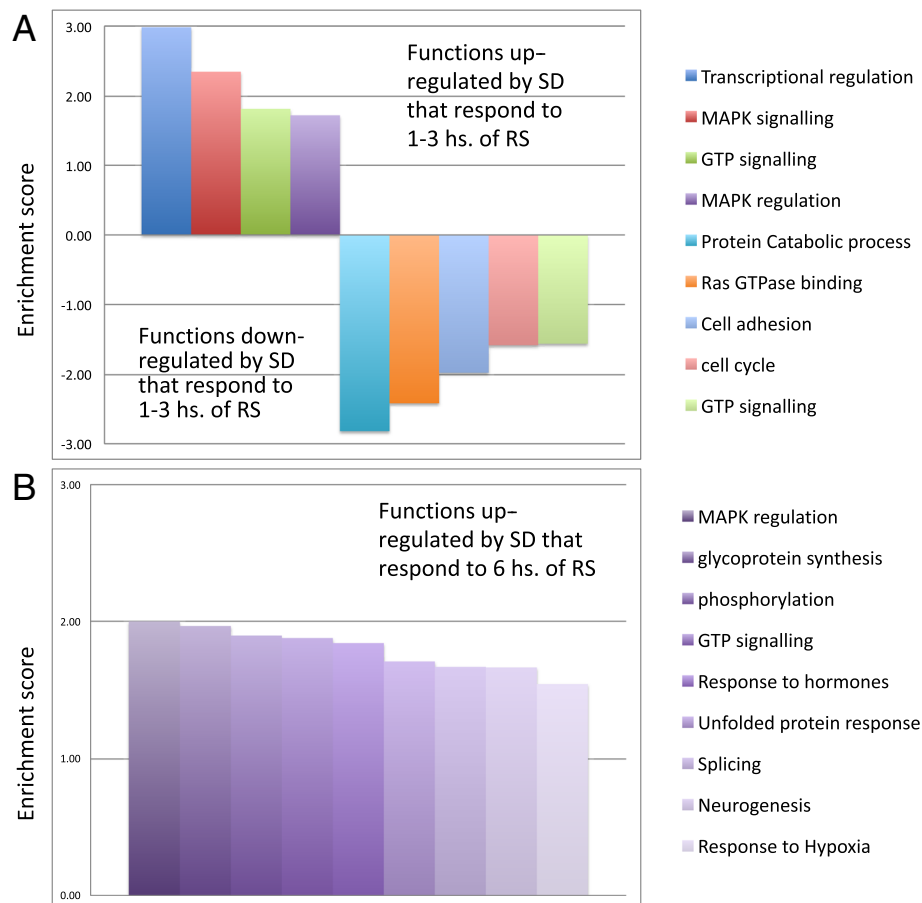


**Fig. 4** Patterns of gene expression regulation during recovery sleep. Plots of log-fold change (logFC) of differential expression relative to controls versus time since sleep deprivation. Color coding corresponds to clusters on Fig. 3b. Plots for genes representative of different expression patterns are shown as dashed lines. The ordering of the gene names within genes of the same cluster (*same color*) reflects the ordering of the plots. The average and standard deviation for each cluster is shown as solid lines and shaded area. Patterns of expression are divided in two classes: ‘fast’ responders (genes that reach basal values within 1–3 h of RS) and ‘slow’ responders (genes that reach basal values >3 h). Time constants for the change during recovery sleep for average of each cluster are shown

of genes fall into two general classes of transcriptional changes: transcripts that return to baseline values quickly (i.e., within the first 3 h: ‘fast responders’) and transcripts that return to baseline values slowly (by 6 h of recovery: ‘slow responders’). There was also a small subset of transcripts not affected by sleep deprivation, but upregulated in the 6<sup>th</sup> hour of the recovery period. Genes that respond “fast” to recovery sleep may constitute the molecular signature underlying the discharge of sleep need based on electrophysiological studies. This is because these molecular changes parallel the discharge of sleep need as measured by changes in NREM EEG delta power. In mice, 6 h of sleep deprivation produces a large increase in delta power which then rapidly declines over the next 3 h of recovery sleep [13, 10]. This pattern is strikingly similar to the time course we find for the ‘fast’ responder genes.

The majority of fast-responders are transcripts initially *downregulated* by sleep deprivation and then upregulated with recovery sleep. Interestingly we find that these specific transcripts do not show time-of-day differences (Additional file 5) suggesting that the biological functions

they serve are more closely tied to sleep homeostasis. These include genes involved in synthesis and degradation of proteins. Examples include ubiquitin-specific-proteases (Usp) and elongation initiation factors (Eifs) (Fig. 4). These results are consistent with previous studies [3, 14–16], supporting the idea that one of the immediate effects of recovery sleep is to influence protein synthesis or turnover. A second class of ‘fast’ responder genes downregulated by sleep deprivation is involved in transcriptional repression linked to histone modification. Histone acetylation and deacetylation modify the structure of chromosomes and influence access of transcription factors to DNA. We find that histone deacetylase 9 (Hdac9) and associated co-repressor Sin3A, together with the GO term “transcriptional co-repression” are downregulated by sleep deprivation (Fig. 4, Additional file 7), as previously shown in the hippocampus [8]. This suggests that part of the compensatory response is a reactivation of transcriptional repression. This may be part of the restorative function of sleep; that it re-establishes a basal level of transcription required for normal neural function. While speculative, it is possible that the cognitive deficits



**Fig. 5** Enriched functional clusters regulated by recovery sleep. **a** Functional clusters regulated by 1–3 h of recovery sleep. **b** Functional clusters regulated by 6 h of recovery sleep. Functional annotation terms from the following databases: Gene Ontology (GO) biological process and molecular function, KEGG pathways and protein information resource keywords, were clustered based on similarity using the Database for Annotation Visualization and Integrated Discovery (see Methods). Clusters of functional terms enriched in down- or upregulated gene lists following SD as compared with the genome as a whole ( $P$  value  $<0.05$ ) are represented as bars. Height of bars represents the enrichment score of each cluster, with the scores of downregulated clusters shown as negative numbers for visualization purposes. Enrichment score was calculated as  $-\log(10)$  of the geometric mean  $p$ -value among all clustered terms. Only clusters with enrichment score  $>1.5$  (average  $p$ -value of functional terms within the cluster  $<0.05$ ) were considered. For details of the functional terms included in these clusters, as well as enriched functional terms that did not cluster with other terms, see Additional file 7

associated with sleep deprivation result from an unchecked expression of certain transcripts; a situation reversed during recovery sleep.

Fast responder genes *upregulated* by SD include immediate early genes previously identified as ‘induced by sleep deprivation’ (e.g. Arc, Egr1, Egr2 and Nr4a1 [2, 3, 14, 15, 17] (Fig. 4 and Additional file 5). The function of immediate early genes is primarily the regulation of transcription, a functional category that is also enriched in this group. We also find that several of the fast-responder genes are traditional ‘circadian’ genes (e.g. Arntl, Dbp, Per1, and Per2 [2]) and show time of day differences in their expression (Additional file 5). The precise role of these immediate early and clock genes in sleep homeostasis is unclear. Deletion of Arntl (BMAL1) in mice alters baseline sleep architecture, increases NREM EEG delta-power

baseline conditions, and attenuates the homeostatic response to sleep deprivation [18]. In addition, Per1 and Per2 brain expression in various inbred mouse strains correlates with changes in NREM EEG delta-power [19–21], suggesting these genes are tied to the sleep homeostat. However, Per mutants have a normal response to sleep deprivation as measured by NREM delta power [22, 23], indicating they play no central role in sleep homeostasis. Similarly, the expression of the immediate early gene Homer1A also tracks sleep need [2], but Homer1A null mice have normal sleep homeostasis [24]. Therefore it is possible that the regulation of these particular clock and immediate early genes may not be as closely linked to sleep homeostasis, as appears to be the case for other clock genes [18, 20]. Instead, circadian rhythms or neural activity may play more influential roles in their expression. This interpretation is consistent with

earlier studies. It has been shown that several immediate early genes, such as *Per1* and *Arc*, are also upregulated following contextual fear conditioning or object location memory for example [15, 25, 26].

The slow responding transcripts represented the majority of all transcripts upregulated by sleep deprivation (Fig. 4). These included genes previously linked to sleep homeostasis including *Homer1*, *Bdnf* and *Npas2* [27, 28, 19]. There appears to be functional overlap between these slow responding transcripts and fast responding transcripts downregulated by sleep. GO terms or functional clusters that overlap include: cell adhesion, neurogenesis, GTP signalling and splicing (Fig. 5, Additional file 7). Adhesion molecules (such as *Neurologin 1*; Fig. 4, Additional file 3) are particularly interesting because they may link early responses to SD (e.g. clock gene expression) with slower responses. This is because sleep deprivation induced changes in *Neurologin1* are dependent on clock transcription factors [29]. A much smaller subset (16 genes) was downregulated by sleep deprivation. The functions of these genes are not well understood.

Lastly, we identified a small number of genes that were unaffected by sleep deprivation but were upregulated in the 6th hour of recovery sleep (pink cluster, Fig. 3b). The function of many of these genes is also obscure. One example is the long non-coding RNA *Neat1*. *Neat1* is retained in the nucleus where it forms the core structural component of the paraspeckle sub-organelle located within the eukaryotic nucleus [11]. *Neat1* has been shown to regulate transcription via protein sequestration within paraspeckles [30]. Paraspeckles are believed to function as a reservoir of mRNA that are released into the cytoplasm under certain conditions (e.g. cellular stress) and/or provide a means of RNA sequestration. The reason for this delayed expression of transcripts is unclear. In mice, 6 h SD not only increases NREM delta power in the first three hours of recovery sleep, but also leads to a delayed 'rebound' in REM sleep time that can occur at or after 6 h [1]. Therefore, it is possible that the expression of this small subset of transcripts is driven by REM sleep [1, 30, 31].

While our study represents a significant improvement over previous studies that set out to identify transcripts associated with sleep homeostasis, there are some limitations. For example, our current methods do not allow for the identification of spatial resolution of specific cortical layers, or regional differences between frontal and parietal cortices. Further, the cell-type specificity for the changes in expression identified in our study is not characterized. In addition, it is difficult to differentiate which gene expression changes are responding to sleep pressure from those that are responding to stress hormones [32]. Future studies using improved techniques, such as the use of TRAP technology [33, 34] will be necessary to

identify cell-type specific changes in transcripts associated with sleep homeostasis.

## Conclusions

This is the first study where RUV normalization has been used to compare multiple genome-wide data sets following sleep deprivation. We also used this approach to examine transcriptional changes during the recovery sleep period. We show that RUV vastly improves the meta-analysis of data generated in different laboratories and reveal novel changes in transcription during recovery sleep. We find that sleep produces two waves of transcription during recovery sleep. Some changes occur rapidly, others more slowly across six hours of sleep. The fast responding transcripts may represent the molecular components of sleep homeostasis as they change with a time constant that is remarkably similar to the time constant for the discharge of sleep need. Further characterization of these genes may reveal sleep function and the biological basis for sleep need.

## Methods

### Subjects

C57BL/6 J adult male mice (2 months of age) were obtained from Jackson laboratories and housed individually for a week in an experimental room on a 12 h./12 h. light/dark schedule with lights on at 7:30 am (Zeitgeber time (ZT) 0). Food and water were available ad libitum throughout the experiment. All experiments were approved by the Institution of Animal Care and Use Committee of the University of Pennsylvania and were carried out in accordance with all National Institutes of Health guidelines.

To examine gene expression in the mouse brain cortex after 5–6 h of sleep deprivation and subsequent recovery sleep we generated a dataset of 96 microarrays (see below). Groups of mice were either sleep deprived or sleep deprived and allowed recovery sleep for 1, 2, 3, or 6 h. Matching control (CC) animals were left undisturbed but sacrificed at the same time of day. Sleep deprivation was achieved by brushing the mice with a soft brush to keep them active. We did not surgically implant EEG electrodes to quantitatively measure sleep. This was done, as true for earlier studies [2, 3], to prevent changes in gene expression that might occur as a result of the surgery. However, analyses of animals implanted with EEG electrodes undergoing the same procedures shows that sleep deprivation procedures similar to ours is effective at maintaining wakefulness in a variety of mouse strains (including c57/bl6) [10, 35]. In addition, when given an opportunity to sleep, mice spend most of the subsequent 6 h period in sleep [10]. The experimental protocol was repeated daily, to obtain one animal per time point per experimental day, to

gather 6–7 mice per experimental group. We also performed a meta-analysis using data from three previously published studies that tested the effect of 6 h of sleep deprivation in the mouse cortex, hippocampus and whole brain [2, 3, 8].

### Microarrays

Cortical dissections were performed by a single experimenter, and tissue was rapidly dissected and immersed in chilled RNAlater (Qiagen), kept overnight at 4 °C, then frozen at –80 °C. RNA extraction was performed using the miRNeasy kit (Qiagen). Biotinylated sense-strand cDNA were prepared from 300 ng total RNA at the UPENN molecular profiling core using the Affymetrix WT Plus Kit. Single stranded cDNA was hybridized to a Mouse Gene 2.1 ST 96-Array Plate using GeneTitan Hybridization, Wash and Stain Kit for WT Array Plates. The array plate was washed and stained in the GeneTitan multi-channel instrument. Gene 1.1 ST Array Plates were scanned using the GeneTitan® Multi-channel Instrument.

### Cross-study data integration

Data from GSE9444 [2], GSE6514 [3] and GSE50423 [8] were obtained using the R/Bioconductor package GEOquery (v. 2.36.0). Data generated in this study are publicly available through GEO (GSE78215). To allow for cross-platform comparison, Affymetrix probeset IDs were mapped to ENSEMBL gene Ids using the R/Bioconductor package biomaRt (v. 2.26.1). Probesets or ENSEMBL genes that showed multiple mappings were excluded from the cross-platform analysis only. Probesets that showed a log expression value > 4 in >50 % of the samples were included.

### Normalization and statistical analysis

We have previously shown that Removal of Unwanted Variation (RUV) [5] is a normalization method that is able to properly correct for batch effects in experimental neuroscience data obtained through RNA sequencing [4]. Any difference between biological replicates can be attributed to unwanted effects. The RUV method exploits the fact that genes that should not be changing in a biological system (negative controls) or differences between replicates, carry in their observed levels patterns of unwanted variation that can be used to adjust for unwanted effects. In this study, we used log-transformed RMA normalized data as input to RUVSeq, a Bioconductor package originally designed to perform RUV analysis on RNA sequencing data [5]. The advantage of doing so is the ease of integration with future or currently available RNA-seq studies, since RNA-seq is now the standard for the quantification of genome-wide gene expression. We compared the results of differential

expression analysis of data normalized using RUV based on replicate samples with data normalized using RMA with Quantile normalization [9], the most commonly used method for microarray data. RUV using negative control samples was used to remove  $k$  factors of unwanted variation before statistical testing was performed. The choice of a proper  $k$  was determined as previously described [4], resulting in  $k = 5$  for the meta-analysis of sleep deprivation (Fig. 1) and  $k = 6$  for the recovery sleep analysis (Fig. 2). Differential expression analysis was performed using R/Bioconductor package limma (v. 3.26.8) comparing the sleep deprivation and recovery sleep samples to time matched circadian controls (SD vs. CC6, RS1 vs. CC7, RS2 vs. CC8, RS3 vs. CC8, RS6 vs. CC11). Comparisons between controls CC0, CC6 and CC1 were used to determine circadian changes in expression. Multiple testing corrections were performed using the method of Benjamini and Hochberg [36]. A cutoff of false discovery rate (FDR) < 0.01 was used to assess significance. To evaluate performance, we assembled sets of independently validated positive control genes that are known to respond to 6 h of sleep deprivation or 2 h of recovery sleep (see Additional file 1 for details). RUV normalization was performed using the R/Bioconductor package RUVSeq (v. 1.0.0). Differential expression analysis was performed using R/Bioconductor package limma (v. 3.26.8).

Estimation of time constants of gene expression changes in response to recovery sleep. Non-linear regression analysis was used to quantify relationships between duration of recovery sleep and gene expression for each cluster. Non-linear least squares implemented in the R package stats (v. 3.3.0) was utilized to estimate parameters by assuming that gene expression varies according to an exponential function following sleep deprivation [10]:

$$\log FC_t = LA + (\log FC_0 - LA) \cdot e^{-(t/\tau d)}$$

$\log FC_0$  and  $\log FC_t$  are the mean logFC at the end of the sleep deprivation period and at time  $t$  respectively.  $\tau d$  is the time constant of the decreasing exponential function that approaches a lower asymptote, LA, approximated as the minimum mean logFC for each cluster. Starting  $\tau d$  values were estimated by visual examination of plots and optimized over subsequent iterations.

### Graphical displays

PCA plots were performed using the R/Bioconductor package EDASeq (v. 2.0.0). The heatmap was prepared using the R package gplots (v. 2.17.0) with modifications to the row dendrogram using the R package dendextend (v. 1.1.8). All other figures were generated using R base graphics and Microsoft Excel.

### Functional enrichment analysis

Affymetrix probeset ID's were mapped to MGI symbol and ensemble gene ID's for downstream analysis using the R/Bioconductor package *mogene21sttranscriptcluster.db* (v. 8.4.0). Functional annotation was based on ENSEMBL Gene IDs and performed using the database for annotation, visualization and integrated discovery v 6.7 (DAVID, <https://david.ncifcrf.gov>). The following functional categories were used: GO Biological Process and Molecular Function, KEGG pathways and Protein Information Resource keywords. Enrichment cutoff relative to background = EASE score <0.05. All genes present in the array were used as background for enrichment. Clustering was used to reduce complexity. Clustering parameters: similarity threshold 0.2, group membership 2.

### Additional files

**Additional file 1:** Positive control genes. (XLSX 44 kb)

**Additional file 2:** The impact of data integration on the detection of differential expression following sleep deprivation (SD). Differentially expressed genes detected, integrating all studies (Fig. 1). (XLSX 98 kb)

**Additional file 3:** Impact of RUV on differential expression. A) Top left. Distribution of unadjusted limma  $p$ -values for tests of differential expression between SD and CC samples following RMA and RUV normalization. RUV increases discovery of differentially expressed genes (genes that have a low  $p$ -value). Bottom. Volcano plot of differential expression ( $-\log_{10} p$ -value vs log fold change) of Quantile and RUV normalized samples. Genes with and FDR <0.01 are highlighted in blue. Positive controls are circled in red. B) Top right. Distribution of unadjusted limma  $p$ -values for tests of differential expression between RS2 and CC samples following RMA and RUV normalization. RUV increases discovery of differentially expressed genes (genes that have a low  $p$ -value). Bottom. Volcano plot of differential expression ( $-\log_{10} p$ -value vs log fold change) of Quantile and RUV normalized samples. Genes with and FDR <0.01 are highlighted in blue. Positive controls are circled in red. C) Bottom left. Distribution of unadjusted limma  $p$ -values for tests of differential expression between RS3 and CC samples following RMA and RUV normalization. RUV increases discovery of differentially expressed genes (genes that have a low  $p$ -value). Bottom. Volcano plot of differential expression ( $-\log_{10} p$ -value vs log fold change) of Quantile and RUV normalized samples. Genes with and FDR <0.01 are highlighted in blue. Positive controls are circled in red. D) Bottom right. Distribution of unadjusted limma  $p$ -values for tests of differential expression between RS6 and CC samples following RMA and RUV normalization. RUV increases discovery of differentially expressed genes (genes that have a low  $p$ -value). Bottom. Volcano plot of differential expression ( $-\log_{10} p$ -value vs log fold change) of Quantile and RUV normalized samples. Genes with and FDR <0.01 are highlighted in blue. Positive controls are circled in red. (PDF 6414 kb)

**Additional file 4:** Differential expression after sleep deprivation and all time point of recovery sleep following RUV normalization. (XLSX 179 kb)

**Additional file 5:** Differential expression due to circadian time following RUV normalization. (XLSX 58 kb)

**Additional file 6:** Clusters of differentially expressed genes as defined on the heatmap in Fig. 3b. (XLSX 105 kb)

**Additional file 7:** Details of functional annotation enrichment and clustering analysis. (XLSX 63 kb)

### Acknowledgements

We thank H. Rodriguez and the Penn Molecular Profiling Facility for the RNA and Microarray work.

### Declarations

#### About this supplement

This article has been published as part of *BMC Genomics* Volume 17 Supplement 8: Selected articles from the Sixth International Conference of the Iberoamerican Society for Bioinformatics on Bioinformatics and Computational Biology for Innovative Genomics. The full contents of the supplement are available online at <https://bmcgenomics.biomedcentral.com/articles/supplements/volume-17-supplement-8>.

#### Funding

Research reported in this article was supported by the National Institute of Mental Health of the National Institutes of Health under award number R01MH099544 to M.G.F. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The publication charges for this article were funded by start-up funds provided by Washington State University to L.P.

#### Availability of data and materials

Data generated in this study is publicly available through GEO (GSE78215).

#### Authors' contributions

Study design by AW and MGF. Data collection by AW and NZ. Data analysis by JK, DR and LP, manuscript preparation by JRG with editing by JK, DR, TPS, MGF and LP. All authors approve of the final version of this manuscript.

#### Competing interests

The authors declare that they have no competing interests.

#### Consent for publication

Not applicable.

#### Ethics approval and consent to participate

The Institutional Animal Care and Use Committee (IACUC) of the University of Pennsylvania and Washington State University approved all animal work conducted during the research presented in this article.

#### Author details

<sup>1</sup>Washington State University, Elson S. Floyd College of Medicine, Spokane, WA 99202, USA. <sup>2</sup>Department of Neuroscience, Perelman School of Medicine University of Pennsylvania, Philadelphia, PA 19104, USA. <sup>3</sup>Division of Biostatistics, School of Public Health, University of California, Berkeley, CA 94720, USA. <sup>4</sup>Department of Statistics, University of California, Berkeley, CA, USA. <sup>5</sup>Department of Mathematics and Statistics, The University of Melbourne, Melbourne, VIC, Australia. <sup>6</sup>Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia.

Published: 25 October 2016

#### References

- Franken P, Malafosse A, Tafti M. Genetic determinants of sleep regulation in inbred mice. *Sleep*. 1999;22(2):155–69.
- Maret S, Dorsaz S, Gurcel L, Pradervand S, Petit B, Pfister C, Hagenbuchle O, O'Hara BF, Franken P, Tafti M. Homer1a is a core brain molecular correlate of sleep loss. *Proc Natl Acad Sci U S A*. 2007;104(50):20090–5.
- Mackiewicz M, Shockley KR, Romer MA, Galante RJ, Zimmerman JE, Naidoo N, Baldwin DA, Jensen ST, Churchill GA, Pack AI. Macromolecule biosynthesis: a key function of sleep. *Physiol Genomics*. 2007;31(3):441–57.
- Peixoto L, Risso D, Poplawski SG, Wimmer ME, Speed TP, Wood MA, Abel T. How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. *Nucleic Acids Res*. 2015;43(16):7664–74.
- Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol*. 2014;32(9):896–902.
- Gagnon-Bartsch JA, Speed TP. Using control genes to correct for unwanted variation in microarray data. *Biostatistics (Oxford, England)*. 2012;13(3):539–52.
- Jacob L, Gagnon-Bartsch JA, Speed TP. Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed. *Biostatistics*. 2016;17(1):16–28.
- Vecsey CG, Peixoto L, Choi JH, Wimmer M, Jaganath D, Hernandez PJ, Blackwell J, Meda K, Park AJ, Hannehalli S, et al. Genomic analysis of sleep deprivation reveals translational regulation in the hippocampus. *Physiol Genomics*. 2012;44(20):981–91.

9. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185–93.
10. Franken P, Chollet D, Tafti M. The homeostatic regulation of sleep need is under genetic control. *J Neurosci*. 2001;21(8):2610–21.
11. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell*. 2009;33(6):717–26.
12. Wang H, Liu Y, Briesemann M, Yan J. Computational analysis of gene regulation in animal sleep deprivation. *Physiol Genomics*. 2010;42(3):427–36.
13. Franken P, Malafosse A, Tafti M. Genetic variation in EEG activity during sleep in inbred mice. *Am J Physiol*. 1998;275(4 Pt 2):R1127–37.
14. Terao A, Wisor JP, Peyron C, Apte-Deshpande A, Wurts SW, Edgar DM, Kilduff TS. Gene expression in the rat brain during sleep deprivation and recovery sleep: an Affymetrix GeneChip study. *Neuroscience*. 2006;137(2):593–605.
15. Cirelli C, Gutierrez CM, Tononi G. Extensive and divergent effects of sleep and wakefulness on brain gene expression. *Neuron*. 2004;41(1):35–43.
16. Naidoo N, Giang W, Galante RJ, Pack AI. Sleep deprivation induces the unfolded protein response in mouse cerebral cortex. *J Neurochem*. 2005;92(5):1150–7.
17. Thompson CL, Wisor JP, Lee CK, Pathak SD, Gerashchenko D, Smith KA, Fischer SR, Kuan CL, Sunkin SM, Ng LL, et al. Molecular and anatomical signatures of sleep deprivation in the mouse brain. *Front Neurosci*. 2010;4:165.
18. Laposky A, Easton A, Dugovic C, Walisser J, Bradfield C, Turek F. Deletion of the mammalian circadian clock gene *BMAL1/Mop3* alters baseline sleep architecture and the response to sleep deprivation. *Sleep*. 2005;28(4):395–409.
19. Wisor JP, Pasumarthi RK, Gerashchenko D, Thompson CL, Pathak S, Sancar A, Franken P, Lein ES, Kilduff TS. Sleep deprivation effects on circadian clock gene expression in the cerebral cortex parallel electroencephalographic differences among mouse strains. *J Neurosci*. 2008;28(28):7193–201.
20. Franken P, Dijk DJ. Circadian clock genes and sleep homeostasis. *Eur J Neurosci*. 2009;29(9):1820–9.
21. Franken P, Thomason R, Heller HC, O'Hara BF. A non-circadian role for clock-genes in sleep homeostasis: a strain comparison. *BMC Neurosci*. 2007;8:87.
22. Kopp C, Albrecht U, Zheng B, Tobler I. Homeostatic sleep regulation is preserved in *mPer1* and *mPer2* mutant mice. *Eur J Neurosci*. 2002;16(6):1099–106.
23. Shiromani PJ, Xu M, Winston EM, Shiromani SN, Gerashchenko D, Weaver DR. Sleep rhythmicity and homeostasis in mice with targeted disruption of *mPeriod* genes. *Am J Physiol Regul Integr Comp Physiol*. 2004;287(1):R47–57.
24. Naidoo N, Ferber M, Galante RJ, McShane B, Hu JH, Zimmerman J, Maislin G, Cater J, Wyner A, Worley P, et al. Role of homer proteins in the maintenance of sleep-wake states. *PLoS One*. 2012;7(4):e35174.
25. Peixoto L, Wimmer ME, Poplawski SG, Tudor JC, Kenworthy CA, Liu S, Mizuno K, Garcia BA, Zhang NR, Giese KP, et al. Memory acquisition and retrieval impact different epigenetic processes that regulate gene expression. *BMC Genomics*. 2015;16(Suppl5):S5.
26. Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA, et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci*. 2008;28(1):264–78.
27. Bachmann V, Klein C, Bodenmann S, Schäfer N, Berger W, Brugger P, Landolt H-P. The BDNF Val66Met polymorphism modulates sleep intensity: EEG frequency- and state-specificity. *Sleep*. 2012;35(3):335–44.
28. Mongrain V, La Spada F, Curie T, Franken P. Sleep loss reduces the DNA-binding of *BMAL1*, *CLOCK*, and *NPAS2* to specific clock genes in the mouse cerebral cortex. *PLoS One*. 2011;6(10):e26622.
29. El Helou J, Belanger-Nelson E, Freyburger M, Dorsaz S, Curie T, La Spada F, Gaudreault PO, Beaumont E, Pouliot P, Lesage F, et al. Neurologigin-1 links neuronal activity to sleep-wake regulation. *Proc Natl Acad Sci U S A*. 2013;110(24):9974–9.
30. Hirose T, Virnichi G, Tanigawa A, Naganuma T, Li R, Kimura H, Yokoi T, Nakagawa S, Benard M, Fox AH, et al. NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies. *Mol Biol Cell*. 2014;25(1):169–83.
31. Nakagawa S, Hirose T. Paraspeckle nuclear bodies—useful uselessness? *Cell Mol Life Sci*. 2012;69(18):3027–36.
32. Mongrain V, Hernandez SA, Pradervand S, Dorsaz S, Curie T, Hagiwara G, Gip P, Heller HC, Franken P. Separating the contribution of glucocorticoids and wakefulness to the molecular and electrophysiological correlates of sleep homeostasis. *Sleep*. 2010;33(9):1147–57.
33. Bellesi M, Pfister-Genskow M, Maret S, Keles S, Tononi G, Cirelli C. Effects of sleep and wake on oligodendrocytes and their precursors. *J Neurosci*. 2013;33(36):14288–300.
34. Bellesi M, de Vivo L, Tononi G, Cirelli C. Effects of sleep and wake on astrocytes: clues from molecular and ultrastructural studies. *BMC Biol*. 2015;13:66.
35. Watson AJ, Henson K, Dorsey SG, Frank MG. The truncated TrkB receptor influences mammalian sleep. *Am J Physiol Regul Integr Comp Physiol*. 2015;308(3):R199–207.
36. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289–300.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

