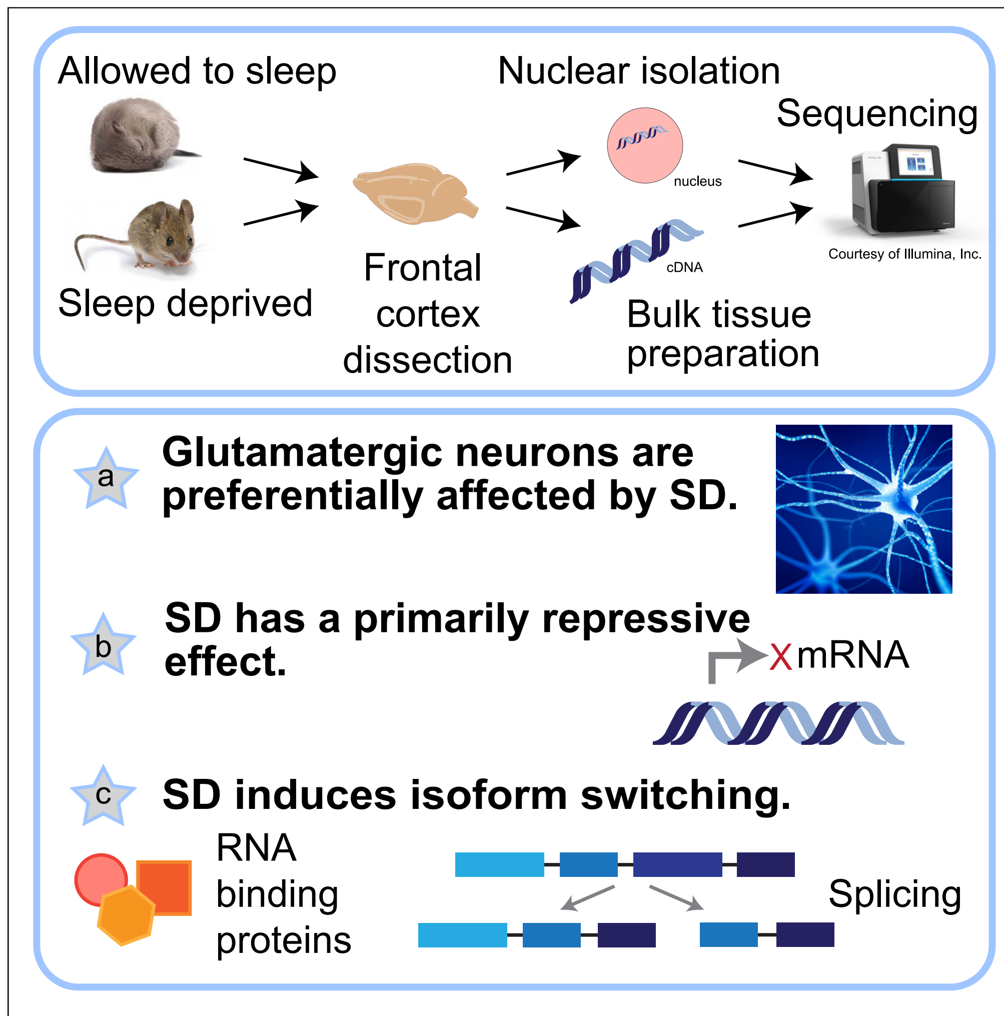
iScience

CellPress
OPEN ACCESS

Article

A global transcriptional atlas of the effect of acute sleep deprivation in the mouse frontal cortex



Kaitlyn Ford, Elena Zuin, Dario Righelli, ..., Stephanie C. Hicks, Davide Risso, Lucia Peixoto

lucia.peixoto@wsu.edu

Highlights

Sleep deprivation (SD) has a strong repressive effect on the cortical transcriptome

SD preferentially affects glutamatergic neurons

Glutamatergic neurons that are Ror-beta positive are preferentially affected by SD

SD causes isoform switching of more than 1500 genes

Ford et al., iScience 27, 110752
September 20, 2024 © 2024 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2024.110752

iScience

CellPress
OPEN ACCESS

Article

A global transcriptional atlas of the effect of acute sleep deprivation in the mouse frontal cortex

Kaitlyn Ford,[1,8] Elena Zuin,[2,3,8] Dario Righelli,[3] Elizabeth Medina,[1] Hannah Schoch,[1] Kristan Singletary,[1] Christine Muheim,[1] Marcos G. Frank,[1] Stephanie C. Hicks,[4,5,6,7] Davide Risso,[3] and Lucia Peixoto[1,9,*]

SUMMARY

Sleep deprivation (SD) has negative effects on brain and body function. Sleep problems are prevalent in a variety of disorders, including neurodevelopmental and psychiatric conditions. Thus, understanding the molecular consequences of SD is of fundamental importance in biology. In this study, we present the first simultaneous bulk and single-nuclear RNA sequencing characterization of the effects of SD in the male mouse frontal cortex. We show that SD predominantly affects glutamatergic neurons, specifically in layers 4 and 5, and produces isoform switching of over 1500 genes, particularly those involved in splicing and RNA binding. At both the global and cell-type specific level, SD has a large repressive effect on transcription, downregulating thousands of genes and transcripts. As a resource we provide extensive characterizations of cell-types, genes, transcripts, and pathways affected by SD. We also provide publicly available tutorials aimed at allowing readers adapt analyses performed in this study to their own datasets.

INTRODUCTION

Sleep is an evolutionary conserved powerful drive, but its function remains a mystery. It is well established that sleep deprivation (SD) has negative effects on brain function and affects a wide array of molecular processes.[1] Sleep problems are widely observed in neurodevelopmental, neurodegenerative, and psychiatric disorders.[2,3] Thus, understanding the molecular consequences of SD is of fundamental importance in neuroscience. We and others have shown that in rodents SD strongly affects the brain transcriptome.[4–13] It was initially thought that there was little agreement on the effects of SD. However, we have shown that if biological and technical noise are properly accounted for, hundreds of genes are differentially expressed after SD in the mouse brain regardless of technology, site or brain region.[14] Currently, studies of the effect of SD on the brain transcriptome are focused on genes. However, in mammals, cells often express multiple transcripts of the same gene (isoforms). Also, most current studies also lack resolution at the cell-type level with sufficient statistical power.

To further our understanding of the molecular consequences of acute SD, in this study we present the first simultaneous bulk and single-nuclear (sn) RNA sequencing (RNA-seq) characterization of the effects of SD in the frontal cortex of adult male mice at high resolution. We chose the frontal cortex based on EEG data in humans (which can only assess cortical areas) as it is the brain region most strongly affected by SD.[15] Using snRNA-seq we show that SD predominantly affects glutamatergic neurons, specifically in layers 4 and 5. Using bulk RNA-seq we performed differential gene and transcript expression (DGE/DTE), as well as differential transcript usage (DTU) analyses. We show that at the bulk level, SD affects half of the frontal cortex transcriptome and produces isoform switching of thousands of genes. Both bulk and snRNA-seq analysis show that SD has a large repressive effect on transcription, downregulating thousands of genes and transcripts both globally and in specific cell-types. This large yet cell-specific effect underscores the importance of controlling or accounting for the effects of sleep in any transcriptome studies of brain function. As a resource to the neuroscience community, we provide extensive characterization of which genes, transcripts and pathways are affected by SD and in which cell types; as well as guided tutorials for reproducible bulk (DGE, DTE, and DTU) and snRNA-seq differential expression analyses.

[1]Department of Translational Medicine and Physiology, Sleep and Performance Research Center, Elson S. Floyd College of Medicine, Washington State University, Spokane, WA 99202, USA
[2]Department of Biology, University of Padova, 35131 Padova, Veneto, Italy
[3]Department of Statistical Sciences, University of Padova, 35121 Padova, Veneto, Italy
[4]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA
[5]Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD 21218, USA
[6]Center for Computational Biology, Johns Hopkins University, Baltimore, MD 21218, USA
[7]Malone Center for Engineering in Healthcare, Johns Hopkins University, Baltimore, MD 21218, USA
[8]These authors contributed equally
[9]Lead contact
*Correspondence: lucia.peixoto@wsu.edu
https://doi.org/10.1016/j.isci.2024.110752

This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

RESULTS

Sleep deprivation preferentially affects gene expression in neurons

It is currently unknown how different cell types respond to SD in the mouse frontal cortex, and which genes and pathways are differentially affected across different types of neurons (e.g., glutamatergic and GABAergic). To address this, we carried out snRNA-seq using 10X Genomics Chromium v3 technology followed by Illumina sequencing of adult male mice, either allowed to sleep in their home cages (HC) or sleep deprived (SD, $n = 3$ per group). After gene abundance quantification and removal of low quality data (Figure S1), we performed cell-type label assignment using a reference dataset obtained from the Brain Initiative Cell Census Network (BICCN) to avoid the lack of reproducibility of cell-type labels that can arise from cluster-based assignment.[16] Nuclei counts for each replicate for each cell type are available in Figure S2. Cell-type nomenclature follows BICCN guidelines as per Yao et al., 2021.[16] Glutamatergic neurons are labeled based on the layer in which they reside (L1-6) as well as where they project. Across the cortex, there are three main classes of projection neurons: IT intratelencephalic tract (IT), pyramidal tract (PT), and corticothalamic tract (CT). GABAergic neurons are labeled based on marker expression (e.g., parvalbumin, somatostatin, and vasoactive intestinal polypeptide). First, we used an independent dataset from the BICCN as a gold standard and show that our pipeline shows good consistency with known cell labels when the Yao et al., 2021 dataset is used as a reference (Figure S3) even when two different methods are used (Azimuth and SingleR). Subsequently we applied Azimuth and the Yao et al. 2021 dataset as reference to assign cell-type labels. Figure 1A depicts a UMAP visualization of the top principal components after transfer of all cell type labels in all biological replicates (3 HC, 3 SD). Figure S4 shows the UMAP visualization per condition (SD vs. HC). This indicates that the reference dataset seems to be the most important aspect when assigning cell-type labels. Subsequently, to further validate the cell-type assignments in our datasets, we created heatmaps using known mouse cortex cell-specific marker expression from the literature, as well as those in the Allen Brain Transcriptomics explorer in each cell type. Both show good consistency between marker expression and cell-type assignment (Figure S5).

To uncover which factors, drive the variance in gene expression across different cell-types, we performed multidimensional scaling (MDS) of the pseudo-bulk sum of all nuclei assigned to a cell-type per independent biological replicate (color-coded by cell type and shaped by condition, HC as circles, SD as triangles). Our data show that SD has a larger effect on glutamatergic (Figure 1B) than GABAergic neurons (Figure 1C). SD does not seem to cause major differences in gene expression in non-neuronal cell types such as Astrocytes, Oligodendrocytes or Microglia (Figure S6A). Differences between conditions cannot be detected if we restrict the MDS plot to genes less likely to be differentially expressed by SD according to publicly available microarray data[14] (Figure S6B).

Sleep deprivation disproportionally affects glutamatergic neurons of the deeper layers of the cortex, particularly layer 4/5

To define differentially expressed genes (DEGs) after SD in each neuronal cell-type, we performed pseudo-bulk analysis after normalization to remove unwanted variation using negative control samples (replicates) and genes (defined using publicly available microarray data).[17,18] Figure 2A summarizes, for each neuronal cell type, the number of nuclei, DEGs, and positive controls recovered. Positive controls were defined based on publicly available microarray data (see STAR methods). Figure 2B shows the proportion of DEGs relative to the abundance (number of nuclei) for each cell-type in ln-scale. Our results are robust to differences in normalization, as they can also be observed when only negative control samples are used to estimate the unwanted variance (Figure S7). Since the number of DEGs is expected to increase linearly with the number of nuclei sequenced due to the distributional properties of count models, cell-types above the line are affected more than expected by SD, while those below the line are affected less than expected. Consistent with what we observed using MDS, glutamatergic neurons are preferentially affected, specifically in neurons in layers 4/5 and 5 that project to the intra-telencephalic tract (L4/5 IT and L5 IT) neurons. Results of differential expression analysis for all cell-types (neuronal and non-neuronal) can be seen in Figure S8. To properly control for the effect of different sample sizes on different cell-types, we ran differential expression analysis on 100 random subsets of 200 nuclei per cell-type, which confirms that SD disproportionally affects glutamatergic neurons (Figure S9). In particular, glutamatergic neurons of L4/5 IT, which express high levels of Ror-beta were the most affected with 1492 DEGs (522 upregulated, 970 downregulated, and 142 positive controls, Figure 2C). In contrast, the most affected of the GABAergic neurons, Pvalb, only had 395 DEGs after SD (129 upregulated, 266 downregulated, 37 positive controls, Figure 2D). Most cell types had more downregulated than upregulated DEGs, thus SD seems to disproportionately repress transcription. Principal component analysis (PCA), histograms of uncorrected *p*-values, and volcano plots for each cell-type are available in Figures S10–S12, respectively. A full list of DEGs per cell-type is available in Table S1.

Sleep deprivation affects distinct pathways and molecular functions in glutamatergic and GABAergic neurons

To better understand which genes and pathways are shared between or unique to glutamatergic or GABAergic neurons, we intersected the union of all DEGs in glutamatergic neurons with the union of all DEGs in GABAergic neurons (Figure 3; Table S3). Glutamatergic neurons contain over 20-fold more DEGs at FDR <0.05 that are unique relative to GABAergic neurons (2239 vs. 108). This enrichment if robust to the choice of FDR, as it can be seen at FDR <0.01 or FDR <0.10, as well (Figure S13). The majority of DEGs in GABAergic neurons are shared with glutamatergic neurons (417) and contain immediate-early genes (IEGs, *Arc*, *Homer1*, *Bdnf*) and stress response genes (*Hspa5*, *Hspa8*, Figure 3A). To further understand which pathways (KEGG, https://www.genome.jp/kegg/pathway.html), molecular functions (MF, https://www.uniprot.org) and biological processes (BP, https://www.uniprot.org) were more affected by SD than by chance in GABAergic versus glutamatergic neurons, we carried out functional enrichment analysis (Figures 3B and 3C; Table S4) of the unique sets of DEGs. When multiple terms had overlapping sets of genes, they were clustered for ease of interpretation (see STAR methods). Neurotransmitter receptors were downregulated in response to SD in both glutamatergic and GABAergic neurons (KEGG pathway: Neuroactive-ligand receptor interaction, *Vip*, *Sst*, *Gria4*, *Grin2d*). However, neurogenesis (Figure 3C Cluster 1 in red, BP and MF), cell adhesion (Figure 3C Cluster 2 in red, BP and

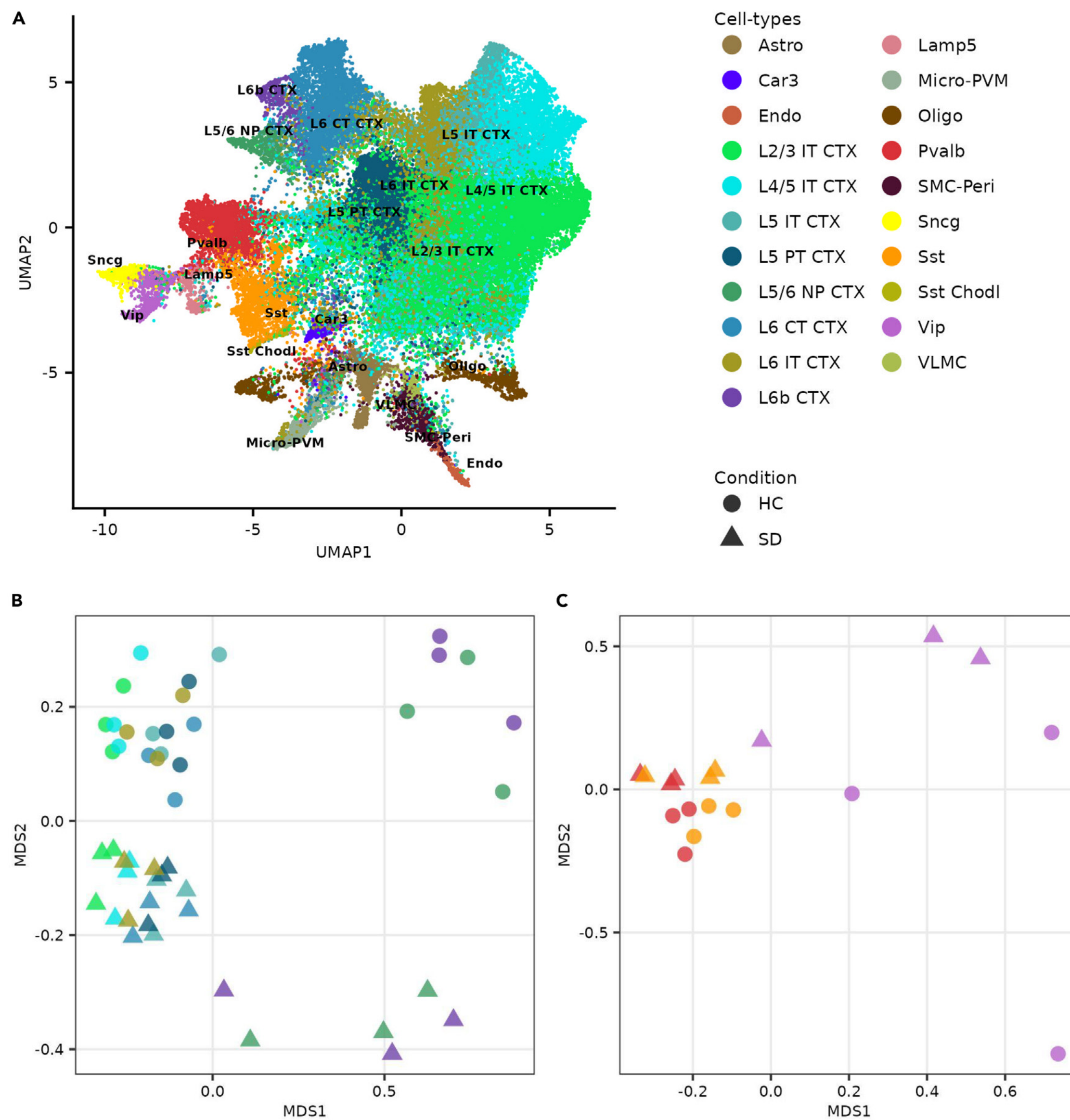


Figure 1. SD has a large detectable effect on glutamatergic and GABAergic cell-types

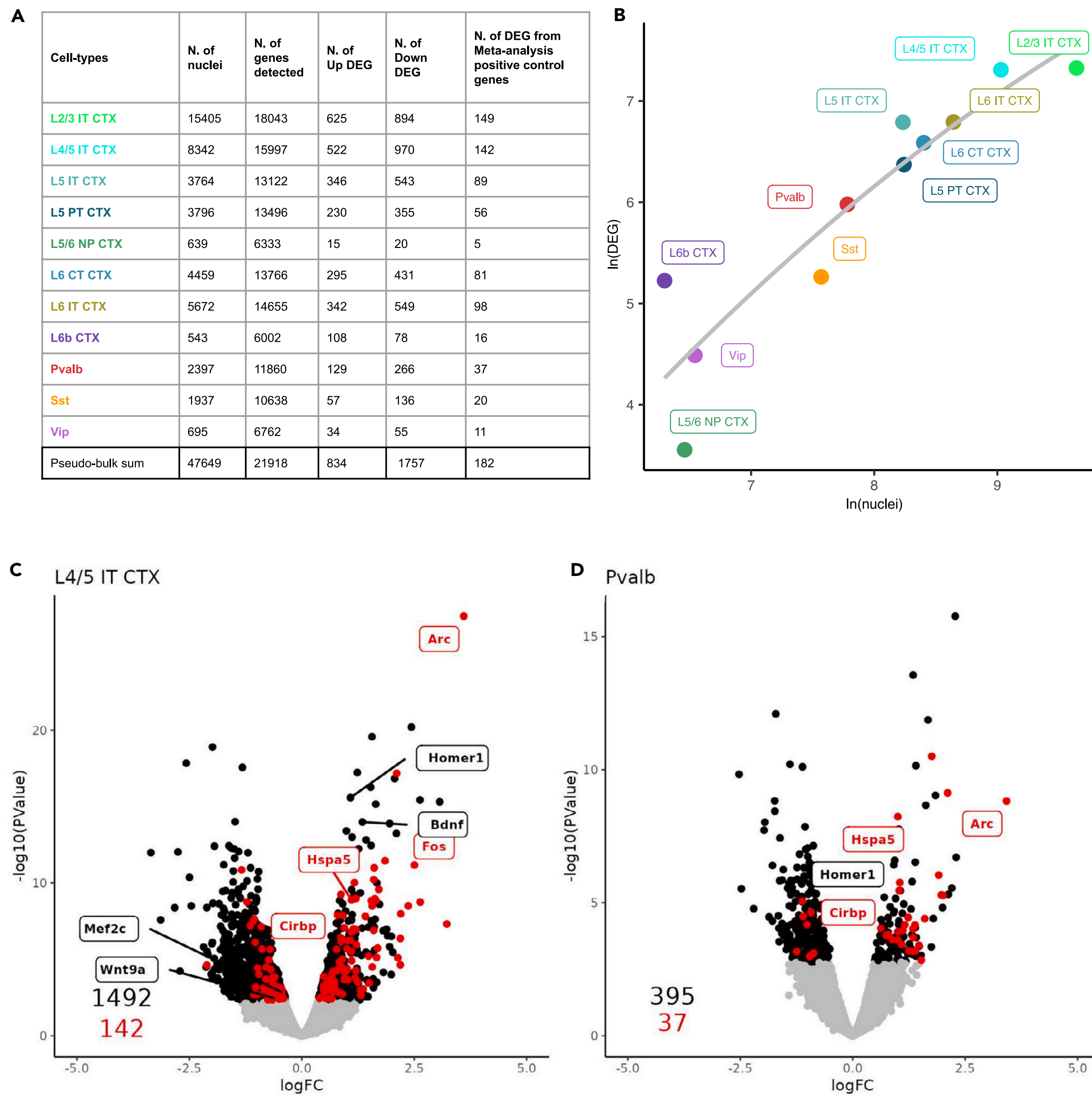
(A) Uniform Manifold Approximation and Projection (UMAP) plot for the 52,651 frontal cortex nuclei from all 6 replicates, annotated with a reference-based Azimuth method,[16] Brain Initiative Cell Census Network, BICCN). Glutamatergic, GABAergic and non-neuronal cell-types are shaded in unique colors.

(B) Multidimensional scaling (MDS) plot for glutamatergic cell-types with more than 500 nuclei shows cell-types are separated along MDS1, accounting for the largest source of dissimilarity in the data, while HC (circles, $N = 3$) and SD (triangles, $N = 3$) are separated along MDS2, accounting for the remainder of dissimilarity in the data.

(C) MDS plot for GABAergic cell-types with more than 500 nuclei shows cell-types are separated along MDS1, accounting for the largest source of dissimilarity in the data, while HC (circles, $N = 3$) and SD (triangles, $N = 3$) are separated along MDS2, accounting for the remainder of dissimilarity in the data. HC, Home Cage Controls. SD, Sleep Deprived.

KEGG), MAPK PI3K-Akt signaling (Figure 3C Cluster 3 in red, MF and KEGG), circadian rhythm (KEGG in red) were enriched in genes upregulated by SD. Genes that belong to development and differentiation (Figure 3C Cluster 1 in blue, BP and MF) were downregulated by SD only in glutamatergic neurons.

We then determined which DEGs were unique to a neuronal cell-type (Figure 4A; Table S5) Glutamatergic neurons had the largest numbers of unique DEGs, specifically L2/3 IT CTX and L4/5 IT CTX. Next, we asked which KEGG, MF, and BP were more affected by SD than expected by chance in some cell-types. Despite similar numbers of DEGs after SD, functional enrichment analysis shows a higher level of specificity for DEGs in L4/5 IT CTX (Figure 4B) relative to L2/3 IT CTX (Figure 4C). Only in L4/5 IT CTX glutamatergic neurons, SD upregulates



Cell-types	N. of nuclei	N. of genes detected	N. of Up DEG	N. of Down DEG	N. of DEG from Meta-analysis positive control genes
L2/3 IT CTX	15405	18043	625	894	149
L4/5 IT CTX	8342	15997	522	970	142
L5 IT CTX	3764	13122	346	543	89
L5 PT CTX	3796	13496	230	355	56
L5/6 NP CTX	639	6333	15	20	5
L6 CT CTX	4459	13766	295	431	81
L6 IT CTX	5672	14655	342	549	98
L6b CTX	543	6002	108	78	16
Pvalb	2397	11860	129	266	37
Sst	1937	10638	57	136	20
Vip	695	6762	34	55	11
Pseudo-bulk sum	47649	21918	834	1757	182

Figure 2. SD predominantly affects deep layers of glutamatergic cell-types

(A) Table shows the total number of nuclei in a neuronal cell-type, the number of genes detected, and up- and downregulated differentially expressed genes (DEGs). Also shown are the recovery of positive control genes from a prior, independent analysis.[14] The bottom row shows the pseudo-bulk sum for the expression analysis.

(B) Scatterplot with the natural log of the total number of nuclei on the x axis and the natural log of the DEGs on the y axis. A line of best fit was drawn through the points, with cell-types that appear to be more affected by the treatment above the line, and cell-types that appear to be less affected by the treatment below the line.

(C) Volcano plot for L4/5 IT CTX shows the log2FC on the x axis and the -log10 of the *p*-value on the y axis. Expressed genes (13,592) are in gray. Significantly differentially expressed genes are in black (2,405), FDR <0.05 and positive controls are shown in red (177).

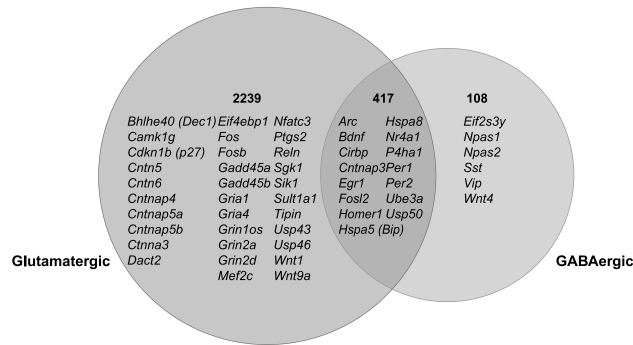
(D) Volcano plot for Pvalb shows the log2FC on the x axis and the -log10 of the *p*-value on the y axis. Expressed genes (11,585) are in gray. Significantly differentially expressed genes are in black (275), FDR <0.05 and positive controls are shown in red (32). A subset of genes from Table S2 are shown in (C) and (D). *N* = 3 per condition. SD, Sleep Deprivation.

certain components of the MAPK, PI3K-Akt, and Ras signaling pathways (Figure 4B, Cluster 1 in red, KEGG), such as *Gadd45a*, *Reln*, *Cdkn1b* (*p27)* and *Sgk1*.

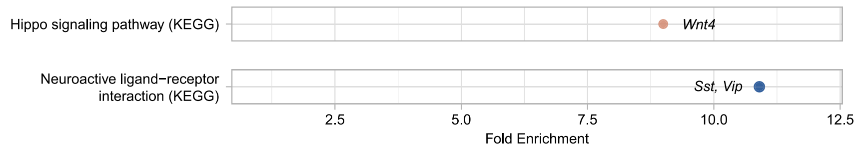
Sleep deprivation affects half of the cortical transcriptome and elicits extensive differential isoform usage

Although the effect of SD on brain gene expression is well-documented, it has never been investigated at the isoform level. We chose to perform isoform-level analyses in bulk RNA-seq, in contrast to 10X Chromium snRNA-seq data which has a 3′ bias and limited our ability

A



B GABAergic Only



C Glutamatergic Only

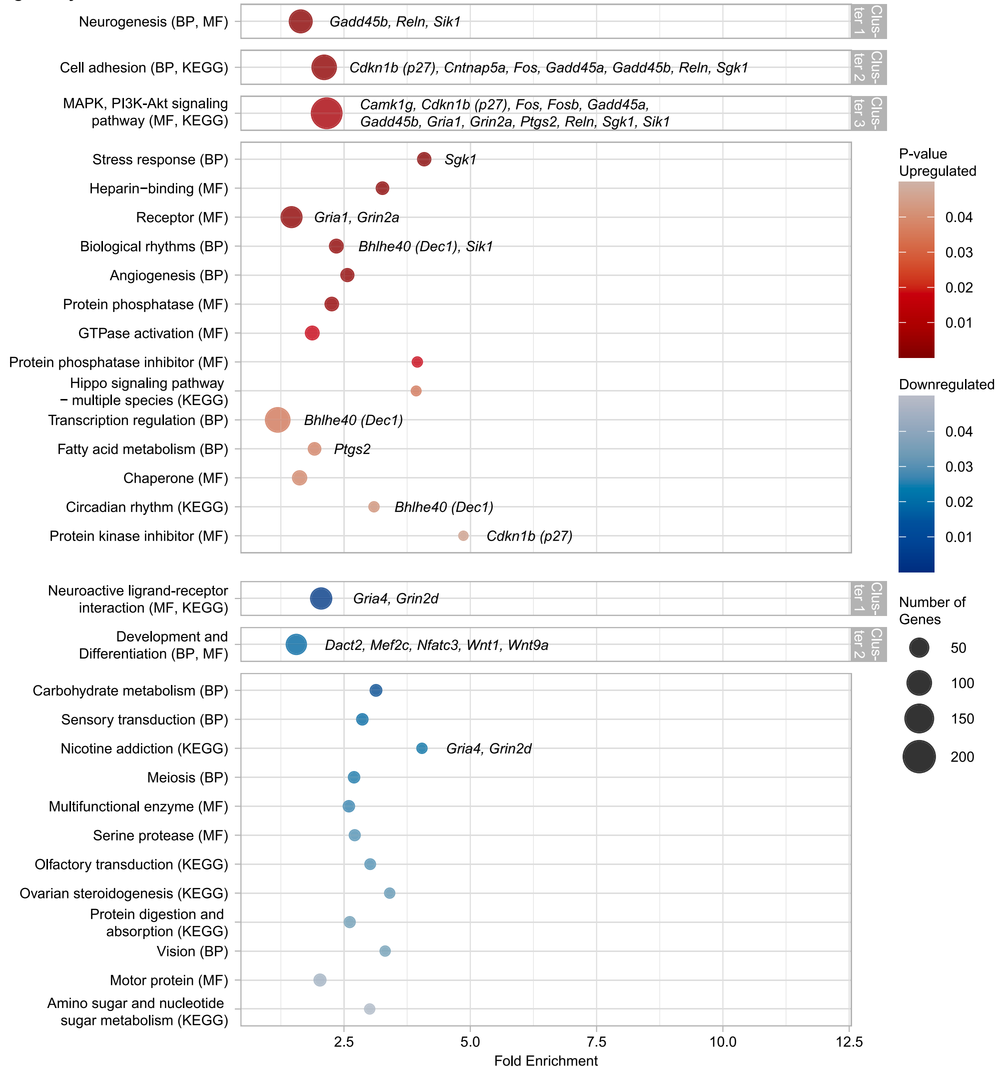


Figure 3. Glutamatergic neurons are preferentially affected by SD

(A) Venn diagram of differentially expressed genes (DEGs) within glutamatergic and GABAergic neuronal cell-types (FDR <0.05, >500 nuclei). Genes from Table S2 are highlighted.

(B and C) Bubbles show enriched terms (modified Fisher's Exact *p*-value <0.05) of upregulated (red) and downregulated (blue) DEGs that are unique to (B) GABAergic or (C) glutamatergic neurons, as compared to the union of expressed genes within the respective category. Bubble size reflects the number of genes per term (minimum of 3). Gray boxes outline clustered terms (similarity threshold >0.2, and enrichment score >1.5). Enrichment scores for each cluster; Cluster 1 Up (3.00), Cluster 2 Up (2.69), Cluster 3 Up (2.16), Cluster 1 Down (3.67), Cluster 2 Down (1.92). Terms were intersected with genes from Table S2 2. *N* = 3 per condition. SD, Sleep Deprivation. BP, Uniprot biological process. MF, Uniprot molecular function. KEGG, Kegg pathways.

to recover all transcripts of a gene. We performed differential transcript and gene expression (DTE and DGE) as well as differential transcript usage (DTU) analysis using nonparametric testing of inferential replicate counts after correcting for unwanted variation.[17–19]

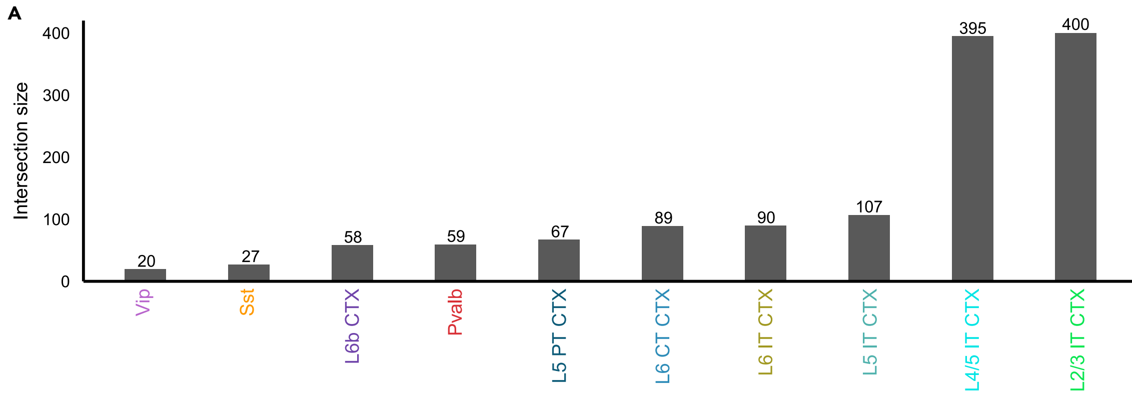
Figure 5 shows that after SD we detect 8,505 DEGs (q-value <0.05, Figures 5A, S14A, and S1C; Table S7) of 18,334 expressed genes, including several genes previously shown to be affected by SD (highlighted in Figure 5A). These include 83% of positive control genes (558/671) we previously detected across multiple published studies (Table S8).[14] We then chose the optimal log2 fold-change threshold based on the balance between the recovery of positive control genes, while simultaneously not recovering genes less likely to be differentially expressed after SD from public data (Figure S15). This resulted in an |log2 fold-change| >0.2 for downstream analysis. Following DGE analysis, we investigated DTE. We show that 15,525 transcripts (from 10,439 genes) are differentially expressed after SD (q-value <0.05, Figures 5B, S14B, and S14D; Table S7), of which 9,709 are downregulated and 5,816 are upregulated after SD. This indicates that transcript level analysis increases our ability to detect DEGs. Given the discrepancy in downregulated transcripts, we investigated which genes were shared between or unique to genes and transcript analyses (Figures 5C and 5D): 3,269 upregulated, and 3,836 downregulated genes were shared between analyses, while an additional 3,117 downregulated genes were detected at the transcript level, but not the gene level. This suggests there is more variation happening at the transcript-level, which is obscured when aggregating to the gene-level. To further explore this we focused on eukaryotic initiation factors, which were previously reported to be repressed by SD and are known to mediate the detrimental effects of SD on learning and memory.[13,20] We show that, several eukaryotic initiation factors were significantly differentially expressed at the transcript level that were not detected at the gene level (Figure S14E; q-value <0.05, |log2FC| > 0.20).

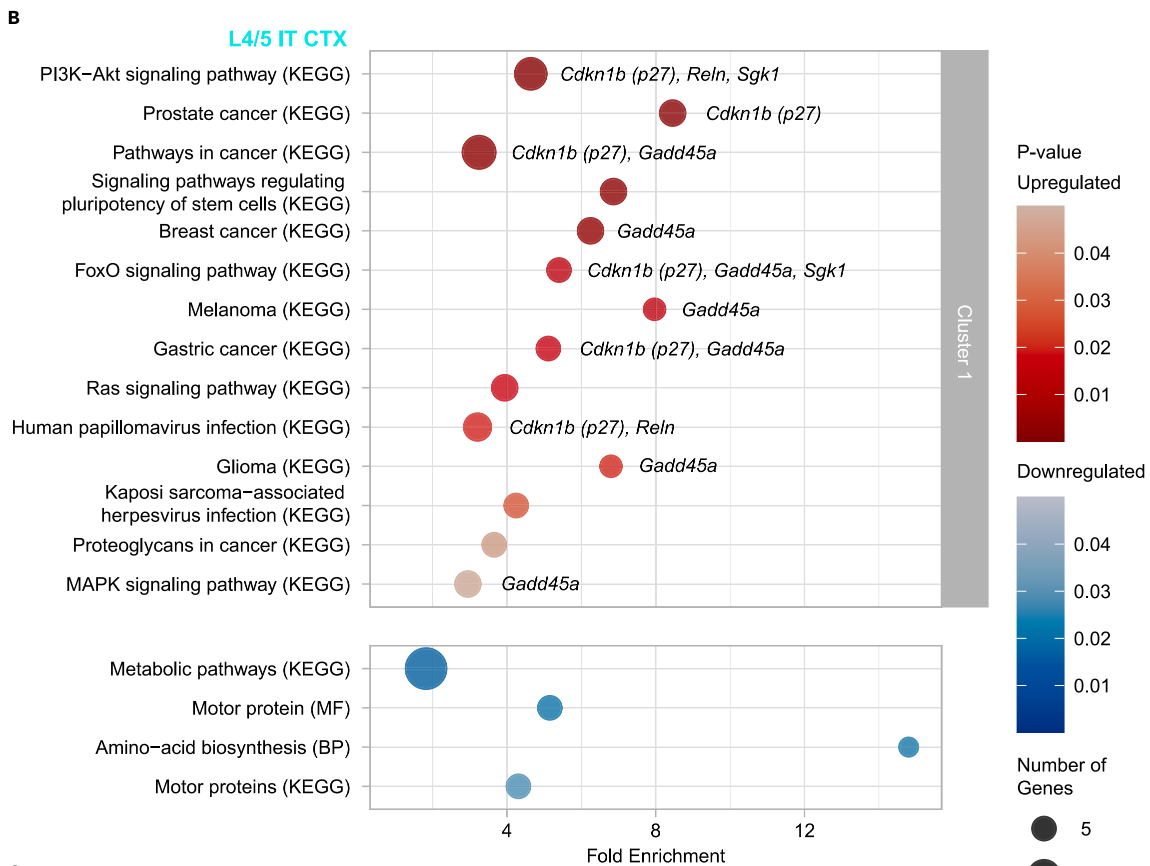
Our transcript analysis also shows that several genes have both upregulated and downregulated transcripts after SD (e.g., *Bdnf*). To investigate potential opposing effects on transcripts of the same gene, we performed differential transcript usage analysis (DTU). Specifically, this allowed us to detect which isoforms of a gene changed the proportion after SD. We detected 2,314 transcripts (corresponding 1,575 genes, Table S9) with significant changes in usage (q-value <0.05) in response to SD (Figure 6A). These include transcripts that were upregulated at both the gene and transcript level, but SD changes which transcript is primarily transcribed (e.g., *Homer1*, Figures 6B and 6C; Table S10), as well as transcripts in which the gene level analysis obscured transcripts being both up and downregulated (e.g., *Bdnf*, Figures 6D, 6E, and 6; Table S10). For *Bdnf* in particular, *Bdnf I* (201, somatic) increases in proportion, while *Bdnf VI* (205, dendritic) decreases in proportion. These examples suggest that SD may influence splicing, RNA-binding and transport of somatic versus dendritic isoforms. To further understand which kinds of processes and pathways were affected by isoform switching after SD, we carried out functional enrichment analysis (Figure 7), which revealed 16 enriched pathways/biological processes with 3 clusters of related pathways (Table S11). Genes that undergo isoform usage switching in response to SD are related to RNA binding/splicing (e.g., *Rbmx*), chromatin regulation (e.g., *Hdac3*), and kinases (e.g., *Camk1*).

DISCUSSION

In this study, we performed for the first time parallel snRNA-seq and bulk RNA-seq with multiple independent biological replicates in response to SD in the adult male mouse frontal cortex. Prior analyses have focused on bulk gene-level analysis,[4–13,24] or do not include independent biological replicates.[25] Thus, to date it was not possible to define what may be occurring at the isoform level or to detect changes specific to cell-types in response to SD. Because SD has a profound effect on brain function, and insufficient sleep is a hallmark of many brain disorders, understanding its molecular impact is not only important to understand the function of sleep, but also to understand how behavioral impairments in response to SD arise.

We focused on the frontal cortex based on EEG data in humans (which can only assess cortical areas) as it is the brain area most affected by acute SD.[15] The frontal cortex plays an essential role in higher-order brain processes, including cognition, attention, reward and emotion processing, all of which are affected by lack of sleep. Our snRNA-seq results indicate that SD has a disproportionate effect on neurons (Figure 1). Surprisingly, we do not detect a strong effect of SD on the transcriptome of glia, despite the documented role of glia such as astrocytes and microglia in sleep homeostasis.[26,27] The lower proportion than expected of glia present in our snRNA-seq data (Figure S2) suggests that to detect the true effect of SD in glia, it may be necessary to first enrich those populations using glia-specific marker sorting. However, in our dataset we have more than 500 nuclei for all the glial types explored, which is a larger number than some neuronal cell-types. This, in conjunction with previous transcriptome studies that find only ~1.4% of the astrocyte transcriptome seems responsive to sleep/wake state,[28] suggests it may be possible that at the transcriptome, glia do not respond to acute SD at the same magnitude that neurons do. Within neurons, the effect of SD is most prominent in glutamatergic neurons (Figure 2), with 2,239 genes exclusively regulated in this neuronal type (Figure 3A). Functional enrichment analysis shows that SD disproportionately affects pathways, molecular functions and biological processes involved in neurogenesis, MAPK PI3K-Akt signaling, circadian rhythms and development in glutamatergic neurons (Figure 3C). Interestingly, downregulation of neurotransmitter receptors by SD is detected in both glutamatergic and GABAergic neurons (KEGG pathway: Neuroactive-ligand receptor interaction, *Vip*, *Sst*, *Gria4*, *Grin2d*). Furthermore, it is important to note that stress response genes (such as those mediating o the


A
Intersection size
400
300
200
100
0
20
27
58
59
67
89
90
107
395
400
Vip
Sst
L6b CTX
Pvalb
L5 PT CTX
L6 CT CTX
L6 IT CTX
L5 IT CTX
L4/5 IT CTX
L2/3 IT CTX


B
L4/5 IT CTX
PI3K−Akt signaling pathway (KEGG)
Prostate cancer (KEGG)
Pathways in cancer (KEGG)
Signaling pathways regulating pluripotency of stem cells (KEGG)
Breast cancer (KEGG)
FoxO signaling pathway (KEGG)
Melanoma (KEGG)
Gastric cancer (KEGG)
Ras signaling pathway (KEGG)
Human papillomavirus infection (KEGG)
Glioma (KEGG)
Kaposi sarcoma−associated herpesvirus infection (KEGG)
Proteoglycans in cancer (KEGG)
MAPK signaling pathway (KEGG)
Cdkn1b (p27), Reln, Sgk1
Cdkn1b (p27)
Cdkn1b (p27), Gadd45a
Gadd45a
Cdkn1b (p27), Gadd45a, Sgk1
Gadd45a
Cdkn1b (p27), Gadd45a
Cdkn1b (p27), Reln
Gadd45a
Gadd45a
Cluster 1
Metabolic pathways (KEGG)
Motor protein (MF)
Amino−acid biosynthesis (BP)
Motor proteins (KEGG)
4
8
12
Fold Enrichment
P-value
Upregulated
0.04
0.03
0.02
0.01
Downregulated
0.04
0.03
0.02
0.01
Number of Genes
5
10
15
20

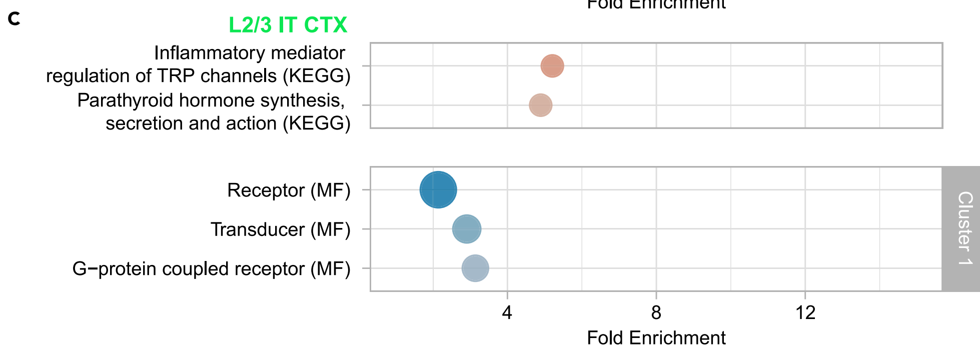

C
L2/3 IT CTX
Inflammatory mediator regulation of TRP channels (KEGG)
Parathyroid hormone synthesis, secretion and action (KEGG)
Receptor (MF)
Transducer (MF)
G−protein coupled receptor (MF)
Cluster 1
4
8
12
Fold Enrichment

Figure 4. L4/5 IT CTX is preferentially affected by SD relative to other cell-types

(A) Upset plot showing the unique differentially expressed genes (DEGs) within neuronal cell-types (FDR <0.05, >500 nuclei).

(B and C) Bubbles show enriched terms (modified Fisher's Exact *p*-value <0.05) of upregulated (red) and downregulated (blue) DEGs and unique to (B) L4/5 IT CTX and (C) L2/3 IT CTX, as compared with the expressed genes within those cell-types. Bubble size corresponds to the number of genes per term (minimum of 3). Gray boxes outline clustered terms (similarity threshold >0.2, and enrichment score >1.5). Enrichment scores for each cluster; L4/5 IT CTX Cluster 1 Up (2.17), L2/3 IT CTX Cluster 1 Down (1.59). *N* = 3 per condition. SD, Sleep Deprivation. BP, Uniprot biological process. MF, Uniprot molecular function. KEGG, Kegg pathways.

unfolded protein response (e.g., *Hspa5)* are shared between both glutamatergic and GABAergic neurons, thus suggesting that the differences in neuronal cell-type response are not related to stress. Nonetheless, the disproportionate effect of SD in glutamatergic versus GABAergic neurons may suggest that by predominantly altering the glutamatergic transcriptome, SD may alter excitatory/inhibitory balance. Indeed, in the visual cortex, excitation and inhibition have been shown to be modulated in a sleep-dependent manner in adult mice.[29]

Although is not clear whether or not the rodent frontal cortex possesses a layer 4 as defined in sensory cortices,[30] our data indicates that ROR-beta positive L4/5 IT glutamatergic neurons are the second most abundant type of neuron in the rodent frontal cortex (Figure 2A) and disproportionately responsive to SD at the transcriptome level (Figure 2B). This includes 522 upregulated and 970 downregulated genes after SD (Figures 2A and 2C), 395 of which are unique to these neurons (Figure 4A). ROR-beta expression is more prominent in frontal brain areas in rodents and primates, and drives the development of cortico-thalamic connectivity.[31] In addition to its role as a genetic marker for glutamatergic neurons of layers 4 and 5 in the cortex, ROR-beta is a key transcription factor controlling brain development and differentiation. ROR-beta expression has also been shown to be circadian and its deletion affects circadian behavior.[32] Furthermore, a previous exome sequencing study showed that ROR-beta variants may be implicated in autism spectrum disorder risk,[33] which is often co-morbid with sleep impairments.[34] Furthermore, it has been suggested that ROR-beta disruption may be linked to thalamocortical axon innervation and circuitry[35] (Jabaudon 2012). In ROR-beta positive neurons, SD uniquely upregulates genes (*Reln*, *Cdkn1b* (*p27)* and *Gadd45a*) and signaling pathways (PI3K, Akt, Ras and MAPK) involved in brain development and neurogenesis and differentiation. The fact that SD specifically affects ROR-beta expressing glutamatergic neurons in the frontal cortex may reflect the importance of sleep in regulating the function of these neurons and thalamocortical circuitry through these pathways.

Using bulk RNA-seq we detect 8,505 differentially expressed genes after SD, while simultaneously recovering 83% of known positive control genes (Figure 5). Although this is not the first study to use bulk RNA-seq to understand the effect of SD on the bulk cortical transcriptome, we detect between 2000 and 7000 more differentially expressed genes driven by sleep/wake than previously published studies in the cortex and hippocampus.[5,6,9,24] This increase in sensitivity is largely due to differences in methodology on RNA-seq data analysis, such as additional normalization to removal unwanted variance (e.g., RUV-seq) and simultaneous transcript and gene level analysis using nonparametric testing and inferential uncertainty (Fishpond/Swish).[19] Incorporating inferential uncertainty alone increases the number of DEGs considerably in our own data compared to our recently published study,[9] while also allowing for differential transcript expression and usage analysis after SD for the first time. We detect 15,525 differentially expressed (Figure 5B) and 2,314 differentially used (corresponding 1,575 genes) transcripts (Figure 6A). The latter indicates that SD can induce many isoform switches. Isoform switching is a phenomenon in which the relative contribution of one or many isoforms to the total expression of a particular gene changes significantly between conditions. Two notable examples of the effect of SD on isoform switching are *Homer1* and *Bdnf* (Figures 6B–6D), two genes with known roles in brain function which we show are altered in all neuronal cell-types after SD (Figure 3A). The upregulation of *Homer1a* after SD is well-known, and it is commonly referred to as a core molecular correlate of sleep loss.[8] Here, however, we show that upregulation of *Homer1a* after SD comes at the expense of a lower proportion of long isoforms of *Homer1*, which are more stably bound to the synapse.[21,22] In our analysis, we also find that *Bdnf I* (201, somatic) increases in proportion, while *Bdnf VI* (205, dendritic) decreases in proportion.[23] These examples suggest that SD may influence splicing or RNA-binding. Those processes were not identified as enriched in our snRNA-seq analysis. This may be because only some isoforms of such genes are affected, and thus when analyzing differential expression only at the gene level, the effect is not detectable or averaged out. Our functional enrichment analysis of genes that are affected by isoform switching after SD, shows that splicing and RNA binding are indeed affected by SD at the isoform level (Figure 7). If SD affects splicing and RNA binding, it may perhaps also affect transport of those isoforms to different neuronal compartments (soma vs. synapse). Our results suggest that the role of sleep and SD on isoform expression and transport, and its implication on brain function, needs to be further explored.

Limitations of our study include the fact that we focused on the adult male frontal cortex in our experiments. Future studies aimed at understanding the effect of both sex and developmental age on the transcriptional response to SD, with cell-type resolution, are needed in different brain regions. Recent spatial-transcriptomic studies suggest that different brain regions may have different responses to SD.[36] However, to achieve a full picture of the effect of SD in different brain areas, it will be necessary to combine different technologies (spatial, single cell and bulk RNA-seq) to balance their strengths and weaknesses. For example, our present study shows that the number of detected genes in snRNA-seq data for most cell-types (Figure 2A) is lower than for bulk RNA-seq. This is because the number of genes detected scales with the number of nuclei sequenced, and even over 15,000 nuclei fall short from detecting the 18,334 expressed genes we obtain using bulk RNA-seq. Thus, cell-type resolution may come at a cost of lower sensitivity for differential gene expression. Sensitivity of spatial transcriptomic experiments for differential expression analysis is likely lower than snRNA-seq. Given the inherent differences in the ability to recover DEGs by different technologies, we need to be cautious of the potential for false negatives to alter our interpretation if using only one technology.

Overall, we present the first global transcriptional atlas of the homeostatic response to SD in the adult male mouse frontal cortex, combining the advantages of both snRNA-seq and bulk RNA-seq with robust and reproducible data analysis pipelines. We show that SD has a large mostly repressive effect on the cortical transcriptome, that this effect is more prominent in glutamatergic neurons, in particular in L4/5 IT ROR-beta positive neurons. We also show that SD can cause isoform switching of thousands of genes. Because sleep and sleep

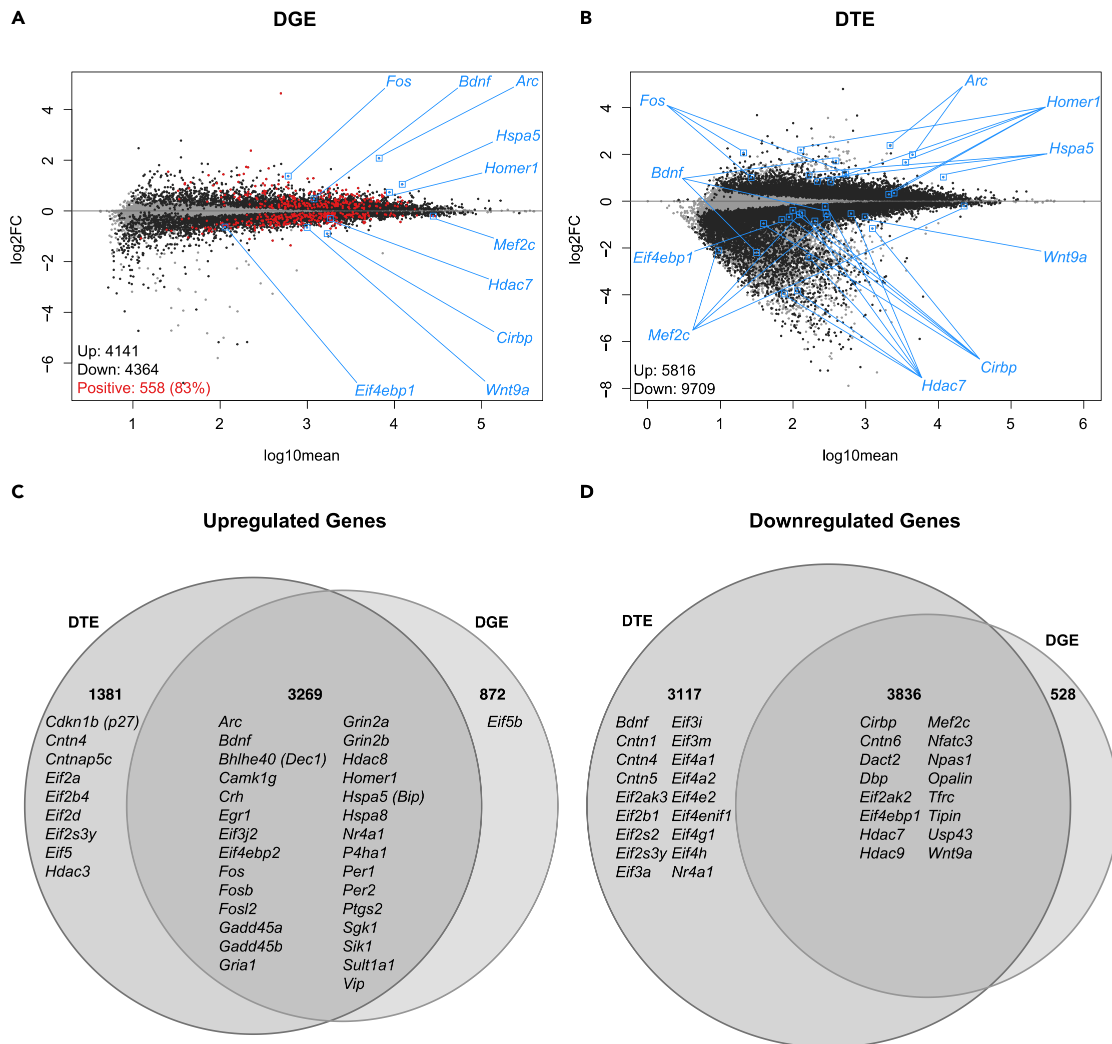


Figure 5. SD predominantly represses transcription at the gene and transcript level

(A) Differential gene expression (DGE) following SD. On the x axis, expression is shown as log10mean, and on the y axis, fold change is shown as log2FC. Expressed genes (18,334) are gray. Significantly differentially expressed genes (8,505), q-value <0.05, are black. Positive control genes previously shown to respond to sleep deprivation are red.[14] 83.16% positive control genes (558/671) were significantly differentially expressed in response to SD, q-value <0.05.

(B) Differential transcript expression (DTE) following SD. On the x axis, expression is shown as log10mean, and on the y axis, fold change is shown as log2FC. Each gray point shows an expressed transcript (54,030). Significantly differentially expressed transcripts (15,525 from 10,439 genes), q-value <0.05, are black. A subset of genes from Table S2 are shown in (A and B).

(C) Venn diagram shows the intersection of upregulated, significantly differentially expressed genes and upregulated, significantly differentially expressed transcripts.

(D) Venn diagram shows the intersection of downregulated, significantly differentially expressed genes and downregulated, significantly differentially expressed transcripts. Genes from Table S2 that have a qvalue <0.05 and |log2FC| > 0.2 are shown as examples in (C and D). *N* = 5 per condition. SD, Sleep Deprivation.

loss are often confounded in rodent studies of brain and behavior, these effects need to be accounted for in *in vivo* transcriptomic studies. As a resource to the community, we provide detailed lists of genes, cells and pathways affected by SD as well as tutorials to reproduce our data analysis. Importantly, we made our analyses completely reproducible by sharing all the code used to generate the results of this article and by providing a docker image to run the code with the exact software setup used for this study. In addition, our tutorials can serve as a starting point for the analysis of bulk and snRNA-seq data generated by future studies.

Limitations of the study

Limitations of our study include the use of adult male mouse frontal cortex tissue only. Therefore, this study cannot draw any conclusions about the effect of SD in other parts of the brain, at different developmental ages or in female mice. In addition, it is important to consider

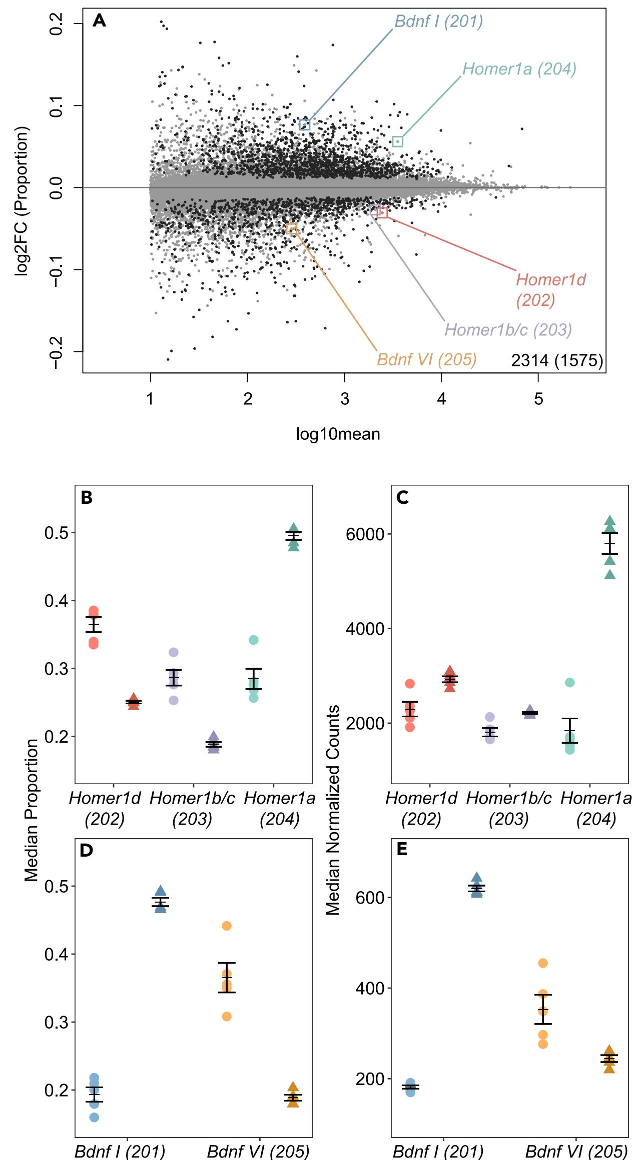


Figure 6. SD promotes differential transcript usage (DTU) for 1,575 genes, including *Homer1* and *Bdnf*

(A) DTU following SD. On the x axis, expression is shown as log10mean, and on the y axis, change in proportion is shown as log2FC. Gray points are expressed transcripts, with a log10mean >1. In black are transcripts that have significant changes in usage, q-value <0.05 and log10mean >1.

(B and C) Dot plot shows the change in (B) transcript proportion (from 0 to 1) and (C) expression levels in normalized transcript counts of "short" (*Homer1a*) and "long" (*Homer1b/c*, *Homer1d*) *Homer1* transcripts.[21,22]

(D and E) Dot plots show the (D) change in proportion (from 0 to 1) and (C) expression levels in normalized transcript counts of "synaptic" (*Bdnf VI*) and "somatic" (*Bdnf I*) *Bdnf* transcripts.[23] For (B–E), Home cage (HC) animals are circles and SD animals are triangles. Mean ± standard error is shown. *N* = 5 per condition. *Homer1* and *Bdnf* transcripts shown have significant DTE and DTU using Swish, q-value <0.05.[19] SD, Sleep Deprivation.

that because number of detected genes in snRNA-seq data for most cell-types (Figure 2A) is lower than for bulk RNA-seq, the cell-type resolution may come at a cost of lower sensitivity for differential gene expression. Thus, many more genes can be detected as differentially expressed after SD in the bulk-analysis, but not in the snRNA-seq analysis.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Lucia Peixoto (lucia.peixoto@wsu.edu).

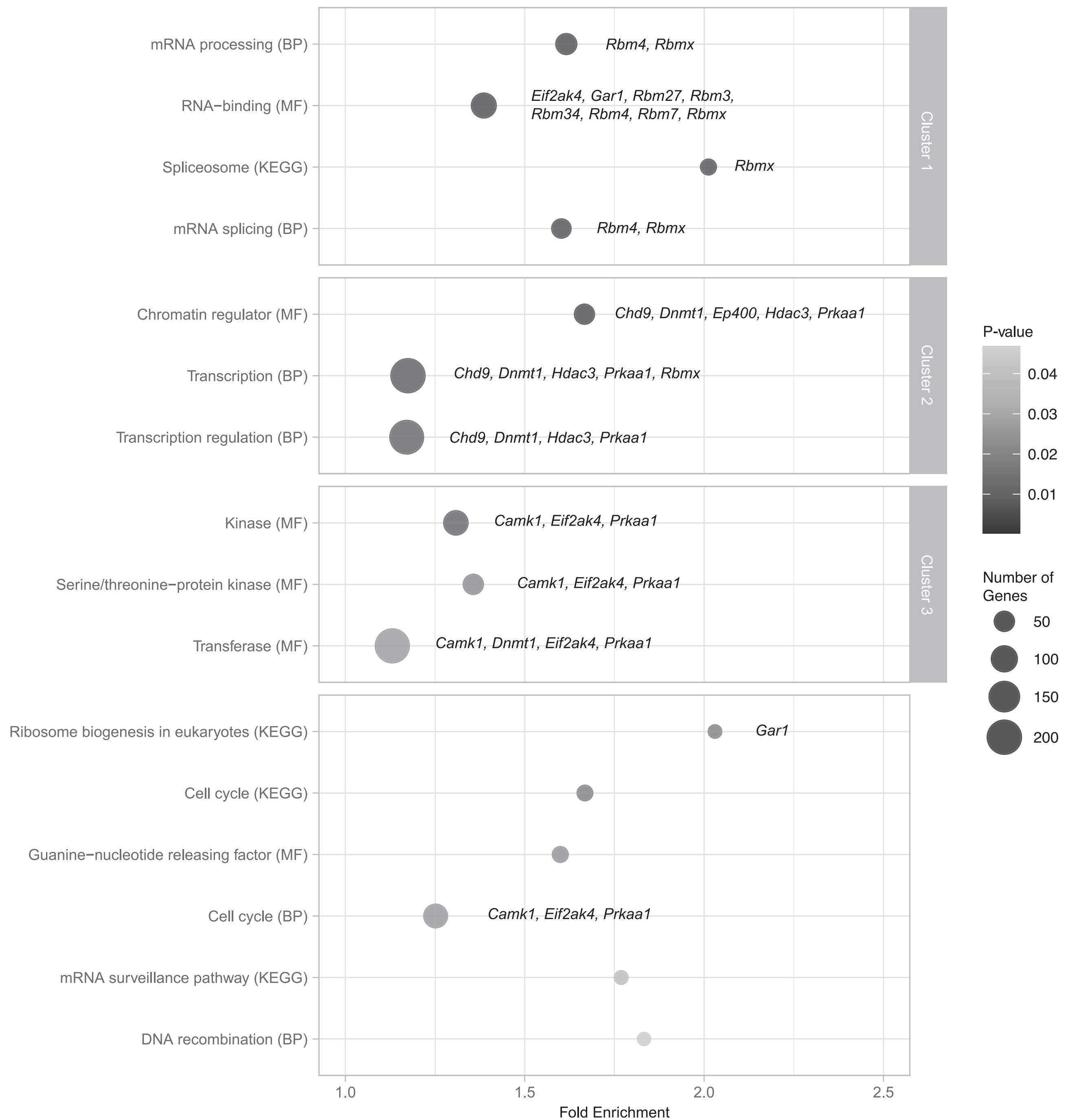


Figure 7. Genes for which SD affects isoform usage are mainly involved in splicing and chromatin regulation

Bubbles show enriched terms (modified Fisher's Exact *p*-value <0.05) following functional enrichment analysis of 1,575 genes that have significant DTU (qvalue <0.05), as compared with the expressed transcript list. Bubble size corresponds to the number of genes per term (minimum of 3), and color gradient represents *p*-values. Gray boxes outline clustered terms (similarity threshold >0.2, and enrichment score >1.5). Enrichment scores for each cluster; Cluster 1 (3.02), Cluster 2 (2.60), Cluster 3 (1.79). Example genes are shown for each cluster. SD, Sleep Deprivation. DTU, Differential Transcript Usage. BP, Uniprot biological process. MF, Uniprot molecular function. KEGG, KEGG pathways.

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Single-cell and bulk RNA-sequencing data have been deposited at GEO and are publicly available as of the data of publication. Accession numbers are listed in the key resources table.
- All original code has been deposited at Github (https://github.com/PeixotoLab/RNAseq_sleep) and is publicly available as of the date of publication. Furthermore, tutorials for reference-based cell-type annotation, differential expression and usage analyses can be found at: https://rissolab.github.io/AtlasCortexSD/index.html. A Docker container is also provided to ensure version control and reproducibility.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Institute of General Medical Sciences (NIGMS) [R35GM147020 to L.P.]; the National Institute of Neurological Disorders and Stroke (NINDS) [R56NS124805 to L.P.]; and the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation [CZF2019-002443 to S.C.H. and D. Ris.]. Funding for open access charge: R35GM147020 to L.P.

AUTHOR CONTRIBUTIONS

K.F.: Methodology, Investigation, Data Curation, Formal Analysis, Software, Writing—Original Draft, Writing—Review and Editing. E.Z.: Formal Analysis, Software, Data Curation, Writing—Original Draft. D.Rig.: Formal Analysis, Software. E.M.: Methodology, Investigation. H.S.: Methodology, Investigation. K.S.: Methodology, Investigation. C.M.: Formal Analysis, Software, Writing—Review and Editing. M.G.F.: Writing—Review and Editing. S.H.: Formal Analysis, Software, Writing—Original Draft, Writing—Review and Editing, Supervision. D.Ris: Writing—Original Draft, Writing—Review and Editing, Supervision. L.P.: Methodology, Investigation, Writing—Original Draft, Writing—Review and Editing, Supervision, Project Administration, Conceptualization, Resources.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animals
- METHOD DETAILS
  - Single nuclear RNA-seq study after SD
  - Bulk RNA-seq gene expression study after SD
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110752.

Received: February 12, 2024
Revised: May 31, 2024
Accepted: August 13, 2024
Published: August 17, 2024

REFERENCES

1. Lyons, L.C., Vanrobaeys, Y., and Abel, T. (2023). Sleep and memory: The impact of sleep deprivation on transcription, translational control, and protein synthesis in the brain. J. Neurochem. *166*, 24–46. https://doi.org/10.1111/jnc.15787.
2. Shen, Y., Lv, Q.K., Xie, W.Y., Gong, S.Y., Zhuang, S., Liu, J.Y., Mao, C.J., and Liu, C.F. (2023). Circadian disruption and sleep disorders in neurodegeneration. Transl. Neurodegener. *12*, 8. https://doi.org/10.1186/s40035-023-00340-6.
3. Veatch, O.J., Maxwell-Horn, A.C., and Malow, B.A. (2015). Sleep in Autism Spectrum Disorders. Curr. Sleep Med. Rep. *1*, 131–140. https://doi.org/10.1007/s40675-015-0012-1.
4. Cirelli, C., Gutierrez, C.M., and Tononi, G. (2004). Extensive and divergent effects of sleep and wakefulness on brain gene expression. Neuron *41*, 35–43. https://doi.org/10.1016/s0896-6273(03)00814-6.
5. Gaine, M.E., Bahl, E., Chatterjee, S., Michaelson, J.J., Abel, T., and Lyons, L.C. (2021). Altered hippocampal transcriptome dynamics following sleep deprivation. Mol. Brain *14*, 125. https://doi.org/10.1186/s13041-021-00835-1.
6. Hor, C.N., Yeung, J., Jan, M., Emmenegger, Y., Hubbard, J., Xenarios, I., Naef, F., and Franken, P. (2019). Sleep–wake-driven and circadian contributions to daily rhythms in gene expression and chromatin accessibility in the murine cortex. Proc. Natl. Acad. Sci. USA *116*, 25773–25783. https://doi.org/10.1073/pnas.1910590116.
7. Mackiewicz, M., Shockley, K.R., Romer, M.A., Galante, R.J., Zimmerman, J.E., Naidoo, N., Baldwin, D.A., Jensen, S.T., Churchill, G.A., and Pack, A.I. (2007). Macromolecule biosynthesis: a key function of sleep. Physiol. Genomics *31*, 441–457. https://doi.org/10.1152/physiolgenomics.00275.2006.
8. Maret, S., Dorsaz, S., Gurcel, L., Pradervand, S., Petit, B., Pfister, C., Hagenbuchle, O., O'Hara, B.F., Franken, P., and Tafti, M. (2007). Homer1a is a core brain molecular correlate of sleep loss. Proc. Natl. Acad. Sci. USA *104*, 20090–20095. https://doi.org/10.1073/pnas.0710131104.
9. Muheim, C.M., Ford, K., Medina, E., Singletary, K., Peixoto, L., and Frank, M.G. (2023). Ontogenesis of the molecular response to sleep loss. Neurobiol. Sleep Circadian Rhyt. *14*, 100092. https://doi.org/10.1016/j.nbscr.2023.100092.
10. Naidoo, N., Giang, W., Galante, R.J., and Pack, A.I. (2005). Sleep deprivation induces the unfolded protein response in mouse cerebral cortex. J. Neurochem. *92*, 1150–1157. https://doi.org/10.1111/j.1471-4159.2004.02952.x.
11. Noya, S.B., Colameo, D., Brüning, F., Spinnler, A., Mircsof, D., Opitz, L., Mann, M., Tyagarajan, S.K., Robles, M.S., and Brown, S.A. (2019). The forebrain synaptic transcriptome is organized by clocks but its proteome is driven by sleep. Science *366*, eaav2642. https://doi.org/10.1126/science.aav2642.

12. Terao, A., Wisor, J.P., Peyron, C., Apte-Deshpande, A., Wurts, S.W., Edgar, D.M., and Kilduff, T.S. (2006). Gene Expression in the Rat Brain during Sleep Deprivation and Recovery Sleep: An Affymetrix GeneChip® Study. Neuroscience *137*, 593–605. https://doi.org/10.1016/j.neuroscience.2005.08.059.
13. Vecsey, C.G., Peixoto, L., Choi, J.H.K., Wimmer, M., Jaganath, D., Hernandez, P.J., Blackwell, J., Meda, K., Park, A.J., Hannenhalli, S., and Abel, T. (2012). Genomic analysis of sleep deprivation reveals translational regulation in the hippocampus. Physiol. Genomics *44*, 981–991. https://doi.org/10.1152/physiolgenomics.00084.2012.
14. Gerstner, J.R., Koberstein, J.N., Watson, A.J., Zapero, N., Risso, D., Speed, T.P., Frank, M.G., and Peixoto, L. (2016). Removal of unwanted variation reveals novel patterns of gene expression linked to sleep homeostasis in murine cortex. BMC Genom. *17*, 727. https://doi.org/10.1186/s12864-016-3065-8.
15. Verweij, I.M., Romeijn, N., Smit, D.J., Piantoni, G., Van Someren, E.J., and van der Werf, Y.D. (2014). Sleep deprivation leads to a loss of functional connectivity in frontal brain regions. BMC Neurosci. *15*, 88. https://doi.org/10.1186/1471-2202-15-88.
16. Yao, Z., van Velthoven, C.T.J., Nguyen, T.N., Goldy, J., Sedeno-Cortes, A.E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., et al. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell *184*, 3222–3241.e26. https://doi.org/10.1016/j.cell.2021.04.021.
17. Peixoto, L., Risso, D., Poplawski, S.G., Wimmer, M.E., Speed, T.P., Wood, M.A., and Abel, T. (2015). How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. Nucleic Acids Res. *43*, 7664–7674. https://doi.org/10.1093/nar/gkv736.
18. Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol. *32*, 896–902. https://doi.org/10.1038/nbt.2931.
19. Zhu, A., Srivastava, A., Ibrahim, J.G., Patro, R., and Love, M.I. (2019). Nonparametric expression analysis using inferential replicate counts. Nucleic Acids Res. *47*, e105. https://doi.org/10.1093/nar/gkz622.
20. Tudor, J.C., Davis, E.J., Peixoto, L., Wimmer, M.E., van Tilborg, E., Park, A.J., Poplawski, S.G., Chung, C.W., Havekes, R., Huang, J., et al. (2016). Sleep deprivation impairs memory by attenuating mTORC1-dependent protein synthesis. Sci. Signal. *9*, ra41. https://doi.org/10.1126/scisignal.aad4949.
21. Clifton, N.E., Trent, S., Thomas, K.L., and Hall, J. (2019). Regulation and Function of Activity-Dependent Homer in Synaptic Plasticity. Mol. Neuropsychiatry *5*, 147–161. https://doi.org/10.1159/000500267.
22. Shiraishi-Yamaguchi, Y., and Furuichi, T. (2007). The Homer family proteins. Genome Biol. *8*, 206. https://doi.org/10.1186/gb-2007-8-2-206.
23. Chiaruttini, C., Sonego, M., Baj, G., Simonato, M., and Tongiorgi, E. (2008). BDNF mRNA splice variants display activity-dependent targeting to distinct hippocampal laminae. Mol. Cell. Neurosci. *37*, 11–19. https://doi.org/10.1016/j.mcn.2007.08.011.
24. Bjorness, T.E., Kulkarni, A., Rybalchenko, V., Suzuki, A., Bridges, C., Harrington, A.J., Cowan, C.W., Takahashi, J.S., Konopka, G., and Greene, R.W. (2020). An essential role for MEF2C in the cortical response to loss of sleep in mice. Elife *9*, e58331. https://doi.org/10.7554/eLife.58331.
25. Jha, P.K., Valekunja, U.K., Ray, S., Nollet, M., and Reddy, A.B. (2022). Single-cell transcriptomics and cell-specific proteomics reveals molecular signatures of sleep. Commun. Biol. *5*, 846. https://doi.org/10.1038/s42003-022-03800-3.
26. Deurveilher, S., Golovin, T., Hall, S., and Semba, K. (2021). Microglia dynamics in sleep/wake states and in response to sleep loss. Neurochem. Int. *143*, 104944. https://doi.org/10.1016/j.neuint.2020.104944.
27. Ingiosi, A.M., and Frank, M.G. (2023). Goodnight, astrocyte: waking up to astroglial mechanisms in sleep. FEBS J. *290*, 2553–2564. https://doi.org/10.1111/febs.16424.
28. Bellesi, M., de Vivo, L., Tononi, G., and Cirelli, C. (2015). Effects of sleep and wake on astrocytes: clues from molecular and ultrastructural studies. BMC Biol. *13*, 66. https://doi.org/10.1186/s12915-015-0176-7.
29. Bridi, M.C.D., Zong, F.-J., Min, X., Luo, N., Tran, T., Qiu, J., Severin, D., Zhang, X.-T., Wang, G., Zhu, Z.-J., et al. (2020). Daily Oscillation of the Excitation-Inhibition Balance in Visual Cortical Circuits. Neuron *105*, 621–629.e4. https://doi.org/10.1016/j.neuron.2019.11.011.
30. Anastasiades, P.G., and Carter, A.G. (2021). Circuit organization of the rodent medial prefrontal cortex. Trends Neurosci. *44*, 550–563. https://doi.org/10.1016/j.tins.2021.03.006.
31. Shibata, M., Pattabiraman, K., Lorente-Galdos, B., Andrijevic, D., Kim, S.-K., Kaur, N., Muchnik, S.K., Xing, X., Santpere, G., Sousa, A.M.M., and Sestan, N. (2021). Regulation of prefrontal patterning and connectivity by retinoic acid. Nature *598*, 483–488. https://doi.org/10.1038/s41586-021-03953-x.
32. André, E., Conquet, F., Steinmayr, M., Stratton, S.C., Porciatti, V., and Becker-André, M. (1998). Disruption of retinoid-related orphan receptor beta changes circadian behavior, causes retinal degeneration and leads to vacillans phenotype in mice. EMBO J. *17*, 3867–3877. https://doi.org/10.1093/emboj/17.14.3867.
33. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell *180*, 568–584.e23. https://doi.org/10.1016/j.cell.2019.12.036.
34. Petruzzelli, M.G., Matera, E., Giambersio, D., Marzulli, L., Gabellone, A., Legrottaglie, A.R., Margari, A., and Margari, L. (2021). Subjective and Electroencephalographic Sleep Parameters in Children and Adolescents with Autism Spectrum Disorder: A Systematic Review. J. Clin. Med. *10*, 3893. https://doi.org/10.3390/jcm10173893.
35. Jabaudon, D., Shnider, S.J., Macklis, J.D., Tischfield, D.J., and Galazo, M.J. (2012). RORβ Induces Barrel-like Neuronal Clusters in the Developing Neocortex. Cereb. Cortex *22*, 996–1006. https://doi.org/10.1093/cercor/bhr182.
36. Vanrobaeys, Y., Mukherjee, U., Langmack, L., Beyer, S.E., Bahl, E., Lin, L.-C., Michaelson, J.J., Abel, T., and Chatterjee, S. (2023). Mapping the spatial transcriptomic signature of the hippocampus during memory consolidation. Nat. Commun. *14*, 6100. https://doi.org/10.1038/s41467-023-41715-7.
37. Ingiosi, A.M., Schoch, H., Wintler, T., Singletary, K.G., Righelli, D., Roser, L.G., Medina, E., Risso, D., Frank, M.G., and Peixoto, L. (2019). Shank3 modulates sleep and expression of circadian transcription factors. Elife *8*, e42819. https://doi.org/10.7554/eLife.42819.
38. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods *14*, 417–419. https://doi.org/10.1038/nmeth.4197.
39. Love, M.I., Soneson, C., Hickey, P.F., Johnson, L.K., Pierce, N.T., Shepherd, L., Morgan, M., and Patro, R. (2020). Tximeta: Reference sequence checksums for provenance identification in RNA-seq. PLoS Comput. Biol. *16*, e1007664. https://doi.org/10.1371/journal.pcbi.1007664.
40. Gaidatzis, D., Burger, L., Florescu, M., and Stadler, M.B. (2016). Erratum: Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat. Biotechnol. *34*, 210. https://doi.org/10.1038/nbt0216-210a.
41. Soneson, C., Srivastava, A., Patro, R., and Stadler, M.B. (2021). Preprocessing choices affect RNA velocity results for droplet scRNA-seq data. PLoS Comput. Biol. *17*, e1008585. https://doi.org/10.1371/journal.pcbi.1008585.
42. Soneson, C., Srivastava, A., Patro, R., and He, D. (2023). alevinQC. Bioconductor. https://bioconductor.org/packages/alevinQC.
43. Rainer, J. (2017). EnsDb.Mmusculus.v79. Bioconductor. http://bioconductor.org/packages/EnsDb.Mmusculus.v79/.
44. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W., and Robinson, M.D. (2021). Doublet identification in single-cell sequencing data using scDblFinder. F1000Res. *10*, 979. https://doi.org/10.12688/f1000research.73600.2.
45. Berg, M., Petoukhov, I., van den Ende, I., Meyer, K.B., Guryev, V., Vonk, J.M., Carpaij, O., Banchero, M., Hendriks, R.W., van den Berge, M., and Nawijn, M.C. (2023). FastCAR: fast correction for ambient RNA to facilitate differential gene expression analysis in single-cell RNA-sequencing datasets. BMC Genom. *24*, 722. https://doi.org/10.1186/s12864-023-09822-3.
46. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
47. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172. https://doi.org/10.1038/s41590-018-0276-y.
48. Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol. *33*, 495–502. https://doi.org/10.1038/nbt.3192.
49. McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186. https://doi.org/10.1093/bioinformatics/btw777.

50. Kolde, R. (2019). Pheatmap: Pretty Heatmaps. Version 1.0.12.
51. Crowell, H.L., Soneson, C., Germain, P.-L., Calini, D., Collin, L., Raposo, C., Malhotra, D., and Robinson, M.D. (2020). muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. Nat. Commun. *11*, 6077. https://doi.org/10.1038/s41467-020-19894-4.
52. Risso, D., Schwartz, K., Sherlock, G., and Dudoit, S. (2011). GC-content normalization for RNA-Seq data. BMC Bioinf. *12*, 480. https://doi.org/10.1186/1471-2105-12-480.
53. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.
54. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940. https://doi.org/10.1093/bioinformatics/btx364.
55. Chen, H. (2022). VennDiagram: Generate High-Resolution Venn and Euler Plots. Version 1.7.3.
56. Wickham, H., Chang, W., Henry, L., Pedersen, T.L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., Dunnington, D., Posit, P.B.C., et al. (2023). ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. Version 3.4.4.
57. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.
58. Srivastava, A., Malik, L., Smith, T., Sudbery, I., and Patro, R. (2019). Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. Genome Biol. *20*, 65. https://doi.org/10.1186/s13059-019-1670-y.
59. Srivastava, A., Malik, L., Sarkar, H., and Patro, R. (2020). A Bayesian framework for inter-cellular information sharing improves dscRNA-seq quantification. Bioinformatics *36*, i292–i299. https://doi.org/10.1093/bioinformatics/btaa450.
60. Amezquita, R.A., Lun, A.T.L., Becht, E., Carey, V.J., Carpp, L.N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Soneson, C., et al. (2020). Orchestrating Single-Cell Analysis with Bioconductor. Nat. Methods *17*, 137–145. https://doi.org/10.1038/s41592-019-0654-x.
61. Hevner, R.F. (2007). Layer-specific markers as probes for neuron type identity in human neocortex and malformations of cortical development. J. Neuropathol. Exp. Neurol. *66*, 101–109. https://doi.org/10.1097/nen.0b013e3180301c06.
62. Lim, L., Mi, D., Llorca, A., and Marín, O. (2018). Development and Functional Diversification of Cortical Interneurons. Neuron *100*, 294–313. https://doi.org/10.1016/j.neuron.2018.10.009.
63. Sun, W., Cornwell, A., Li, J., Peng, S., Osorio, M.J., Aalling, N., Wang, S., Benraiss, A., Lou, N., Goldman, S.A., and Nedergaard, M. (2017). SOX9 Is an Astrocyte-Specific Nuclear Marker in the Adult Brain Outside the Neurogenic Regions. J. Neurosci. *37*, 4493–4507. https://doi.org/10.1523/JNEUROSCI.3199-16.2017.
64. Tremblay, R., Lee, S., and Rudy, B. (2016). GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. Neuron *91*, 260–292. https://doi.org/10.1016/j.neuron.2016.06.033.
65. Xin, W., Mironova, Y.A., Shen, H., Marino, R.A.M., Waisman, A., Lamers, W.H., Bergles, D.E., and Bonci, A. (2019). Oligodendrocytes Support Neuronal Glutamatergic Transmission via Expression of Glutamine Synthetase. Cell Rep. *27*, 2262–2271.e5. https://doi.org/10.1016/j.celrep.2019.04.094.
66. Gautier, O., Blum, J.A., Maksymetz, J., Chen, D., Schweingruber, C., Mei, I., Hermann, A., Hackos, D.H., Hedlund, E., Ravits, J., et al. (2023). Human Motor Neurons Are Rare and Can Be Transcriptomically Divided into Known Subtypes. Neuroscience. Preprint at bioRxiv. https://doi.org/10.1101/2023.04.05.535689.
67. Bullard, J.H., Purdom, E., Hansen, K.D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinf. *11*, 94. https://doi.org/10.1186/1471-2105-11-94.
68. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. Roy. Stat. Soc. B *57*, 289–300.
69. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57. https://doi.org/10.1038/nprot.2008.211.
70. Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T., and Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. *50*, W216–W221. https://doi.org/10.1093/nar/gkac194.
71. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA *100*, 9440–9445. https://doi.org/10.1073/pnas.1530509100.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
Nuclei PURE Prep kit	Millipore Sigma	Cat#: NUC201-1KT
Single Cell 3′ Reagent Kits v3 Chemistry Kit	10× Genomics	Cat#: PN-1000269
KAPA Library Quantification Kit	Kapabiosystems	Cat#: KK4824
NextSeq 500/550 High Output Kit v2.5	Illumina	Cat#: 20024907
RNAeasy kit	Qiagen	Cat#: 74104
High Sensitivity RNA Analysis Kit	Advanced Analytical Technologies, Inc.	Cat#: DNF-472
TruSeq Stranded mRNA Library Prep Kit	Illumina	Cat#: RS-122–2101
High Sensitivity NGS Fragment Analysis Kit	Advanced Analytical Technologies, Inc.	Cat#: DNF-474
HiSeq Cluster Kit v4	Illumina	No longer manufactured.
HiSeq SBS kit v4	Illumina	No longer manufactured.
Deposited data		
Single-nuclear raw sequencing data	This paper	GEO: GSE211088
Bulk single-nuclear raw sequencing data	Ingiosi et al.[37]	GEO: GSE113754
GRCm38 (M25) primary genome assembly, reference transcriptome and GTF coordinates	Gencode	https://www.gencodegenes.org/mouse/release_M25.html
Allen Whole Cortex & Hippocampus - 10× genomics data	Yao et al.[16]	http://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-10x
10× Nuclei v3 Broad (Reference dataset)	Macosko Lab	http://data.nemoarchive.org/biccn/lab/zeng/transcriptome/sncell/10x_v3/mouse/processed/analysis/10X_nuclei_v3_Broad/ RRID:SCR_015820
10× Nuclei v2 AIBS (Reference dataset)	Allen Institute for Brain Science Yao et al.[16]	http://data.nemoarchive.org/biccn/lab/zeng/transcriptome/sncell/10x_v2/mouse/processed/analysis/10X_nuclei_v2_AIBS/RRID:SCR_015820
Code for single-nuclear and bulk RNA-seq pipelines	This paper	https://github.com/PeixotoLab/RNAseq_sleep.
Publicly accessible website for the following tutorials: cell-type annotation, pseudobulk analysis, differential gene and transcript expression analyses and differential transcript usage analysis	This paper	https://rissolab.github.io/AtlasCortexSD/
Experimental models: organisms/strains		
Wild-Type C57BL6/J Mice	Bred in house after obtaining from The Jackson Laboratory	Strain#: 000664; RRID: SCR_004633
Software and algorithms		
R	R	https://www.r-project.org; RRID: SCR_001905
RStudio	RStudio	https://posit.co; RRID: SCR_000432
Illustrator 2023	Adobe	https://www.adobe.com; RRID: SCR_010279
Microsoft Excel	Microsoft	https://www.microsoft.com/en-gb/; RRID: SCR_016137

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DAVID	DAVID	https://david.ncifcrf.gov/; RRID: SCR_001881
HiSeq Control Software (v. 2.2.68)	Illumina	https://support.illumina.com/sequencing/sequencing_instruments/hiseq_2500/downloads.html
bcl2fastq (v. 2.17.1.14)	Illumina	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html
Cellranger mkfastq.	10x Genomics	https://www.10xgenomics.com/support/software/cell-ranger/downloads/eula?closeUrl=%2Fsupport%2Fsoftware%2Fcell-ranger&lastTouchOfferName=Cell%20Ranger&lastTouchOfferType=Software%20Download&product=chromium&redirectUrl=%2Fsupport%2Fsoftware%2Fcell-ranger%2Fdownloads
Salmon (v. 1.3.0)	Patro et al.[38]	https://github.com/COMBINE-lab/salmon
Tximeta (v. 1.15.2)	Love et al.[39]	https://bioconductor.org/packages/release/bioc/html/tximeta.html
eisaR	Gaidatzis et al.,[40] Soneson et al.[41]	https://bioconductor.org/packages/release/bioc/html/eisaR.html
alevinQC	Soneson et al.[42]	https://bioconductor.org/packages/release/bioc/html/alevinQC.html
EnsDb.Mmusculus.v79 (v.2.99)	Rainer et al.[43]	https://bioconductor.org/packages/release/data/annotation/html/EnsDb.Mmusculus.v79.html
scDblFinder (v. 1.12.0)	Germain et al.[44]	https://bioconductor.org/packages/release/bioc/html/scDblFinder.html
FastCAR (v. 0.1.0)	Berg et al.[45]	https://github.com/LungCellAtlas/FastCAR
AllenInstituteBrainData (v.0.99.1)	Dario Righelli	https://github.com/drighelli/AllenInstituteBrainData
Azimuth (v. 0.4.6)	Hao et al.[46]	https://github.com/satijalab/azimuth
SingleR (v. 2.0.0)	Aran et al.[47]	https://bioconductor.org/packages/release/bioc/html/SingleR.html
Seurat (v. 4.3.0)	Satija et al.[48]	https://github.com/satijalab/seurat/releases
scuttle (v. 1.8.4)	McCarthy et al.[49]	https://bioconductor.org/packages/release/bioc/html/scuttle.html
pheatmap (v. 1.0.12)	Kolde[50]	https://cran.rproject.org/web/packages/pheatmap/index.html
Muscat (v. 1.12.1)	Crowell et al.[51]	https://bioconductor.org/packages/release/bioc/html/muscat.html
EDASeq (v. 2.32.0)	Risso et al.[52]	https://bioconductor.org/packages/release/bioc/html/EDASeq.html
RUVSeq (v. 1.32.0)	Risso et al.[18]	https://bioconductor.org/packages/release/bioc/html/RUVSeq.html
edgeR (v. 3.40.2)	Robinson et al.[53]	https://bioconductor.org/packages/release/bioc/html/edgeR.html
UpSetR (v. 1.4.0)	Conway et al.[54]	https://cran.r-project.org/web/packages/UpSetR/index.html
VennDiagram (v.1.7.3)	Chen[55]	https://cran.rproject.org/web/packages/VennDiagram/index.html
ggplot2 (v. 3.4.2)	Wickham et al.[56]	https://cran.rproject.org/web/packages/ggplot2/index.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Fishpond (v. 2.4.0)	Zhu et al.[19]	https://bioconductor.org/packages/release/bioc/html/fishpond.html
Limma (v. 3.54.0)	Ritchie et al.[57]	https://bioconductor.org/packages/release/bioc/html/limma.html

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animals

Adult, male wild-type (WT) C57BL6/J mice (8–10-week-old) were housed in standard cages at 24 ± 1°C on a 12:12 h light:dark cycle with food and water *ad libitum.* All experimental procedures were approved by the Institutional Animal Care and Use Committee of Washington State University under ASAF 6841 and conducted in accordance with National Research Council guidelines and regulations for experiments in live animals.

METHOD DETAILS

Single nuclear RNA-seq study after SD

Tissue collection

Adult male (8–10-week-old) WT C57BL6/J mice were divided into 2 groups ($n = 3$ independent animals per group): sleep deprived (SD5) and home cage controls (HC5). All mice were individually housed. Home cage control mice were left undisturbed and sacrificed 5 h after light onset (ZT5). Mice in the sleep-deprived group were sleep-deprived for 5 h via gentle handling starting at light onset and then sacrificed upon completion of sleep deprivation (ZT5) without allowing for recovery sleep. Mice were sacrificed by live cervical dislocation (alternating between home cage controls and sleep-deprived mice), decapitated, and the frontal cortex was swiftly dissected on a cold block. Tissue was flash frozen in liquid nitrogen and stored at −80°C until processing.[9,37]

RNA isolation, library preparation and sequencing

Nuclei were extracted from mouse frontal cortical tissue using the Nuclei PURE Prep kit (NUC201-1KT, Millipore Sigma, Burlington MA, USA) with volumes reduced to a quarter of the recommended amount. Briefly, frontal cortex tissue (~0.035 g per sample/mouse) were lysed using 2 mL glass dounce homogenizers (Kimble-Chase, Vineland, NJ USA) in cold phosphate buffered saline and RNAase inhibitor (03335399001, Roche, Basel, Switzerland). Nuclei was then isolated using a sucrose gradient and ultracentrifugation at 13,000 rcf for 45 min at 4°C (Sorvall WX-100, F65L-6 x 13.5 rotor, Thermo Fisher, Waltham, MA USA). Isolated nuclei were resuspended in Nuclei PURE Prep kit storage buffer and RNAase inhibitor. Nuclei count and integrity was assessed prior to library preparation.

Single nuclear RNA-seq libraries were generated using the Chromium Single Cell 3′ Solution microfluidics platform (10× Technologies, Pleasonton, CA USA). Single-nuclei libraries were generated from the nuclei suspensions using the 10× Genomics Chromium Controller and Single Cell 3′ Reagent Kits v3 Chemistry following manufacturer's instructions. Briefly, we targeted capture of 10,000 single nuclei through generation of gel beads in emulsion (GEMs) which allowed partitioning of an individual nuclei along with a bead containing barcoded oligonucleotides. Reverse transcription and barcoding occurred within this emulsion resulting in transcripts from an individual nucleus having a unique molecular identifier (UMI). After barcoding, the emulsion was broken, and the cDNA processed in bulk. The barcoded cDNA was first amplified to generate sufficient mass for library construction and then sample index, P5 and P7 adapters were added for Illumina sequencing. The sizes of 10× cDNA libraries were assessed by Fragment Analyzer with the High Sensitivity NGS Fragment Analysis Kit. The concentrations were measured by StepOnePlus Real-Time PCR System (ThermoFisher Scientific, San Jose, CA) with the KAPA Library Quantification Kit (Kapabiosystems, Wilmington, MA). The libraries were diluted to 2 nM with RSB (10 mM Tris-HCl, pH8.5) and denatured with 0.1 N NaOH. Three pM libraries were loaded onto NextSeq 500 (Illumina, San Diego, CA) for sequencing using the NextSeq 500/550 High Output Kit v2.5 (150 Cycles). The libraries were sequenced from both ends with 28 + 8+0 + 91 cycles (read length 100 bp) at an average depth of 40 million paired-end reads per sample.

Quantification of raw sequencing reads

We processed the raw sequencing reads in FASTQ files using the Salmon (v. 1.3.0) package using the 'mapping mode' that runs in two phases: (i) the indexing step and (ii) the quantification step.[38] To prepare to create the index, we downloaded Gencode (release M25) reference genome ('GRCm38.primary_assembly.genome.fa.gz') and reference transcriptome ('gencode.vM25.transcripts.fa.gz)', along with GTF coordinates ('gencode.vM25.annotation.gtf.gz'). To improve the accuracy of quantification estimates from Salmon, we built an index that incorporated a set of genome targets as decoys (https://combine-lab.github.io/alevin-tutorial/2019/selective-alignment). Using the concatenated list of transcriptome targets along with genome targets, we used 'salmon index' function to build the index with the flags: '–gencode', '–threads 4', and '-k 31'. Next, we used 'salmon quant' to perform quantification with the flags '–threads 6' and '–numBootstraps 30'. Using this index, we used 'salmon alevin' to quantify reads to the gene level with flags '–chromiumV3', '–threads 6', '–forceCells 10000', '–dumpFeatures –dumpBfh', and '–numCellBootstraps 30'[58,59]. Next, we created an R/Bioconductor SingleCellExperiment object[60] with the tximeta (v. 1.15.2) R/Bioconductor package,[39] where

we quantified counts for both spliced mRNA and introns using the 'getFeatureRanges()' function from the eisaR R/Bioconductor package.[40,41] Also, we used the alevinQC R/Bioconductor package to calculate QC metrics for each sample processed with 'salmon alevin'.[42]

Data setting, quality control, normalization and doublets removal

We summed the UMI counts of spliced mRNA and introns sharing the same Ensembl Gene ID. To identify mitochondrial genes, we retrieved the chromosome location of each Ensembl Gene ID with the EnsDb.Mmusculus.v79 (v. 2.99.0) R/Bioconductor package.[43] We then split the data into six SingleCellExperiment objects, one for each mouse.

For each sample, we used the scuttle (v. 1.8.4) R/Bioconductor package to detect low-quality and damaged droplets.[49] Particularly, we computed per-cell (nuclei) quality-control metrics with the *perCellQCMetrics* function. Furthermore, we employed the *isOutlier* function and set "type = lower," which removed any nuclei that were more than three median absolute deviations from the median. The sum of UMI counts, the number of detected genes, and the percentage of mitochondrial counts were visualized with violin plots for each biological replicate (Figure S1). Lastly, for each sample, we removed potential doublets with the scDblFinder (v. 1.12.0) R/Bioconductor package,[44] using the *computeDoubletDensity* function to calculate the scores and the *doubletThresholding* function to set the doublet scores threshold (0.5) with the *griffiths* method. To correct for ambient RNA, the FastCAR (v. 0.1.0) R package[45] was employed. After reviewing the results, we found that the corrected matrix after removing the ambient RNA was the same as our expression matrix.

Overall, our quality control procedure retained 52,651 high-quality nuclei, with an average of 8,775 nuclei per mouse.

Cell-type annotation and validation of cell-type labels

To identify cell types, we used the Allen Whole Cortex & Hippocampus - 10x genomics data (http://portal.brain-map.org/atlases-and-data/rnaseq/mouse-whole-cortex-and-hippocampus-10x) as a reference dataset.[16] This dataset was imported with the *AllenInstituteBrainData* function of the AllenInstituteBrainData (v. 0.99.1) package (available at https://github.com/drighelli/AllenInstituteBrainData). We then selected the "Non-Neuronal", "Glutamatergic'' and "GABAergic" clusters coming from the Visual Cortex (VIS, VISl, VISm, VISp) to annotate our dataset. For computational issues, we selected a random subset of 100,000 cortical cells.

To identify the best cell annotation method, we used two datasets of primary motor cortex tissue. The first dataset, "10× Nuclei v3 Broad," (http://data.nemoarchive.org/biccn/lab/zeng/transcriptome/sncell/10x_v3/mouse/processed/analysis/10X_nuclei_v3_Broad/) was from the Broad (Macosko Lab), while the second dataset, "10× Nuclei v2 AIBS," (http://data.nemoarchive.org/biccn/lab/zeng/transcriptome/sncell/10x_v2/mouse/processed/analysis/10X_nuclei_v2_AIBS/) was from the Allen Institute for Brain Science.[16]

We then annotated these datasets using two methods: Azimuth and SingleR. For Azimuth, the reference data was converted into a Seurat object and into an Azimuth compatible object, using the *AzimuthReference* function of the Azimuth (v. 0.4.6) package https://satijalab.github.io/azimuth/articles/run_azimuth_tutorial.html.[46] Then query samples were merged and converted into a Seurat object. Cell annotation was computed using the *RunAzimuth* function of the Azimuth package. The t-SNE and the UMAP embeddings were computed using the *RunTSNE* and *RunUMAP* functions of the Seurat (v. 4.3.0) package,[48] https://cran.r-project.org/web/packages/Seurat/index.html with *seed.use = 1*. For SingleR, the reference dataset was aggregated across groups of cell types and was normalized, using the *aggregateAcrossCells* and the *logNormCounts* functions of the scuttle (v. 1.8.4) package. Then, cell annotation was computed using the *SingleR* function of the SingleR (v. 2.0.0) R/Bioconductor package,[47] https://bioconductor.org/packages/release/bioc/html/SingleR.html. We found that Azimuth was the best-performing method on these already annotated datasets and hence we chose this annotation method for the annotation of our rodent PFC snRNA-seq dataset.

In addition, to evaluate the cell-type assignments in our dataset, we visualized cell-type specific markers based on ref.,[61–65] with a heatmap of the log2-normalized count average in each group. We used the *pheatmap* function of the pheatmap (v. 1.0.12) package (https://cran.r-project.org/web/packages/pheatmap/index.html).[50] Furthermore, we used the Seurat package to visualize cell-type specific markers as a bubble plot, with the size of the bubbles corresponding to the percentage of cells expressing the marker. Additional cell-type markers from the Allen Brain Atlas Transcriptomics Explorer (https://celltypes.brain-map.org/rnaseq/mouse_ctx-hpf_10x?selectedVisualization=Heatmap&colorByFeature=Gene+Expression&colorByFeatureValue=Cux2) were visualized with the pheatmap package. If a gene was differentially expressed in a given cell-type, a black box surrounding the respective square was added.

As an additional quality control, we checked if there were cell types with a low proportion of intronic reads, as this could be a sign of cytoplasmic RNA (likely from cell debris) and assigned incorrectly to nuclei. All cell types had a high proportion of intronic reads, as expected in single-nuclear RNA-seq.[66] To visualize the assigned cell-type labels in two dimensions, the UMAP embeddings were computed using the *DimPlot* function of the Seurat package, with option *reduction = "integrated_dr"*, where "integrated_dr" is the supervised principal component analysis obtained by the Azimuth method. Finally, a pseudo-bulk level Multidimensional Scaling (MDS) plot was created with the *pbMDS* function of the muscat (v. 1.12.1) R/Bioconductor package.[51] Each point represents one subpopulation-sample instance; points are colored by subpopulation and shaped by treatment. A tutorial for mouse brain reference-based cell-type assignment is available through GitHub and at the following website: https://rissolab.github.io/AtlasCortexSD/articles/1_ct_anno.html.

Pseudo-bulk differential expression analysis for snRNA-seq data

For each neuronal cell type with more than 500 nuclei, differential gene expression analysis was carried out with a negative binomial generalized linear model (GLM) on pseudo-bulk samples. Specifically, we created the pseudo-bulk samples with the function *aggregateAcrossCells* of the scuttle package by summing the counts of each gene for each cell type and mouse combination.

Then, we normalized the raw counts for each cell type with the upper-quartile method, using the *betweenLaneNormalization* function of the EDASeq (v. 2.32.0) R/Bioconductor package with option *which = "upper"*.[52] To account for latent confounders, we computed the factors of unwanted variation on the normalized data, using the *RUVs* function of the RUVSeq R/Bioconductor (v. 1.32.0) package with *k = 2*,[17,18] and using a list of genes previously characterized as non-differential in sleep deprivation in a large microarray meta-analysis,[14] herein referred to as "negative control genes." Specifically, 10% of negative control genes were randomly selected to be used for evaluation and the remaining control genes were used to fit RUV normalization. Analysis and visualization was also repeated while using all rownames as negative control genes during RUVseq analysis.

We then used the edgeR R/Bioconductor (v. 3.40.2) package to perform differential expression after filtering the lowly expressed genes with the *filterByExpr* function (with parameters: *min.count = 10, min.total.count = 15, large*.n *= 10, min.prop = 0.7)*.[53] The raw counts were normalized with the upper-quartile method, using the function *calcNormFactors*.[67] The factors of unwanted variation were added to the design matrix. The differential gene expression analysis was performed with the function *glmLRT* by specifying "SD-HC" (Sleep Deprived vs. Home Cage Control) as contrast. We used the Benjamini-Hochberg procedure to control for the false discovery rate (FDR), i.e., we considered as differentially expressed those genes that had an adjusted *p*-value less than 5%.[68]

For each cell type, we visualized differentially expressed genes (DEGs) with volcano plots and assessed the model's goodness-of-fit by visualizing the *p*-value histograms. We incorporated cross-study, cross-brain tissue positive controls (Additional File 2 from Gerstner et al., 2016 to evaluate the performance of our differential gene expression pipeline.

To understand which cell-types may be preferentially affected by treatment, the natural log of the DEGs were plotted against the natural log of the number of nuclei. With this approach, the number of DEGs were expected to increase linearly with the number of nuclei sequenced. Due to the distributional properties of count models, cell-types above the line can be said to be affected more than expected by SD, while those below the line are affected less than expected. If there were no differences between SD and HC control animals in a given cell-type, the number of DEGs would be zero. To further the effect of sample size differences, differential expression was run on 100 random subsets of 200 nuclei in each cell-type. The results of this analysis were depicted on a scatterplot with the cell-types on the x axis, and the number of DEGs on the y axis.

For glutamatergic and GABAergic neurons, we used the *upset* function of the UpSetR (v. 1.4.0) package to compare the lists of differentially expressed genes within each cell type with more than 500 nuclei.[54] A minimum of 20 unique genes were required. To better understand which genes were shared between glutamatergic and GABAergic cell types, the union of all glutamatergic DEGs and GABAergic DEGs was determined. To visualize shared and unique genes, a Venn Diagram was generated using the *venn.diagram* function of the VennDiagram package (v. 1.7.3, https://cran.r-project.org/web/packages/VennDiagram/index.html).[55] Select genes were highlighted.

A tutorial for pseudo-bulk differential gene expression analysis in response to treatment of sc/snRNA-seq data is available through GitHub and at the following website: https://rissolab.github.io/AtlasCortexSD/articles/2_pb_dgea.html.

Functional enrichment analysis of snRNA-seq data

Additionally, functional enrichment analysis of genes that were shared between glutamatergic and GABAergic cell types, or unique to a given category (glutamatergic only or GABAergic only) were subjected to functional annotation using the Database for Annotation, Visualization and Integrated Discovery v2021 (DAVID).[69,70] Prior to the analysis, genes were separated by fold change to obtain one list of upregulated and one list of downregulated genes per category. Genes that were upregulated in one cell type, but downregulated in another were excluded from analysis. The following categories were used: Uniprot Biological Process, Uniprot Molecular Function (https://www.uniprot.org) and KEGG Pathways (https://www.genome.jp/kegg/pathway.html). Enrichment was relative to the union of all expressed genes within a category: unique to glutamatergic cell types or unique to GABAergic cell types. An EASE Score <0.05 and similarity threshold >0.20 were used to allow for inclusive clustering. Clustered and unclustered terms were visualized with a bubble plot using the *ggplot* function from the ggplot2 (v. 3.4.2, https://cran.r-project.org/web/packages/ggplot2/index.html) package,[56] with the size of the bubbles corresponding to the number of genes within a term. For glutamatergic and GABAergic bubble plots, clustered terms were reduced to one bubble, with the size of the bubble corresponding to the union of genes within all terms in that category. Duplicated genes were removed. Fold enrichment is visualized along the x axis. For clustered terms, the geometric mean of the fold enrichments was determined and plotted along the x axis. P-values are shown as a color gradient, red for upregulated and blue for downregulated. For clustered terms, the geometric mean of the *p*-values was plotted.

Functional enrichment of genes that were differentially expressed, and unique to a cell type, was performed using DAVID. DGE lists were separated by fold change to obtain one list of upregulated, and one list of downregulated genes per cell type. The same categories were used as detailed above: Uniprot Biological Process (https://www.uniprot.org), Uniprot Molecular Function (https://www.uniprot.org) and KEGG Pathways (https://www.genome.jp/kegg/pathway.html). Again, an EASE Score <0.05 and similarity threshold >0.20 were used to allow for inclusive clustering. For L2/3 IT CTX and L4/5 IT CTX, clustered and unclustered terms were visualized with a bubble plot using the *ggplot* function from the ggplot2 (v. 3.4.2, https://cran.r-project.org/web/packages/ggplot2/index.html) package, with the size of the bubbles corresponding to the number of genes within a term. The fold enrichment is visualized along the x axis. P-values are shown as a color gradient, red for upregulated and blue for downregulated.

Bulk RNA-seq gene expression study after SD

Tissue collection

Adult male (8–10-week-old) WT C57BL6/J mice were divided into 2 groups (*n* = 5 independent animals per group): sleep deprived (SD5) and home cage controls (HC5). All mice were individually housed. Home cage control mice were left undisturbed and sacrificed 5 h after light

onset (ZT5). Mice in the sleep deprived group were sleep deprived for 5 h via gentle handling starting at light onset and then sacrificed upon completion of sleep deprivation (ZT5) without allowing for recovery sleep. Mice were sacrificed by live cervical dislocation (alternating between home cage controls and sleep deprived mice), decapitated, and the frontal cortex was swiftly dissected on a cold block. Tissue was flash frozen in liquid nitrogen and stored at −80°C until processing.[9,37] This protocol was repeated over a 5 day period, and all tissue was collected within the first 15 min of the hour.

RNA isolation, library preparation and sequencing

Frontal cortex tissue was homogenized in Qiazol buffer (Qiagen, Hilden, Germany) using a TissueLyser (Qiagen) and all RNA was extracted using the Qiagen RNAeasy kit (Qiagen) on the same day. The integrity of total RNA was assessed using Fragment Analyzer (Advanced Analytical Technologies, Inc., Ankeny, IA) with the High Sensitivity RNA Analysis Kit (Advanced Analytical Technologies, Inc.). RNA Quality Numbers (RQNs) from 1 to 10 were assigned to each sample to indicate its integrity or quality. "10" stands for a perfect RNA sample without any degradation, whereas "1" marks a completely degraded sample. RNA samples with RQNs ranging from 8 to 10 were used for RNA library preparation with the TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA). Briefly, mRNA was isolated from 2.5 μg of total RNA using poly-T oligo attached to magnetic beads and then subjected to fragmentation, followed by cDNA synthesis, dA-tailing, adaptor ligation and PCR enrichment. The sizes of RNA libraries were assessed by Fragment Analyzer with the High Sensitivity NGS Fragment Analysis Kit (Advanced Analytical Technologies, Inc.). The concentrations of RNA libraries were measured by StepOnePlus Real-Time PCR System (ThermoFisher Scientific, San Jose, CA) with the KAPA Library Quantification Kit (Kapabiosystems, Wilmington, MA). The libraries were diluted to 2 nM with Tris buffer (10 mM Tris-HCl, pH8.5) and denatured with 0.1 N NaOH. Eighteen pM libraries were clustered in a high-output flow cell using HiSeq Cluster Kit v4 on a cBot (Illumina). After cluster generation, the flow cell was loaded onto HiSeq 2500 for sequencing using HiSeq SBS kit v4 (Illumina). DNA was sequenced from both ends (paired-end) with a read length of 100 bp. The average depth for all samples was 52 million read pairs.

Quantification of raw sequencing reads

To process the raw sequencing reads from the bulk RNA-seq experiments, we again used Salmon along with the same index built with decoys, which is detailed in the 'Quantification of raw sequencing reads' in the 'single nuclear RNA-seq study after SD' section. Again, we used the tximeta to create a SummarizedExperiment object at the transcript level and a second object summarized to the gene level using the 'summarizeToGene()' from tximeta.

Differential gene and transcript expression

Following tximeta, inferential replicates were scaled and filtered using the default parameters in the fishpond R/Bioconductor package (v. 2.4.0) so that only features with a minimum of 3 samples with a minimum count of 10 reads remained, herein referred to as "expressed".[19] To correct for unwanted variation, raw counts were first normalized with the upper-quartile method, using the *betweenLaneNormalization* function of the EDASeq (v. 2.32.0) R/Bioconductor package with option *which* = "*upper*".[52] Next, *RUVs* (k = 4) from the RUVseq (v. 1.32.0) R/Bioconductor package was used to generate 'W,' the factors of unwanted variation.[17,18] For RUVs at the gene level, we implemented the same list of genes less likely to be affected by sleep deprivation according to microarrays detailed previously, herein referred to as "negative control genes".[14] For RUVs at the transcript level, we used all expressed transcripts as controls. We then used 'removeBatchEffect' from the limma R/Bioconductor package (v. 3.54.0) to remove the variation from the inferential replicates.[57] After correcting for unwanted variation, differential expression was performed using *Swish* from the fishpond package. Genes and transcripts with a q-value <0.05 (multiple test corrected *p*-value)[71] were deemed to be significantly differentially expressed. We incorporated the same cross-study, cross-brain tissue positive controls detailed previously (Additional File 2 from Gerstner et al., 2016 to evaluate the performance of our differential gene expression pipeline. We recovered 83.2% (558/671) of the positive control genes detected in the matrix at the gene level. To determine which genes were only detected with expression analysis at the gene or transcript level, the *venn.diagram* function of the VennDiagram package (v. 1.7.3, https://cran.r-project.org/web/packages/VennDiagram/index.html) was implemented.

A tutorial to perform DGE and DTE analysis is available through GitHub and at the following website: https://rissolab.github.io/AtlasCortexSD/.

Differential transcript usage

During our differential expression analysis, we discovered genes that had both upregulated and downregulated transcripts. Therefore, we decided to perform differential transcript usage (DTU) analysis, to detect which genes had transcripts with differential proportion in response to sleep deprivation. To do so, we incorporated 'isoformProportions' from the fishpond package, to convert the counts of inferential replicates to proportions before proceeding with *Swish*. To increase the reproducibility of the results presented, a secondary filter was immediately applied following the initial filtering which kept transcripts with a minimum of 10 reads across 3 samples. Only transcripts that had a log10mean >1 were kept, removing transcripts with low counts that passed the initial filtering.

To better visualize the change in the proportion of transcripts within genes of particular interest (*Homer1* and *Bdnf*), dot plots were generated using the *ggplot* function from the ggplot2 (v. 3.4.2, https://cran.r-project.org/web/packages/ggplot2/index.html) package. Briefly, for each biological replicate, the median of the inferential replicates was determined to obtain one value per transcript per animal. Transcripts

were only included if they had significant changes in both proportion and expression, q-value <0.05. Additionally, the mean and standard error within each condition was determined for each transcript and are shown.

In addition to plotting the proportions of transcripts that had significant changes in proportion, dot plots were also generated to show the normalized counts of the same transcripts following DTE analysis. To generate these plots, the median of the inferential replicates was determined following swish to obtain one value per transcript per animal. The normalized counts were only plotted for transcripts that had significant changes in proportion and expression, q-value <0.05. Additionally, the mean and standard error within each condition was determined for each transcript and are shown.

A tutorial to perform DTU analysis is available through GitHub and at the following website: https://rissolab.github.io/AtlasCortexSD/.

Functional enrichment analysis of bulk RNA-seq data

Functional enrichment analysis of the 1,575 genes with significant differential transcript usage (q-value <0.05) was performed using DAVID as detailed in the 'Functional enrichment analysis of snRNA-seq data' section within the 'single nuclear RNA-seq study after SD'. The same categories were used as detailed previously: Uniprot Biological Process (https://www.uniprot.org), Uniprot Molecular Function (https://www.uniprot.org) and KEGG Pathways (https://www.genome.jp/kegg/pathway.html). Enrichment was relative to the expressed genes after the initial filter preserving transcripts with a minimum of 10 reads across 3 samples, and before the additional log10mean filter. A *p*-value threshold for gene enrichment analysis (EASE Score) < 0.05 was used. A similarity threshold >0.20 was used to allow for inclusive clustering. Both clustered and unclustered terms were visualized with a bubble plot using the *ggplot* function from the ggplot2 (v. 3.4.2, https://cran.r-project.org/web/packages/ggplot2/index.html) package. For functional annotation of genes with significant changes in expression in response to sleep deprivation, please see Muheim et al., 2023.[9]

QUANTIFICATION AND STATISTICAL ANALYSIS

R was used to perform statistical analyses pertaining to differential expression analyses. For single-nuclear analysis with edgeR, *p*-values were adjusted using the Benjamini-Hochberg procedure, and significant when the FDR <0.05. For bulk analysis, *p*-values were adjusted using the Storey Tibshirani method, and significant when the q-value <0.05. For functional enrichment analysis using DAVID, a *p*-value <0.05 was considered significantly enriched, and a similarity threshold >0.20 was required for clustering.