nature neuroscience

Corresponding Author:	X. William Yang	# Main Figures:	8
Manuscript Number:	NN-RS54102A	# Supplementary Figures:	5
Manuscript Type:	Resource	# Supplementary Tables:	20
		# Supplementary Videos:	0

Reporting Checklist for Nature Neuroscience

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. For more information, please read Reporting Life Sciences Research.

Please note that in the event of publication, it is mandatory that authors include all relevant methodological and statistical information in the manuscript.

Statistics reporting, by figure

- Please specify the following information for each panel reporting quantitative data, and where each item is reported (section, e.g. Results, & paragraph number).
- Each figure legend should ideally contain an exact sample size (n) for each experimental group/condition, where n is an exact number and not a range, a clear definition of how n is defined (for example x cells from x slices from x animals from x litters, collected over x days), a description of the statistical test used, the results of the tests, any descriptive statistics and clearly defined error bars if applicable.
- For any experiments using custom statistics, please indicate the test used and stats obtained for each experiment.
- Each figure legend should include a statement of how many times the experiment shown was replicated in the lab; the details of sample collection should be sufficiently clear so that the replicability of the experiment is obvious to the reader.
- For experiments reported in the text but not in the figures, please use the paragraph number instead of the figure number.

Note: Mean and standard deviation are not appropriate on small samples, and plotting independent data points is usually more informative. When technical replicates are reported, error and significance measures reflect the experimental variability and not the variability of the biological process; it is misleading not to state this clearly.

		TEST USED		n			DESCRIPTIVE STATS (AVERAGE, VARIANCE)		P VALUE		DEGREES OF FREEDOM & F/t/z/R/ETC VALUE	
	FIGURE NUMBER	WHICH TEST?	SECTION & PARAGRAPH #	EXACT VALUE	DEFINED?	SECTION & PARAGRAPH #	REPORTED?	SECTION & PARAGRAPH #	EXACT VALUE	SECTION & PARAGRAPH #	VALUE	SECTION & PARAGRAPH #
example	1a	one-way ANOVA	Fig. legend	9, 9, 10, 15	mice from at least 3 litters/group	Methods para 8	error bars are mean +/- SEM	Fig. legend	p = 0.044	Fig. legend	F(3, 36) = 2.97	Fig. legend
example	results, para 6	unpaired t-test	Results para 6	15	slices from 10 mice	Results para 6	error bars are mean +/- SEM	Results para 6	p = 0.0006	Results para 6	t(28) = 2.808	Results para 6

		TEST USED		n			DESCRIPTIVE STATS (AVERAGE, VARIANCE)		P VALUE		DEGREES OF FREEDOM & F/t/z/R/ETC VALUE	
	FIGURE NUMBER	WHICH TEST?	SECTION & PARAGRAPH #	EXACT VALUE	DEFINED?	SECTION & PARAGRAPH #	REPORTED?	SECTION & PARAGRAPH #	EXACT VALUE	SECTION & PARAGRAPH #	VALUE	SECTION & PARAGRAPH #
+ -	1b,c	The figure does not report individual statistical tests; reports numbers of FDR-significant associations in Wald tests from NB GLM	Methods, heading "Statistical testing"	Too many to list	Numbers of sequenced distinct animals at each genotype	Supplementary Table S20	Number of significant associations	Figure legend	NA		NA	
+ -	2c-e	Robust correlation test and meta-analysis	Methods, heading "Statistical testing"	Striatum modules: 140; cortex modules: 142	Numbers of sequenced samples in each tissue	Results, heading "CAG length- and age-dependent gene co-expression modules"; Supplementary Table 20	Mean and s.e.m	Figure legend	NA			Z values are indicated within each panel; all test statistics are contained in Supplementary Table 3
+ -	3a-d	Permutation test	Figure legend and Methods, heading "Preservation of module-genotype association in independent data"	too many to list	Numbers of independently profiled animal and patient samples	Methods, heading "Preservation studies in independent human and mouse data"; Supplementary table 8	Mean	Figure legend	Too many to list;	Supplementary Table S8	NA	
+ -	3e	Fisher's exact test	Figure legend and Methods, heading "Statistical testing"	NA			-log(p-value) and numbers of overlapping genes	Figure legend	Too many to list	Figure 3e	NA	

+ -	5c-e	Permutation test	Figure legend and Methods, heading "Preservation of module-genotype association in independent data"	too many to list	Numbers of independently profiled animal and patient samples	Supplementary table S9	Mean	Figure legend	Too many to list	Figure 5c-e and legend; Supplementary table S12	NA	
+ -	7b	Figure reports numbers of significant associations in a robust correlation test and Fisher exact test	Figure legend and Methods, heading "Statistical testing "	14, 15, 15, 15, 46	Numbers of independently profiled samples	Supplementary Table 18	Number of significant associations	Figure legend	Too many to list	Figure 7b	NA	
+ -	7c	Correlation test	Figure legend	7039	Number of matched profiled proteins and mRNAs	Figure legend	NA	NA	<1e-200	Figure 7c	t = 48, n = 7037	
+ -	7d	Fisher exact test	Figure legend	NA			Numbers of overlap proteins and overlap p-values	Figure legend	Too many to list	Figure 7d	NA	
+ -	7e-i	Panels do not report statistical testing		46	Number of profiled independent animal samples	Supplementary Table S20	Standard boxplot statistics: median, confidence interval of the median, interquartile range,	NA	NA		NA	
+ -	7j	Fisher exact (hypergeometric) test	Figure legend	NA			Overlap count and p-value	Figure legend	Too many to list	Figure 7j	NA	
+ -	8b	ANOVA followed by Dunnett's test	Figure legend	20 per genotype and time point	2 replicates of 10 trials per genotype per time point	Methods, heading "Screening in Drosophila"; Figure 8 legend	Mean and standard error	Figure legend	Too many to list	Indicated by star symbols in table in panel 8a; Supplementary Table S19	Supplementary Figure 6	
+ -	4b,c	Permutation test	Figure legend	46	mean of sets of genes across 46 samples	Supplementary Table 18	Significance (-log(p-value))	Figure legend	Too many to list	Figures 4b,c	NA	
+ -	4d,e	Hypergeometric test	Figure legend	NA			Significance (-log(p-value))	Figure legend	Too many to list	Figures 4d,e	NA	

Representative figures

1. Are any representative images shown (including Western blots and immunohistochemistry/staining) in the paper?

   If so, what figure(s)?

   There are no representative images.

2. For each representative image, is there a clear statement of how many times this experiment was successfully repeated and a discussion of any limitations in repeatability?

   If so, where is this reported (section, paragraph #)?

Statistics and general methods

1. Is there a justification of the sample size?

   If so, how was it justified?

   Where (section, paragraph #)?

   Even if no sample size calculation was performed, authors should report why the sample size is adequate to measure their effect size.

   This study was not designed to validate an effect of a known size.

2. Are statistical tests justified as appropriate for every figure?

   Where (section, paragraph #)?

   The statistical tests are described in Methods , heading "Statistical testing".

   a. If there is a section summarizing the statistical methods in the methods, is the statistical test for each experiment clearly defined?

      Methods , heading "Statistical testing".

   b. Do the data meet the assumptions of the specific statistical test you chose (e.g. normality for a parametric test)?

      Where is this described (section, paragraph #)?

      We have not tested assumptions of statistical test on our data. Our tests on RNA counts use NB GLM methods with explicitly modeled variance; test on other types of data mostly use robust semi-parametric methods on approximately variance-stabilized data.

   c. Is there any estimate of variance within each group of data?

      Is the variance similar between groups that are being statistically compared?

      Where is this described (section, paragraph #)?

      See Methods, heading "Statistical testing". For RNA counts we use methods that explicitly model variance differences between groups; for non-count data we use approximate variance-stabilizing transformations.

   d. Are tests specified as one- or two-sided?

      Methods, heading "Statistical testing": association tests are two-sided; enrichment tests are one-sided.

   e. Are there adjustments for multiple comparisons?

      Yes (Methods, heading "Statistical testing").

3. Are criteria for excluding data points reported?

   Was this criterion established prior to data collection?

   Where is this described (section, paragraph #)?

   Yes, Methods, heading "Data preprocessing"; reference 64.

4. Define the method of randomization used to assign subjects (or samples) to the experimental groups and to collect and process data. If no randomization was used, state so. Where does this appear (section, paragraph #)?	NA
5. Is a statement of the extent to which investigator knew the group allocation during the experiment and in assessing outcome included? If no blinding was done, state so. Where (section, paragraph #)?	NA
6. For experiments in live vertebrates, is a statement of compliance with ethical guidelines/regulations included? Where (section, paragraph #)?	Yes, Methods, heading "Ethical statement"
7. Is the species of the animals used reported? Where (section, paragraph #)?	Yes, Methods, heading "Animal breeding and husbandry"
8. Is the strain of the animals (including background strains of KO/ transgenic animals used) reported? Where (section, paragraph #)?	Yes, Methods, heading "Animal breeding and husbandry"
9. Is the sex of the animals/subjects used reported? Where (section, paragraph #)?	Methods, heading "Tissue selection"; Supplementary Table 18
10. Is the age of the animals/subjects reported? Where (section, paragraph #)?	Yes
11. For animals housed in a vivarium, is the light/dark cycle reported? Where (section, paragraph #)?	Yes, Methods, heading "Animal breeding and husbandry", paragraph 4
12. For animals housed in a vivarium, is the housing group (i.e. number of animals per cage) reported? Where (section, paragraph #)?	Yes, Methods, heading "Animal breeding and husbandry"
13. For behavioral experiments, is the time of day reported (e.g. light or dark cycle)? Where (section, paragraph #)?	NA
14. Is the previous history of the animals/subjects (e.g. prior drug administration, surgery, behavioral testing) reported? Where (section, paragraph #)?	NA

a. If multiple behavioral tests were conducted in the same group of animals, is this reported?

Where (section, paragraph #)?

15. If any animals/subjects were excluded from analysis, is this reported?

Where (section, paragraph #)?

Yes, Methods, headings "Animal breeding and husbandry" and "Data pre-processing".

a. How were the criteria for exclusion defined?

Where is this described (section, paragraph #)?

Methods, heading "Data preprocessing"; reference 64

b. Specify reasons for any discrepancy between the number of animals at the beginning and end of the study.

Where is this described (section, paragraph #)?

Only the aforementioned outlier removal

Reagents

1. Have antibodies been validated for use in the system under study (assay and species)?

NA

a. Is antibody catalog number given?

Where does this appear (section, paragraph #)?

b. Where were the validation data reported (citation, supplementary information, Antibodypedia)?

Where does this appear (section, paragraph #)?

2. Cell line identity

NA

a. Are any cell lines used in this paper listed in the database of commonly misidentified cell lines maintained by ICLAC and NCBI Biosample?

Where (section, paragraph #)?

b. If yes, include in the Methods section a scientific justification of their use--indicate here in which section and paragraph the justification can be found.

c. For each cell line, include in the Methods section a statement that specifies:

- the source of the cell lines
- have the cell lines been authenticated? If so, by which method?
- have the cell lines been tested for mycoplasma contamination?

Where (section, paragraph #)?

Data deposition

Data deposition in a public repository is mandatory for:

a. Protein, DNA and RNA sequences
b. Macromolecular structures
c. Crystallographic data for small molecules
d. Microarray data

Deposition is strongly recommended for many other datasets for which structured public repositories exist; more details on our data policy are available here. We encourage the provision of other source data in supplementary information or in unstructured repositories such as Figshare and Dryad.

We encourage publication of Data Descriptors (see Scientific Data) to maximize data reuse.

1. Are accession codes for deposit dates provided?

   Where (section, paragraph #)?

Yes, Methods, heading "Accession number"

Computer code/software

Any custom algorithm/software that is central to the methods must be supplied by the authors in a usable and readable form for readers at the time of publication. However, referees may ask for this information at any time during the review process.

1. Identify all custom software or scripts that were required to conduct the study and where in the procedures each was used.

Most of our analyses were performed in R and involve custom-created R scripts. We have not used any custom algorithms except for our own implementation of previously-described algorithms; our analysis scripts rely on functions from published and publicly available R packages.

2. If computer code was used to generate results that are central to the paper's conclusions, include a statement in the Methods section under "**Code availability**" to indicate whether and how the code can be accessed. Include version information as necessary and any restrictions on availability.

We have included a code availability statement under heading "Code availability".

Human subjects

1. Which IRB approved the protocol?

   Where is this stated (section, paragraph #)?

NA

2. Is demographic information on all subjects provided?

   Where (section, paragraph #)?

3. Is the number of human subjects, their age and sex clearly defined?

   Where (section, paragraph #)?

4. Are the inclusion and exclusion criteria (if any) clearly specified?

   Where (section, paragraph #)?

5. How well were the groups matched?

   Where is this information described (section, paragraph #)?

6. Is a statement included confirming that informed consent was obtained from all subjects?

   Where (section, paragraph #)?

7. For publication of patient photos, is a statement included confirming that consent to publish was obtained?

   Where (section, paragraph #)?

fMRI studies

For papers reporting functional imaging (fMRI) results please ensure that these minimal reporting guidelines are met and that all this information is clearly provided in the methods:

1. Were any subjects scanned but then rejected for the analysis after the data was collected? NA

   a. If yes, is the number rejected and reasons for rejection described?

      Where (section, paragraph #)?

2. Is the number of blocks, trials or experimental units per session and/or subjects specified?

   Where (section, paragraph #)?

3. Is the length of each trial and interval between trials specified?

4. Is a blocked, event-related, or mixed design being used? If applicable, please specify the block length or how the event-related or mixed design was optimized.

5. Is the task design clearly described?

   Where (section, paragraph #)?

6. How was behavioral performance measured?

7. Is an ANOVA or factorial design being used?

8. For data acquisition, is a whole brain scan used?

   If not, state area of acquisition.

   a. How was this region determined?

9. Is the field strength (in Tesla) of the MRI system stated?

    a. Is the pulse sequence type (gradient/spin echo, EPI/spiral) stated?

    b. Are the field-of-view, matrix size, slice thickness, and TE/TR/flip angle clearly stated?

10. Are the software and specific parameters (model/functions, smoothing kernel size if applicable, etc.) used for data processing and pre-processing clearly stated?

11. Is the coordinate space for the anatomical/functional imaging data clearly defined as subject/native space or standardized stereotaxic space, e.g., original Talairach, MNI305, ICBM152, etc? Where (section, paragraph #)?

12. If there was data normalization/standardization to a specific space template, are the type of transformation (linear vs. nonlinear) used and image types being transformed clearly described? Where (section, paragraph #)?

13. How were anatomical locations determined, e.g., via an automated labeling algorithm (AAL), standardized coordinate database (Talairach daemon), probabilistic atlases, etc.?

14. Were any additional regressors (behavioral covariates, motion etc) used?

15. Is the contrast construction clearly defined?

16. Is a mixed/random effects or fixed inference used?

    a. If fixed effects inference used, is this justified?

17. Were repeated measures used (multiple measurements per subject)?

    a. If so, are the method to account for within subject correlation and the assumptions made about variance clearly stated?

18. If the threshold used for inference and visualization in figures varies, is this clearly stated?

19. Are statistical inferences corrected for multiple comparisons?

    a. If not, is this labeled as uncorrected?

20. Are the results based on an ROI (region of interest) analysis?

    a. If so, is the rationale clearly described?

    b. How were the ROI's defined (functional vs anatomical localization)?

21. Is there correction for multiple comparisons within each voxel?

22. For cluster-wise significance, is the cluster-defining threshold and the corrected significance level defined?

Additional comments

Additional Comments