nature research

Corresponding author(s): Timothy E. Reddy

Last updated by author(s): Dec 14, 2018

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ A description of all covariates tested |
| ☒ | ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | We used standard Illumina software to demultiplex and generate FASTQs. |
| Data analysis | Alignments and region calling<br>STARR-seq input library and output libraries were individually aligned to the human genome assembly hg38 with Bowtie2 (version 2.2.4), using the following parameters: bowtie2 -X 2000 --sensitive. Only properly-paired alignments with a MAPQ score ≥ 30 outside hg38 centromeres, gap and blacklist regions were retained in downstream analyses. Regions were called individually for each sample using merged STARR-seq input alignments as controls with the MACS2 package using the following parameters: -f BAMPE -g hs --ratio --keep-dup all -q 0.10. The custom scaling ratio (--ratio) provided to MACS was generated for each sample using the NCIS algorithm. We generated a union set of called regions after merging any overlapping regions with bedtools (v.2.25.0). We tested for differential STARR-seq activity across the union region set by fitting negative binomial models and performing quasi-likelihood F-tests with edgeR (version 3.8.6). RPKM normalized STARR-seq read density was computed at single bp resolution using deepTools utility bamCoverage.<br><br>Comparisons to other genomic datasets<br>We extensively compared our whole genome reporter assay results to Hi-C, RNA-seq, DNase-seq, ATAC-seq, and ChIP-seq datasets from a time course study of the same dex response in A549 cells. Differentially expressed genes and topological associated domains were previously identified. Analysis of ChIP-seq, ATAC-seq and DNase-seq signals were performed using the bwtools package. ChIP-seq read density (RPKM) values represent input control RPKM subtracted values, truncating the difference as zero. Post-dex averages of genomic signals represent the mean across the full dex timecourse as reported in reference #21 (0.5, 1, 2, 3, 4, 5, 6, 7, 8, 10, and 12 hr of dex treatment). ChIP-seq peak set intersections were visualized using the UpSetR package.<br><br>Motif enrichment analysis<br>We tested the set of eight-hour DREs for enrichment of RSAT-clustered JASPAR motifs. To do so we first binned the dynamic regulatory elements by the magnitude of their response and controlled for baseline activity. Relative motif enrichment analysis was performed with the AME tool from the MEME suite using randomly generated dinucleotide shuffled sequences as a comparator. Motif enrichment was similarly performed on steady-state regulatory elements after separating elements into quintiles according to reporter activity. Motif |

analysis of the rs10505411 alleles was performed with the MAST tool using a first-order Markov background model.

Allele-specific dex-responsive regulatory activity analysis
To identify allele-specific dex-responsive regulatory activity we used the mpileup tool in the samtools package to identify alignments overlapping heterozygous SNPs in the genome of GM12878 cells. We further parsed variants requiring that they overlapped a regulatory region identified in this study and were covered by at least five fragments in at least five samples. In total we tested 10,669 heterozygous alleles for differential regulatory activity in response to dex using a negative binomial model model implemented with the edgeR that included main effect and interaction terms for treatment and time. The DSS package was used to estimate dispersion.

Allele-specific ChIP analysis
We used bowtie to perform allele-specific analysis of GR, EP300 and cJUN ChIP-seq reads. Reads were aligned (bowtie -v 0 -m 1) to either the reference or alternate alleles of rs10505411, requiring no mismatches or multiple matches.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Whole genome STARR-seq sequencing data are publicly available through the National Center for Biotechnology Information Gene Expression Omnibus (Series GSE114063). Published ChIP-seq, RNA-seq, ATAC-seq, and DNase-seq as well as Hi-C datasets are available through the NCBI BioProject (PRJNA356880) and the ENCODE DCC portal (http://www.encodeproject.org).

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Several pilot studies were conducted to determine the number of cultured cells needed to generate a high quality STARR-seq library. We then collected five samples per replicate time point for a total of 25 samples. This was limit of our tissue culture capacity. |
| Data exclusions | No data was excluded |
| Replication | Sample replication is extensively address in our supplemental analyses. In example, sequencing coverage of peak calls in replicate samples was highly correlated. |
| Randomization | To control for batch effects, samples were randomly grouped prior to constructing sequencing libraries. These assignments were controlled for during differential analysis and no effect was observed. |
| Blinding | Blinding was not performed because we carried out a time course experiment. Blinding would confound our interpretation of the progression of effects. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☐ | ☐ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☐ | ☐ Palaeontology |
| ☐ | ☐ Animals and other organisms |
| ☐ | ☐ Human research participants |
| ☐ | ☐ Clinical data |

## Methods

| n/a | Involved in the study |
|---|---|
| ☐ | ☐ ChIP-seq |
| ☐ | ☐ Flow cytometry |
| ☐ | ☐ MRI-based neuroimaging |

# Antibodies

Antibodies used — *Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*

Validation — *Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

# Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) — Human A549 cells was obtained from Duke University Cell Culture Facility (obtained ultimately from ATCC)

Authentication — The seed cell line culture was not authenticated

Mycoplasma contamination — The seed cell line culture was not tested for mycoplasma

Commonly misidentified lines (See ICLAC register) — None

# Palaeontology

Specimen provenance — *Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*

Specimen deposition — *Indicate where the specimens have been deposited to permit free access by other researchers.*

Dating methods — *If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

# Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals — *For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*

Wild animals — *Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*

Field-collected samples — *For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*

Ethics oversight — *Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Human research participants

Policy information about studies involving human research participants

Population characteristics — *Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic*

| | |
|---|---|
| Population characteristics | *information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."* |
| Recruitment | *Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.* |
| Ethics oversight | *Identify the organization(s) that approved the study protocol.* |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

| | |
|---|---|
| Clinical trial registration | *Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.* |
| Study protocol | *Note where the full trial protocol can be accessed OR if not available, explain why.* |
| Data collection | *Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.* |
| Outcomes | *Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.* |

## ChIP-seq

### Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

| | |
|---|---|
| Data access links<br>*May remain private before publication.* | *For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.* |
| Files in database submission | *Provide a list of all files available in the database submission.* |
| Genome browser session<br>(e.g. UCSC) | *Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.* |

### Methodology

| | |
|---|---|
| Replicates | *Describe the experimental replicates, specifying number, type and replicate agreement.* |
| Sequencing depth | *Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.* |
| Antibodies | *Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.* |
| Peak calling parameters | *Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.* |
| Data quality | *Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.* |
| Software | *Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.* |

## Flow Cytometry

### Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| | |
|---|---|
| Sample preparation | *Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.* |
| Instrument | *Identify the instrument used for data collection, specifying make and model number.* |
| Software | *Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.* |
| Cell population abundance | *Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.* |
| Gating strategy | *Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.* |

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

# Magnetic resonance imaging

## Experimental design

| | |
|---|---|
| Design type | *Indicate task or resting state; event-related or block design.* |
| Design specifications | *Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.* |
| Behavioral performance measures | *State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).* |

## Acquisition

| | |
|---|---|
| Imaging type(s) | *Specify: functional, structural, diffusion, perfusion.* |
| Field strength | *Specify in Tesla* |
| Sequence & imaging parameters | *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.* |
| Area of acquisition | *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.* |

Diffusion MRI ☐ Used ☐ Not used

## Preprocessing

| | |
|---|---|
| Preprocessing software | *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).* |
| Normalization | *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.* |
| Normalization template | *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.* |
| Noise and artifact removal | *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).* |
| Volume censoring | *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.* |

## Statistical modeling & inference

| | |
|---|---|
| Model type and settings | *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).* |
| Effect(s) tested | *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.* |

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

| | |
|---|---|
| Statistic type for inference (See Eklund et al. 2016) | *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.* |

Correction

*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

## Models & analysis

| n/a | Involved in the study |
|---|---|
| ☐ | ☐ Functional and/or effective connectivity |
| ☐ | ☐ Graph analysis |
| ☐ | ☐ Multivariate modeling or predictive analysis |

Functional and/or effective connectivity

*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis

*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis

*Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*