nature research

Corresponding author(s): Rickard Sandberg

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

   Describe how sample size was determined.

   The cell numbers were sufficient for the inference of kinetic parameters for most genes as we obtained relatively small confidence intervals. Identification of significant differences between genotypes and cell types had good power for 4-fold differences in the most densely populated areas in parameter space.

2. Data exclusions

   Describe any data exclusions.

   We excluded sequencing libraries that had low technical qualities, i.e. either a low percentage of reads mapping uniquely to the genome or a low percentage of reads aligning within genes, as detailed in the methods section. These criteria are standard although exact thresholds depend on the cell types investigated. It is important to remove faulty sequence libraries as these otherwise would lead to the underestimation of burst frequencies.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   The general inference of kinetic parameters were done on several datasets, including fibroblasts sequenced from different members of the lab. We also determine that the two genotypes generally agree. Important findings (e.g. that core promoter element dictate burst size) was demonstrated using both genotypes, and that enhancers dictate burst frequencies were demonstrated using both SNP densities in enhancers and by CRISPR-mediated enhancer deletion. All the findings above were also replicated at both RPKM and molecule-level of Smart-seq2 data. Altogether, all replications were successful.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   The cells were all F1 offspring of fully inbred mouse strains. Analyzed cells were grouped by their type (fibroblasts and mESCs) and by the presence or absence of the Sox2-enhancer. Samples were not randomized across experiments.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Blinding is not applicable to the experiments made (single-cell analyses of defined cell types and genotypes).

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Code for inference and the analysis performed within this study will be deposited at https://github.com/sandberg-lab/txburst. We additionally used Python3, scphaser (https://github.com/edsgard/scphaser), liftOver and MACS (v 1.4).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Segmentation of cells was facilitated by adding Vimentin antibody to tail fibroblasts (ThermoFisher Scientific, Cat# MA5-11883-A647, 1:50 dilution, Alexa Fluor 647) or SSEA-1 antibody to embryonic fibroblasts (BD Pharmingen, Cat#: 560120, 1:20 dilution, Alexa Fluor 647). These antibodies are reported by the manufacturers as suitable to characterize the two respective cell types in mouse.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	The F1X S129/SvJAE Sox2 distal enhancer CRISPR deletion cell line was generated in the Ren lab.
b. Describe the method of cell line authentication used.	The Ren Lab developed the cell line. They have the gene expression patterns described for mouse embryonic stem cells.
c. Report whether the cell lines were tested for mycoplasma contamination.	The CASTx129SvEv cell lines (with and without Sox2 enhancer deletion) was not tested for Mycoplasma. For primary male fibroblasts and mESCs (both of C57 and CAST genotype) DAPI staining during single-molecule RNA FISH confirmed their negative status.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No commonly misidentified cell line were used.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

Primary tail fibroblasts were derived from adult male and female CAST/EiJ x C57BL/6J or C57BL/6J x CAST/EiJ mice. mESCs were derived from pre-implantation embryonic development for male CAST/EiJ x C57BL/6J or C57BL/6J x CAST/EiJ mice.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study did not involve human research participants.