nature research

Corresponding author(s): Liana F Lareau

☐ Initial submission ☐ Revised version ☒ Final submission

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.

Experimental design

1. Sample size

   Describe how sample size was determined.

   We grew up strains from 8 yeast colonies for each version of eCitrine, representing 8 independent integrations of the eCitrine reporter into a fixed locus.

2. Data exclusions

   Describe any data exclusions.

   qPCR: We measured mRNA levels in three biological replicates (independent integrations of the eCitrine reporter) for each version of eCitrine. One integrant of eCitrineMAX was excluded because the RNA pellet was aspirated during the experiment, leaving two biological replicates of that strain.
   Flow cytometry: Biological replicates of yeast strains were excluded for the following reasons, documented in our analysis scripts on GitHub:
   1. eCitrine333: we discovered a polymorphism in the donor plasmid and tested each integrant for the mutation. We discarded 4 of 8 integrants because they were not wildtype at that position.
   2. eCitrine000, eCitrine333, yECitrine: one integrant of each eCitrine variant was discarded because the FACS data did not pass pre-determined filters for number of cells or had bimodal fluorescence.
   3. eCitrine999: one integrant was discarded because genomic copy number qPCR gave results inconsistent with a single copy integration.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   All attempts at replication were successful.

4. Randomization

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

   For neural network training, genes were randomly allocated into the training set (2/3) or test set (1/3).

5. Blinding

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

   Blinding was not relevant to our study.

Note: all studies involving animals and/or human research participants must disclose 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 (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals 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

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.

All iXnos software, transcriptome files used for alignments, and a complete script to reproduce our entire analysis including figures are available at: https://github.com/lareaulab/iXnos
Linkers were trimmed from ribosome footprints with custom code as documented in the paper and on GitHub.
Alignments were performed with bowtie v. 1.2.1.1 and RSEM v. 1.2.31.
The neural network method, iXnos, was implemented in python and is fully documented in the paper and on GitHub. It relied on python packages Lasagne v. 0.2.dev1 and Theano v. 0.9.0.
Flow cytometry was analyzed with custom R scripts fully documented on GitHub. They incorporated bioconductor packages flowCore v. 1.40.3, flowStats v. 3.32.0, and flowViz v. 1.38.0.
All methods used to create the figures, including statistics and qPCR analsysis, are included as R scripts on GitHub.

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 for-profit company.

All materials and strains are readily available. Yeast strains will be provided on request. Commercial sources:
Direct-zol RNA MiniPrep kit (Zymo Research)
CircLigase I and II (Epicentre)
DyNAmo HS SYBR Green qPCR Kit (Thermo Scientific)
GoTaq DNA polymerase (Promega)
Turbo DNase I (Ambion)
Protoscript II (NEB)

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

S. cerevisiae strains were constructed from frozen stocks of BY4741 and BY4742 obtained from Thermo Fisher.

b. Describe the method of cell line authentication used.

Mating type and marker genotype were authenticated by growth on selective plates.

c. Report whether the cell lines were tested for mycoplasma contamination.

Yeast are not susceptible to mycoplasma contamination.

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 lines 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 details on animals and/or animal-derived materials used in the study.

No animals were used.

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.

The study did not involve human research participants.