

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 our [Editorial Policies](#) 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
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ 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
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ 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

The qPCR analysis software used for Figures 1b, 1d, 1e, and Supplementary Table 1 was Bio-Rad CFX Maestro 4.1.2434.0124. The software used for Figures 2, 3e, Supplementary Figures 1, 2b, 3, 4b, and Supplementary Table 2 was Fragment Analyzer Version 1.2.0.11 and Prosize 3.0 version 3.0.1.5.

Data analysis

The software algorithms we developed to perform the reported analyses for Figure 3b, 3c, 3f, 3g, and 3i are available at <https://doi.org/10.5281/zenodo.4747693> and <https://github.com/dna-storage/ncomm-file-preview> under a permissive open-source license with instructions for installation. We implemented code in python using many standard open-source packages tested for compatibility with python versions 3.6 to 3.9. The dependences are documented in the form of a python requirements.txt file that guides installation of additional dependent software packages. NUPACK 3.0 was used to develop the primer hybridization model. We used python-Levenshtein 0.12.0 for edit distance calculations. The open-source sequencing data clustering software Starcode Algorithm was used to aid in the process of determining read counts for strands. The version used in analysis is the master branch that can be accessed at <https://github.com/gui11aume/starcode>. A docker file is available to make setup on a wide variety of systems easier.

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 and 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

The source data for the figures presented in this manuscript and the Supplementary Information are available in the Source Data file provided with this paper and at <https://github.com/dna-storage/ncomm-file-preview/releases/tag/v0.1-alpha> along with the next-generation sequencing data for Figures 3b, 3c, 3f, 3g, 3h, and Supplementary figures 3c and 4b. Any other data are available upon reasonable request.

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](https://www.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	Sample sizes were chosen based on previous experience and precedent set in the field of DNA-based data storage. Triplicate sample sizes were chosen for select experiments by default and found in early experiments to be sufficient given the large and nearly binary effect sizes in this work.
Data exclusions	No data were excluded.
Replication	Experimental reactions resulting in data shown in Figures 3e-h and Supplementary Figures 3b, 3c, and 4b were replicated three times. All replications successfully showed the same effect.
Randomization	N/A as samples were all in vitro DNA samples.
Blinding	N/A as samples were all in vitro DNA samples.

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 and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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