eLife's transparent reporting form

We encourage authors to provide detailed information *within their submission* to facilitate the interpretation and replication of experiments. Authors can upload supporting documentation to indicate the use of appropriate reporting guidelines for health-related research (see EQUATOR Network), life science research (see the BioSharing Information Resource), or the ARRIVE guidelines for reporting work involving animal research. Where applicable, authors should refer to any relevant reporting standards documents in this form.

If you have any questions, please consult our Journal Policies and/or contact us: editorial@elifesciences.org.

Sample-size estimation
- You should state whether an appropriate sample size was computed when the study was being designed
- You should state the statistical method of sample size computation and any required assumptions
- If no explicit power analysis was used, you should describe how you decided what sample (replicate) size (number) to use

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn't apply to your submission:

Sample sizes for DNN model training is found in the Result section. All figures include a label "n=1234" where relevant. For training and evaluation of the DNN, if not otherwise specified we used the full 150.000 trace dataset.
No explicit method was used for sample size computation. The data sets of ~150.000 traces should be large enough to cover all behaviors expected in FRET traces. See Methods, Synthetic smFRET data generation

Replicates
- You should report how often each experiment was performed
- You should include a definition of biological versus technical replication
- The data obtained should be provided and sufficient information should be provided to indicate the number of independent biological and/or technical replicates
- If you encountered any outliers, you should describe how these were handled
- Criteria for exclusion/inclusion of data should be clearly stated
- High-throughput sequence data should be uploaded before submission, with a private link for reviewers provided (these are available from both GEO and ArrayExpress)

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn't apply to your submission:

Replicability was ensured by reporting all parameters (See Methods, Synthetic smFRET data generation, of Fig. S6).

For comparison of human selection of traces in Fig. S15 and Methods 3 technical replicates were performed. 1000 ground traces were simulated out of which 46 were FRET . Human participants were not informed on underslyign distribution or fraction of correct traces.

All data simulated and extracted are used as is. No outliers are excluded from analysis.

eLIFE

1st Floor
24 Hills Road
Cambridge CB2 1JP, UK

P 01223 855340
W elifesciences.org
T @elife

Statistical reporting

- Statistical analysis methods should be described and justified
- Raw data should be presented in figures whenever informative to do so (typically when N per group is less than 10)
- For each experiment, you should identify the statistical tests used, exact values of N, definitions of center, methods of multiple test correction, and dispersion and precision measures (e.g., mean, median, SD, SEM, confidence intervals; and, for the major substantive results, a measure of effect size (e.g., Pearson's r, Cohen's d)
- Report exact p-values wherever possible alongside the summary statistics and 95% confidence intervals. These should be reported for all key questions and not only when the p-value is less than 0.05.

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn't apply to your submission:

Raw data, plotting, and statistical nomenclature used is extensively discussed in the Methods section. Classification metrics (precision, recall) are reported in figure labels with True/False Positive/Negative rates when relevant.

(For large datasets, or papers with a very large number of statistical tests, you may upload a single table file with tests, Ns, etc., with reference to sections in the manuscript.)

Group allocation

- Indicate how samples were allocated into experimental groups (in the case of clinical studies, please specify allocation to treatment method); if randomization was used, please also state if restricted randomization was applied
- Indicate if masking was used during group allocation, data collection and/or data analysis

Please outline where this information can be found within the submission (e.g., sections or figure legends), or explain why this information doesn't apply to your submission:

We used a classical 80/20 split for validation of our data during training and validation of the DNN, as discussed in Methods > **Synthetic smFRET data generation.**

Additional data files ("source data")

- We encourage you to upload relevant additional data files, such as numerical data that are represented as a graph in a figure, or as a summary table
- Where provided, these should be in the most useful format, and they can be uploaded as "Source data" files linked to a main figure or table
- Include model definition files including the full list of parameters used
- Include code used for data analysis (e.g., R, MatLab)
- Avoid stating that data files are "available upon request"

Please indicate the figures or tables for which source data files have been provided:

Data used producing plots are provided when relevant.
Code for producing all plots as well as a link to the datafiles used can be found at
https://github.com/hatzakislab/DeepFRET-plots