

Deep learning based local feature classification to automatically identify single molecule fluorescence events

Corresponding Author: Dr Chunlai Chen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors developed and trained neural networks to make fragmented categories of fluorescence time traces. Their neural networks were established and tested on one- and two-color time trajectories. They compared the results from their neural network determined using ~1500 time traces with that two expert users who analyzed the data manually. They demonstrate various applications of their deep-learning analysis approach on five different experimental systems. Although potentially interesting, there are various issues that need to be addressed before the paper can be considered for publication.

Major points

As the algorithm was trained using simulated data, the authors should also test the algorithm on additional simulated traces that the network has not yet seen. In this case, the ground truth is known and an appropriate confusion matrix can be determined. The comparison with expert uses is also very important. However, for experimental data, there is no ground truth. The confusion plot should be added to Figure 2.

The algorithm currently does not include additional information that would be available from experiments performed using Alternating Laser EXcitation (ALEX) (Kapanidis et al, 2005, Accounts of chemical research, 38(7), 523-533). Hence, their program cannot distinguish between molecules where both dyes are present and photoactive but the complex does not undergo FRET with molecules where the acceptor is not photoactive. The authors should at least mention this limitation to their software.

In analyzing FRET data, determination of the correction factors like spectral crosstalk, direct excitation of acceptor and the detection efficiency (gamma factor) are often desired. Considering their ability to detect photobleaching steps, can their approach be utilized to determine the different correction factors?

What are the advantages of DEBRIS compared to other published networks, for instance in Deep-FRET and Deep-LASI? Many assays use fluorescence quenching, which can lead to dynamic single-color data. Is there software capable of analyzing dynamic single-color traces? Figure 4a, middle panel could, theoretically, be a dynamic trace that is then discarded.

The authors mentioned user-defined criteria when using the neural networks specifically for their data. Could the authors elaborate about these criteria, at least in the material and methods? To what extent the user can adjust the neural network's performance or confidence levels?

Page 3: "or variable intensities with anti-correlation behaviors, as their total intensities remain close to constant over time". The latter criteria can be a problem when the experimental system has differing sensitivities to the donor and acceptor dyes. Can the authors address this concern? Can one include the various correction factors and loosen the criteria so that this does not lead to throwing out good traces?

Page 4: The authors include data from AsCas12a, FnCas12a and LbCas12a. For non experts, the authors should introduce these different systems.

Figures: Since the authors are referring to five different experimental systems in this paper, they should add a simple schematic of the different systems to help readers follow the various assays easier at the appropriate positions in the different figures.

Minor points

Many experimental details are missing from the paper. The frame rate of data collection, laser powers used, etc.

Page 2: "models trained using equilibrium datasets are unsuitable for the single-molecule traces taken under non-equilibrium conditions" This statement is not always true. Some of the previously mentioned software packages would be

able to analysis this type of dynamics. Please reword.

Page 3: "which closely matches the reported value of 100 ± 3 nM." The two values do not overlap within the error bars. Please reword to something such as 'which is similar to ...'

Figure 3d: There is a low/no FRET peak showing up around 5 s after binding for the LbCas12a and FnCas12a systems. The authors should at least mention this peak and what it is due to.

Figure S8. Panels a, b: There appears to be a problem here. In panel a, categorization of donor appearance and disappearance correlates with what is observed in the smFRET trace. In panel b, the categorization is shifted by several seconds from the event in the smFRET trace.

Figure S11, Panel a: It is not clear from the figure (nor from the text) how the SNR was determined. Please correct.

Figure S13: Figure is confusing. According to the schematic (panel a), the DNA is cleaved from the surface. In panel b, multiple binding events are observed. Also, what is labeled as "appearance time" in panel b is the time between binding events, not the time from injection until binding (with the exception of the first event), which is the definition given in the text. The authors talk about "bleaching". This should be referred to as "photobleaching".

Reviewer #2

(Remarks to the Author)

In their submitted manuscript, Zhou et al. present a deep-learning model to classify single molecule FRET trajectories by a sliding window classification. The model is suitable for single color and two color FRET trajectories. It is worth mentioning, that the model only works for single color FRET experiments and does not support the acceptor excitation and acceptor emission analysis from alternating laser excitation experiments like other neural network models do. New is the feature of recognizing emerging fluorescence events in an experiments do to binding events. The authors compare the classification of experimental trajectories of their model to the evaluation of experts.

The manuscript is well written, the classification criteria of the model are explained in detail and sound. The software is published and a manual is available which makes the software easy to use in other labs. I suggest the publication in Communications Biology after addressing my questions:

1. The authors test their model on experimental data. The assays are only depicted in the SI but single molecule trajectories are shown in the main manuscript. A sketch of the assays (holliday junction, Cas12 and PRE) is missing in all figures. I highly emphasize to add a sketches of each assays in order to make the interpretation of the trajectories easier.
2. What is the advantage of the sliding time-window approach? I guess the model can be faster trained as only single features (e.g. acceptor bleaching) must be recognized and not combinations of them. If so, please state it in the manuscript.
3. Why did the authors choose the sliding time-window length to be 10 frames?
4. Why was the model trained for single color excitation? The alternating laser excitation approach offers additional filter criteria. ALEX excitation gives access to the dye-stoichiometry and monitors the acceptor fluorescence to detect photophysics.
5. The enzyme concentrations (glucose oxidase catalase) are usually given in units per mL and not μg per mL as the activity of every batch is different.

Reviewer #3

(Remarks to the Author)

The manuscript is exceptionally well-written, with clear and concise language throughout. The quality of the figures is notably high, effectively enhancing the presentation of the research findings. Additionally, the text is well-organized and structured, contributing to a seamless reading experience. However, there are technical and general issues that need to be clarified before publication.

The authors introduce and apply a neural network, DEBRIS, for single molecule trace analysis in MATLAB. They provide MATLAB scripts for generating training data and classifying frames in single molecule traces. The neural network automates the classification of regions in these traces, significantly simplifying the analysis process typically performed manually by experts. Processing experiments, including FRET, Holliday junction, Cas complex, and beta2AR analysis. They compared the results obtained from their neural network (NN) to those derived from expert analysis, demonstrating the effectiveness and reliability of their automated approach very similar to the more comprehensive work previously published (Deep-LASI).

As noted by the authors, several neural networks and software tools perform similar tasks in single-molecule trace analysis. Examples include DeepFRET, AutoSIM, and Deep-LASI. In my comparison of Deep-LASI and AutoSIM with the authors' MATLAB scripts, the existing software generally offers more comprehensive functionality. For instance, Deep-LASI can read images to create traces within a single software suite. The authors' scripts primarily focus on the classification task within single molecule traces. While they mention a custom ImageJ plugin, the separation of tasks for trace computation and trace analysis is appreciated. However, the authors do not provide the software for these tasks. Could the authors provide raw images and software (full stack) to process image traces along with a manual? Providing a full analysis pipeline could significantly increase the impact of their work.

The authors claim that their neural network (DEBRIS) is less opaque compared to published systems for neural network-based FRET and single molecule trace analysis. However, I did not find evidence to support this claim, as other software also allows for training. While the authors developed DEBRIS and naturally view it as less opaque, an outsider's perspective suggests that DEBRIS is similarly opaque compared to other software like AutoSIM, which also provides training capabilities. I cannot follow this claim, as the authors' software is similarly opaque compared to other available software.

The authors claim, "DEBRIS is the first method to achieve automatic classification of dynamically emerging events by classifying short-fragment traces instead of long-duration ones." Please specify this claim. While DEBRIS works on short fragment traces, it is unclear how this is better compared to other software. Did you compare your approach to other existing approaches? If so, please present the results of your comparison. Additionally, how does the trace length impact your classification? When using your scripts, I only found a "final" classification result. Can you provide a confidence measure for the classification? How does the confidence depend on the trace length?

Technical Issues:

The program requires MATLAB dependencies that are not included in a standard MATLAB installation. Please list the required toolboxes and packages in your installation manual. Additionally, I require the following toolboxes to use your programs:

- Signal Processing Toolbox
- Deep Learning Toolbox
- Parallel Computing Toolbox(TM)

File: Netv230712_Training.mlx in Chinese

Missing example data for three-color FRET.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have done a commendable job in addressing the majority of the reviewers' comments. There are a couple minor corrections I would suggest, upon which the paper would be acceptable for publication.

1) The intent of asking for schematic drawings of the different systems was to add them next to the data (i.e. in Figure 2c or 2d, Figure 3d, and Figure 4c and 4e).

2) I am still unclear about the details of the assay used in Figure 4c. I believe DNA is attached to Cy3, but what is it binding too? Please explain.

3) The authors were asked to compare their programs with others in the field. This they have nicely done in the rebuttal (although the accuracy calculations are not always correct (e.g. Figure R1b, HJ-low salt)). I fully understand that one needs to be careful when discussing other people's work in a publication. Still, I would recommend mentioning that DEBRIS was also tested against AutoSIM and DeepFRET and found that DEBRIS performed better, in particular when dealing with traces that have emerging dynamics, for which DEBRIS was designed. (Certainly, the performed test are not a complete benchmarking of the various softwares, so the phrasing needs to be done carefully).

Reviewer #2

(Remarks to the Author)

The authors have adequately addressed my comments. As final comment, I suggest inserting the sketches of Figure S4 into the main text figures to facilitate fluent reading of the manuscript. The manuscript should then be ready for publication.

Reviewer #3

(Remarks to the Author)

The authors have thoroughly addressed all the concerns raised in the previous review. The revisions have been carefully considered and implemented, ensuring that all issues are satisfactorily resolved. Given the improvements made, the manuscript is now of high quality and suitable for publication in its current form without the need for further modifications.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>



Response to reviewer 1

Reviewer #1:

Comments for the Authors:

The authors developed and trained neural networks to make fragmented categories of fluorescence time traces. Their neural networks were established and tested on one- and two-color time trajectories. They compared the results from their neural network determined using ~1500 time traces with that two expert users who analyzed the data manually. They demonstrate various applications of their deep-learning analysis approach on five different experimental systems. Although potentially interesting, there are various issues that need to be addressed before the paper can be considered for publication.

Response #1: We sincerely appreciate the reviewer's positive comments and valuable feedback.

Major points

As the algorithm was trained using simulated data, the authors should also test the algorithm on additional simulated traces that the network has not yet seen. In this case, the ground truth is known, and an appropriate confusion matrix can be determined. The comparison with expert uses is also very important. However, for experimental data, there is no ground truth. The confusion plot should be added to Figure 2.

Response #2: We are grateful to the reviewer for this valuable suggestion. To evaluate the efficacy of the neural network, we conducted analysis using 10,000 novel fragments with known ground truths. The classification outcomes have been represented in a confusion matrix, now included in Figure S2. The neural network demonstrated an accuracy of 98% in predicting these fragments. This information has been added to the manuscript on **page 2**, and the confusion matrix is presented in **Figure S2** of the supplementary file.

Page 2: “The trained neural network was tested using newly simulated fragments as the validation dataset, whose prediction accuracy was ~98% (Figure S2).”

Figure S2

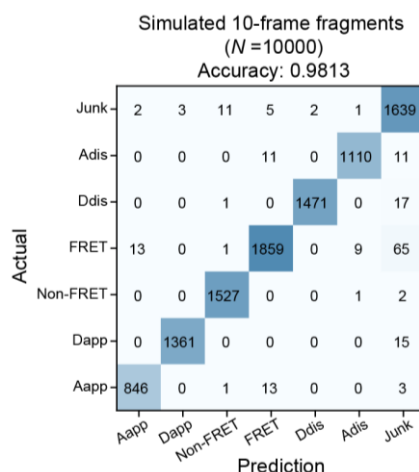


Figure S2 legend: “Confusion matrix of neural network classification based on the ground



truth of simulated 10-frame fragments.”

The algorithm currently does not include additional information that would be available from experiments performed using Alternating Laser EXcitation (ALEX) (Kapanidis et al, 2005, Accounts of chemical research, 38(7), 523-533). Hence, their program cannot distinguish between molecules where both dyes are present and photoactive, but the complex does not undergo FRET with molecules where the acceptor is not photoactive. The authors should at least mention this limitation to their software.

Response #3: We are appreciative of the reviewer's perspicacious recommendation. DEBRIS currently has limitations in handling single-molecule fluorescence experiments using the ALEX mode. We have highlighted this limitation in the **Discussion** section of the manuscript, **page 6**: *“It should be noted that the current version of DEBRIS is trained to analyze two-color fluorescence traces under single laser excitation, which is also fully compatible with one-color traces. However, the deep-learning model needs to be retrained to analyze three or more colors traces and single-molecule traces under Alternating Laser Excitation (ALEX) mode using two or more lasers. Also the current DEBRIS does not account for quenching and protein-induced fluorescence enhancement (PIFE) phenomena. The traces displaying these phenomena were discarded by our stringent criteria and can be selected if criteria are adjusted. To maximize the accessibility of the DEBRIS model, we offer MATLAB source codes for fragment simulation, deep learning network training, and single-molecule fluorescence event identification for four different scenarios presented in our study. Thus, users can optimize the DEBRIS deep learning model or adjust user defined criteria for their unique experimental conditions and designs.”*

In analyzing FRET data, determination of the correction factors like spectral crosstalk, direct excitation of acceptor and the detection efficiency (gamma factor) are often desired. Considering their ability to detect photobleaching steps, can their approach be utilized to determine the different correction factors?

Response #4: We are grateful to the reviewer for this valuable suggestion. To enhance the utility of the DEBRIS classification results, we have replenished a new code, ‘FRET_correction.mat’, for the calculation of the gamma factor and spectral crosstalk based on the classifications. This code will be provided to users and accompanied by detailed instructions on its use in the ‘**User manual of DEBRIS**’ document, **page 17**.

Using the ‘acceptor photobleaching’ category of the steady two-color traces, the gamma factor¹ is defined as the ratio of the change in acceptor intensity to the change in donor intensity around the acceptor photobleaching time point. Spectral crosstalk is defined as the ratio of the acceptor channel intensity to the donor channel intensity after acceptor photobleaching.

In the ‘DEBRIS.mat’ file, the user can input the gamma factor from known value or by roughly estimating the ratio of the intensity changes of the acceptor and donor around the acceptor photobleaching time point or the FRET conversion time point. We want to emphasize that



DEBRIS can tolerate to the variation of the input gamma factor value to a certain extent (Examples shown in the current Fig. S17), which is considered in our deep-learning model. Based on traces classifications, the 'FRET_correction.mat' file provides the calculated gamma factor for the calculated ratio of intensity change around the receptor photobleaching time point. This value can be used to perform subsequent iterations, if needed.

What are the advantages of DEBRIS compared to other published networks, for instance in Deep-FRET and Deep-LASI?

Response #5: We appreciate the reviewer's inquiry regarding the advantages of DEBRIS compared to other published networks such as DeepFRET² and AutoSiM³.

We conducted tests on steady single- and two-color data using DeepFRET and AutoSiM and observed that DEBRIS outperformed both in these scenarios, at least when applied to these datasets. For dynamically emerging signals, DeepFRET and AutoSiM were less effective, as they struggled to accurately discern the traces, making them less capable of identifying emerging events.

To test the efficacy of DeepFRET and AutoSiM, we selected five datasets (Figure 2b) for steady two-color trace test sets and three datasets (Figure 4b) for steady one-color trace test sets. For dynamically emerging two-color signals, we selected 2075 traces from AsCas12a-WT 0MM (Figure 3d) as test set, and 1425 traces from 11 bp dsDNA extension in the PAM downstream (Figure 4e) were selected as test set for dynamically emerging one-color signals. DEBRIS identified 1713 events from the 2075 dynamic two-color traces and 6478 events from the 1425 dynamic one-color traces. Single-color traces were replicated to mimic a two-channel input.

Performance of DeepFRET on steady traces: When applying DeepFRET's recommended thresholds (confidence level = 0.8) to steady traces, the classification accuracy for steady two-color traces, using expert manual labeling (Expert #1, Expert #2) as the ground truth, was approximately 0.5 (**Figure R1a-1b**). For steady single-color traces, the classification accuracy was around 0.43 (**Figure R1c**). Notably, we identified some differences in criteria between DeepFRET and DEBRIS: 1) DeepFRET accepts traces where donor does not undergo photobleaching (**Figure R2a**), whereas DEBRIS and experts require photobleaching of donor. 2) DeepFRET appears to accept traces with abnormal signal changes after acceptor photobleaching (**Figure R2b**), but DEBRIS and experts require steady signals before photobleaching of all fluorophores (both acceptor and donor).

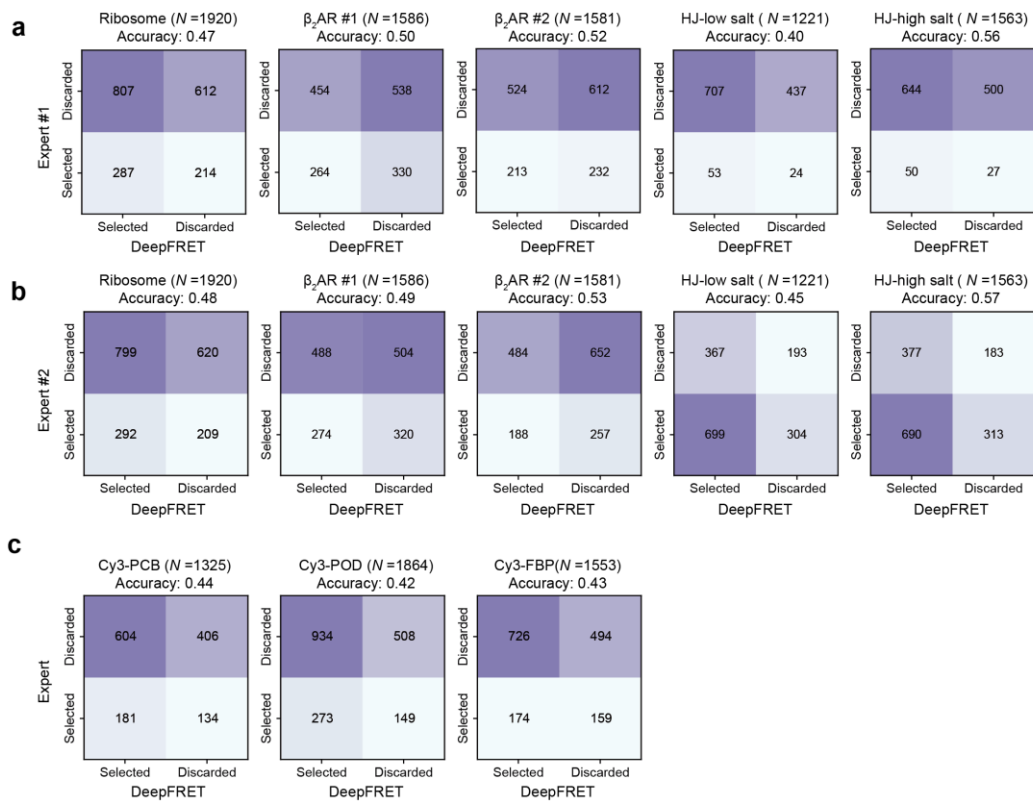


Figure R1. Confusion matrix for DeepFRET classification based on manually labeled ground truth. (a) Confusion matrix for DeepFRET classification on steady two-color traces with expert #1 manual label as ground truth. (b) Confusion matrix for DeepFRET classification on steady two-color traces with expert #2 manual label as ground truth. (c) Confusion matrix for DeepFRET classification on steady one-color traces with expert manual label as ground truth.

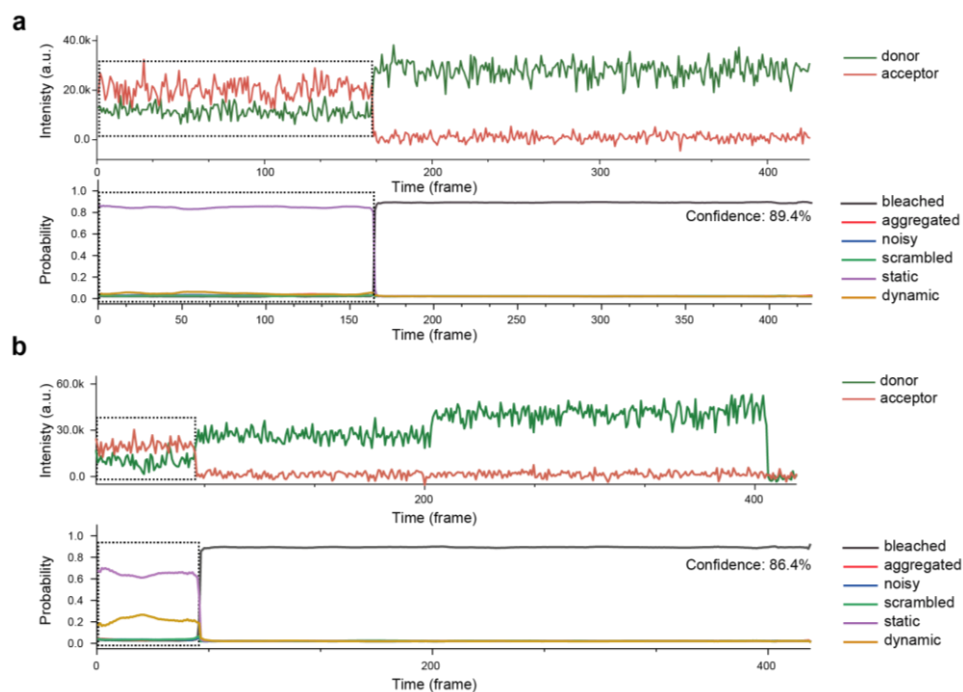




Figure R2. Examples of selected steady two-color traces selected by DeepFRET but rejected by DEBRIS and experts. (a) Steady two-color trace without donor photobleaching. (b) Steady two-color trace with significant noise before fluorophores photobleaching. The upper panel is an intensity-time trace. The lower panel shows the corresponding category-time traces predicted by DeepFRET. Black dashed boxes indicate range for DeepFRET selection.

Performance of DeepFRET on dynamically emerging signals: when applying DeepFRET's recommended threshold (confidence = 0.8) to dynamically emerging signals, no traces were selected. Upon lowering the threshold to 0.5, only 24 dynamic two-color traces and 16 dynamic single-color traces were selected. As shown in **Figure R3**, DeepFRET tends to predict the "bleaching" category when signals emerge in the middle time points of traces, resulting in a low confidence value for the trace. This leads to the rejection of the entire trace, which is the primary reason for the majority of rejected traces.

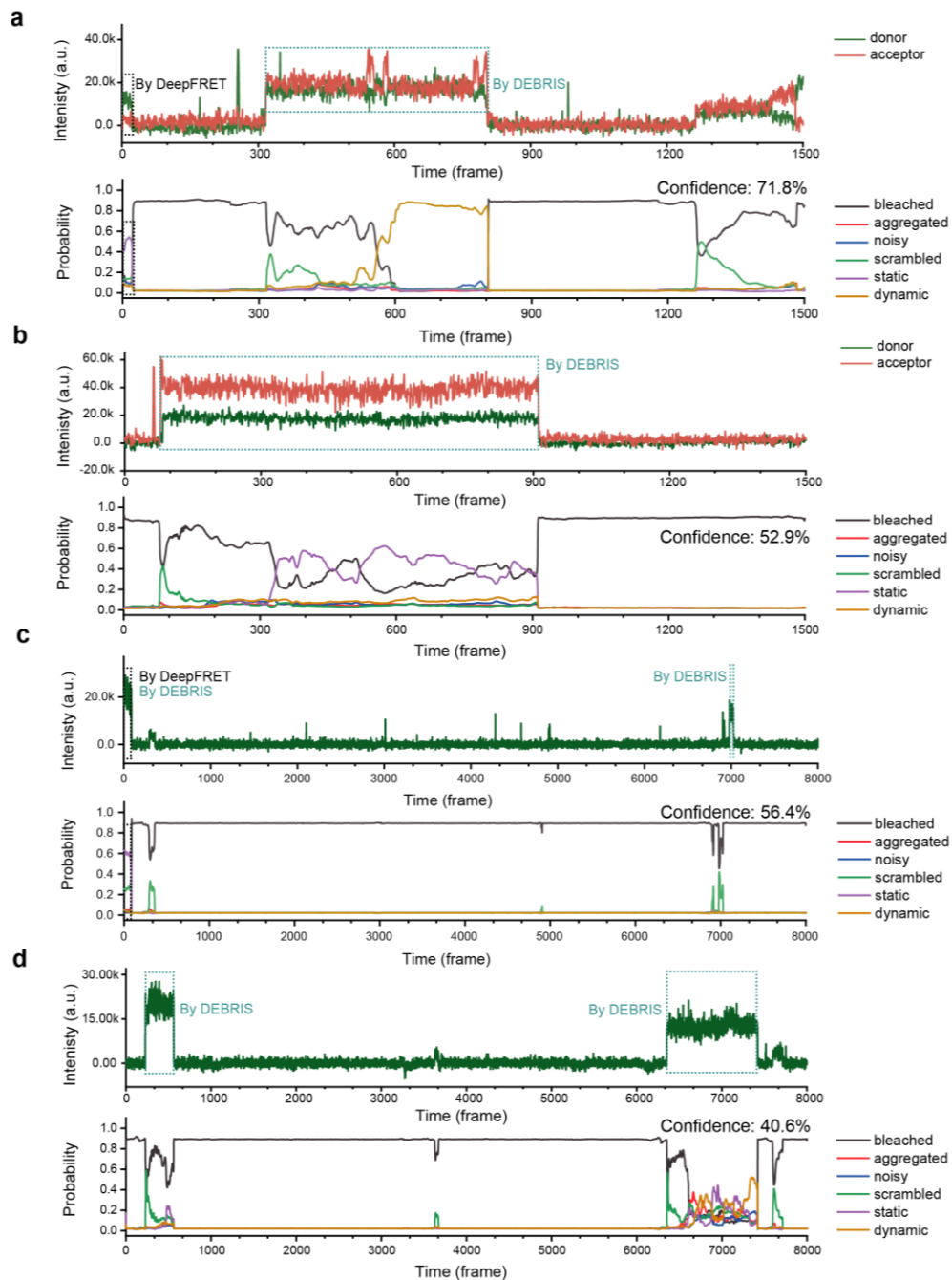


Figure R3. Examples of dynamically emerging one- and two-color traces and prediction results by DeepFRET. (a) Dynamically emerging two-color trace selected by DeepFRET. (b) Dynamically emerging two-color trace rejected by DeepFRET. (c) Dynamically emerging one-color trace selected by DeepFRET. (d) Dynamically emerging one-color trace rejected by DeepFRET. The upper panel is an intensity-time trace. The lower panel shows the corresponding category-time traces predicted by DeepFRET. Black dashed boxes indicate range for DeepFRET selection. Cyan dashed boxes indicate a valid range for DEBRIS selection.

Performance of AutoSiM on steady traces: following AutoSiM's recommended strategy for new data, we selected 500 manually labeled traces for transfer learning to achieve more accurate



classification results. The classification accuracy of steady two-color traces, after using expert manual labeling (Expert #1, Expert #2) as the ground truth for transfer learning, was approximately 0.75 (**Figure R4a-4b**). For steady single-color traces, the classification accuracy was about 0.86 (**Figure R4c**). However, we observed that AutoSiM easily made errors in the following aspects: 1) AutoSiM incorrectly stated the time point for photobleaching (**Figure R5a**). 2) AutoSiM selected the intermediate segment of trace but with incorrect start points (**Figure R5b**).

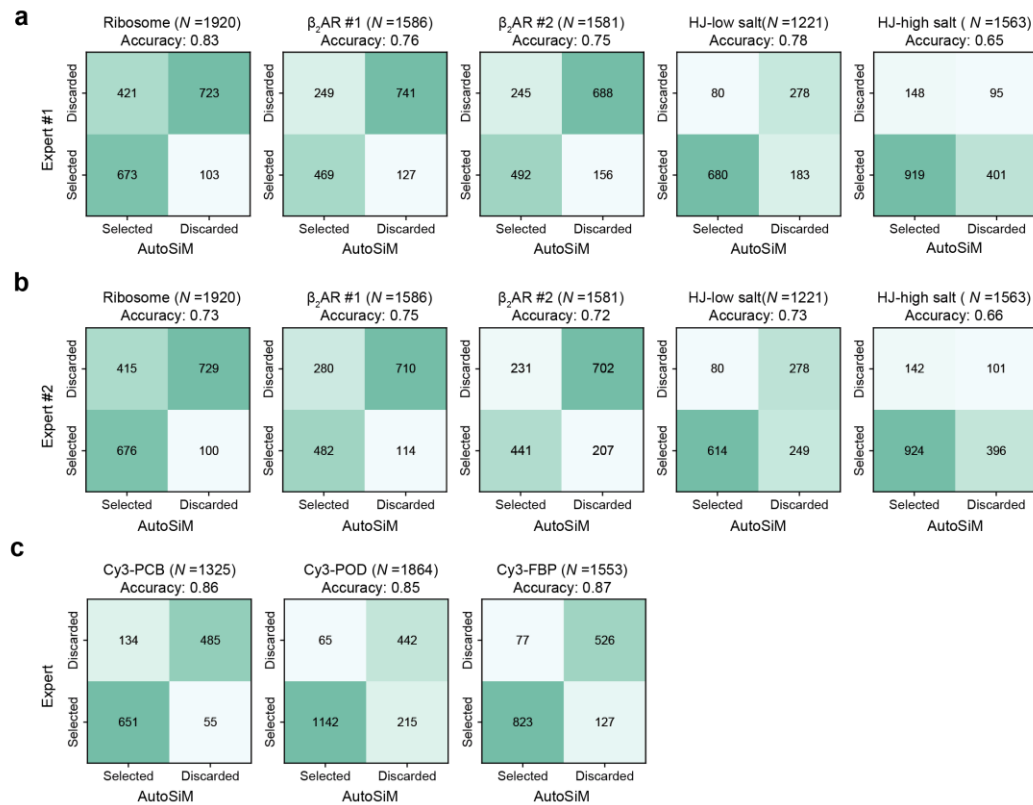


Figure R4. Confusion matrix for AutoSiM classification based on manually labeled ground truth. (a) Confusion matrix for AutoSiM classification on steady two-color traces with expert #1 manual label as ground truth. (b) Confusion matrix for AutoSiM classification on steady two-color traces with expert #2 manual label as ground truth. (c) Confusion matrix for AutoSiM classification on steady one-color traces with expert manual label as ground truth.

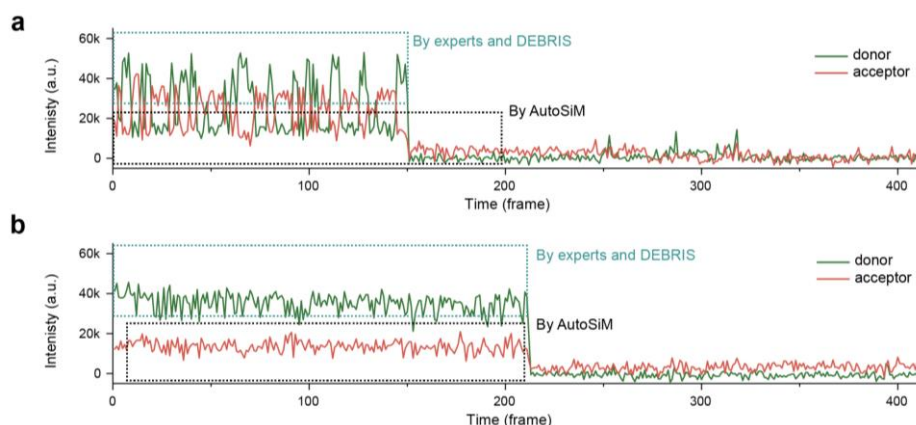


Figure R5. Examples of selected steady two-color traces identified differently by AutoSiM and experts and DEBRIS. (a) A steady two-color trace, with experts selecting 1 to 150 frames and AutoSiM selecting 1 to 191 frames. (b) A steady two-color trace, with experts selecting 1 to 212 frames and AutoSiM selecting 15 to 211 frames. Cyan dashed boxes indicate a valid range for experts and DEBRIS selection, black dashed boxes indicate a range for AutoSiM selection.

Performance of AutoSiM on dynamically emerging signals: in the absence of manual labeling datasets for transfer learning, we directly analyzed the dynamic traces using AutoSiM with the provided neural networks. AutoSiM only identified a total of 55 events for dynamically emerging two-color signals and 44 events for dynamically emerging one-color signals, which are ~ 20 folds less than DEBRIS. A detailed review of these traces revealed that AutoSiM is prone to errors in determining the start and end points of events. Additionally, AutoSiM struggled to accurately identify multiple dynamic events within a single trace (**Figure. R6**).

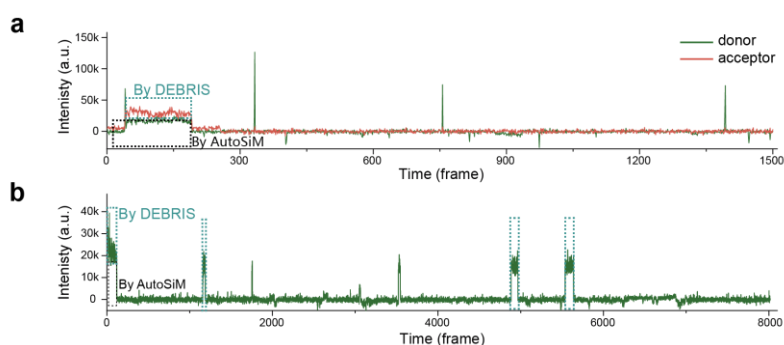


Figure R6. Examples of dynamically emerging one- and two-color traces predicted by AutoSiM. (a) Dynamically emerging two-color trace predicted by AutoSiM. (b) Dynamically emerging one-color trace predicted by AutoSiM. Cyan dashed boxes indicate a valid range for DEBRIS selection, black dashed boxes indicate a range for AutoSiM selection.

Regarding Deep-LASI⁴, we cannot get it work on our raw movies and processed traces after following their instructions. Their paper stated that the simulated data used for training Deep-LASI's network is derived from the DeepFRET simulated training set. Moreover, the Deep-



LASI paper does not describe how to process dynamically emerging signals. Thus, we expected that the performance of Deep-LASI on dynamically emerging signals would be close to the DeepFRET.

Many assays use fluorescence quenching, which can lead to dynamic single-color data. Is there software capable of analyzing dynamic single-color traces?

Response #6: The current DEBRIS neural network does not account for quenching and protein-induced fluorescence enhancement (PIFE)⁵ phenomena. There are several existing methods can address this issue, including Hidden Markov Models (HMM)^{4,5}, Bayesian approach⁶ and Machine learning approaches⁷.

In the **Discussion** section of the manuscript **page 6** we added: "*Also the current DEBRIS does not account for quenching and protein-induced fluorescence enhancement (PIFE) phenomena. The traces displaying these phenomena were discarded by our stringent criteria and can be selected if criteria are adjusted. Thus, users can optimize the DEBRIS deep learning model or adjust user defined criteria for their unique experimental conditions and designs.*"

Figure 4a, middle panel could, theoretically, be a dynamic trace that is then discarded.

Response #7: We apologize for any confusion caused. We presume that your concern pertains to the phenomenon of protein-induced fluorescence enhancement (PIFE)⁵ in fluorescence experiments, which can lead to a sudden increase or decrease in the fluorescence signal. This might have led you to believe that the signal in the middle panel of Figure 4a is acceptable as a dynamic trace.

In fact, in our study, photobleaching was defined as a complete decrease in fluorescence intensity to background value. Based on our stringent criteria, any fluctuations or enhancements in fluorescence intensity (aside from the anticorrelation seen in two-color fluorescence signals) were classified as the 'Junk' category. The data in Figure 4a was specifically used to test the photostability of dye molecules, which were expected to maintain stable fluorescence signals and exhibit single-step photobleaching. The trace in the middle panel of Figure 4a shows significant noise fluctuations prior to photobleaching, which does not align with our criteria for steady one-color traces. Therefore, it was discarded.

Although the current DEBRIS neural network does not account for multi-step quenching and PIFE phenomena, user defined criteria can be modified to choose trace as shown in Figure 4a, middle panel for further analysis. Users can optimize the DEBRIS deep learning model or adjust user defined criteria for their unique experimental conditions and designs.

The authors mentioned user-defined criteria when using the neural networks specifically for their data. Could the authors elaborate about these criteria, at least in the material and methods? To what extent the user can adjust the neural network's performance or confidence levels?

Response #8: We appreciate the reviewer's suggestions. We have added detailed criteria for



user-defined criteria in the ‘**User-defined criteria for classification**’, **Methods** section of manuscript, **page 7**. The parameters we describe in the manuscript are flexible, allowing users to adjust them according to their specific needs.

Page 7:

" User-defined criteria for classification

The gamma factor¹⁸ value has been used to compensate variations in sensitivity between different detection channels (REF). Our standard datasets (Figure S17) demonstrate that DEBRIS can tolerate to the variation of the input gamma factor value to a certain extent. In addition, DEBRIS calculate the gamma factor from the 'acceptor photobleaching' traces.

To analyze steady two-color traces, a threshold of 3 frames is established for the 'Ddis' feature, indicating that only 'Ddis' feature lasting 3 frames or longer will be classified as reliable donor photobleaching events. Similarly, the 'Adis' feature also has a threshold of 3 frames, while the 'Junk' feature is set at 5 frames, and the 'Non-FRET' feature threshold is 2 frames. For the analysis of steady one-color traces, the 'Sdis' category retains a threshold of 3 frames, and the 'Junk' feature is set at 5 frames.

In the assessment of dynamically emerging events in two-color traces, the threshold for the 'Dapp/Aapp' feature is established at 5 frames, and 'Ddis' is also set to 5 frames. The 'Adis' feature maintains a threshold of 3 frames, with the 'Junk' feature set at 5 frames, and the 'Non-FRET' feature at 2 frames. For dynamically emerging events in one-color traces, the thresholds are set as follows: 'Sapp' at 5 frames, 'Sdis' at 5 frames, 'Junk' at 5 frames, and 'Non-Sig' at 2 frames."

Page 3: "or variable intensities with anti-correlation behaviors, as their total intensities remain close to constant over time". The latter criteria can be a problem when the experimental system has differing sensitivities to the donor and acceptor dyes. Can the authors address this concern? Can one include the various correction factors and loosen the criteria so that this does not lead to throwing out good traces?

Response #9: We sincerely appreciate the reviewer’s insightful suggestions and apologize for the inaccurate description. During the simulation of fragments, differences in the detection efficiencies of different fluorescence channels have been considered. Therefore, a slight fluctuation in total intensity is acceptable for the DEBRIS network. We have revised the manuscript in **page 3**: “1) Before donor photobleaching, donor and acceptor have to exhibit either steady intensities or variable intensities with anti-correlation behaviors. 4) Acceptor photobleaching causes the disappearance of acceptor intensity accompanied by an increase in donor intensity, which remains close to constant before photobleaching. A one-step donor photobleaching must occur after the acceptor photobleaching to confirm the presence of only one donor fluorophore.”

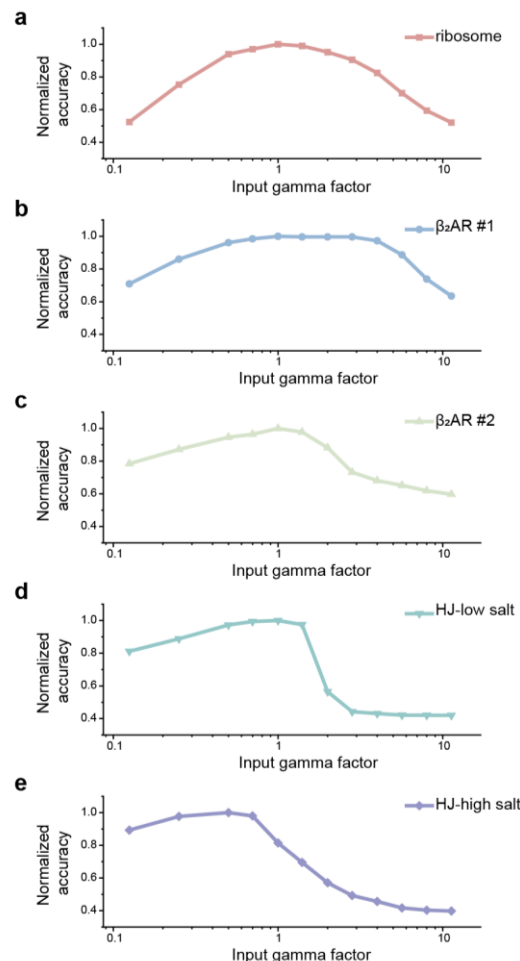
Furthermore, the ‘**DEBRIS.mat**’ file offers an adjustable parameter, the input gamma correction factor, which enables users to compensate for variations in sensitivity between donor and acceptor dyes.

As shown in the current Fig. S17, DEBRIS can tolerate to the variation of the input gamma



factor value to a certain extent. In addition, we described in our methods section, that the 'FRET_correction.mat' file provides the calculated gamma factor for the calculated ratio of intensity change around the receptor photobleaching time point. This value can be used to perform subsequent iterations, if needed.

Figure S17



Current Figure S17: “Input gamma factor dependent accuracy of ‘ribosome’ (a), ‘ β_2AR #1’ (b), ‘ β_2AR #2’ (c), ‘Holliday Junction low salt’ (d) and ‘Holliday Junction high salt’ (e).”

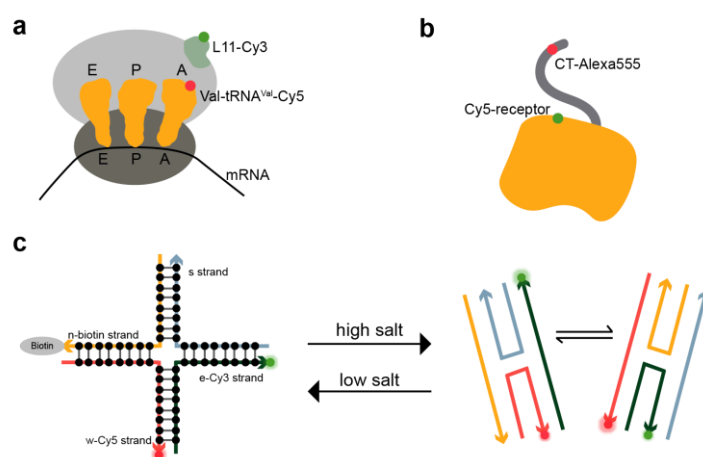
Page 4: The authors include data from *AsCas12a*, *FnCas12a* and *LbCas12a*. For non experts, the authors should introduce these different systems.

Response #10: We appreciate the reviewer's suggestions and have included a brief introduction to Cas12a in the manuscript, **page 4**. Additionally, we have cited relevant references to offer further details on these Cas12a proteins for readers who wish to explore them in more depth.

Page 4: " This design employs Cas12a effector proteins, including *Acidaminococcus* sp. Cas12a (*AsCas12a*)^{42,43}, *Lachnospiraceae* bacterium ND2006 Cas12a (*LbCas12a*)⁴⁴ and *Francisella novicida* U112 Cas12a (*FnCas12a*)^{42,45} , which can be guided by crRNA to recognize DNA target sites adjacent to PAM sites to initiate R-loop formation and double-stranded DNA cleavage "

Figures: Since the authors are referring to five different experimental systems in this paper, they should add a simple schematic of the different systems to help readers follow the various assays easier at the appropriate positions in the different figures.

Response #11: We appreciate the reviewer's suggestions and have added a schematic diagram of the ribosome, β_2 AR, and Holliday junction in **Figure S4** of supplementary file, **page 5**. These illustrations are designed to help readers better follow and understand the different experimental systems.



Current Figure S4. "Schematic models of the ribosome, β_2 adrenergic receptor (β_2 AR) and Holliday junction system. (a) Ribosome pre-translocation (PRE) complex containing cyanine3 (Cy3) labeled L11 protein and cyanine5 (Cy5) labeled Val-tRNA^{Val} in the A-site. Grey indicates ribosomes, and orange indicates tRNAs. (b) β_2 AR complex labeled with Alexa555 in the C-terminal (CT, solid grey line) and Cy5 in the receptor core domain (orange part). (c) DNA Holliday junction, a four-way stranded DNA, complex labeled with Cy3 in the e-strand and Cy5 in the w-strand."

Minor points

Many experimental details are missing from the paper. The frame rate of data collection, laser powers used, etc.

Response #12: We appreciate the reviewer's comments. We have included detailed information on laser power, exposure time, and frame rate of data collection for each experiment in the "Single-molecule measurements", **Methods** section of manuscript, **page 8**.

Page 2: "models trained using equilibrium datasets are unsuitable for the single-molecule traces taken under non-equilibrium conditions" This statement is not always true. Some of the previously mentioned software packages would be able to analysis this type of dynamics. Please reword.

Response #13: We apologize for the inaccurate and overly definitive description. We have



revised the wording in the manuscript, **page 2**: " *Thus, models trained on limited datasets are unlikely to fit all user's needs, due to their unique experimental conditions and designs.* "

Page 3: "*which closely matches the reported value of 100 ± 3 nM.*" The two values do not overlap within the error bars. Please reword to something such as '*which is similar to ...*'"

Response #14: We apologize for the inaccurate description. We have revised the wording in the manuscript **on page 3**: "*We determined the apparent EC_{50} value of 88 ± 1 nM (Figure 3b), compared to the reported value of 100 ± 3 nM³⁹.*"

Figure 3d: *There is a low/no FRET peak showing up around 5 s after binding for the LbCas12a and FnCas12a systems. The authors should at least mention this peak and what it is due to.*

Response #15: We appreciate the reviewer's suggestions and have included the statements of FRET peaks FRET peak showing up around 5 s after binding for the LbCas12a and FnCas12a systems in manuscript, **page 4**: "*In addition, the FRET dwell times of LbCas12a and FnCas12a were shorter than those of AsCas12a, probably due to their fast cleavage product release rate, resulting in the zero FRET population appearing ~ 5 s after binding of LbCas12a and FnCas12a (Figure 3d).*"

Figure S8. Panels a, b: *There appears to be a problem here. In panel a, categorization of donor appearance and disappearance correlates with what is observed in the smFRET trace. In panel b, the categorization is shifted by several seconds from the event in the smFRET trace.*

Response #16: We sincerely apologize for the oversight. We have corrected this by updating the pattern-time traces in **Figure S10b** (previously Figure S8b) to align with the smFRET trace accurately.

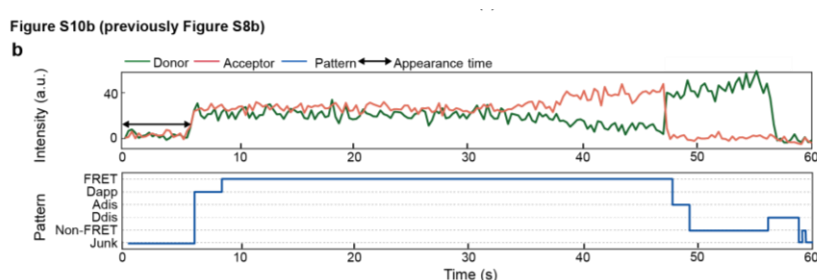


Figure S11, Panel a: *It is not clear from the figure (nor from the text) how the SNR was determined. Please correct.*

Response #17: We apologize for the missing definition of SNR and have included it in the legend of **Figure S13** (previously Figure S11) in the supplementary file, **page 14**: "*SNR = $\bar{I}/\sigma$, where $\bar{I}$ represents the mean intensity before photobleaching, σ represents the standard deviation of intensity value before photobleaching.*"



Figure S13: Figure is confusing. According to the schematic (panel a), the DNA is cleaved from the surface. In panel b, multiple binding events are observed. Also, what is labeled as "appearance time" in panel b is the time between binding events, not the time from injection until binding (with the exception of the first event), which is the definition given in the text.

Response #18: For first-order binding events under equilibrium, "appearance time" is defined as the time from one event disappearing to the next event appearing. Thus, by definition, "appearance time" is the time between binding events. However, a unique feature of the first-order reaction is that the time from observation starting (no matter when it starts) until the first event appearing can also be defined as "appearance time" to calculate the first-order reaction rate. A commonly known case is that the half-life of a radioactive isotope is independent of the start time of measurement. To verify our model, the appearance times calculated using times from injection until binding and using times between binding events are close to each other (Figure R7).

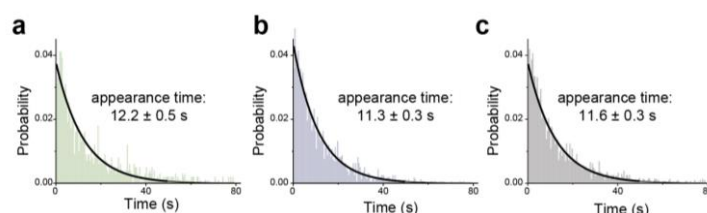


Figure R7. Appearance time fitting results using times from injection until binding (a), times between binding events (b), or all dwell times (c).

The authors talk about "bleaching". This should be referred to as "photobleaching".

Response #19: We appreciate the reviewer's feedback. We have updated the terminology throughout the figures and descriptions in the revised manuscript to use "photobleaching" instead of "bleaching".

Response to reviewer 2

Reviewer #2:

Comments for the Authors:

In their submitted manuscript, Zhou et al. present a deep-learning model to classify single molecule FRET trajectories by a sliding window classification. The model is suitable for single color and two-color FRET trajectories. It is worth mentioning, that the model only works for single color FRET experiments and does not support the acceptor excitation and acceptor emission analysis from alternating laser excitation experiments like other neural network models do. New is the feature of recognizing emerging fluorescence events in an experiments do to binding events. The authors compare the classification of experimental trajectories of their model to the evaluation of experts.

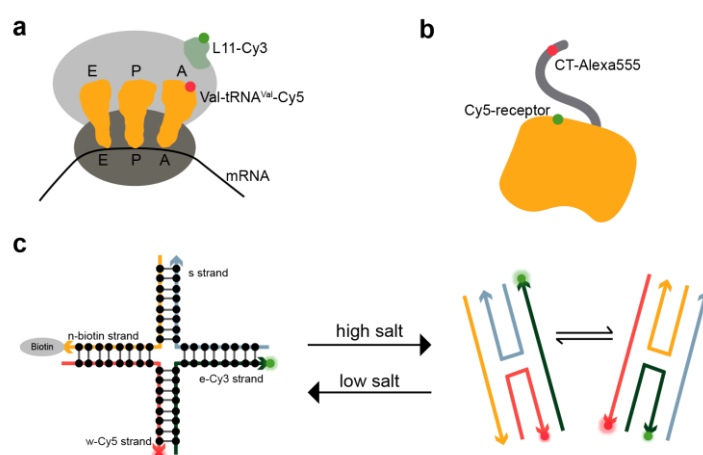


The manuscript is well written, the classification criteria of the model are explained in detail and sound. The software is published, and a manual is available which makes the software easy to use in other labs. I suggest the publication in Communications Biology after addressing my questions:

Response #20: We thank the reviewer for the positive comments.

1. The authors test their model on experimental data. The assays are only depicted in the SI but single molecule trajectories are shown in the main manuscript. A sketch of the assays (holliday junction, Cas12 and PRE) is missing in all figures. I highly emphasize to add sketches of each assay in order to make the interpretation of the trajectories easier.

Response #21: We appreciate the reviewer's suggestions and have added a schematic diagram of the ribosome, β_2 AR, and Holliday junction in **Figure S4** of supplementary file, **page 5**. These illustrations are designed to help readers better follow and understand the different experimental systems.



Current Figure S4. "Schematic models of the ribosome, β_2 adrenergic receptor (β_2 AR) and Holliday junction system. (a) Ribosome pre-translocation (PRE) complex containing cyanine3 (Cy3) labeled L11 protein and cyanine5 (Cy5) labeled Val-tRNA^{Val} in the A-site. Grey indicates ribosomes, and orange indicates tRNAs. (b) β_2 AR complex labeled with Alexa555 in the C-terminal (CT, solid grey line) and Cy5 in the receptor core domain (orange part). (c) DNA Holliday junction, a four-way stranded DNA, complex labeled with Cy3 in the e-strand and Cy5 in the w-strand. "

2. What is the advantage of the sliding time-window approach? I guess the model can be faster trained as only single features (e.g. acceptor bleaching) must be recognized and not combinations of them. If so, please state it in the manuscript.

Response #22: We appreciate the reviewer's suggestions and have added the following explanation on manuscript, **page 2**: "The DEBRIS model employs a sliding time-window classification approach during its application, which facilitates the rapid prediction of



fragment-based neural networks and applies to traces of varying lengths for the precise classification of local features in a sequential manner."

3. Why did the authors choose the sliding time-window length to be 10 frames?

Response #23: We appreciate the reviewer's question and would like to clarify our reasoning. The choice of a 10-frame sliding time window corresponds to the fragment length used during network training. In fluorescence imaging, a signal persisting for more than 3 frames is typically deemed reliable, and we aimed to simulate a broad range of potential short fragment categories. After evaluating the neural network trained on 8-, 10-, and 12-frame fragments, we found that fragments length had little effect on the classification accuracy of our network structure (**Figure R8**). Therefore, the 10-frame fragments is not unique, 8- and 12-frame fragments based networks perform equally well.

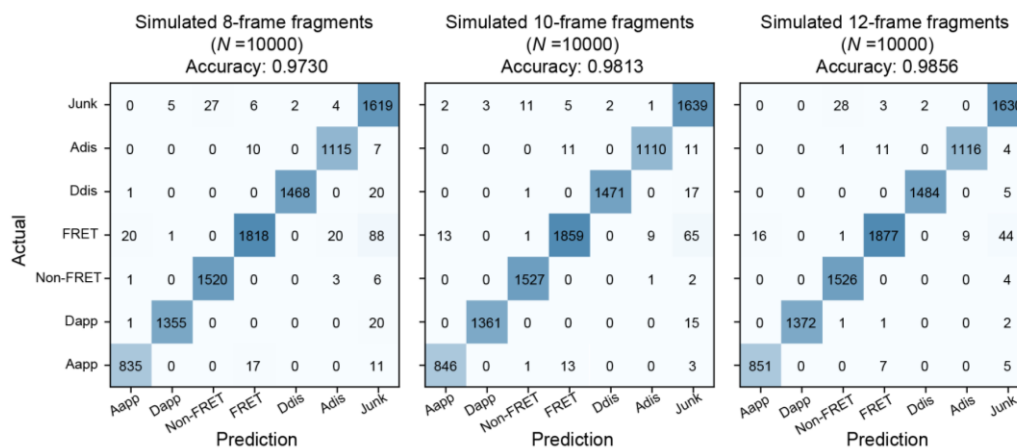


Figure R8. Confusion matrix of a neural network trained to classify 8-frame (left), 10-frame (middle), and 12-frame (right) fragments based on the ground truth of the simulated fragments.

4. Why was the model trained for single color excitation? The alternating laser excitation approach offers additional filter criteria. ALEX excitation gives access to the dye-stoichiometry and monitors the acceptor fluorescence to detect photophysics.

Response #24: We are appreciative of the reviewer's perspicacious recommendation. The current version of DEBRIS is trained to analyze two-color fluorescence traces under single laser excitation, under which conditions most of our experiments were performed, and current DEBRIS version cannot handle single-molecule fluorescence experiments using the ALEX mode. We have highlighted this issue in the **Discussion** section of the manuscript, **page 6**: "It should be noted that the current version of DEBRIS is trained to analyze two-color fluorescence traces under single laser excitation, which is also fully compatible with one-color traces. However, the deep-learning model needs to be retrained to analyze three or more colors traces and single-molecule traces under Alternating Laser Excitation (ALEX) mode using two or more lasers. Also the current DEBRIS does not account for quenching and protein-induced fluorescence enhancement (PIFE) phenomena. The traces displaying these phenomena were discarded by our stringent criteria and can be selected if criteria are adjusted. To maximize the



accessibility of the DEBRIS model, we offer MATLAB source codes for fragment simulation, deep learning network training, and single-molecule fluorescence event identification for four different scenarios presented in our study. Thus, users can optimize the DEBRIS deep learning model or adjust user defined criteria for their unique experimental conditions and designs."

5. The enzyme concentrations (glucose oxidase catalase) are usually given in units per mL and not μg per mL as the activity of every batch is different.

Response #25: We appreciate the reviewer's suggestion. Upon review, we found that the glucose oxidase (Sigma-Aldrich) used in our experiments has an activity of 100,000 units/g, meaning that 100 $\mu\text{g}/\text{mL}$ glucose oxidase corresponds to 10 units/mL. Similarly, the catalase (Sigma-Aldrich) used has an activity of 30,000 units/mg, so 40 $\mu\text{g}/\text{mL}$ catalase corresponds to 1,200 units/mL. We have revised the enzyme concentrations in the **Methods** section of manuscript, **page 8**.

Response to reviewer 3

Reviewer #3:

Comments for the Authors:

The manuscript is exceptionally well-written, with clear and concise language throughout. The quality of the figures is notably high, effectively enhancing the presentation of the research findings. Additionally, the text is well-organized and structured, contributing to a seamless reading experience. However, there are technical and general issues that need to be clarified before publication.

The authors introduce and apply a neural network, DEBRIS, for single molecule trace analysis in MATLAB. They provide MATLAB scripts for generating training data and classifying frames in single molecule traces. The neural network automates the classification of regions in these traces, significantly simplifying the analysis process typically performed manually by experts. Processing experiments, including FRET, Holliday junction, Cas complex, and beta2AR analysis. They compared the results obtained from their neural network (NN) to those derived from expert analysis, demonstrating the effectiveness and reliability of their automated approach very similar to the more comprehensive work previously published (Deep-LASI).

Response #26: We thank the reviewer for the positive comments.

As noted by the authors, several neural networks and software tools perform similar tasks in single-molecule trace analysis. Examples include DeepFRET, AutoSIM, and Deep-LASI. In my comparison of Deep-LASI and AutoSIM with the authors' MATLAB scripts, the existing software generally offers more comprehensive functionality. For instance, Deep-LASI can read images to create traces within a single software suite. The authors' scripts primarily focus on the classification task within single molecule traces. While they mention a custom ImageJ plugin, the separation of tasks for trace computation and trace analysis is appreciated. However,



the authors do not provide the software for these tasks. Could the authors provide raw images and software (full stack) to process image traces along with a manual? Providing a full analysis pipeline could significantly increase the impact of their work.

Response #27: We appreciate the reviewer's insightful comments. Unlike fully integrated workflows that handle everything from movies to FRET histograms, DEBRIS is primarily focused on the classification of traces, a process that can be subject to user bias. While we regret that we are unable to provide the custom ImageJ plugins at this time. Because we do not own its copyright. We are developing a complete workflow to analyze images/movies and traces (via DEBRIS) to provide a more comprehensive solution for single-molecule fluorescence analysis.

The authors claim that their neural network (DEBRIS) is less opaque compared to published systems for neural network-based FRET and single molecule trace analysis. However, I did not find evidence to support this claim, as other software also allows for training. While the authors developed DEBRIS and naturally view it as less opaque, an outsider's perspective suggests that DEBRIS is similarly opaque compared to other software like AutoSIM, which also provides training capabilities. I cannot follow this claim, as the authors' software is similarly opaque compared to other available software.

Response #28: Thank you for your suggestion. We have removed this sentence and modified the following sentence as “*In this study, we introduce DEBRIS (Deep lEarning Based fRagmentatIon approach for Single-molecule fluorescence event identification), a deep-learning model focusing on classifying local features and capable of automatically identifying steady fluorescence signals and dynamically emerging signals of different patterns.*” in **Abstract** of the manuscript, **page 1**.

The authors claim, “DEBRIS is the first method to achieve automatic classification of dynamically emerging events by classifying short-fragment traces instead of long-duration ones.” Please specify this claim. While DEBRIS works on short fragment traces, it is unclear how this is better compared to other software. Did you compare your approach to other existing approaches? If so, please present the results of your comparison.

Response #29: We appreciate the reviewer's inquiry regarding the advantages of DEBRIS compared to other published networks such as DeepFRET² and AutoSiM³.

We conducted tests on steady single- and two-color data using DeepFRET and AutoSiM and observed that DEBRIS outperformed both in these scenarios, at least when applied to these datasets. For dynamically emerging signals, DeepFRET and AutoSiM were less effective, as they struggled to accurately discern the traces, making them less capable of identifying emerging events.

To test the efficacy of DeepFRET and AutoSiM, we selected five datasets (Figure 2b) for steady two-color trace test sets and three datasets (Figure 4b) for steady one-color trace test sets. For dynamically emerging two-color signals, we selected 2075 traces from AsCas12a-WT 0MM (Figure 3d) as a test set, and 1425 traces from 11 bp dsDNA extension in the PAM downstream



(Figure 4e) were selected as a test set for dynamically emerging one-color signals. DEBRIS identified 1713 events from the 2075 dynamic two-color traces and 6478 events from the 1425 dynamic one-color traces. Single-color traces were replicated to mimic a two-channel input.

Performance of DeepFRET on steady traces: When applying DeepFRET's recommended thresholds (confidence level = 0.8) to steady traces, the classification accuracy for steady two-color traces, using expert manual labeling (Expert #1, Expert #2) as the ground truth, was approximately 0.5 (**Figure R9a-9b**). For steady single-color traces, the classification accuracy was around 0.43 (**Figure R9c**). Notably, we identified some differences in criteria between DeepFRET and DEBRIS: 1) DeepFRET accepts traces where donor does not undergo photobleaching (**Figure R10a**), whereas DEBRIS and experts require photobleaching of donor. 2) DeepFRET appears to accept traces with abnormal signal changes after acceptor photobleaching (**Figure R10b**), but DEBRIS and experts require steady signals before photobleaching of all fluorophores (both acceptor and donor).

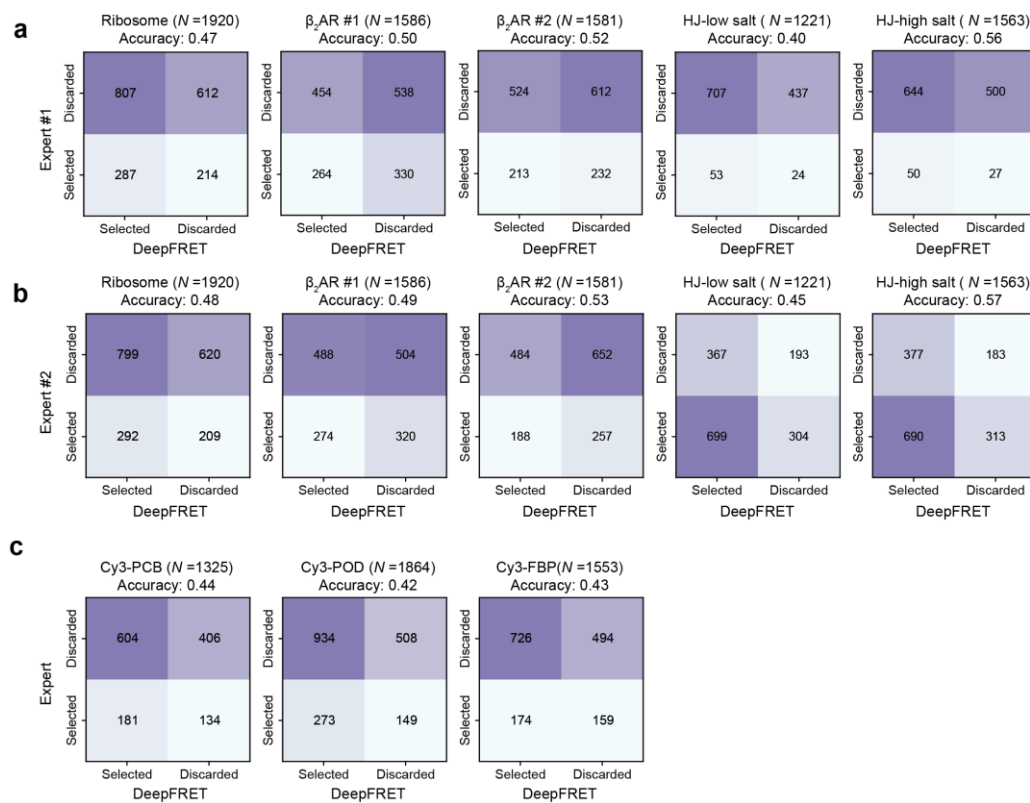


Figure R9. Confusion matrix for DeepFRET classification based on manually labeled ground truth. (a) Confusion matrix for DeepFRET classification on steady two-color traces with expert #1 manual label as ground truth. (b) Confusion matrix for DeepFRET classification on steady two-color traces with expert #2 manual label as ground truth. (c) Confusion matrix for DeepFRET classification on steady one-color traces with expert manual label as ground truth.

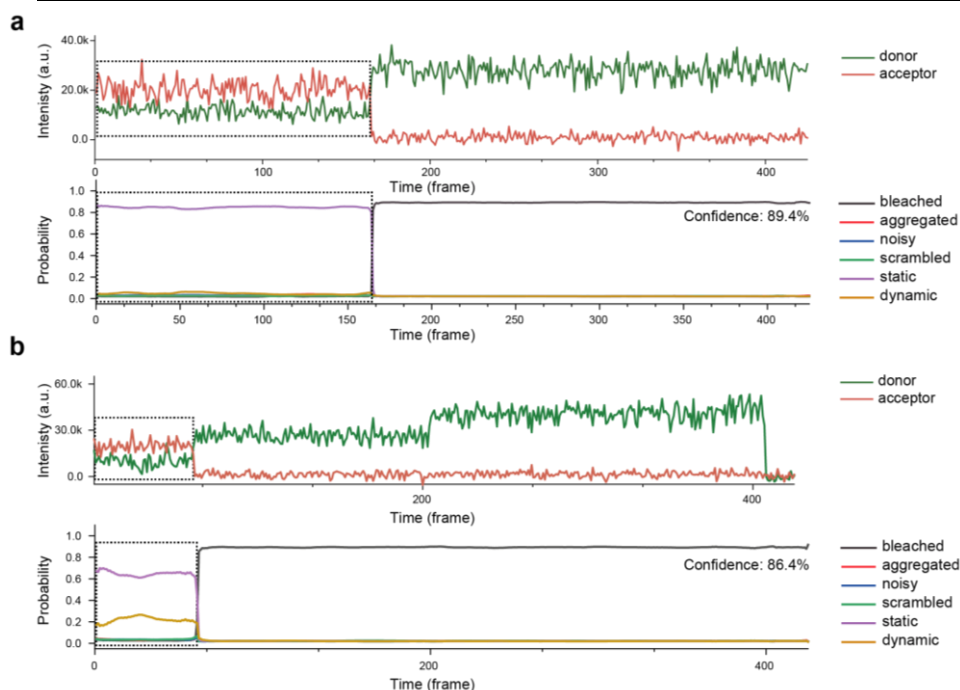


Figure R10. Examples of selected steady two-color traces selected by DeepFRET but rejected by DEBRIS and experts. (a) Steady two-color trace without donor photobleaching. (b) Steady two-color trace with significant noise before fluorophores photobleaching. The upper panel is an intensity-time trace. The lower panel shows the corresponding category-time traces predicted by DeepFRET. Black dashed boxes indicate range for DeepFRET selection.

Performance of DeepFRET on dynamically emerging signals: when applying DeepFRET's recommended threshold (confidence = 0.8) to dynamically emerging signals, no traces were selected. Upon lowering the threshold to 0.5, only 24 dynamic two-color traces and 16 dynamic single-color traces were selected. As shown in **Figure R11**, DeepFRET tends to predict "bleaching" when signals emerge in the middle time points of traces, resulting in a low confidence value for the trace. This leads to the rejection of the entire trace, which is the primary reason for the majority of rejected traces.

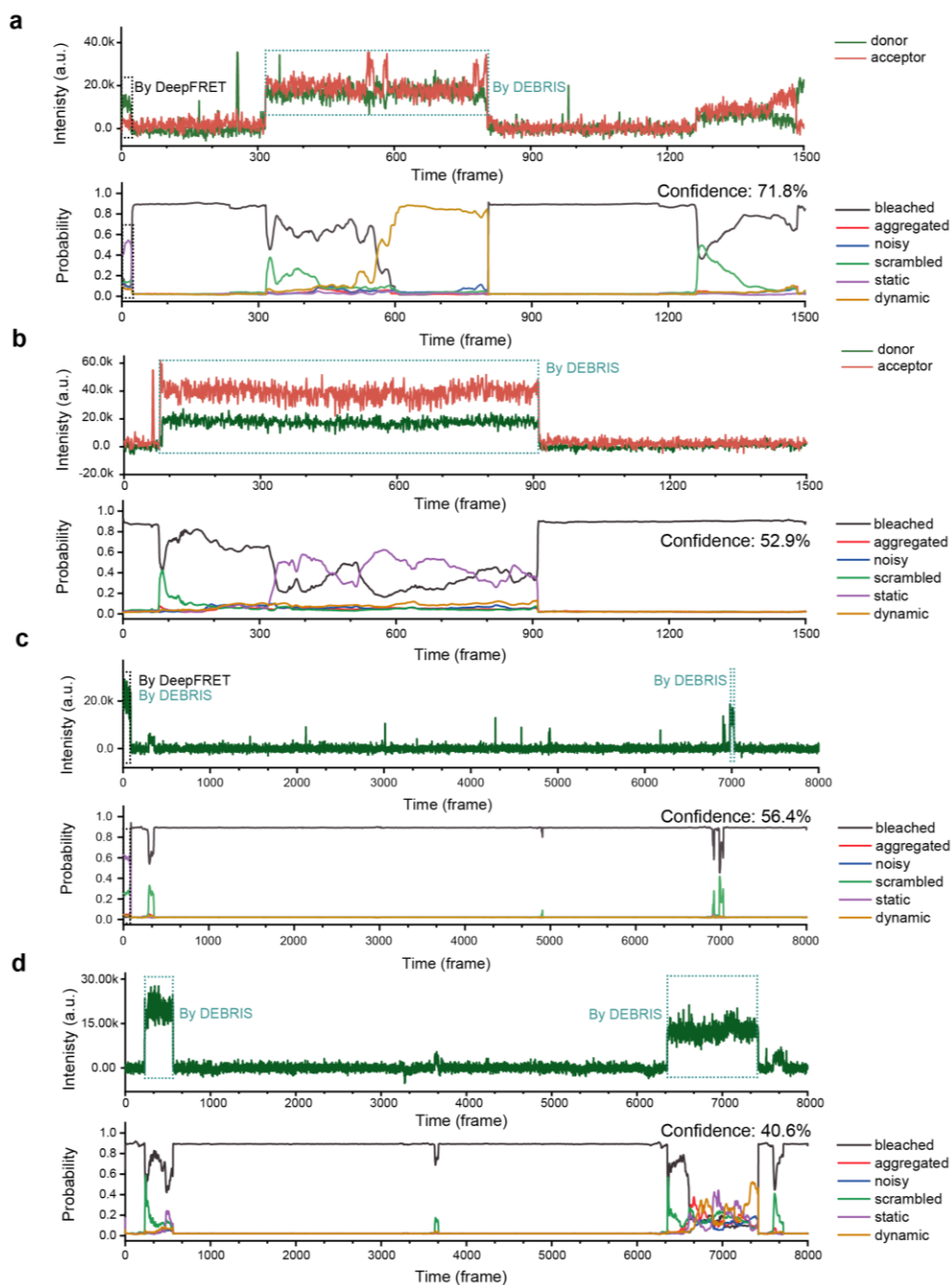


Figure R11. Examples of dynamically emerging one- and two-color traces and prediction results by DeepFRET. (a) Dynamically emerging two-color trace selected by DeepFRET. (b) Dynamically emerging two-color trace rejected by DeepFRET. (c) Dynamically emerging one-color trace selected by DeepFRET. (d) Dynamically emerging one-color trace rejected by DeepFRET. The upper panel is an intensity-time trace. The lower panel shows the corresponding category-time traces predicted by DeepFRET. Black dashed boxes indicate range for DeepFRET selection. Cyan dashed boxes indicate a valid range for DEBRIS selection.

Performance of AutoSiM on steady traces: following AutoSiM's recommended strategy for new data, we selected 500 manually labeled traces for transfer learning to achieve more accurate



classification results. The classification accuracy of steady two-color traces, after using expert manual labeling (Expert #1, Expert #2) as the ground truth for transfer learning, was approximately 0.75 (**Figure R12a-12b**). For steady single-color traces, the classification accuracy was about 0.86 (**Figure R12c**). However, we observed that AutoSiM easily made errors in the following aspects: 1) AutoSiM incorrectly stated the time point for photobleaching (**Figure R13a**). 2) AutoSiM selected the intermediate segment of trace but with incorrect start points (**Figure R13b**).

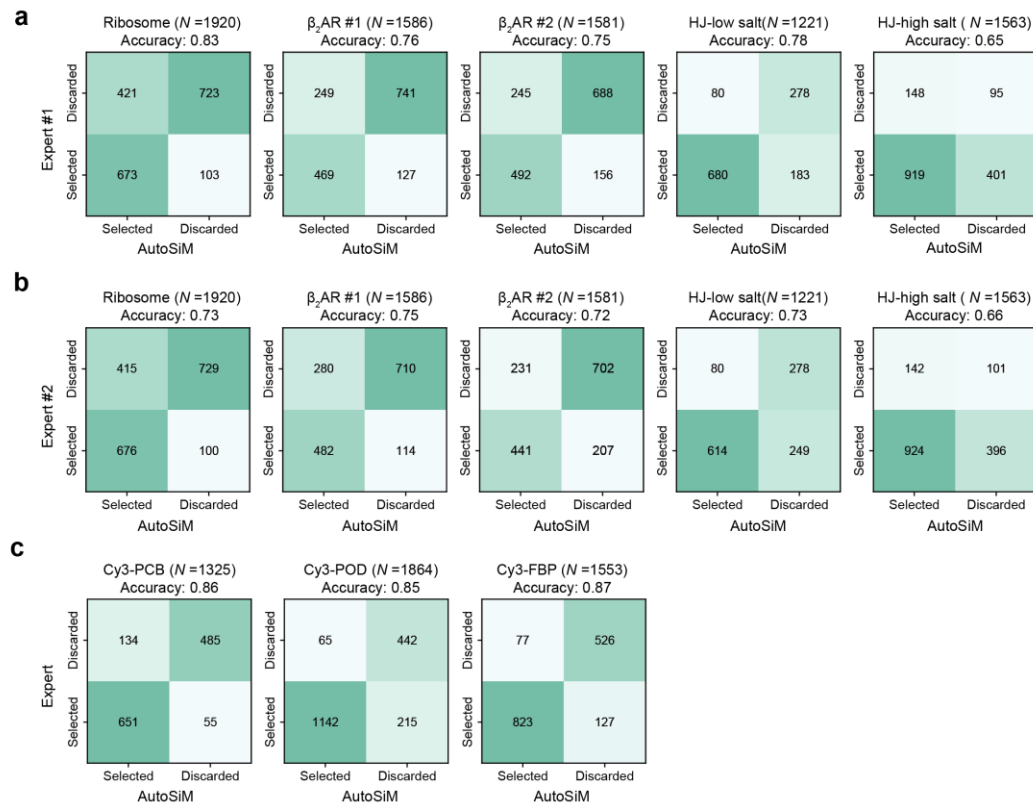


Figure R12. Confusion matrix for AutoSiM classification based on manually labeled ground truth. (a) Confusion matrix for AutoSiM classification on steady two-color traces with expert #1 manual label as ground truth. (b) Confusion matrix for AutoSiM classification on steady two-color traces with expert #2 manual label as ground truth. (c) Confusion matrix for AutoSiM classification on steady one-color traces with expert manual label as ground truth.

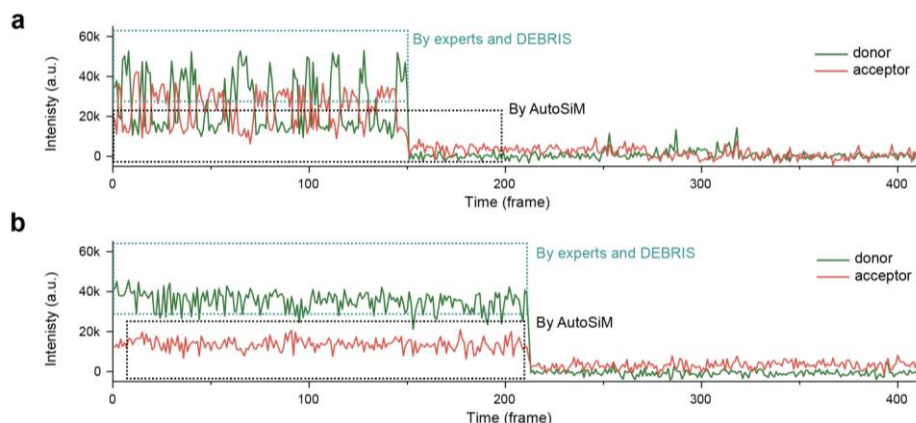


Figure R13. Examples of selected steady two-color traces identified differently by AutoSiM and experts and DEBRIS. (a) A steady two-color trace, with experts selecting 1 to 150 frames and AutoSiM selecting 1 to 191 frames. (b) A steady two-color trace, with experts selecting 1 to 212 frames and AutoSiM selecting 15 to 211 frames. Cyan dashed boxes indicate a valid range for experts and DEBRIS selection, black dashed boxes indicate a range for AutoSiM selection.

Performance of AutoSiM on dynamically emerging signals: in the absence of manual labeling datasets for transfer learning, we directly analyzed the dynamic traces using AutoSiM with the provided neural networks. AutoSiM only identified a total of 55 events for dynamically emerging two-color signals and 44 events for dynamically emerging one-color signals, which are ~20 folds less than DEBRIS. A detailed review of these traces revealed that AutoSiM is prone to errors in determining the start and end points of events. Additionally, AutoSiM struggled to accurately identify multiple dynamic events within a single trace (**Figure. R14**).

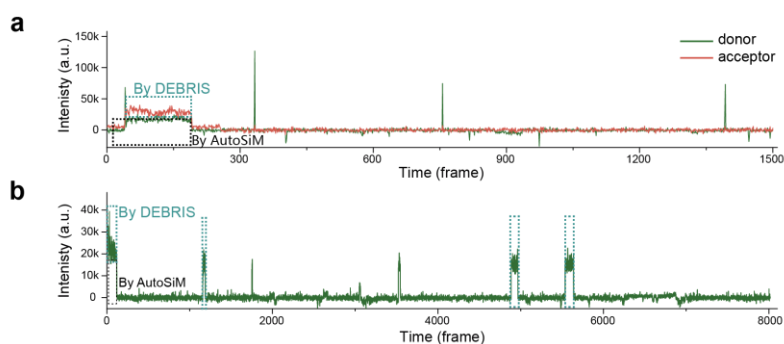


Figure R14. Examples of dynamically emerging one- and two-color traces predicted by AutoSiM. (a) Dynamically emerging two-color trace predicted by AutoSiM. (b) Dynamically emerging one-color trace predicted by AutoSiM. Cyan dashed boxes indicate a valid range for DEBRIS selection, black dashed boxes indicate a range for AutoSiM selection.

Additionally, how does the trace length impact your classification? When using your scripts, I only found a “final” classification result. Can you provide a confidence measure for the



classification? How does the confidence depend on the trace length?

Response #30: The DEBRIS method is designed to handle intensity-time traces of varying lengths by converting them into pattern-time traces using sliding windows. This approach ensures that the method can adapt to traces of different lengths without compromising classification accuracy. To investigate the effect of trace length on classification results, we conducted an analysis using modified steady single- and two-color data. Specifically, we spliced three traces in sequence, creating new traces that are three times longer than the original. This transformation effectively converts an original dataset containing n traces of length m into a new dataset with $(n-2)$ traces of length $3m$ (**Figure R15**).

Our findings indicate that, according to the classification criteria for steady traces, the classification results and FRET distributions for the spliced traces remain consistent. The confusion matrix (**Figure R16a and 16c**) demonstrates that the classification accuracy of the spliced traces by DEBRIS is not affected their extended length and comparable to that of expert #1 and expert #2. Furthermore, the FRET histograms (**Figure R16b and 16d**) align closely, indicating that the trace length does not significantly impact classification accuracy.

In our workflow, user-defined criteria were used to classify traces into different categories based on the pattern-time trace predicted by the deep learning model. Currently, we have no idea how to add a confidence measure at this stage.

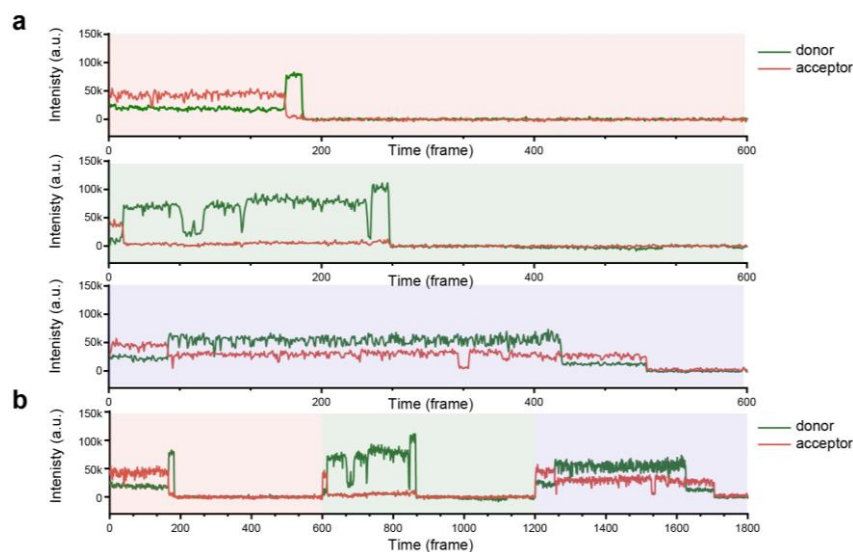


Figure R15. Steady two-color traces splicing. Three 600-frame-long traces (a) are sequentially spliced into a single 1800-frame-long trace (b).

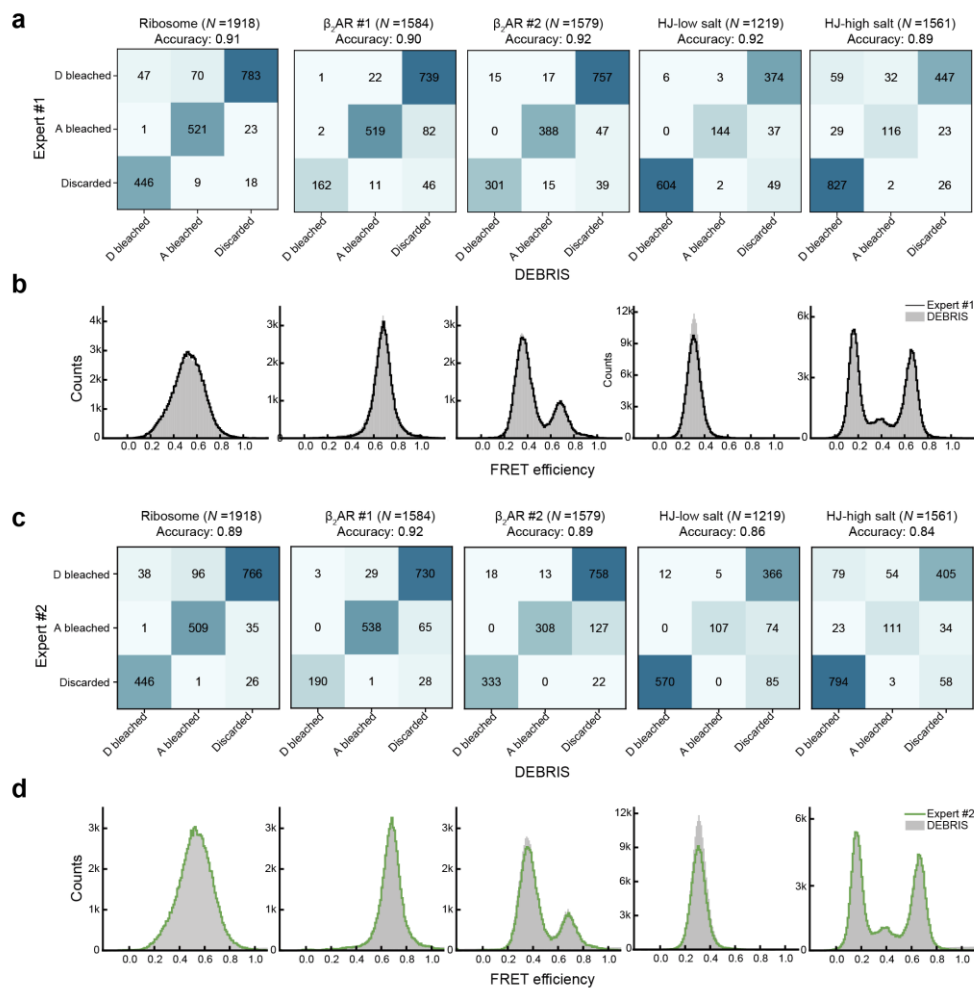


Figure R16. Comparison of spliced traces classified by DEBRIS and experts. (a) Confusion matrix indicating difference between DEBRIS and expert #1. (b) FRET distributions extracted from traces classified by DEBRIS and expert #1. (c-d) Confusion matrix (c) and FRET distributions (d) of traces classified by the DEBRIS and expert #2. D bleached: donor photobleaching; A bleached: acceptor photobleaching. N stands for the number of classified traces.

Technical Issues:

The program requires MATLAB dependencies that are not included in a standard MATLAB installation. Please list the required toolboxes and packages in your installation manual. Additionally, I require the following toolboxes to use your programs:

- Signal Processing Toolbox
- Deep Learning Toolbox
- Parallel Computing Toolbox(TM)

File: Netv230712_Training.mlx in Chinese

Missing example data for three-color FRET.



Response #31: We appreciate the reviewer bringing these technical issues to our attention. We have included a comprehensive list of the required MATLAB toolboxes necessary to run DEBRIS in the 'User manual of DEBRIS' document, **page 2**. Additionally, we have revised the 'Netv230712_Training.mlx' file to ensure all content is in English, eliminating any Chinese text that may cause confusion.

The current version of DEBRIS can only analyze one-color and two-color traces. Thus, no three-color FRET is provide as an example data.



References:

1. Rahul, R., Sungchul, H. & Taekjip, H. A practical guide to single molecule FRET. *Nat Methods* **4**, 507–16 (2008).
2. Thomsen, J. *et al.* Deepfret, a software for rapid and automated single-molecule fret data classification using deep learning. *Elife* **9**, 1–21 (2020).
3. Li, J., Zhang, L., Johnson-Buck, A. & Walter, N. G. Automatic classification and segmentation of single-molecule fluorescence time traces with deep learning. *Nat Commun* **11**, 5833 (2020).
4. Wanninger, S. *et al.* Deep-LASI: deep-learning assisted, single-molecule imaging analysis of multi-color DNA origami structures. *Nat Commun* **14**, 6564 (2023).
5. Stennett, E. M. S., Ciuba, M. A., Lin, S. & Levitus, M. Demystifying PIFE: The Photophysics behind the Protein-Induced Fluorescence Enhancement Phenomenon in Cy3. *Journal of Physical Chemistry Letters* **6**, (2015).
6. Tibbs, J. *et al.* KERA: Analysis tool for multi-process, multi-state single-molecule data. *Nucleic Acids Res* **49**, (2021).
7. Xu, J. *et al.* Automated Stoichiometry Analysis of Single-Molecule Fluorescence Imaging Traces via Deep Learning. *J Am Chem Soc* **141**, (2019).



Response to reviewer #1

Reviewer #1:

The authors have done a commendable job in addressing the majority of the reviewers' comments. There are a couple minor corrections I would suggest, upon which the paper would be acceptable for publication.

Response #1: We would like to extend our sincere gratitude to the commentators for their positive remarks.

1) The intent of asking for schematic drawings of the different systems was to add them next to the data (i.e. in Figure 2c or 2d, Figure 3d, and Figure 4c and 4e).

Response #2: We are very grateful to the reviewer for the valuable suggestions. We have provided the corresponding schematic diagrams of the different systems in the corresponding figures to facilitate the reading and understanding of the DEBRIS application.

2) I am still unclear about the details of the assay used in Figure 4c. I believe DNA is attached to Cy3, but what is it binding too? Please explain.

Response #3: We apologize for the confusion and have added a schematic diagram of the system to **Figure 5a** on page **18** of the manuscript. In practice, double-stranded DNA is labeled with Cy3 on one strand and biotin on the other. To immobilize the DNA, polyethylene glycol (PEG) biotin-modified slides are first incubated with streptavidin and then with biotin-DNA. In brief, the biotin-DNA is attached to the PEG biotin-modified slides by streptavidin.

3) The authors were asked to compare their programs with others in the field. This they have nicely done in the rebuttal (although the accuracy calculations are not always correct (e.g. Figure R1b, HJ-low salt)). I fully understand that one needs to be careful when discussing other people's work in a publication. Still, I would recommend mentioning that DEBRIS was also tested against AutoSIM and DeepFRET and found that DEBRIS performed better, in particular when dealing with traces that have emerging dynamics, for which DEBRIS was designed. (Certainly, the performed test are not a complete benchmarking of the various softwares, so the phrasing needs to be done carefully).

Response #4: We are grateful to the reviewer for this valuable suggestion. We have added the following sentence to highlight the better performance of DEBRIS in dynamically emerging signals of the manuscript, **page 4**: “An evaluation of DEBRIS against existing deep learning models showed that DEBRIS was superior in analysing our own dynamically emerging signals.”

Response to reviewer #2

The authors have adequately addressed my comments. As final comment, I suggest inserting the sketches of Figure S4 into the main text figures to facilitate fluent reading of the manuscript.



The manuscript should then be ready for publication.

Response #5: We are very grateful to the reviewer for the valuable suggestions. We have provided the corresponding schematic diagrams of the different systems in the corresponding figures to facilitate the reading and understanding of the DBERIS application.

Response to reviewer #3

The authors have thoroughly addressed all the concerns raised in the previous review. The revisions have been carefully considered and implemented, ensuring that all issues are satisfactorily resolved. Given the improvements made, the manuscript is now of high quality and suitable for publication in its current form without the need for further modifications.

Response #6: We sincerely thank the reviewer for the positive comments.