

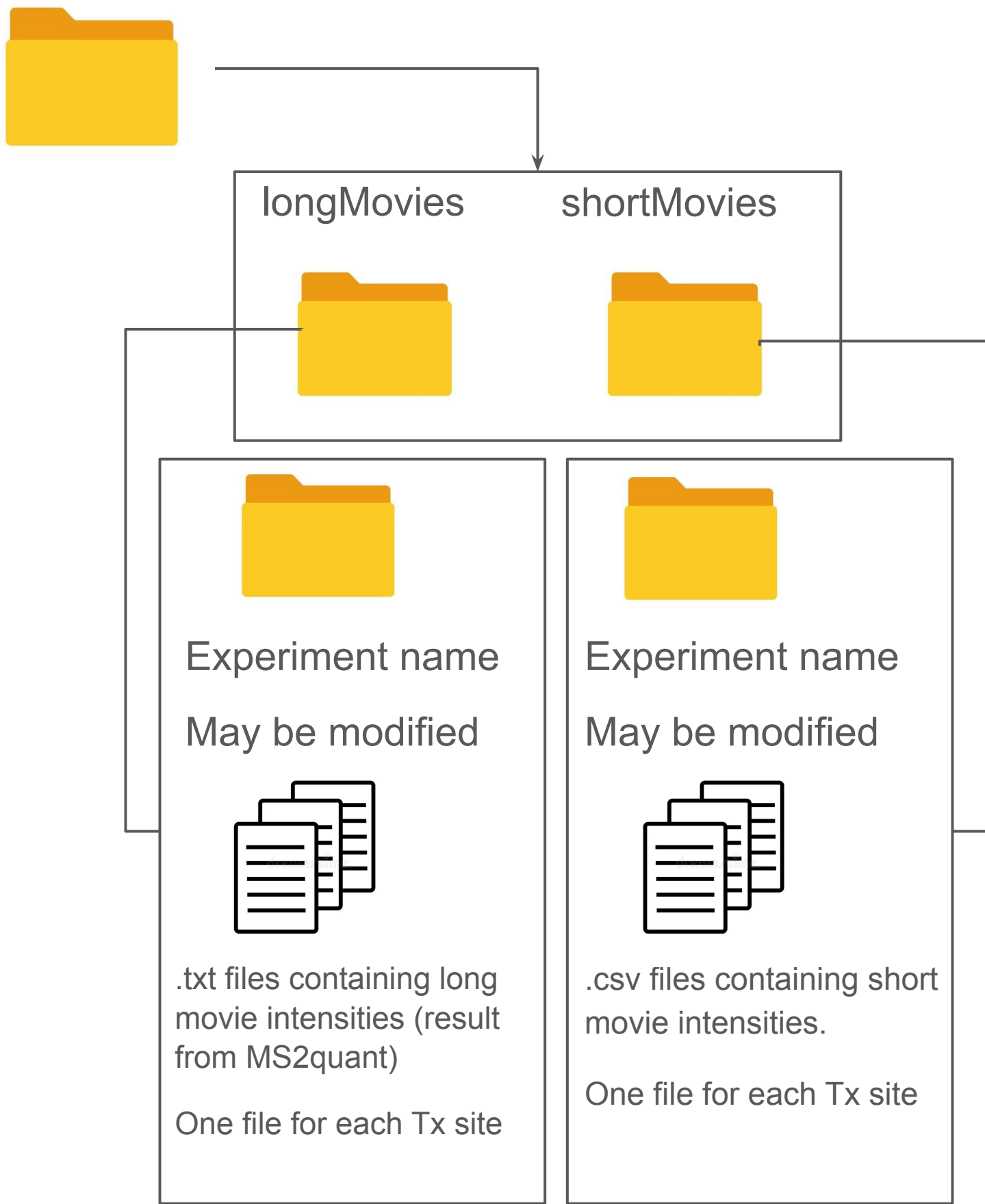
BurstDECONV

A machine learning procedure to
analyse MS2 data.

Directory structure

Mixed Resolution Movies

Parent Folder (may be modified)



Directory structure

Short Resolution Movies

Parent Folder (may be modified)

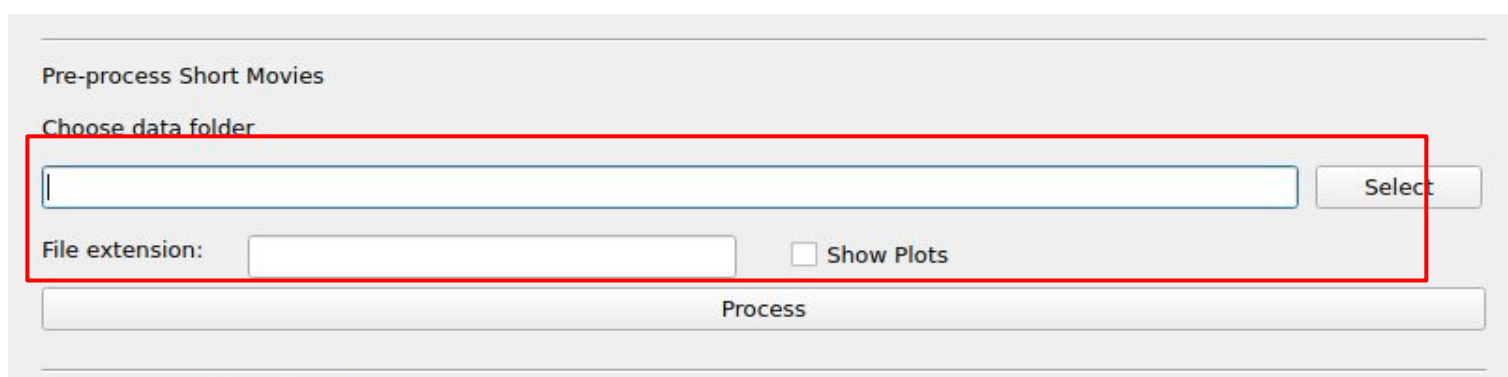
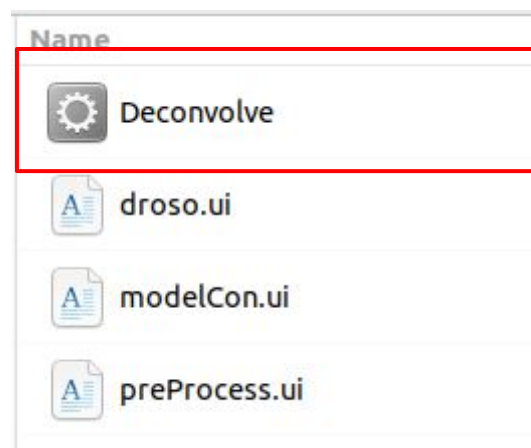


.xls files containing
short movie
intensities

Pre-process txt/xls files (short Movies)

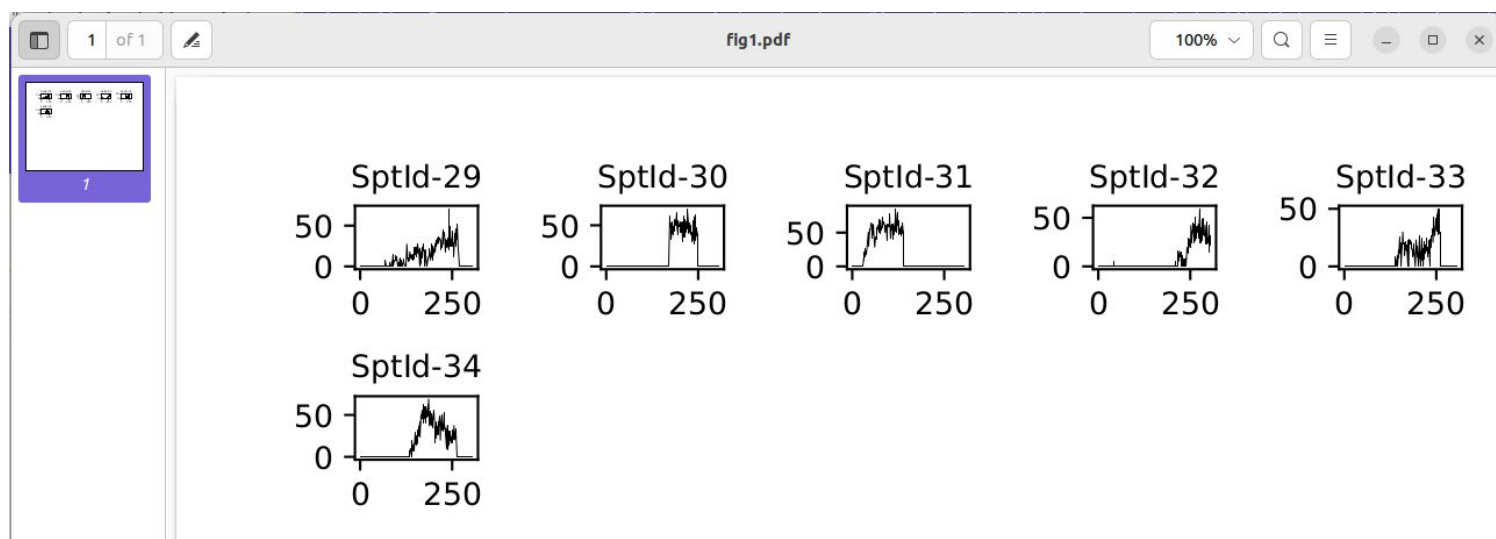
1. Download the file “BurstDECONV_GUI.zip” and extract the contents into a folder. You can use `unzip filename.zip` in the command line.
2. Open the application by double clicking on the icon called Deconvolve.

(Depending on your OS it may look different)
3. Click on the Pre-Process Raw Files button. A window should appear with option to choose files/folders.
4. Select the Parent folder containing the xls/xlsx files
5. Also indicate the file extension (whether its an xls,xlsx,csv file)
6. Tick the show plots option in case you want to see the intensity traces
7. Click on Process



Once the files have been processed, you will find two folder inside the directory, “shortMovies” called ‘npzdatafiles’ and “images”

Folder “npzdatafiles” has .npz files for each Experiment that can be used for deconvolution.
Folder “images” has figures with intensity traces. Example shown below



Deconvolve short movies

Next step is to find the sequence of polymerase positions using genetic algorithm.

1. Choose the Parent folder which contains the “npzdatafile” folder in the section “Choose data folder”
2. Fill in the details for the signal from a single polymerase according to the MS2 system used. When you click “Save and Load Parameters” an “.npz” file called “nameParameters.npz” is saved in the data folder selected in previous step and loaded in the application. Next time you can simply load the saved file.
3. Make sure the “Perform deconvolution” option is checked.
4. Indicate the number of thread to use in order to perform parallel processing. (It is recommended to use less than the total number of threads available when running on non-server machines.
5. Maximum number of Generation (default value of 100)
This is used to set an upperbound to the number of iterations the genetic algorithm runs for.
6. Select the type of fit to be performed. (It is recommended to perform both kinds of fits and then evaluate the quality of the fit.

Deconvolve

Choose data folder

hel/Documents/deconvolutionAppdemo/sna/

Select

☐ Include long movies

Set Parameters

Polymerase Speed

45

×

Length of Null signal (in bp)

41

Length of MS2 signal (in bp)

1292

Length of stable signal (in bp)

4526

Minimum distance between two polymerase (in bp)

30

Frame length (in seconds)

3.9

Pick a name for saving these parameters

droso

Save and Load Parameters

Load saved parameters

Select

☒ Perform deconvolution

Choose Deconvolution Parameters

Choose number of threads

Check Maximum Threads Available

8

Maximum number of Generation

100

Perform Fit

☒ fit 2 state model ☒ fit 3 state model

Analyse!

0%

Back

Deconvolution Outputs

The deconvolution+fit output includes two folders called “resultDec” and “Results”

1. Contents of the “resultDec” folder consist of the output of the genetic algorithm. The .npz bundle includes the experimental trace, the predicted trace and the polymerase positions.

2. The contents of the “Results” folder include the following

a. result_M1080919SnaPrxMCPHomoem1CFCalibratedTraces_result

It contains the plots for various fits of the survival function, the heatmap of raw intensity traces, heatmap of polymerase positions.

b. fit2_results.xlsx

c. fit3M1_results.xlsx

d. fit3M2_results.xlsx

b,c,d, contain the value of the kinetic parameters obtained from the survival function as well as the results from the kolmogorov smirnov test, objective function etc.

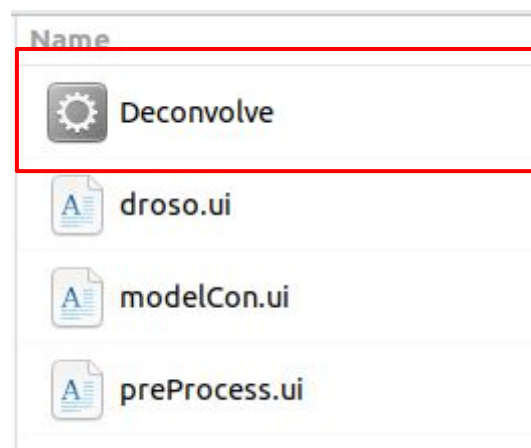
Pre-process txt/xls files (Mixed resolution Movies)

1. Open the application by double clicking on the icon called Deconvolve.

(Depending on your OS it may look

different)

2. Click on the Pre-Process Raw Files button. A window should appear with option to choose files/folders.
3. Select the folder containing Parent Folder containing folders longMovie and shortMovie
4. Also indicate the threshold value for long movie intensities (See help to determine the value)
5. Click on Process



Pre-process Short Movies

Choose data folder

Select

File extension: ☐ Show Plots

Process

Pre-process Mix Resolution Movies

Choose Long Movie data folder

Select

Choose Short Movie data folder

Select

Threshold:

Process

Deconvolve mixed resolution movies

Next step is to find the sequence of polymerase positions using genetic algorithm.

1. Choose the Parent folder which contains the “npzdatafile” folder in the section “Choose data folder”
2. Make sure to tick the check box “include long movies”. This will look for a long movie file matching the name of the short movie npz data file.
3. Fill in the details for the signal from a single polymerase according to the MS2 system used. When you click “Save and Load Parameters” an “.npz” file called “nameParameters.npz” is saved in the data folder selected in previous step and loaded in the application. Next time you can simply load the saved file.
4. Make sure the “Perform deconvolution” option is checked.
5. Indicate the number of thread to use in order to perform parallel processing. (It is recommended to use not more than half the total number of threads available when running on non-server machines.
6. Maximum number of Generation (default value of 100)
This is used to set an upperbound to the number of iterations the genetic algorithm runs for.
7. Select the type of fit to be performed. (It is recommended to perform both kinds of fits and then evaluate the quality of the fit.

Deconvolve

Choose data folder

demo/hiv/example/demoData/shortDataFile/

Select

☒ Include long movies

Set Parameters

Polymerase Speed

67

✕

Length of Null signal (in bp)

700

Length of MS2 signal (in bp)

2900

Length of stable signal (in bp)

8300

Minimum distance between two polymerase (in bp)

30

Frame length (in seconds)

3

Pick a name for saving these parameters

hiv

Save and Load Parameters

Load saved parameters

Select

☒ Perform deconvolution

Choose Deconvolution Parameters

Choose number of threads

Check Maximum Threads Available

8

Maximum number of Generation

100

Perform Fit

☒ fit 2 state model

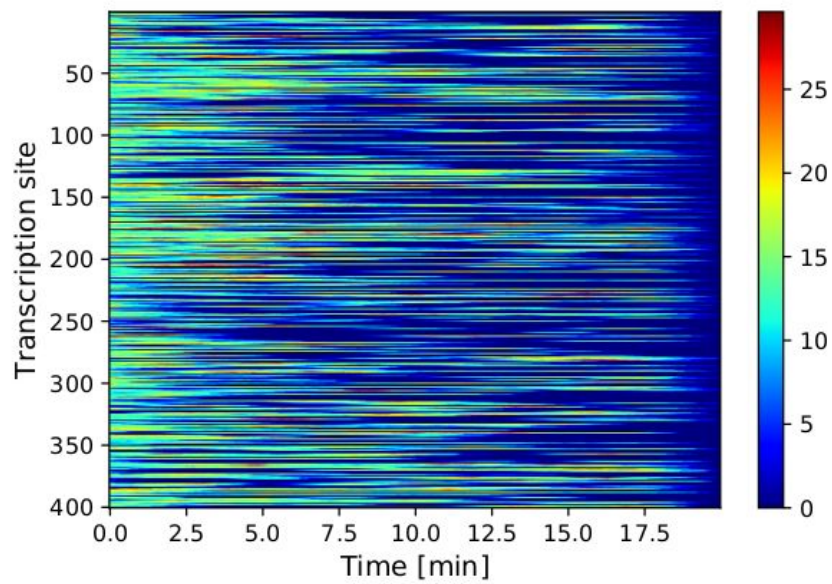
☒ fit 3 state model

Analyse!

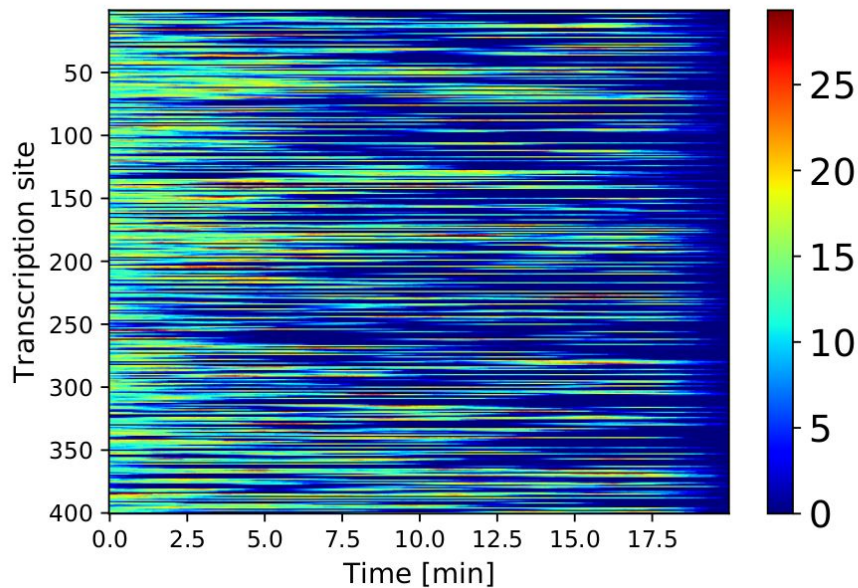
0%

Back

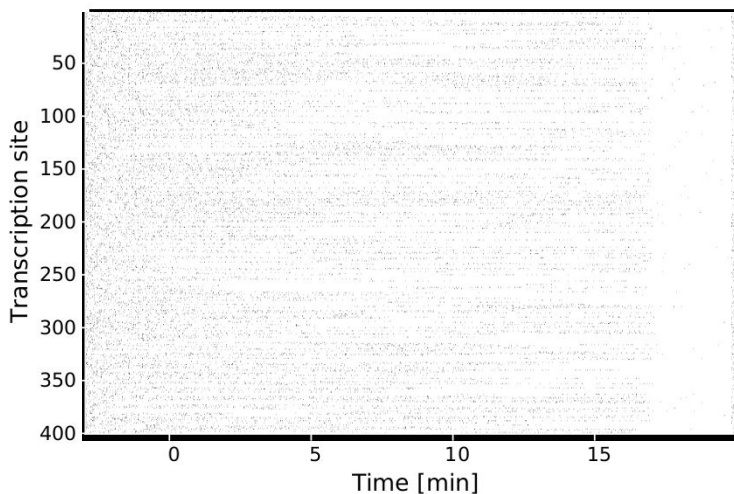
Output examples for artificial dataset (D13)



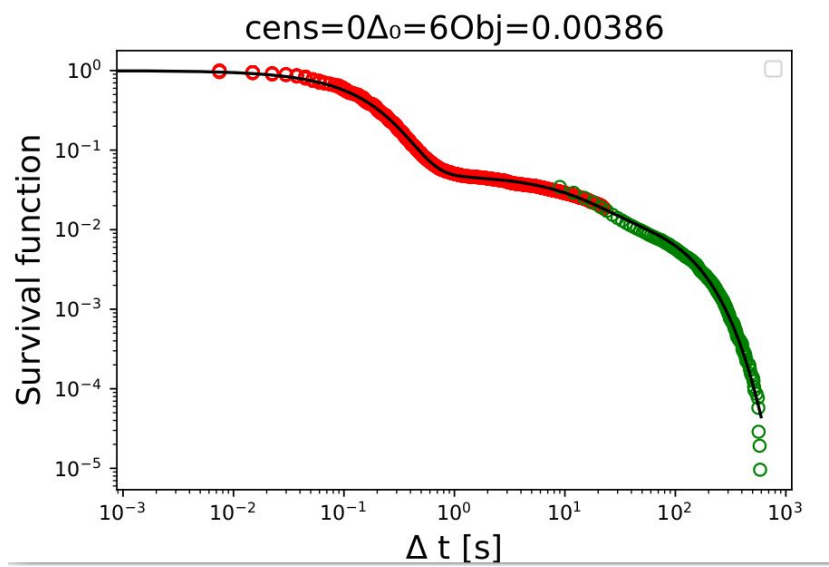
Intensity traces
(Experimental data)



Intensity traces (Predicted
signal)



Reconstructed polymerase
positions



Optimal survival function.

Table of kinetic parameters

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Data	k1p	k1m	k2p	k2m	k3	lambda1	lambda2	lambda3	A1	A2	A3	Obj	mRNA	shift	alpha	cens	samples	Frames
HIVo	0.00012	0.00126	0.00118	0.00347	0.09458	-0.09936	-0.00114	-0.00012	0.95146	0.03521	0.01333	0.00057	17.3	6	0.9	1	400	400
	0.00012	0.00076	0.00118	0.00301	0.09458	-0.09989	-0.00172	-0.00012	0.95146	0.0309	0.00787							
	0.00012	0.00126	0.00178	0.00347	0.09607	-0.09936	-0.00114	-0.00012	0.96124	0.03521	0.01333							
HIVno	4.07E-05	0.00162	0.00108	0.00344	0.05743	-0.06255	-0.00102	-4E-05	0.91723	0.0552	0.02757	0.00024	3.52341	6	0.9	1	396	400
	4.07E-05	0.00162	0.00108	0.00344	0.05743	-0.06255	-0.00102	-4E-05	0.91723	0.0552	0.02757							
	4.07E-05	0.00162	0.00108	0.00344	0.05743	-0.06255	-0.00102	-4E-05	0.91723	0.0552	0.02757							