Multiplexing

Project organization.

Meeting Organization

1. The process to analyze microscope images and to generate intensity data.

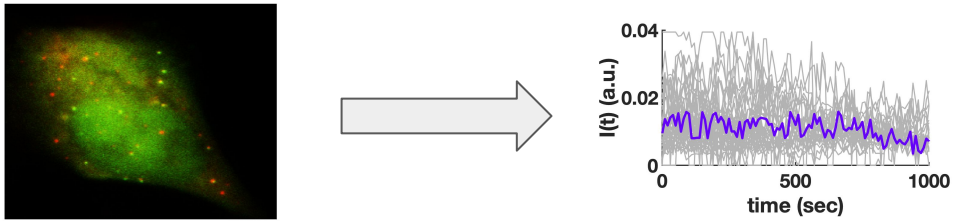


Image acquisition

Data generation

2. How much data do we have?

3. Planning next steps.

Experimental Data Acquisition Pipeline

1) Image acquisition. -> Images/Videos for 3 channels. Files are Gb in size and stored in format .tif.
    - The image acquisition can be taken at different frame rates for the 3 different channels.
        - For example the RNA (red channel) is taken at 1FPS and the Proteins green channel at 2 FPS.

2) Image preprocessing -> ImageJ/Fiji
    - Maximum projection. Transforms 3D video (13 slices in the Z-axis) to 2D images.
    - Photobleaching correction.
    - Correcting for bleed-through artifacts.
    - Background removal

3) Mathematica codes.
    - Spot Tracking:
        - Region of interest. Manual mask creation.
        - Fluorophore position correction. Transformation function. Using tetraspeck beads.
        - Running z-average. Smooth the images. (Optional; have never used).
        - Particle detection for the red channel (mRNA). Detect spot position with Gaussian fit.
        - Linking spots for different time points. (Select/Reject by length).
    - Intensity measurement:
        - Create trims. Sections of interest.
        - Measure intensity information for the two other channels.

Experimental Data Acquisition Pipeline. General Comments.

"The devil is in the details"

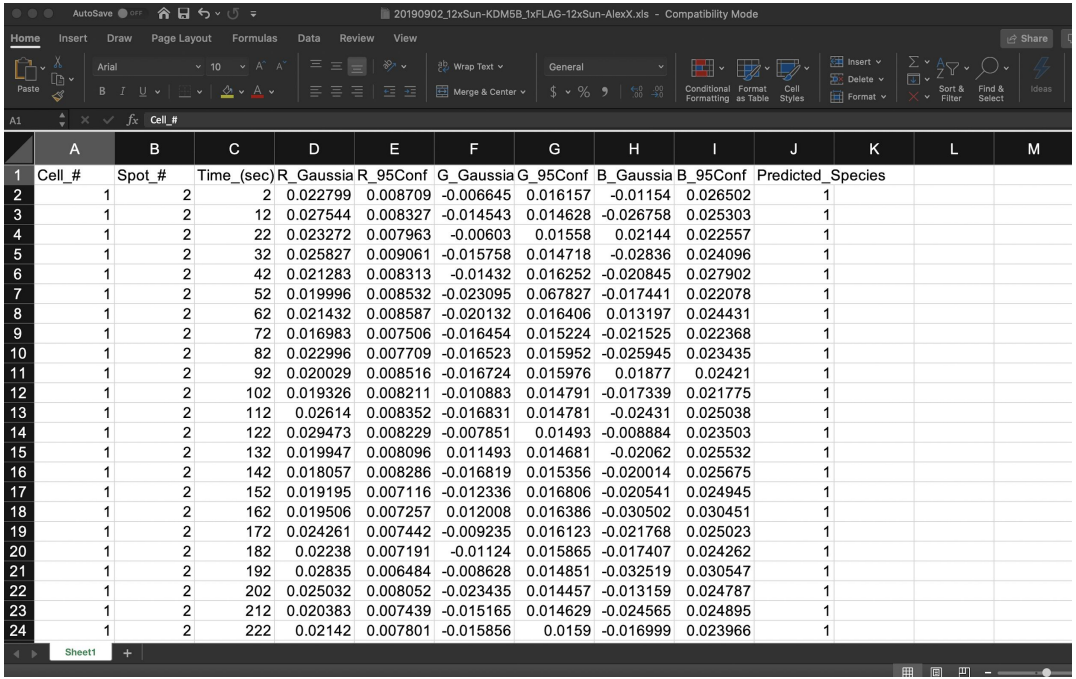
- The Mathematica code has been developed and optimized during years.

- Every cell is different. (Intensity, spot density, shape, background). For this reason, the code requires manual inputs and visual corroboration to reject false positives and to rescue false negatives.

- For example. Small variations in the selection of thresholds can lead to big differences in the final reported intensities.

- Need to develop another part of the code for tracking RNA in the "Green" (JF549) channel when using PCP + MCP

Experimental data (Format)



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Cell_#	Spot_#	Time_(sec)	R_Gaussia	R_95Conf	G_Gaussia	G_95Conf	B_Gaussia	B_95Conf	Predicted_Species			
2	1	2	2	0.022799	0.008709	-0.006645	0.016157	-0.01154	0.026502	1			
3	1	2	12	0.027544	0.008327	-0.014543	0.014628	-0.026758	0.025303	1			
4	1	2	22	0.023272	0.007963	-0.00603	0.01558	0.02144	0.022557	1			
5	1	2	32	0.025827	0.009061	-0.015758	0.014718	-0.02836	0.024096	1			
6	1	2	42	0.021283	0.008313	-0.01432	0.016252	-0.020845	0.027902	1			
7	1	2	52	0.019996	0.008532	-0.023095	0.067827	-0.017441	0.022078	1			
8	1	2	62	0.021432	0.008587	-0.020132	0.016406	0.013197	0.024431	1			
9	1	2	72	0.016983	0.007506	-0.016454	0.015224	-0.021525	0.022368	1			
10	1	2	82	0.022996	0.007709	-0.016523	0.015952	-0.025945	0.023435	1			
11	1	2	92	0.020029	0.008516	-0.016724	0.015976	0.01877	0.02421	1			
12	1	2	102	0.019326	0.008211	-0.010883	0.014791	-0.017339	0.021775	1			
13	1	2	112	0.02614	0.008352	-0.016831	0.014781	-0.02431	0.025038	1			
14	1	2	122	0.029473	0.008229	-0.007851	0.01493	-0.008884	0.023503	1			
15	1	2	132	0.019947	0.008096	0.011493	0.014681	-0.02062	0.025532	1			
16	1	2	142	0.018057	0.008286	-0.016819	0.015356	-0.020014	0.025675	1			
17	1	2	152	0.019195	0.007116	-0.012336	0.016806	-0.020541	0.024945	1			
18	1	2	162	0.019506	0.007257	0.012008	0.016386	-0.030502	0.030451	1			
19	1	2	172	0.024261	0.007442	-0.009235	0.016123	-0.021768	0.025023	1			
20	1	2	182	0.02238	0.007191	-0.01124	0.015865	-0.017407	0.024262	1			
21	1	2	192	0.02835	0.006484	-0.008628	0.014851	-0.032519	0.030547	1			
22	1	2	202	0.025032	0.008052	-0.023435	0.014457	-0.013159	0.024787	1			
23	1	2	212	0.020383	0.007439	-0.015165	0.014629	-0.024565	0.024895	1			
24	1	2	222	0.02142	0.007801	-0.015856	0.0159	-0.016999	0.023966	1			

RNA tracked for more than 41 frames (5 seconds imaging interval x 41 = 200 seconds)
Average length 100s to 1000 of seconds
Translation signals sampled every ~10 seconds (every 2 frames)

Experimental data

Construct	No of cells	No of Total Measurements
1xFLAG-12xSun-AlexX-MS2	3 (imaged with 12xSun-KDM5B-MS2)	206 RNA tracked
12xSun-KDM5B-MS2	3 (imaged with 1xFLAG-12xSun-AlexX-MS2)	206 RNA tracked
1xSun-12xFLAG-Alex-MS2	0	0
12xFLAG-KDM5B-MS2	0	0
smFLAG-KDM5B-PP7 (Priority)	0	0
smFLAG-KDM5B-MS2	4	322 RNA tracked
smFLAG-ActB-MS2	3	288 RNA tracked
smFLAG-H2B-MS2	5	286 RNA tracked
smHA-FSS-MF-Alex/XXL-MS2	4	Not tracked yet

File organization. (Single-plasmid measurements).

File organization. (Multi-plasmid measurements).

Experimental data

Multiplexing constructs

3 Plasmids:
1x-FLAG-12XSun-Alex-MS2
12xSun-KDM5B-MS2
12xFLAG-H2B-MS2

4 Plasmids:
24xSun-Kif18b-Xbp1-PP7
12xFLAG-KDM5B-PP7
1xSun-12FLAG-Alex-MS2
12xFLAG-H2B-MS2

Simulation Approach.

Multiplexing constructs

(Training data: 1st half-data. Test data: 2nd Half-data)

1) Pure simulation using Luis's parameters and construct designs. -> M_SIM-0.

2) Retrain simulations with new data to get new {ki} and {ke}. M_SIM-1.
(Include positions x(t), and y(t))

3) Learn ML model from 1st data set. M_ML-1.

4) Mix of real and simulated data.

Zach's Previous Work

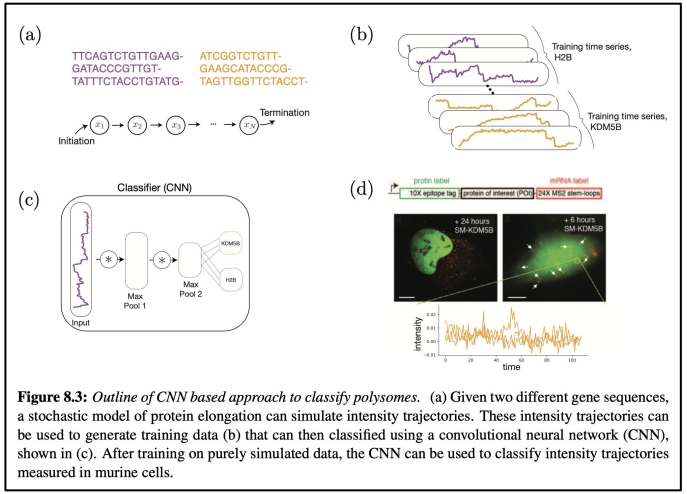


Figure 8.3: *Outline of CNN based approach to classify polysomes.* (a) Given two different gene sequences, a stochastic model of protein elongation can simulate intensity trajectories. These intensity trajectories can be used to generate training data (b) that can then classified using a convolutional neural network (CNN), shown in (c). After training on purely simulated data, the CNN can be used to classify intensity trajectories measured in murine cells.

Machine Learning Codes

DISSERTATION

INTEGRATING DISCRETE STOCHASTIC MODELS WITH SINGLE-CELL AND SINGLE-MOLECULE EXPERIMENTS

Submitted by

Zachary R. Fox

Graduate Degree Program in Bioengineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

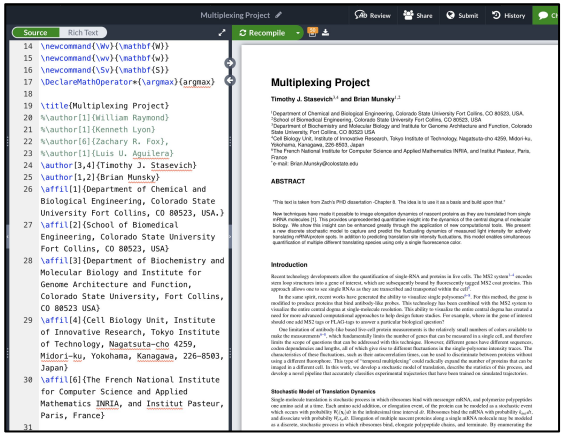
Colorado State University

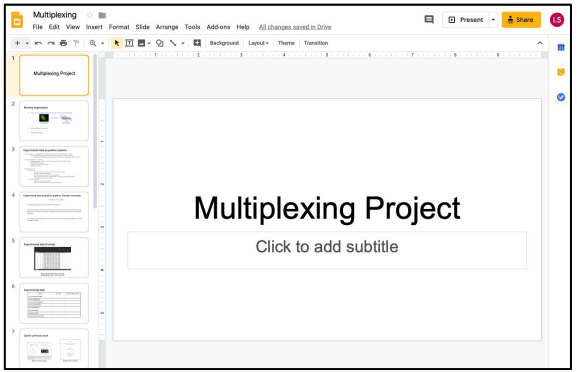
Fort Collins, Colorado

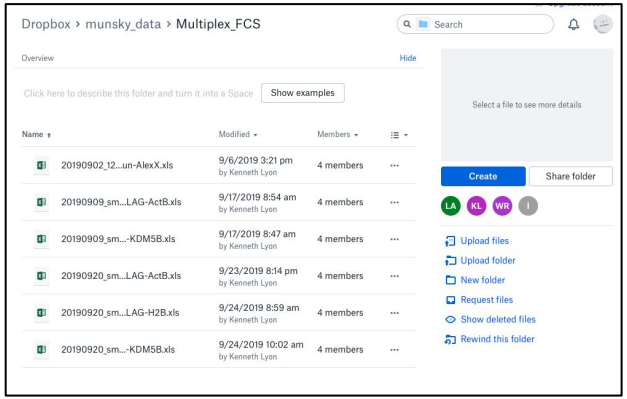
Summer 2019

Chapter 8 from his thesis.

Overall Data Organization







Overleaf- Draft document

This presentation

Dropbox with Excel files.

The big picture

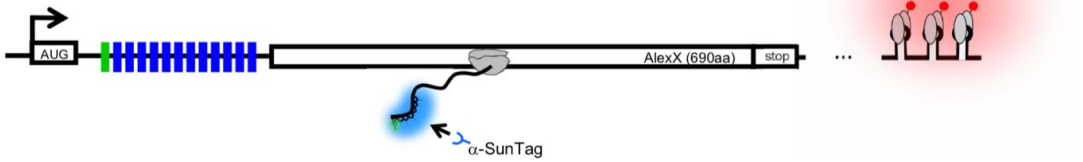
Images -> Tracking -> Trajectory Data -> Parameter Estimation.

->Classification

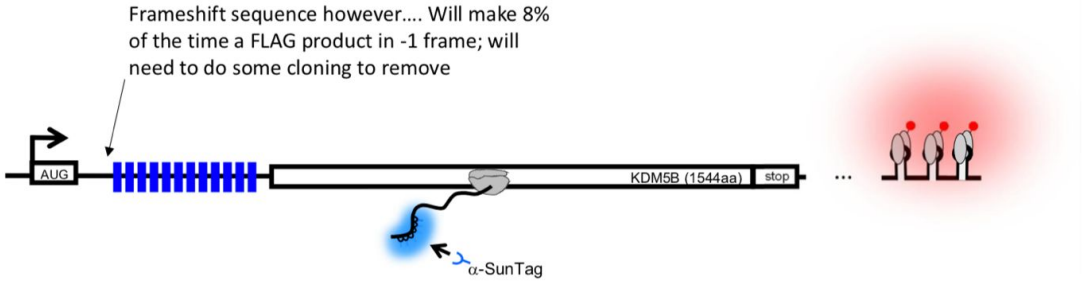
Thank you!

Gene sequences

12xSun-KDM5B (1942)



1xFLAG-12xSun-AlexX (1061)



Code that uses .dna files

Expected dwell time assuming ke = 3aa/sec

Minimal time (total frames) needed to observe the dwell time.

12xSun-KDM5B (1942 aa).

Av_dwell = 1942 / 3 = 647 sec (Minimal time)

1xFLAG-12xSun-AlexX (1061 aa).

Av_dwell = 1061 /3 = 354 sec (Minimal time)

Calculating the dwell times for the two sequences under different experimental conditions.

Conditions.

- Sampling rate. 7,14,28,35 seconds.
- No Spots. 10,20,50,100.
- Total no frames is fixed to 300, 100.

Assumptions:

- Elongation rate is 3 aa/sec.
- Initiation rate 1/30 sec.
- No pause in the system.

Report results.

Table of dwell times.

Results - dwell times.

12xSun-KDM5B min frames(647s)

	1 sec	5 sec	10 sec	20 sec
10 spots	79	330	520	480
20 spots	89	340	500	500
50 spots	87	360	480	500
100 spots	89	330	470	520

No Spots = 500 rep, SR = 0.5 s, Frames = 4000

Dwell time = 610s

1xFLAG-12xSun-AlexX min frames(354s)

	1 sec	5 sec	10 sec	20 sec
10 spots	71	240	300	340
20 spots	84	260	310	320
50 spots	84	230	290	320
100 spots	84	250	320	320

No Spots = 500 rep, SR = 0.5 s, Frames = 4000

Dwell time = 324s

Machine learning

ML - classify SSA trajectories.

Gene Sequences