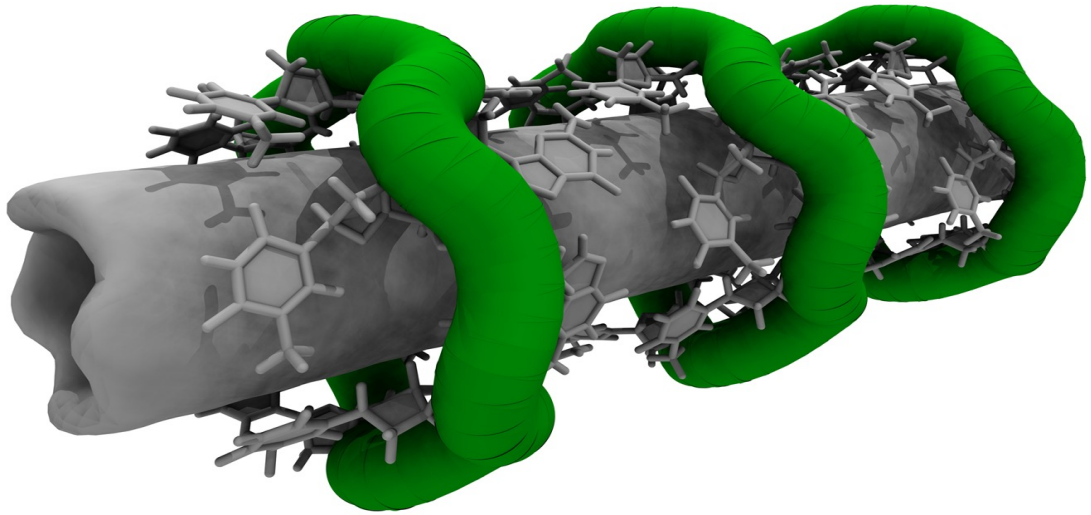
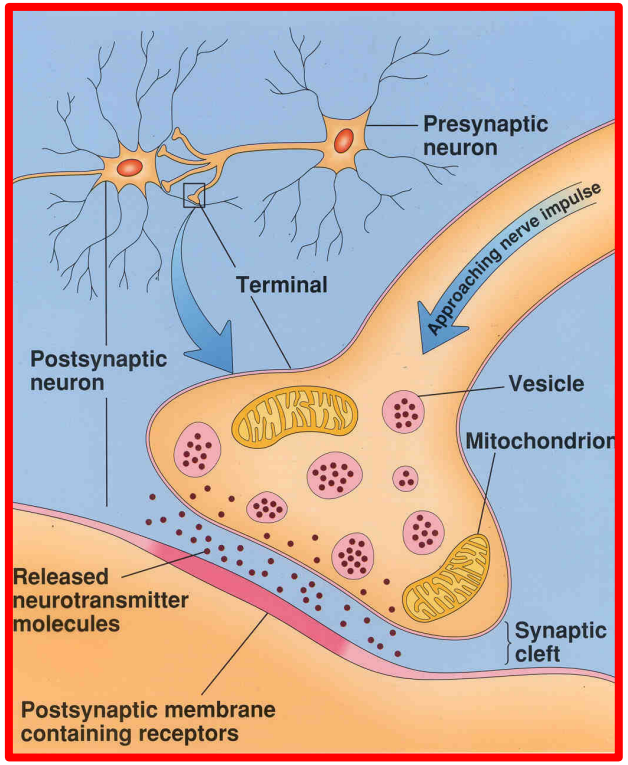
Learning and predicting DNA sequences in DNA-nanotube conjugates with high response to serotonin

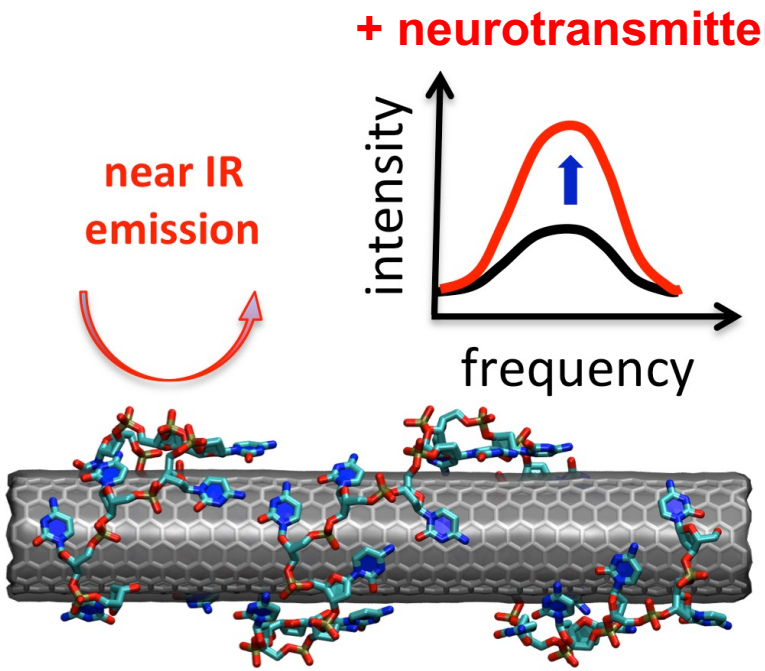
Payam Kelich
Department of Chemistry and Biochemistry, University of Texas at El Paso

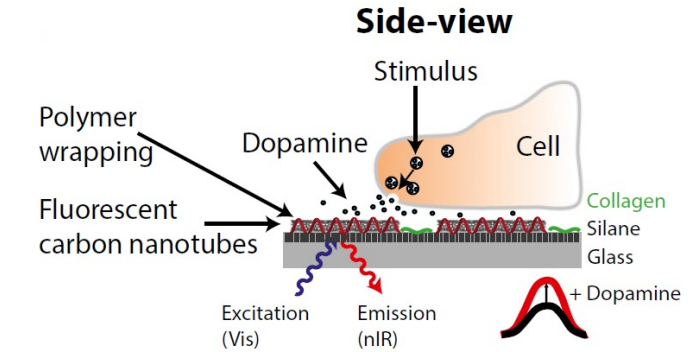


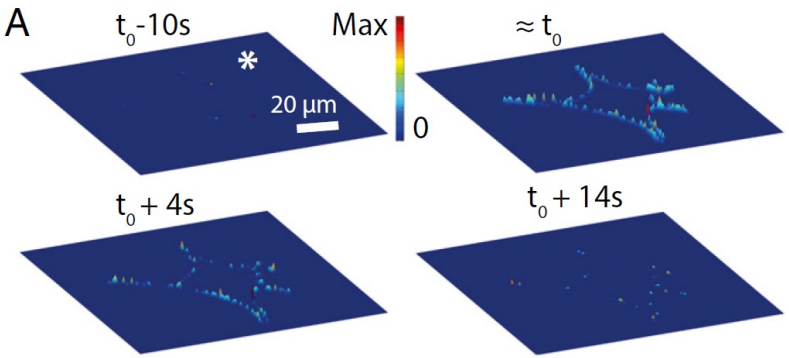
23 September 2021, TACC

Development of carbon nanotube-based optical sensors of neurotransmitters





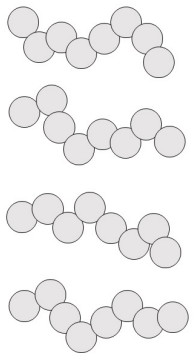




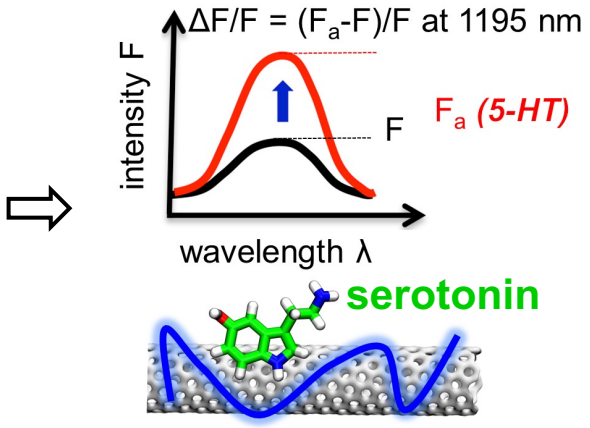
- **neurotransmitter** release is the basis of neurotransmission in chemical synapses in the brain
- new sensors are needed to detect neurotransmitters
- nanomaterials made of DNA-wrapped carbon nanotubes can emit light and detect neurotransmitters

A. Beyene+, A. Alizadeh-Mojarad+,…L. Vuković*, M. Landry*, *Nano Letters (2018)*.

How can we predict DNA sequences to detect other neurotransmitters?



10^{10} DNA candidates



Challenges:

- vast sequence space

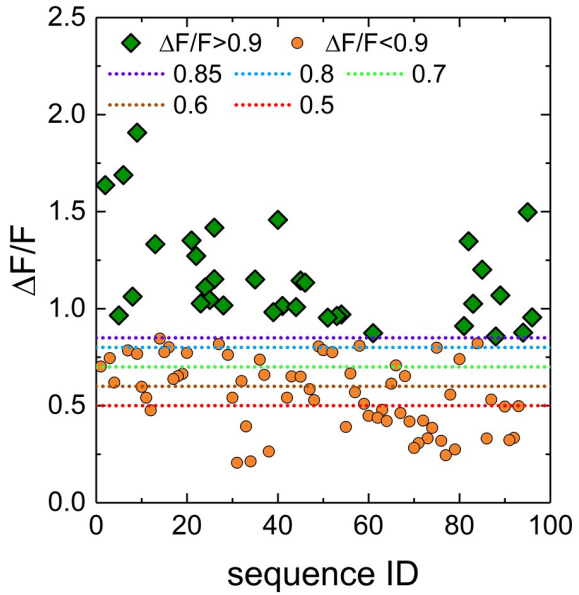
- number of systems that can be optically measured is small (≤100)

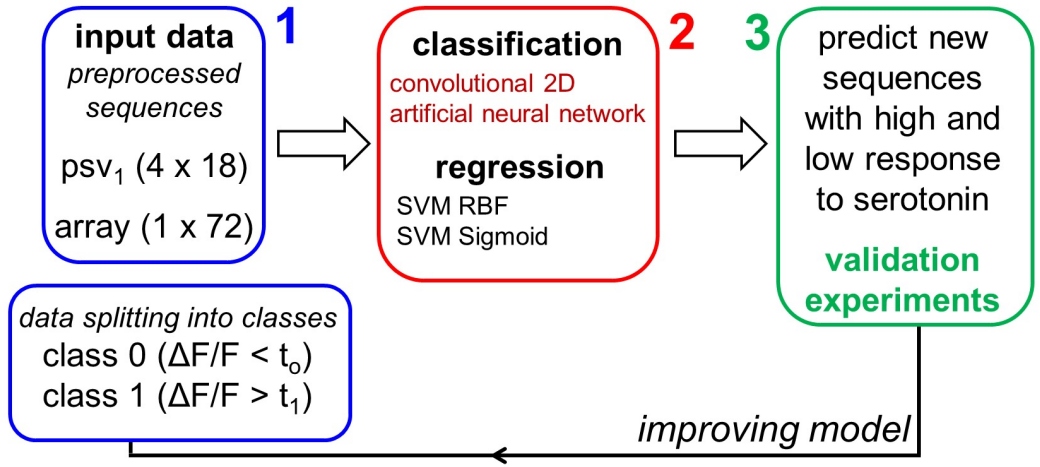
- datasets containing DNA sequences and their measured optical signals ΔF/F can be used to train artificial intelligence models

class 1

class 0

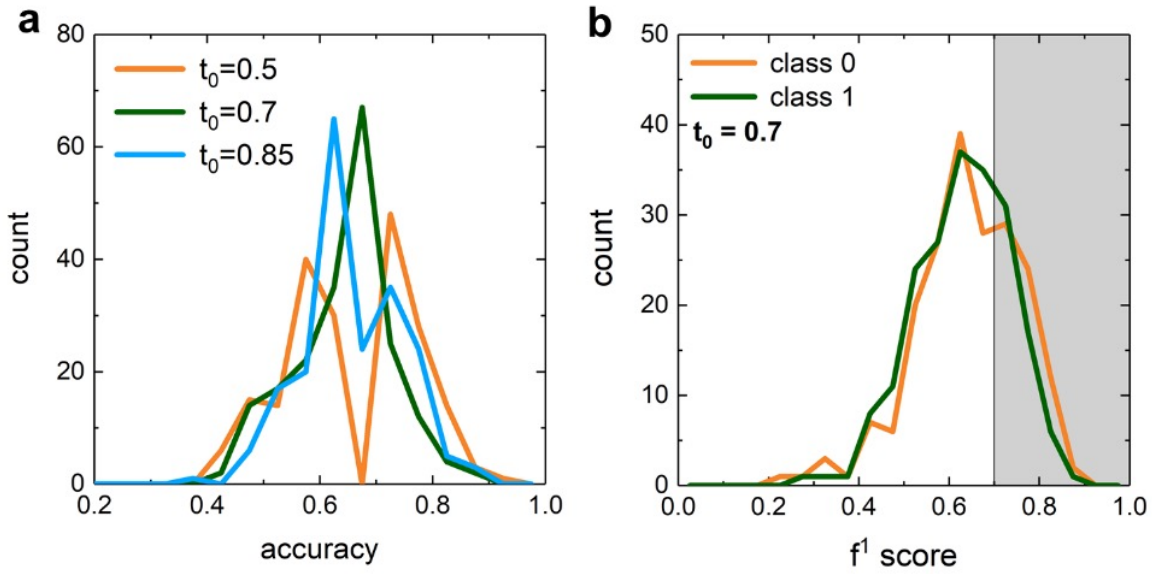
Landry lab (UC Berkeley)





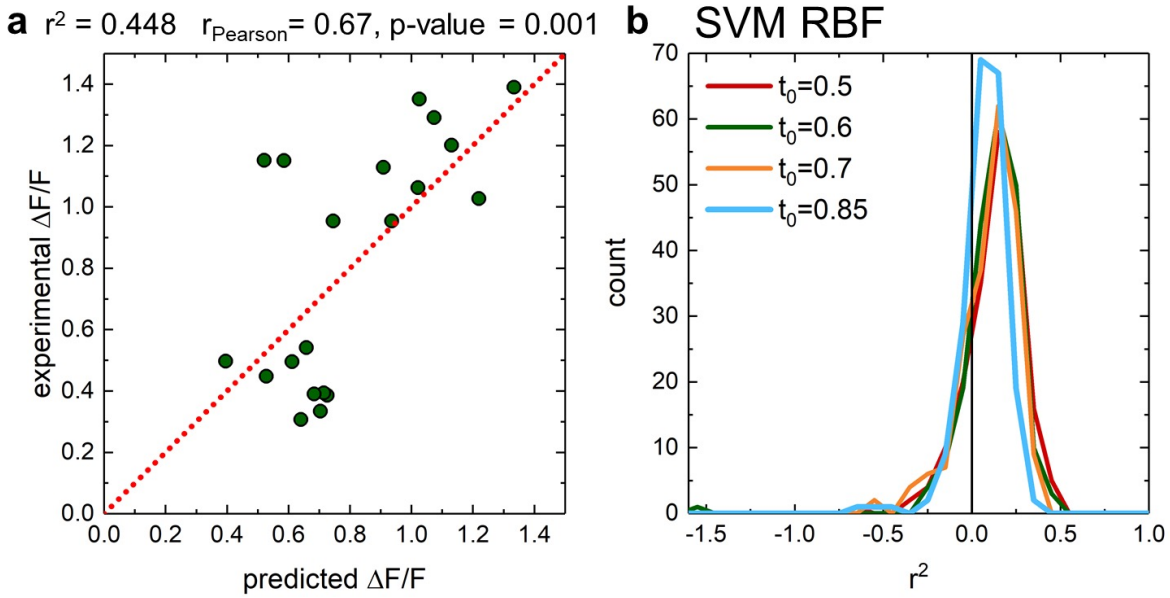
P. Kelich+, S. Jeong+, N. Navarro+, J. Adams, X. Sun, H. Zhao, M. Landry*, L. Vuković*. Machine learning enables discovery of DNA-carbon nanotube sensors for serotonin, *bioRxiv* (2021).

Neural network classifiers and support vector machine regression models are stochastic



- classification models are trained on sparse data and are stochastic (variable accuracy)

- SVM regression models were trained to predict ΔF/F values based on sequence input
- models are also stochastic, some of acceptable quality

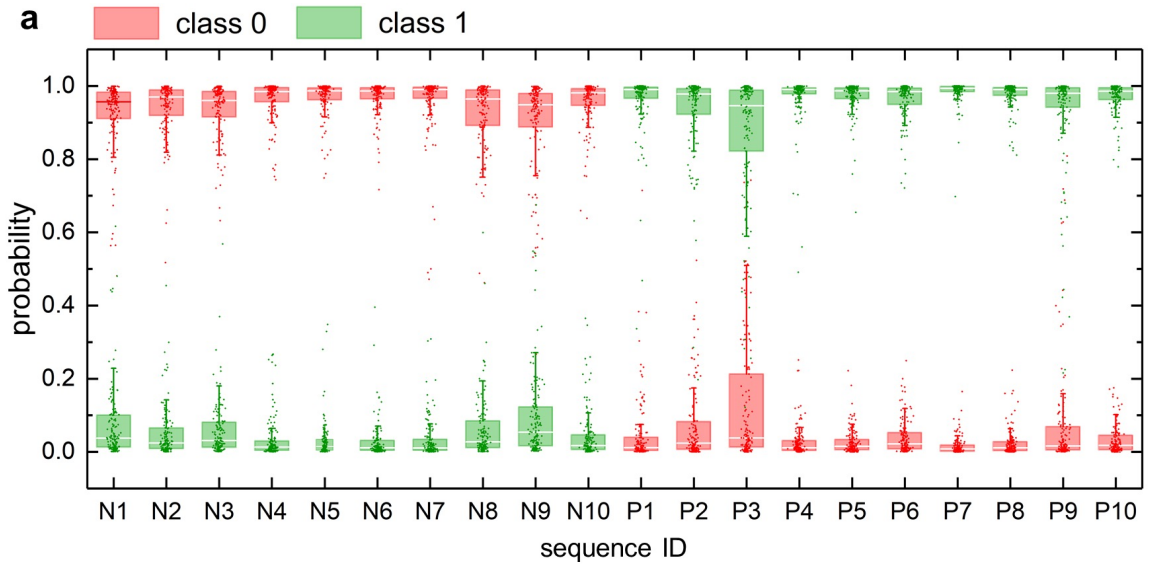


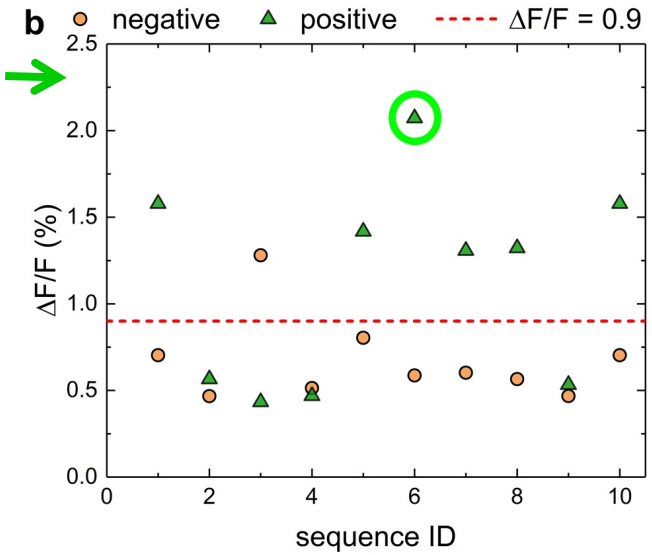
Predictions of ensembles of best classification and regression models

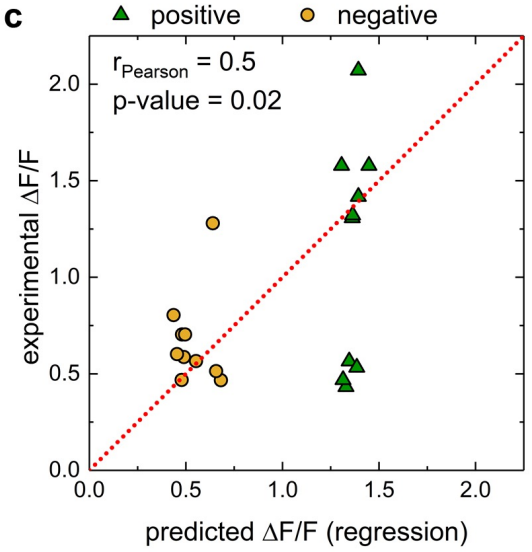
rank 3,000 experimental sequences (SELEC) by the best regression models

examine their probabilities to be high/low responders by the best classification models

experimental validation of high/low responders







Correct predictions:

- 90% for low response sequences
- 60% for high response sequences

- 5 new sensors discovered in all the validation experiments

Groups

UT El Paso

Prof. Lela Vukovic

H. Zhao

UC Berkeley Landry Lab

Prof. Markita Landry

S. Jeong

N. Navarro

J. Adams

X. Sun

M. Landry

NSF-CBET (2106587)

Texas Advanced Computing Center (TACC)

Thank you!