

# Open science in drug design : Analysis and visualization of an open experimental dataset of FDA approved cancer drugs

A hand-drawn diagram of a DNA double helix. The structure consists of two interlocking spiral chains. Several labels are written in red ink across the diagram:

- Python**: Located near the top left of the left chain.
- gitHub**: Located near the bottom left of the left chain.
- xm1**: Located near the middle right of the right chain.
- Yope**: Located near the bottom right of the right chain.

The DNA structure itself is drawn with black lines and features red ovals representing nucleotides at the junctions where the two chains meet.

Sunday 1-2 pm  
Open Science  
9<sup>th</sup> Floor 902

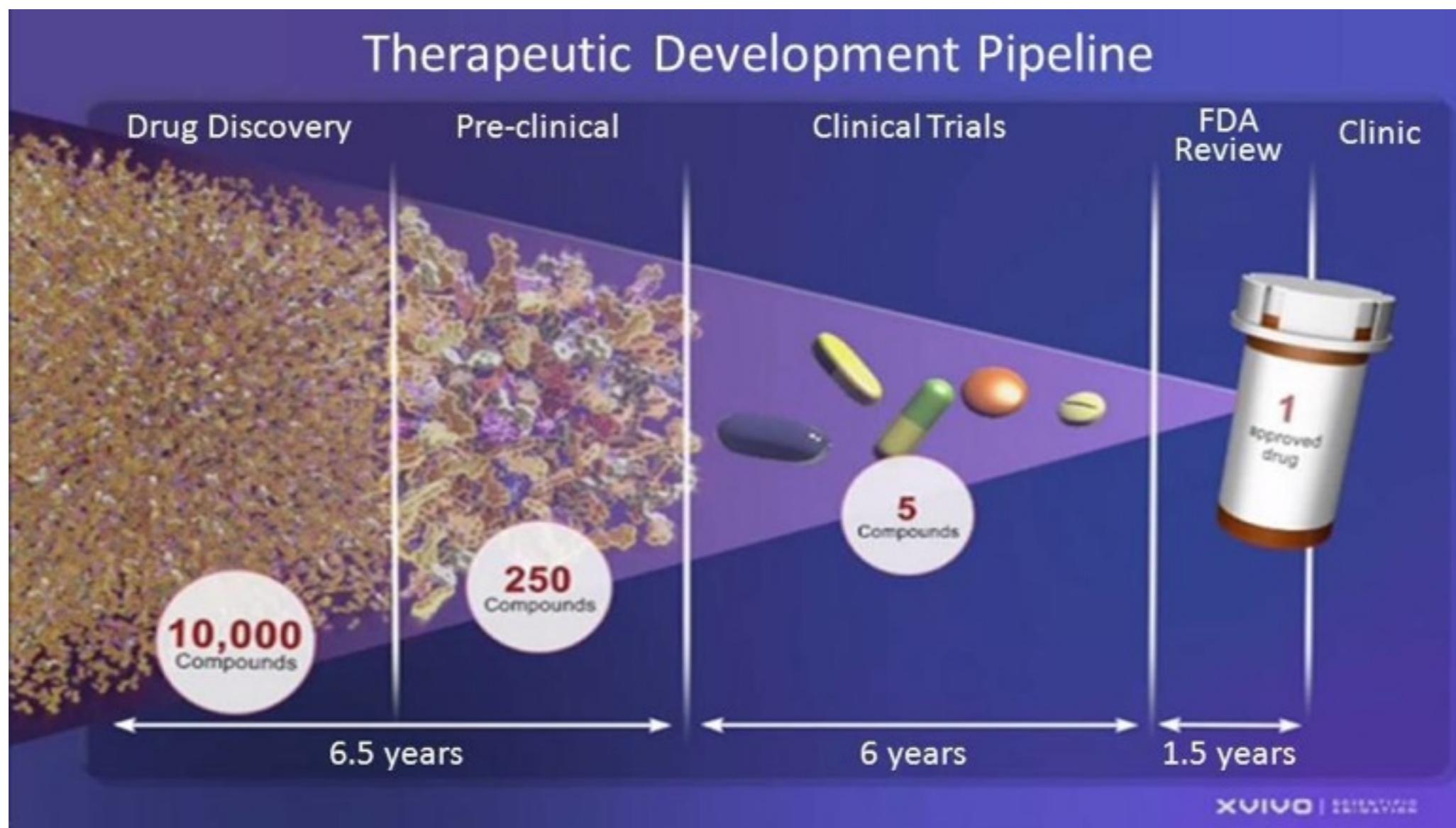
by @sonyahans

# Goals of this learning lab:

- **Explore a** real world case of open data **and open source code in wetlab biology.**
- **Discuss** difficulties of open data/analysis **(despite long-term benefits to the community) even in a case where standard software development best practices are easy to implement.**
- **Analyze** simulated and real data **with an open analysis pipeline, and understand differences between real and simulated data.**
- **Learn about awesome Python tools!**

# How does ‘Drug Discovery’ work?

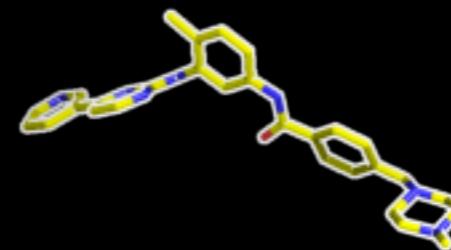
# How does ‘Drug Discovery’ work?



# How does ‘Drug Discovery’ work?



Imatinib (Gleevec)  
FDA approved 2001



# How does ‘Drug Discovery’ work?



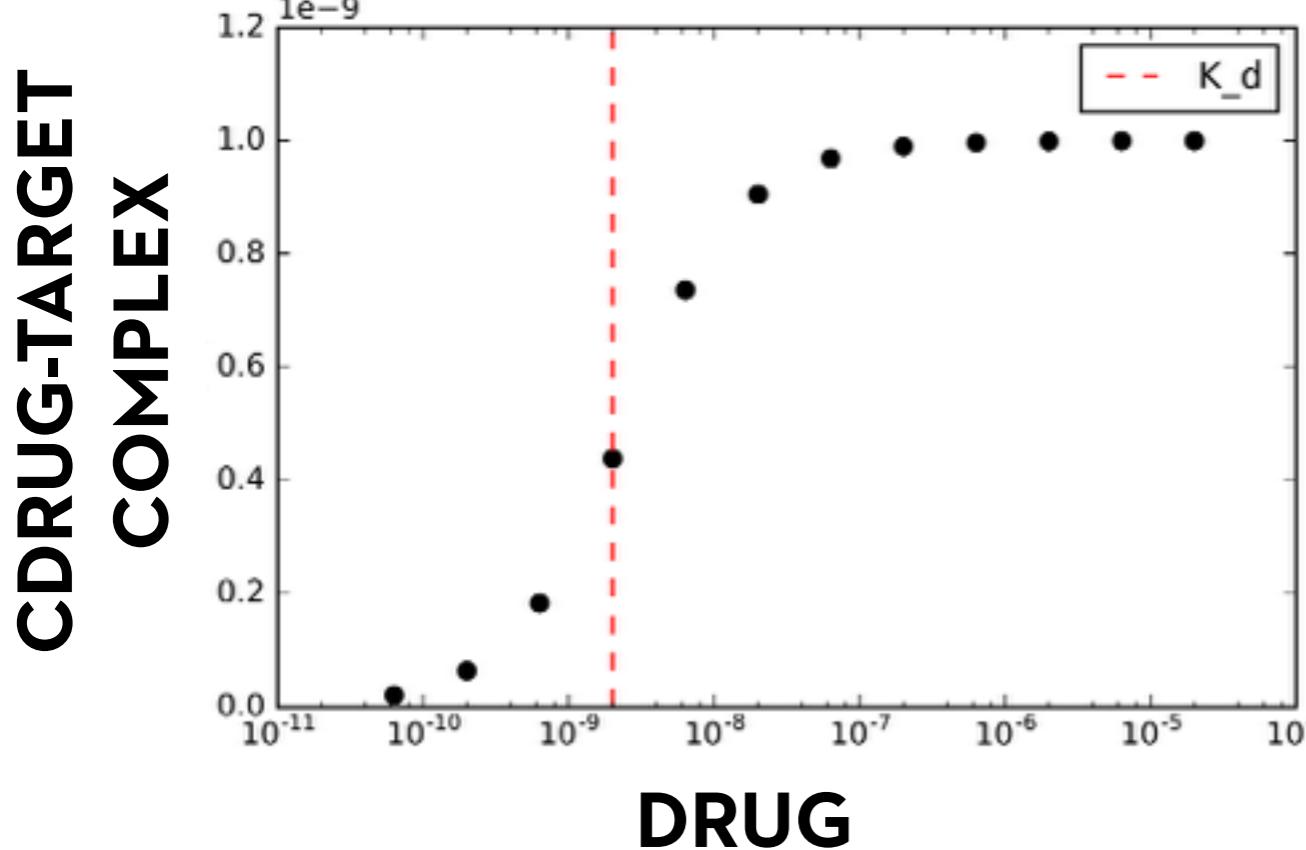
Imatinib (Gleevec)  
FDA approved 2001



A bit of an introduction  
to our dataset.

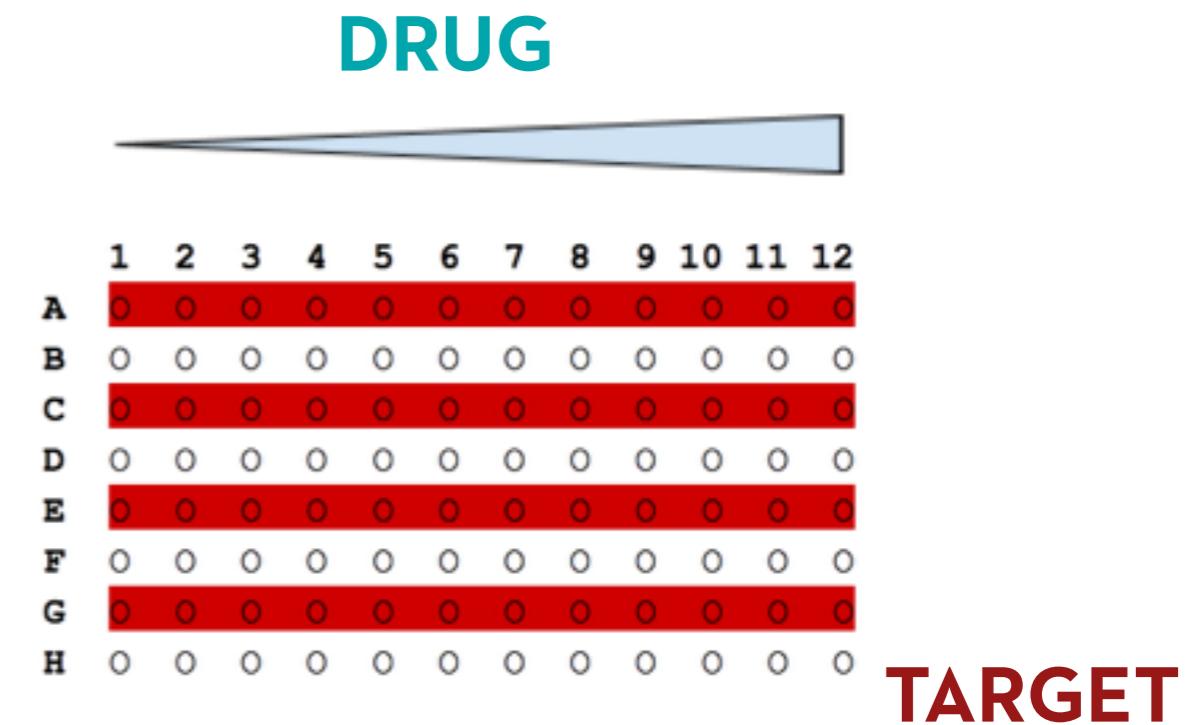
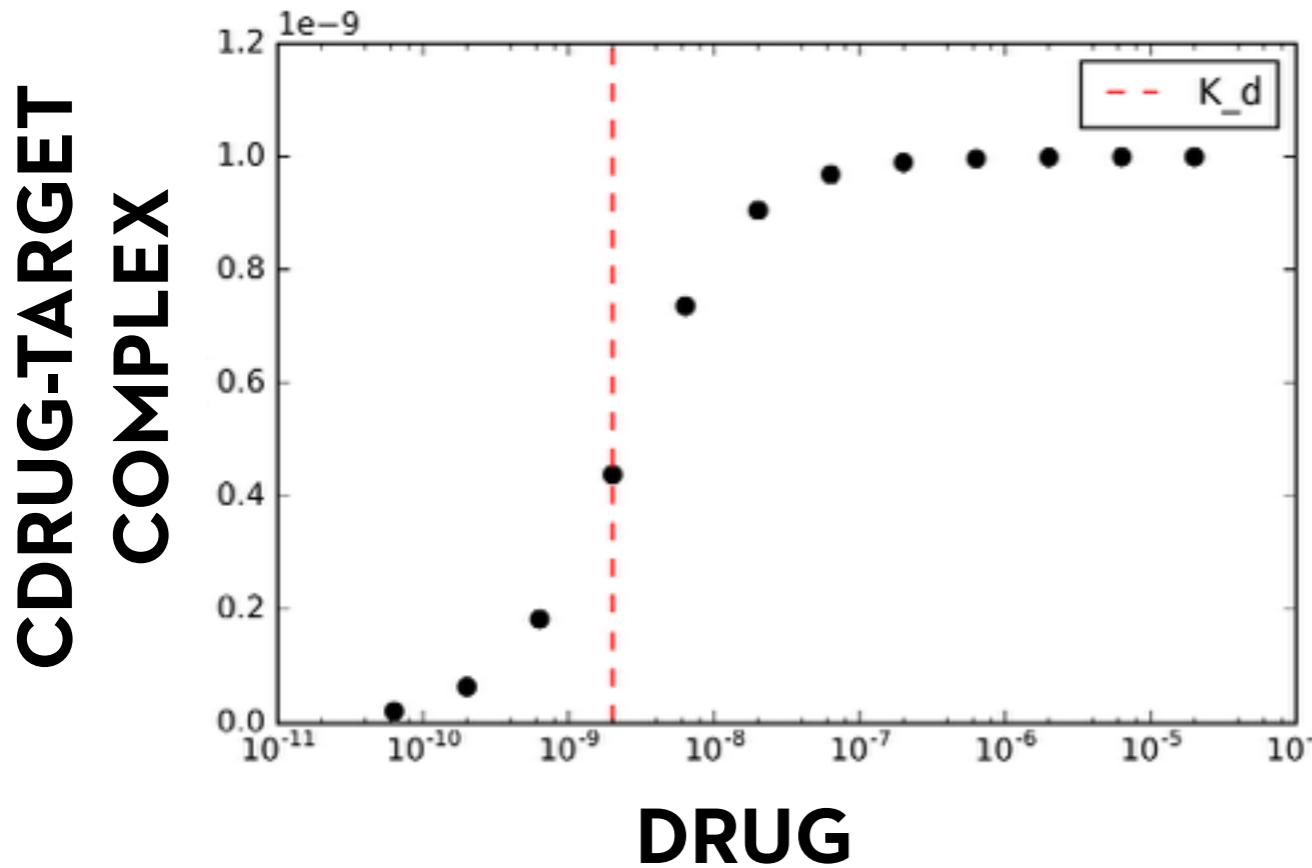
# The ‘ligand binding assay’: A standard experiment in drug design.

**Measuring the concentration of our drug-target complex as a function of drug concentration, tells us the affinity of that drug for that target!**



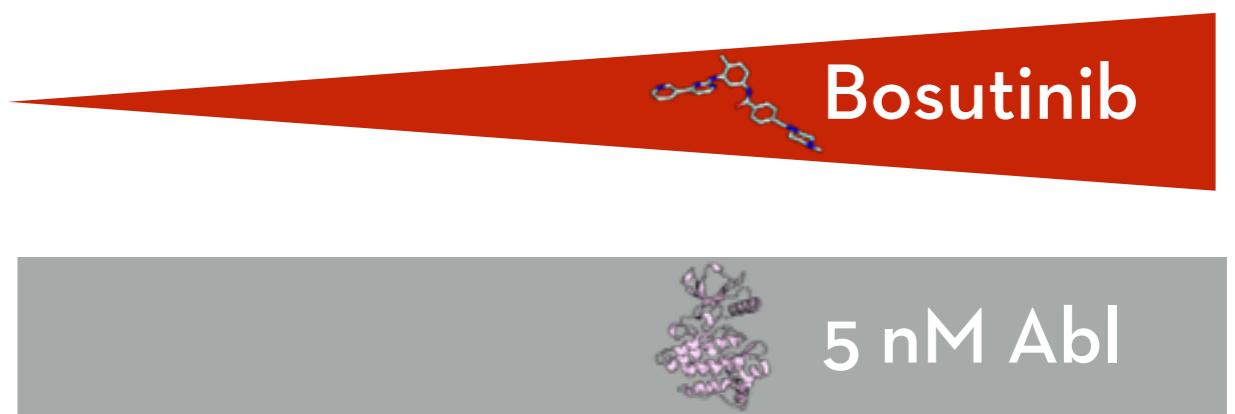
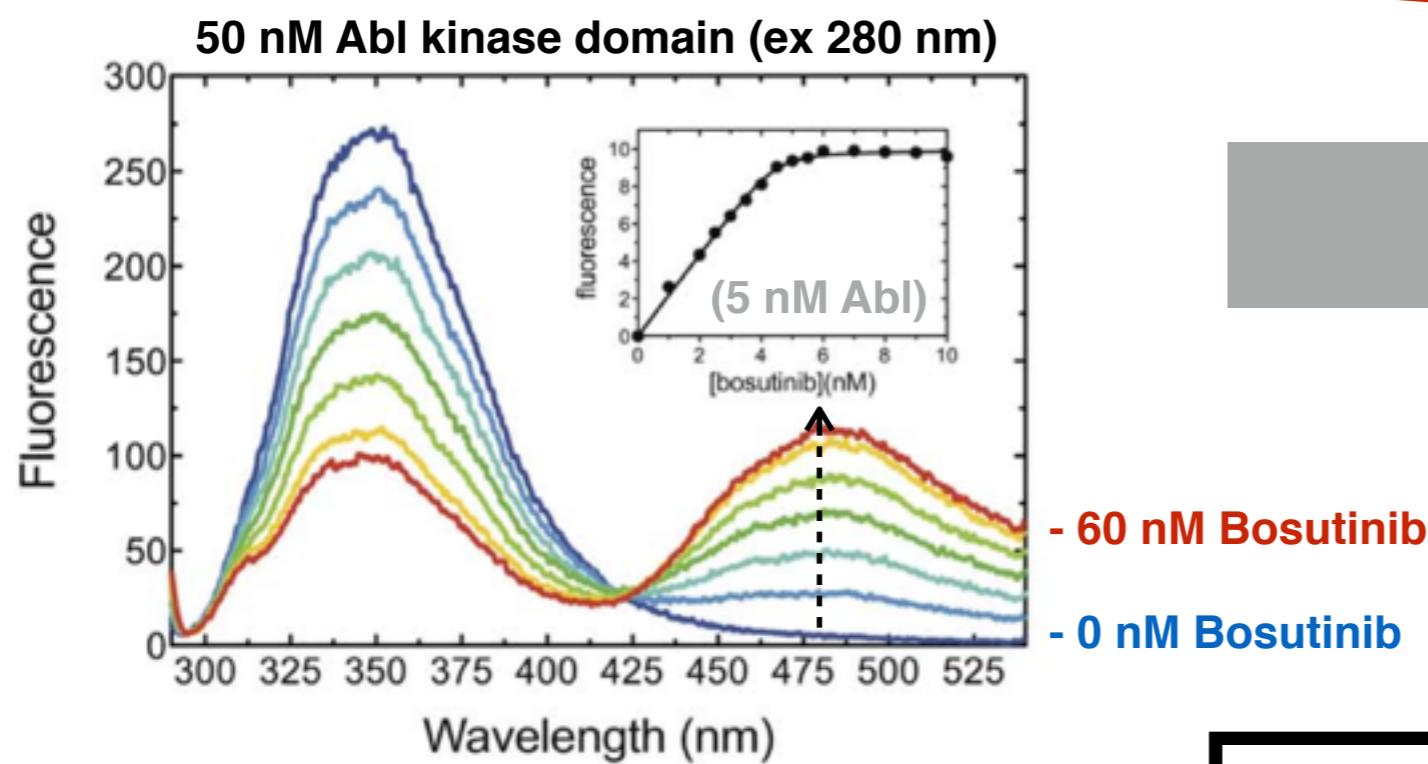
# The ‘ligand binding assay’: A standard experiment in drug design.

**Measuring the concentration of our drug-target complex as a function of drug concentration, tells us the affinity of that drug for that target!**



TARGET

# Nick Levinson discovers **bosutinib** greatly increases fluorescence upon binding kinase.

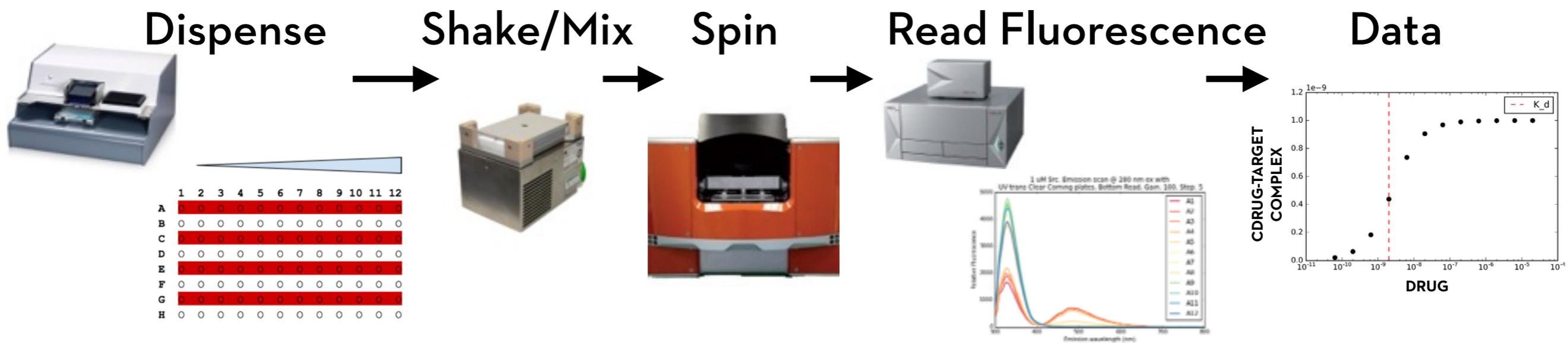


$$K_d = \sim 200 \text{ pM}$$





# A robotic binding assay.



# Using HP printing technology to ‘print’ small volumes:



# Questions?

**Let's simulate the data that  
we would expect from this!**

[https://github.com/sonyahanson/mozfest-  
assaytools/](https://github.com/sonyahanson/mozfest-assaytools/)

More in depth into our actual  
experimental pipeline.

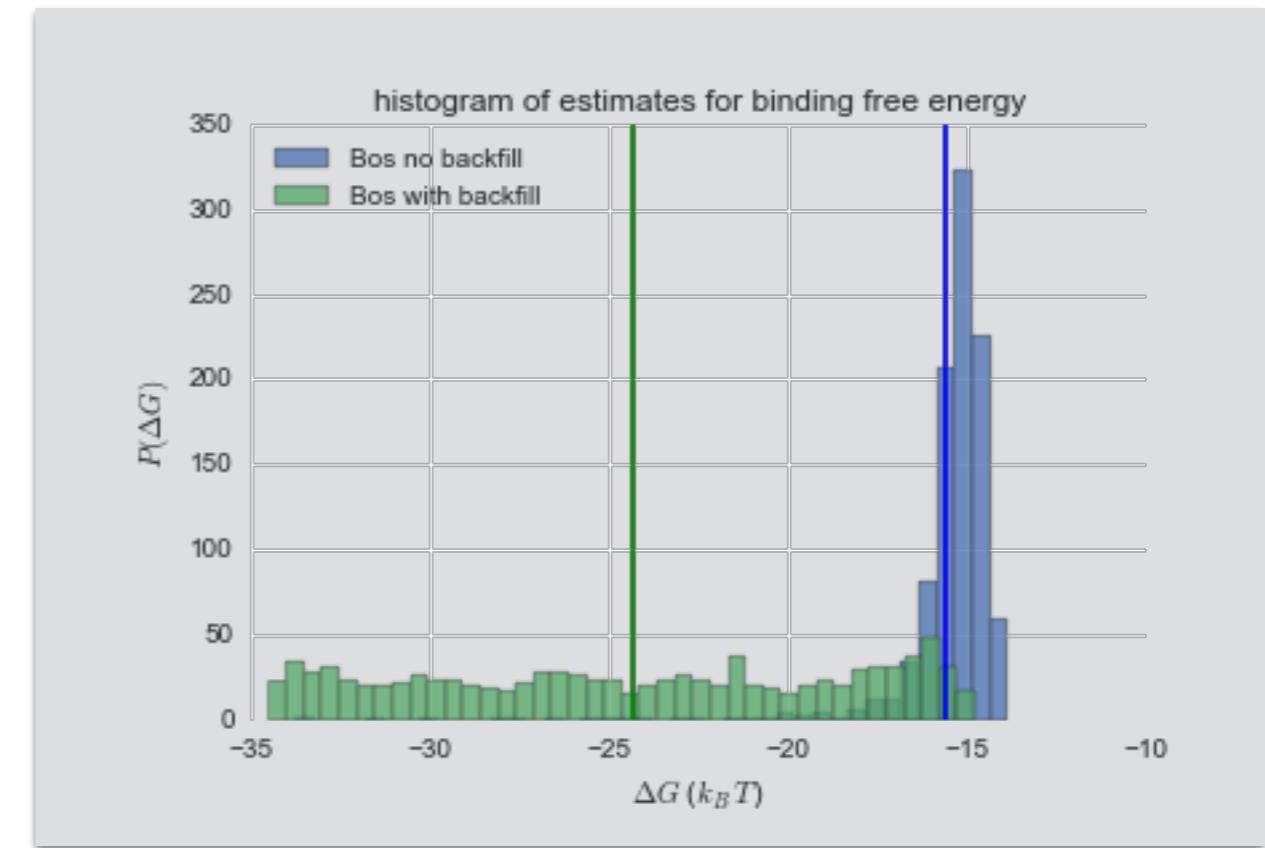
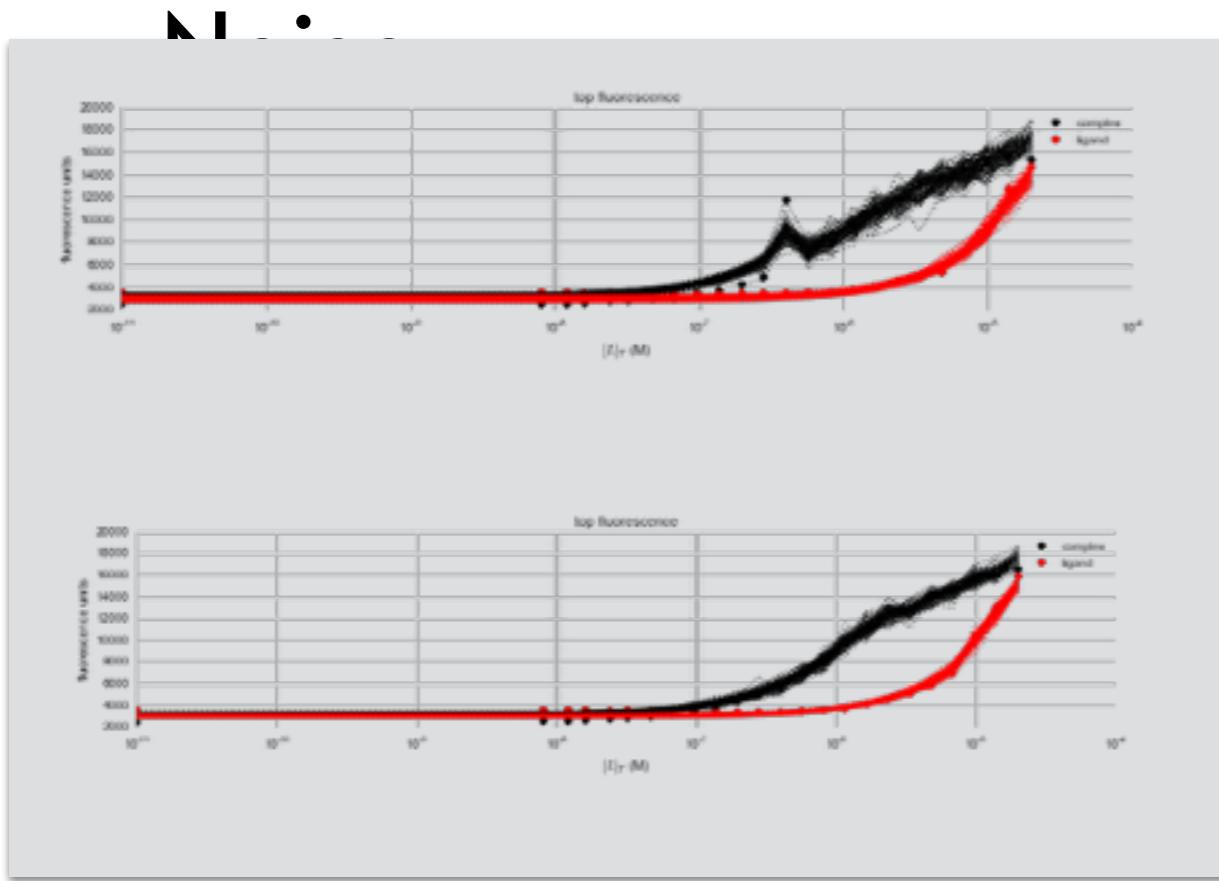
Assaytools.  
Manuscript repo.

Example xml file.

# Some of the unfortunate realities of real experimental datasets.

- Noise
- Outlier rejection

# Some of the unfortunate realities of real experimental datasets.



# Some of the unfortunate realities of real experimental datasets.

- Noise
- Outlier rejection
- ‘OVER’s
- Data file formats

# Some of the unfortunate realities of real experimental analysis.

- Instruments each have their own output file formats.
- Experimentalists are generally most comfortable doing data analysis in programs like Excel and Prism.
- Anything else?

How can we make open data  
and analysis easier for the  
experimental sciences?

# Robots and open hardware! THE FUTURE?

- **Transcriptic**  
<https://www.transcriptic.com/>
- **OpenTrons**  
<https://opentrons.com/>
- **Autoprotocol**  
<https://github.com/autoprotocol/autoprotocol-python>