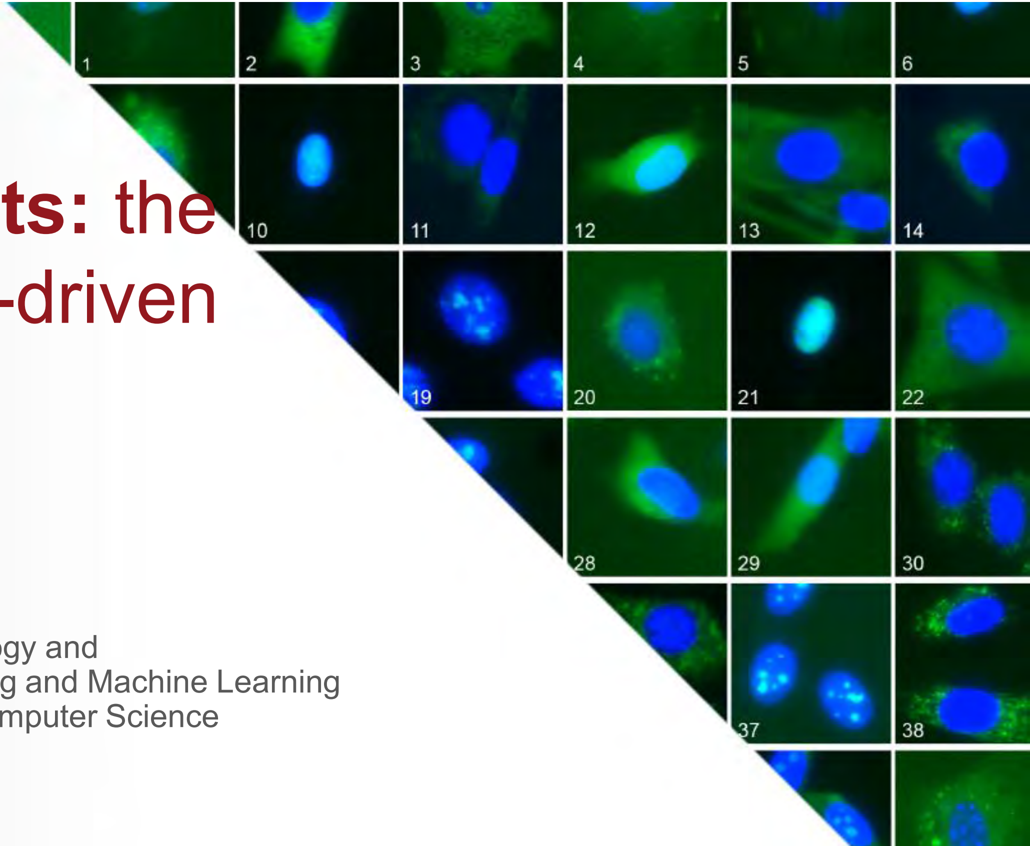
Computational Biology Department | Carnegie Mellon University School of Computer Science



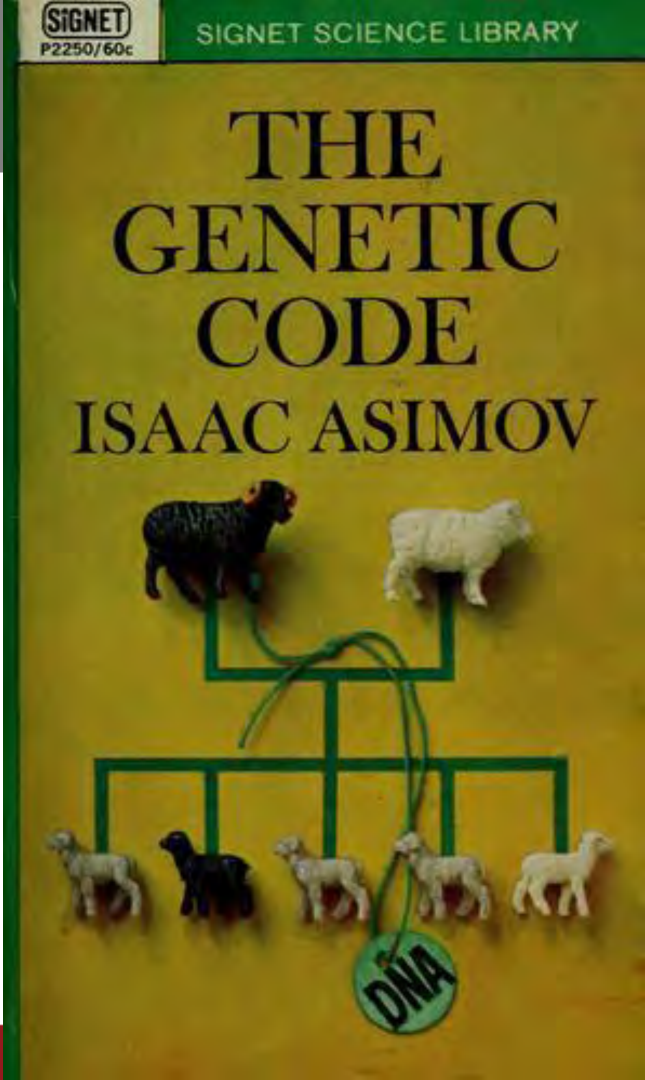
Self-driving instruments: the need for closed loop, AI-driven biomedical research

Robert F. Murphy
Ray & Stephanie Lane Professor of Computational Biology and
Professor of Biological Sciences, Biomedical Engineering and Machine Learning
Head, Computational Biology Department, School of Computer Science

March 2020

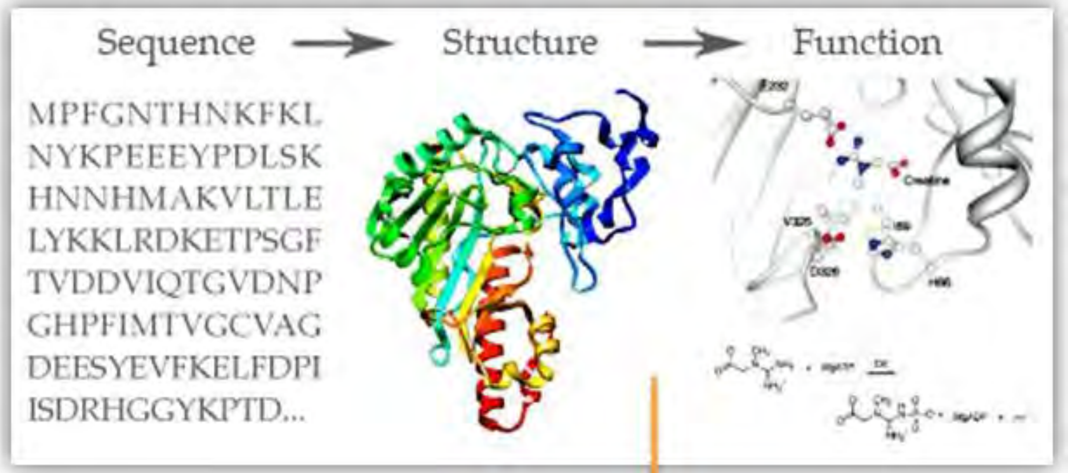
The glory days of reductionism

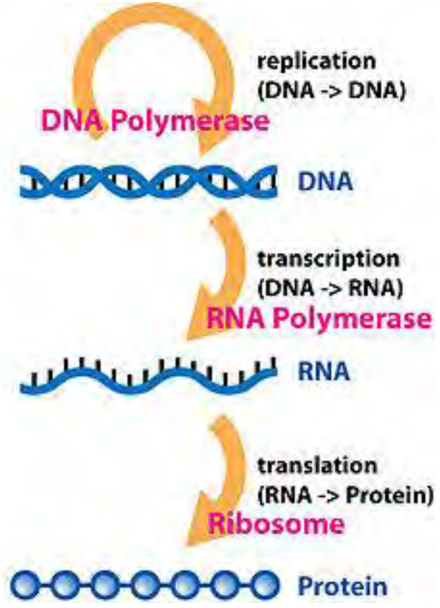
- *Drosophila* Genetics
- Watson-Crick DNA structure
- The Genetic Code
- Biochemistry of Metabolism
- Receptors, channels
- Organelle structure/function
- DNA recombination



Biological laws and principles

- The Central Dogmas
  - DNA >> RNA >> protein
  - Sequence >> Structure >> Function





By the 1970’s, exceptions begin to rule

- Horizontal gene transfer
- Transposons
- Reverse transcriptase
- Introns and alternative splicing
- Prions
- Disordered proteins
- Chaperones
- RNA interference / Non-coding RNA
- Epigenetic modification

Biomedical research: What doesn't work

- Reductionism fails because cells, tissues, and organisms are **complex systems**
- Engineering simplified biological systems to study in order to learn rules fails because **there are no rules**
- Systems biology approaches that do a block of experiments, build a predictive model and try to validate it fail because **you can't prove an empirical model**

Why we need AI-driven biomedical research

- We need to accept
  - biological complexity
  - absence of rules
  - inability to measure everything
- Which means
  - human understanding is not possible
  - we can only build predictive, empirical models
  - we can’t expect to be able to prove, interpret or explain them
  - need continuous AI-driven experimentation to *improve* models not test them
  - hypotheses are useless but specific predictions can be
  - humans can’t be relied upon to decide what experiments to do

Solution?

- Use ***active*** machine learning/iterative design of experiments
- Choose experiments not to ***prove*** model but to ***improve*** model

- **Illustration:**
  Drug Discovery

NATURE CHEMICAL BIOLOGY | VOL 7 | JUNE 2011

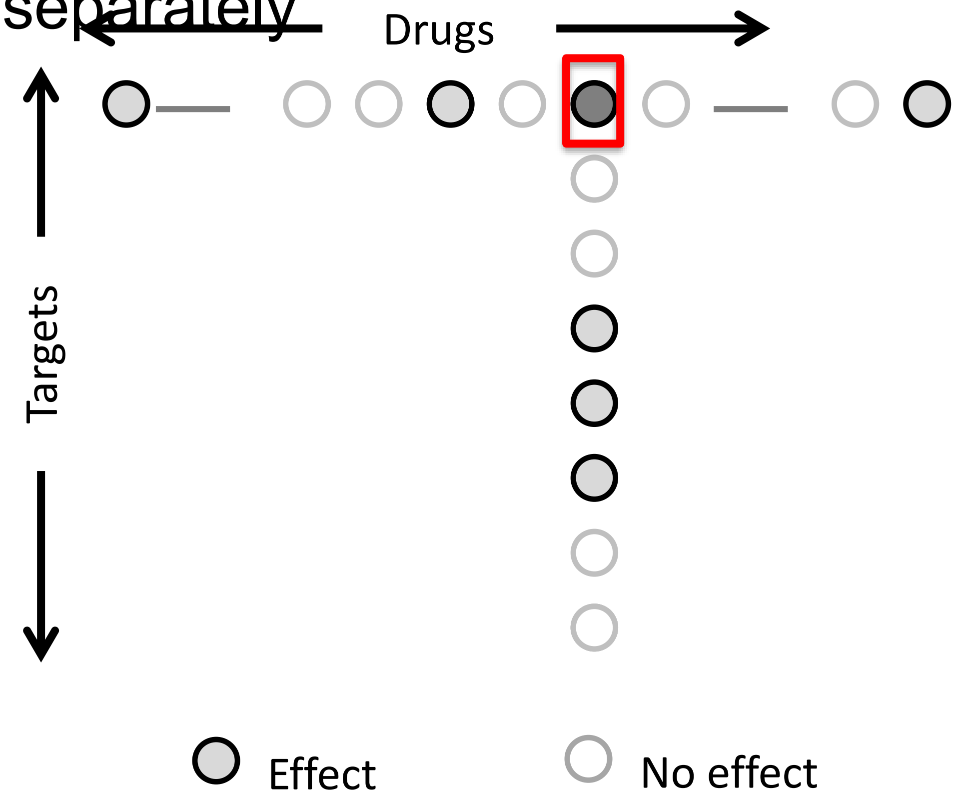
commentary

An active role for machine learning in drug development

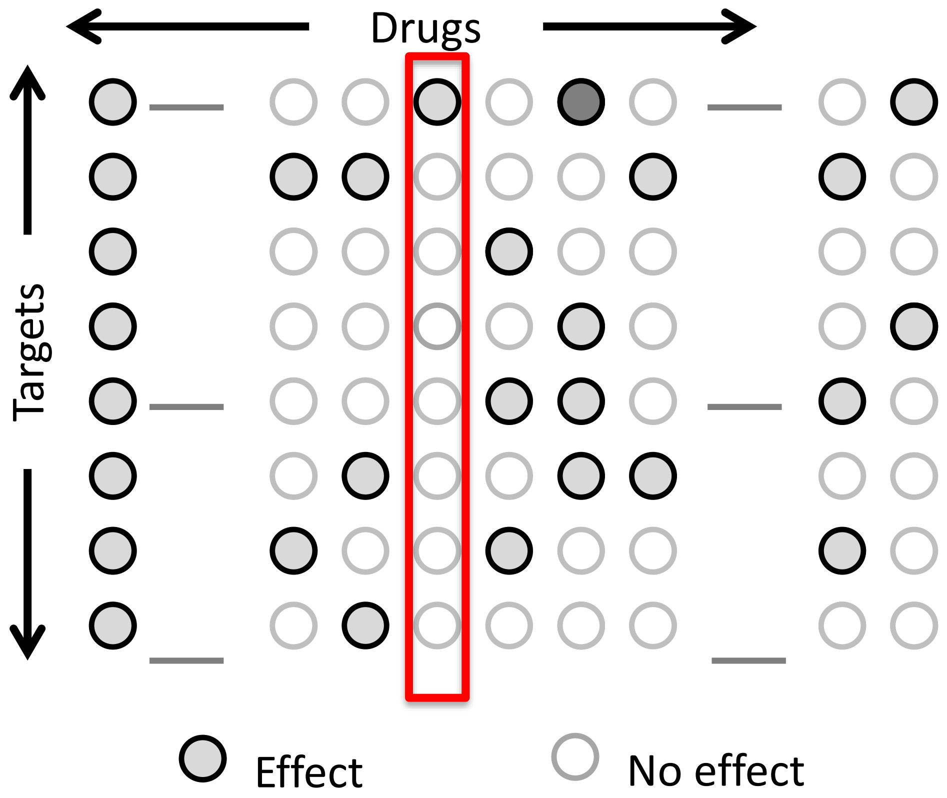
Robert F Murphy

Because of the complexity of biological systems, cutting-edge machine-learning methods will be critical for future drug development. In particular, machine-vision methods to extract detailed information from imaging assays and active-learning methods to guide experimentation will be required to overcome the dimensionality problem in drug development.

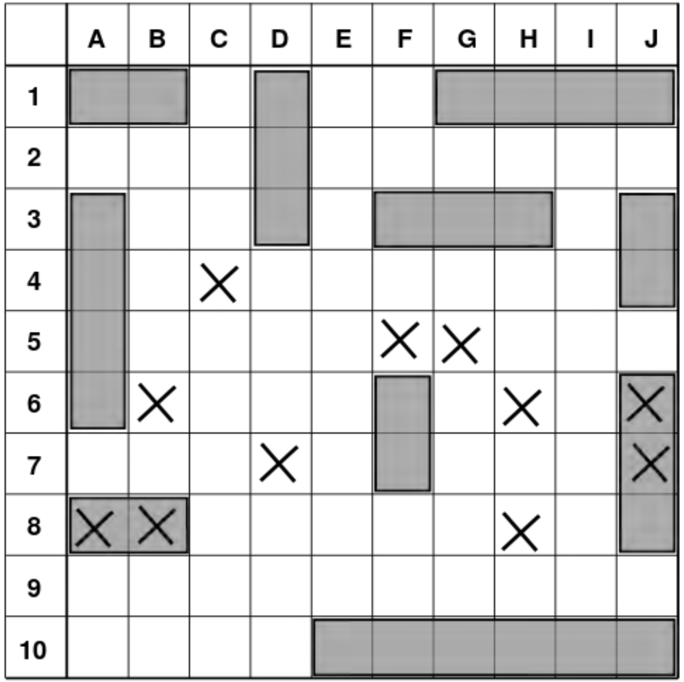
Typical drug development: consider each target separately



Where we'd like to be: measure all drugs for all targets



Playing Battleship with Drugs and Cells

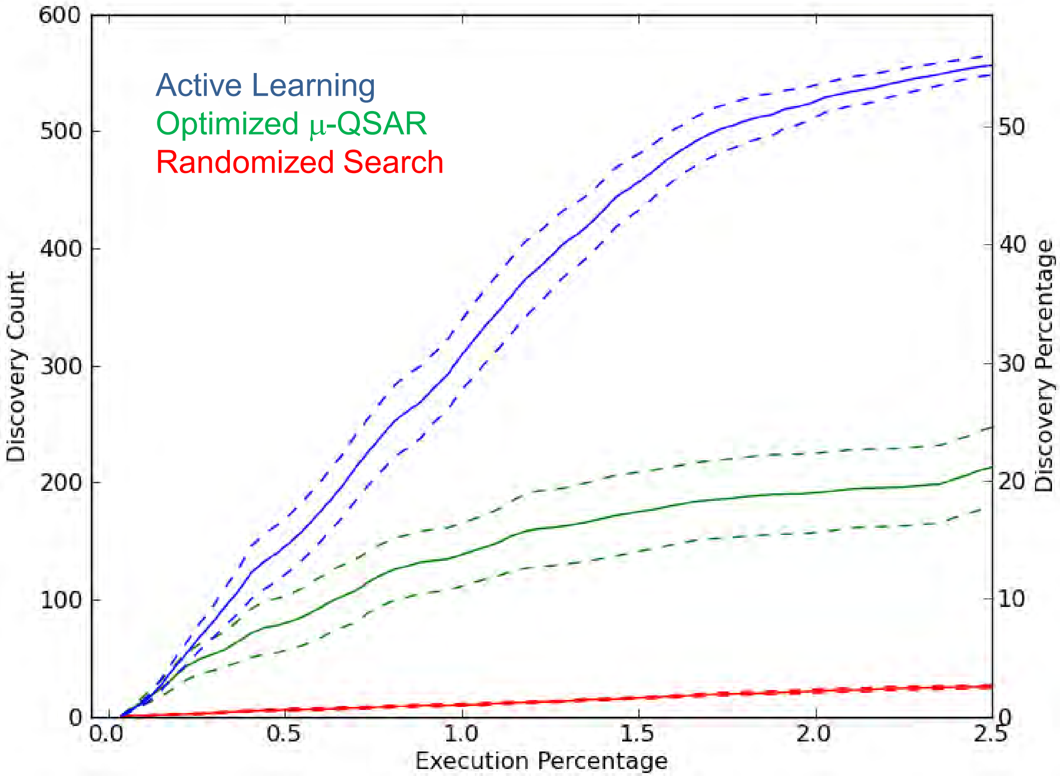


Retrospective Test: Use data from PubChem

- Assays: 177
- Unique Protein Targets: 133
- Compounds: 20,000
- Experiments: ~1,000,000 (30% coverage)
- Use features to measure similarity between drugs and between targets
- Compare discovery rate across different methods

With only **2.5%** of the matrix covered, we can identify **57%** of the active compounds using active learning

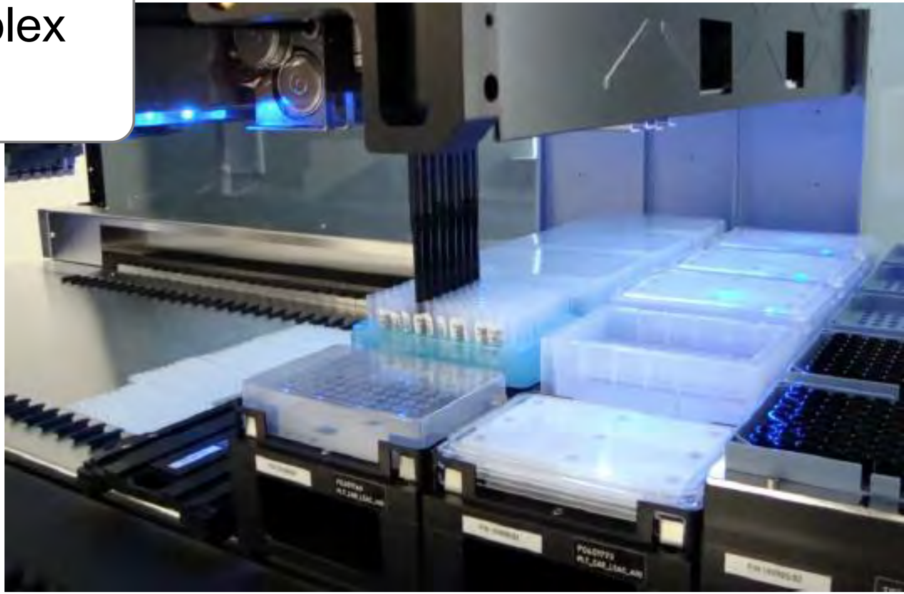
Kangas, Naik, Murphy, BMC Bioinformatics 2014



First *Prospective* Use of Active Learning for Complex System

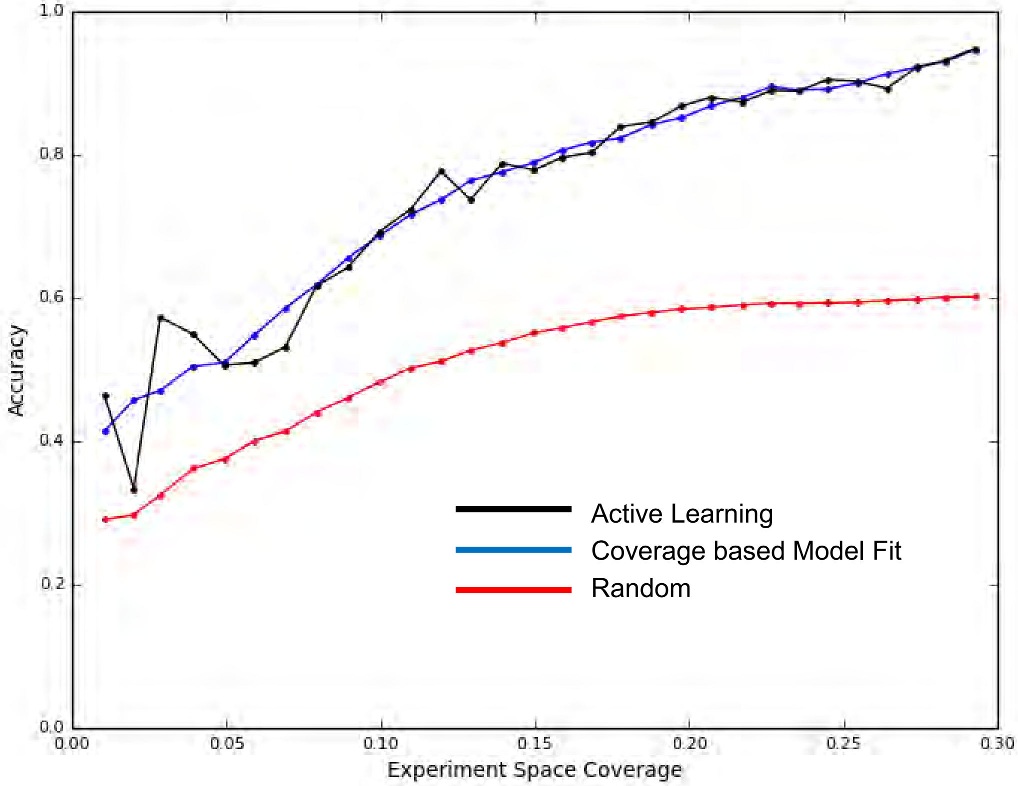
Use **laboratory automation** to execute experiments chosen by **AI** to model complex phenotypes

Try to learn the effects of 96 drugs upon 96 GFP-tagged proteins: have to learn phenotypes on the fly



After doing 28% of possible experiments, model is 92% accurate and 40% more accurate than would have been obtained by random choice of experiments

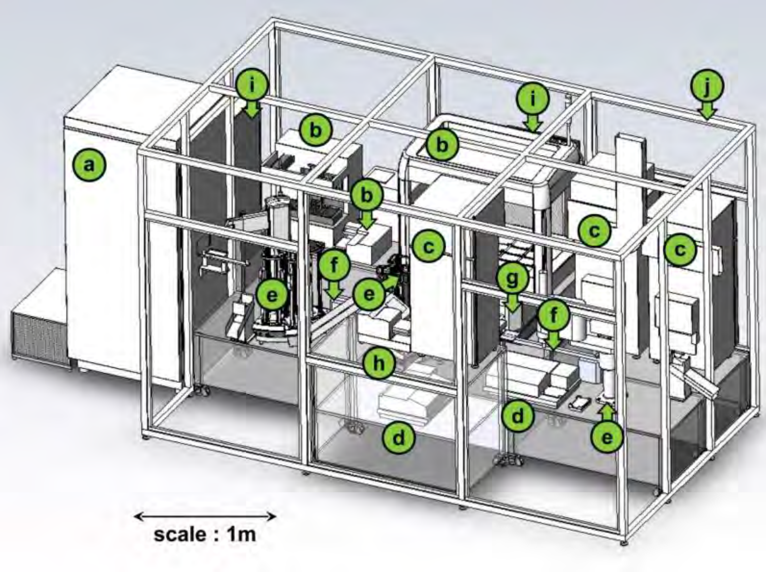
Naik, Kangas, Sullivan, Murphy, eLife 2016



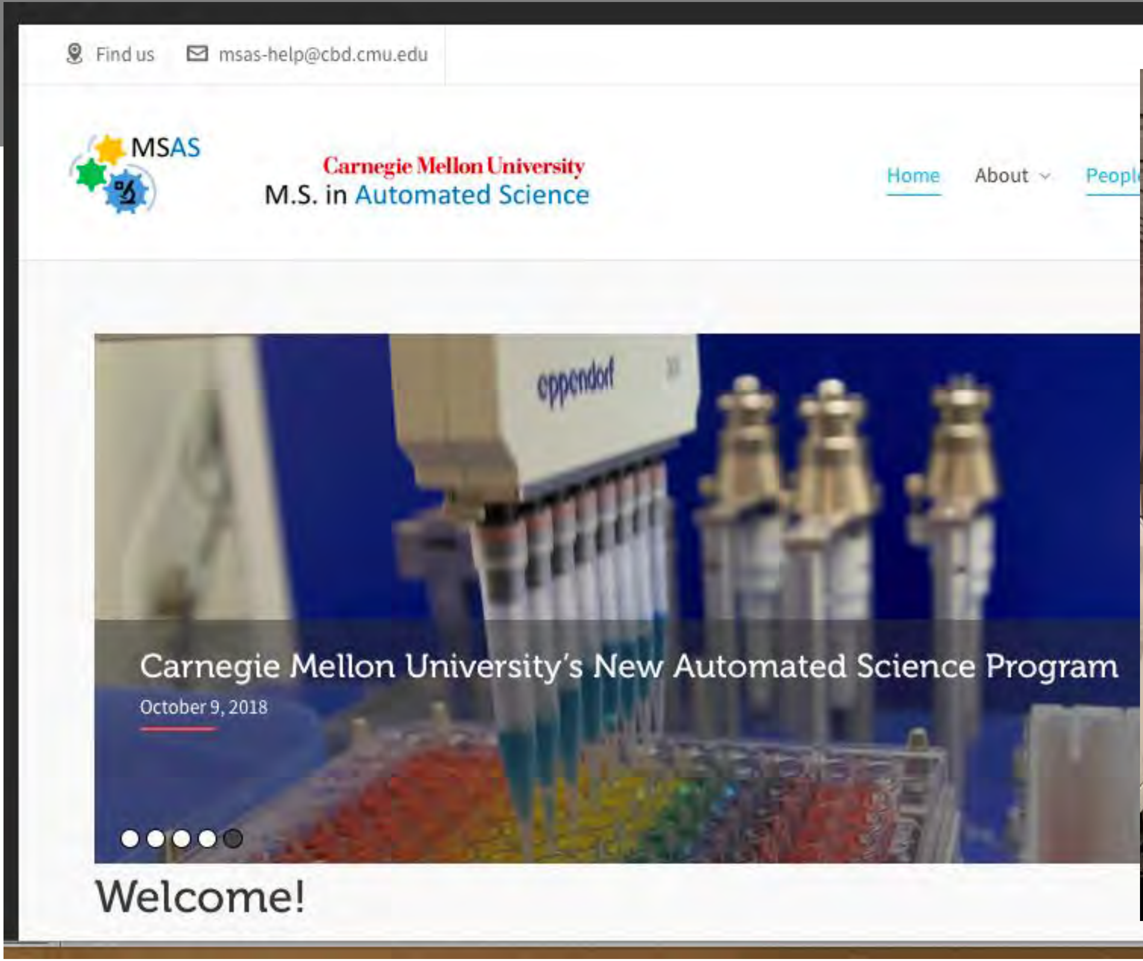
Automated Science

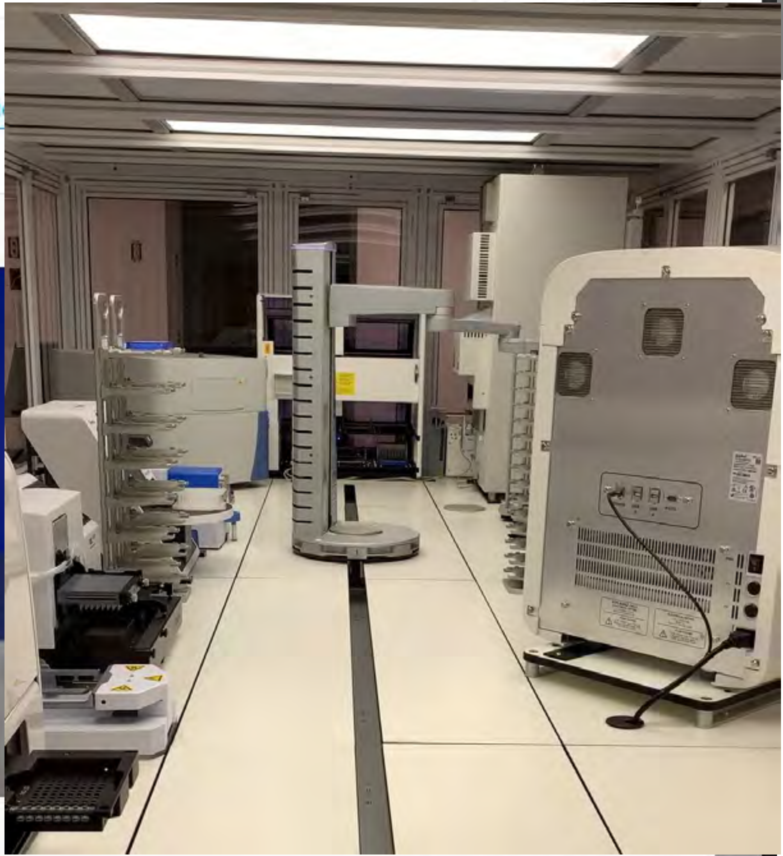
- The combination of *laboratory automation* plus *artificial intelligence* creates unique opportunities to expand research capabilities
- Analogous to self-driving cars: self-driving instruments

- Need better standards for laboratory automation
- Need trained experts



King et al, Science, 2009


Find us
msas-help@cbd.cmu.edu
MSAS
Carnegie Mellon University
M.S. in Automated Science
Home
About
Peopl
Carnegie Mellon University's New Automated Science Program
October 9, 2018
Welcome!



The future?

- **Autonomous** laboratory automation
- Real-time primary data sharing (and standardization)
- National research resources that execute particular types of automated experiments on request – enabling cost-sensitive, proactive learning
- Training of researchers to
  - **Choose** goals and define campaigns combining diverse technologies
  - **Invent** new measurement technologies
  - Develop and improve AI methods

Research supported by:
NSF
NIH National Institute of General Medical Sciences