Abstract
Collaborations Pharmaceuticals, Inc. (CPI) aims to streamline the development of drugs for rare and neglected tropical diseases. This is a big endeavor for a large company, and yet it is feasible for a small company to tackle. Using our core technology, Assay Central, and combining forces with many academic collaborators we have leveraged a broad, collective expertise to identify treatments for parasites (T. cruzi), bacteria (M. tuberculosis), and viruses (Ebola, HIV etc), progressing to in vivo models. CPI bridges industry and academia to combine many disciplines to accomplish our goals. As a small business we are able to apply for non-dilutive funding from the NIH and DOD. By targeting rare and neglected diseases rather than bigger diseases we are a part of an important niche group that remains relatively untouched by large pharmaceutical companies. As the primary investigator, we are able to open new avenues for our collaborators, and by partnering with academic groups we can be included in larger, longer-lasting grants with our collaborators. Using this strategy, we are able to limit our costs to consulting, computation, and compound purchase, while reaping the benefits of wet-lab counterparts. We have also recently started to build our own in-house laboratory for performing some wet-lab experiments. To date CPI has been awarded 9 NIH and DOD grants, two FDA orphan-designations, and we have filed multiple patents. Our business model is both scalable and replicable, and herein we present several of our case studies in order to encourage others to start small drug discovery companies for neglected diseases.

Company Overview
- Founded in 2015 to focus on drug discovery for rare and neglected diseases
- ~5M of funding from NIH and DOD to date, with additional ~3M pending
- We currently have ~1,400 ft² of laboratory and office space at the NC State University Centennial Campus
- More information at www.collaborationspharma.com

Services
- Expertise
  - Consulting
  - Grant writing, access to non-academic funding opportunities
  - Drug discovery and commercialization
- Immediately actionable (short projects)
  - Consulting or partnership to repurpose our current assets
  - Access literature data to pursue drug discovery projects
  - Develop machine learning models from public data (ChEMBL)
  - Comprehensive analysis of compounds (ADME/Tox, similar analogs)
  - Generate virtual libraries
- Longer-term engagement (multi-month to year-long projects)
  - Licensing of Assay Central or other software
  - Validate machine learning models (predictions with experiments)
  - Develop machine learning models from private data

Figure 1. The CPI Team.

Figure 2. Repurposed Compounds Discovered with Machine Learning at CPI.

Figure 3. Assay Central Overview. A) Iterative loop of workflow. B) Machine learning model performance summary of a whole cell Bayesian (left) and Random Forest (right) model of T. cruzi amastigotes. C) Atom highlighting of active features of Yellow Fever Virus (left) and HHERG (right). D) Honeycomb images of predicted compounds (black background) in relation to model compounds, with actives highlighted in green. E) Prescription example before and after editing.

Figure 4. The CPI Pipeline of Funded Projects.

Would you like to collaborate with us?

Ongoing Neglected Disease Collaborations
- Schistosomiasis (S. mansoni)
- Zika virus
- Dengue virus
- Malaria (P. falciparum)
- ESFAPE pathogens
- Yellow fever – NS3 protease, NSS RdRp
- Chikungunya – NSP1, NSP2, NSP4

Desired Collaborations
- Leishmaniasis – target experiments, lesser known strains (brazil**)
- African trypanosomiasis
- Chagas disease – Cruzain targeted experiments
- Cholera
- Yellow fever – whole cell experiments
- Chikungunya – whole cell experiments
- Animal health – rabies, hookworm, ***
- YOU DISEASE HERE!

Acknowledgments
We kindly acknowledge funding: NIAID R41AI108003-01, R41AI134561-01A1, NIGMS R43GM122196, R44GM122196-02A1, NCAATS R21TR001718, 1UH2TR002084-01, NINDS R01NS102164-01, R43NS107079-01, 3R43NS107079-01S1, DOD W81XWH1810802, One NC Small Business program grant.

Publications on Assay Central and Neglected Diseases
- Mol Pharm. 2019 Feb 26. doi: 10.1021/acs.molpharmaceut.8b01297
- Mol Pharm. 2019 Feb 4;16(2):898-906
- ACS Omega. 2019 Jan 31;4(1):2353-2361
- Mol Pharm. 2018 Apr 26. doi: 10.1021/acs.molpharmaceut.8b00083

A tool used for the iterative building and sharing of machine learning models
- Built from bioactivity data against disease targets, ADME/Tox liabilities, and molecular properties from public (ChEMBL, PubChem) and private sources
- Bayesian and Random Forest algorithms
- ECFP6 descriptors
- Iterative loop (Figure 3A)
  - Curation to generate a QSAR-friendly dataset
  - GitHub for seamless cross-machine collaboration
  - Generate predictions across a variety of targets to select lead compounds
- Visualization
  - Model performance summaries (Figure 3B)
- Users can generate predictions for new molecules and view features contributing to predicted activity in different formats (Figure 3C)
- Honeycombs for data visualization (Figure 3D)
- Prescriptions to minimize compounds with liabilities (Figure 3E)
- Future considerations include adding different algorithms and descriptors
- Public distribution available at www.assaycentral.org

MegaTrans
MegaPredict
MegaTox

• MegaTrans: A recently funded grant to build models for known human drug uptake transporters, this follows on from our published work on OCT2 & MATE1
• MegaPredict: A method of evaluating a single molecule across thousands of auto-curated (by target and measurement) ChEMBL datasets
• MegaTox: A subset of machine learning datasets that focuses on ADME/Tox liability datasets. Models include Vero cell cytotoxicity, LD50 in rats, hERG, cytochrome P450 inhibition etc.