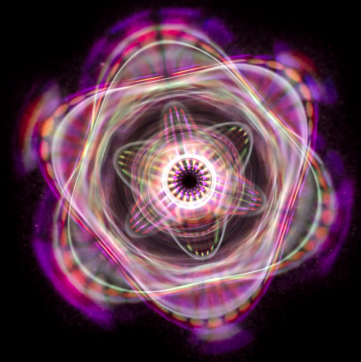
ATOM

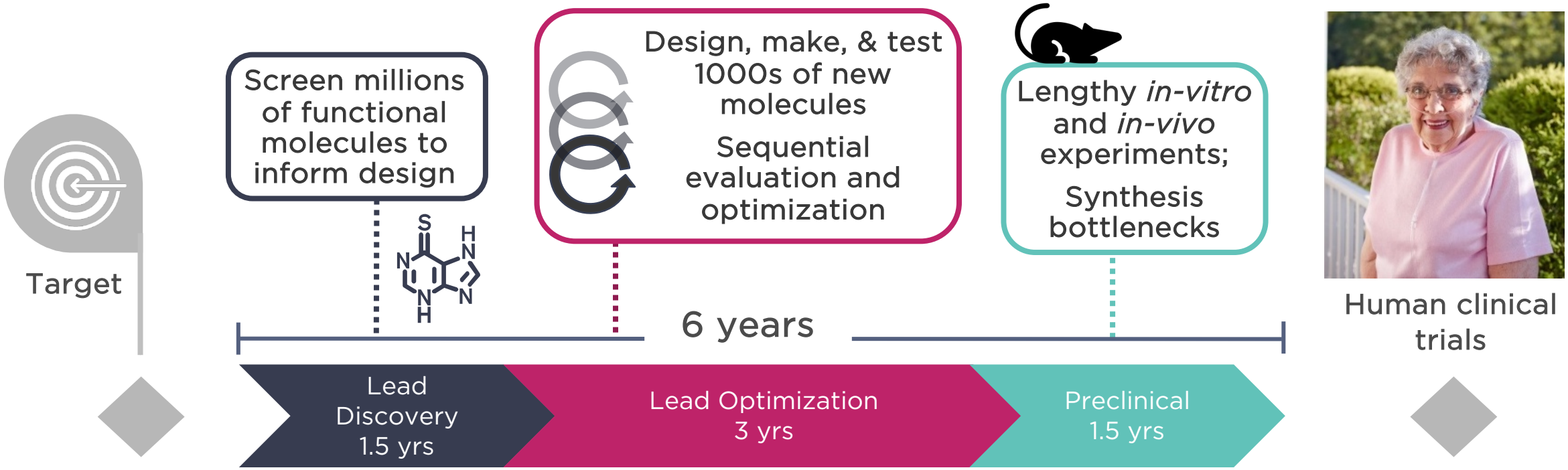


06 June 2019

Third HPC Applications in Precision Medicine Workshop

Current drug discovery: long, costly, high failure

Is there a better way to get medicines to patients?

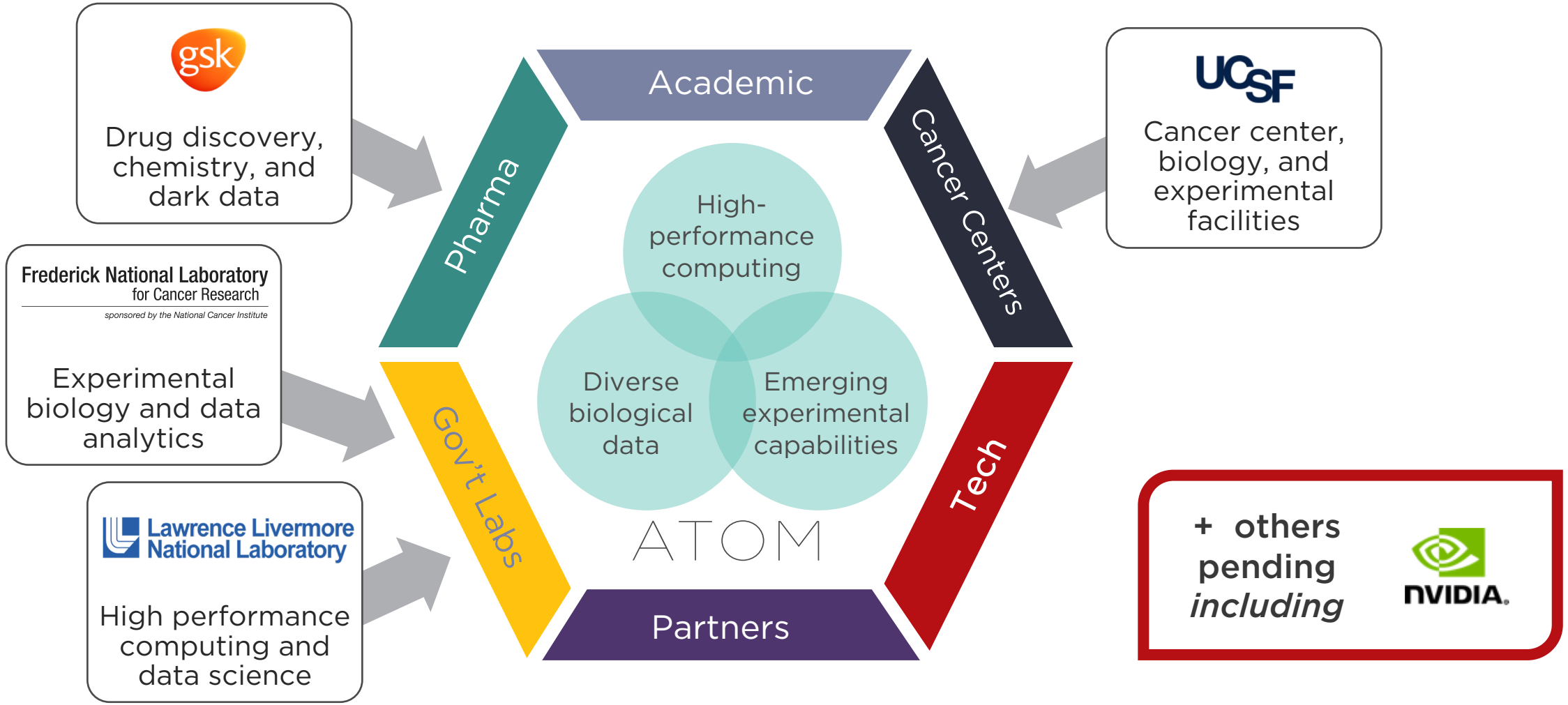


- 33% of total cost of medicine development
- Clinical success only ~12%, indicating poor translation in patients

Source: http://www.nature.com/nrd/journal/v9/n3/pdf/nrd3078.pdf

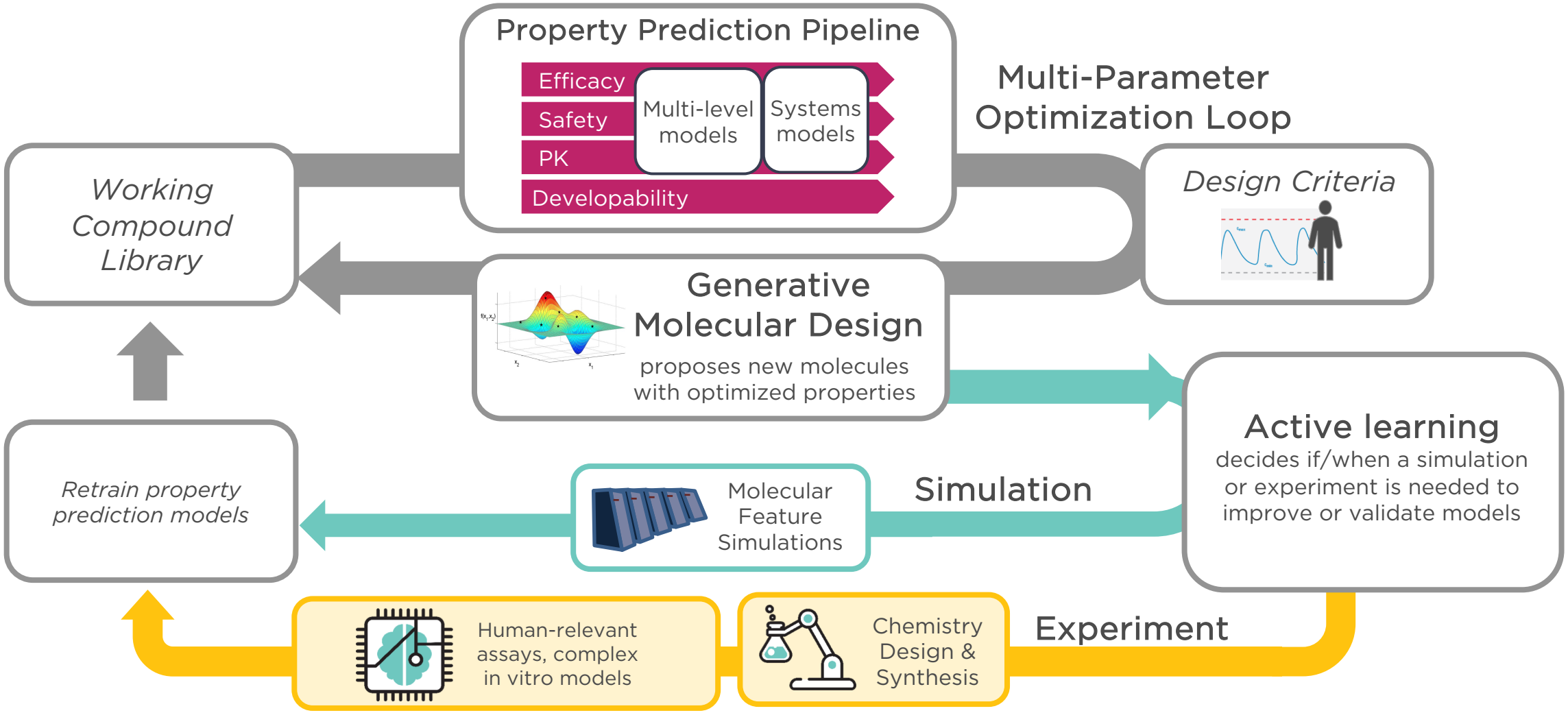
Accelerating Therapeutics for Opportunities in Medicine

A consortium building an integrated precompetitive platform

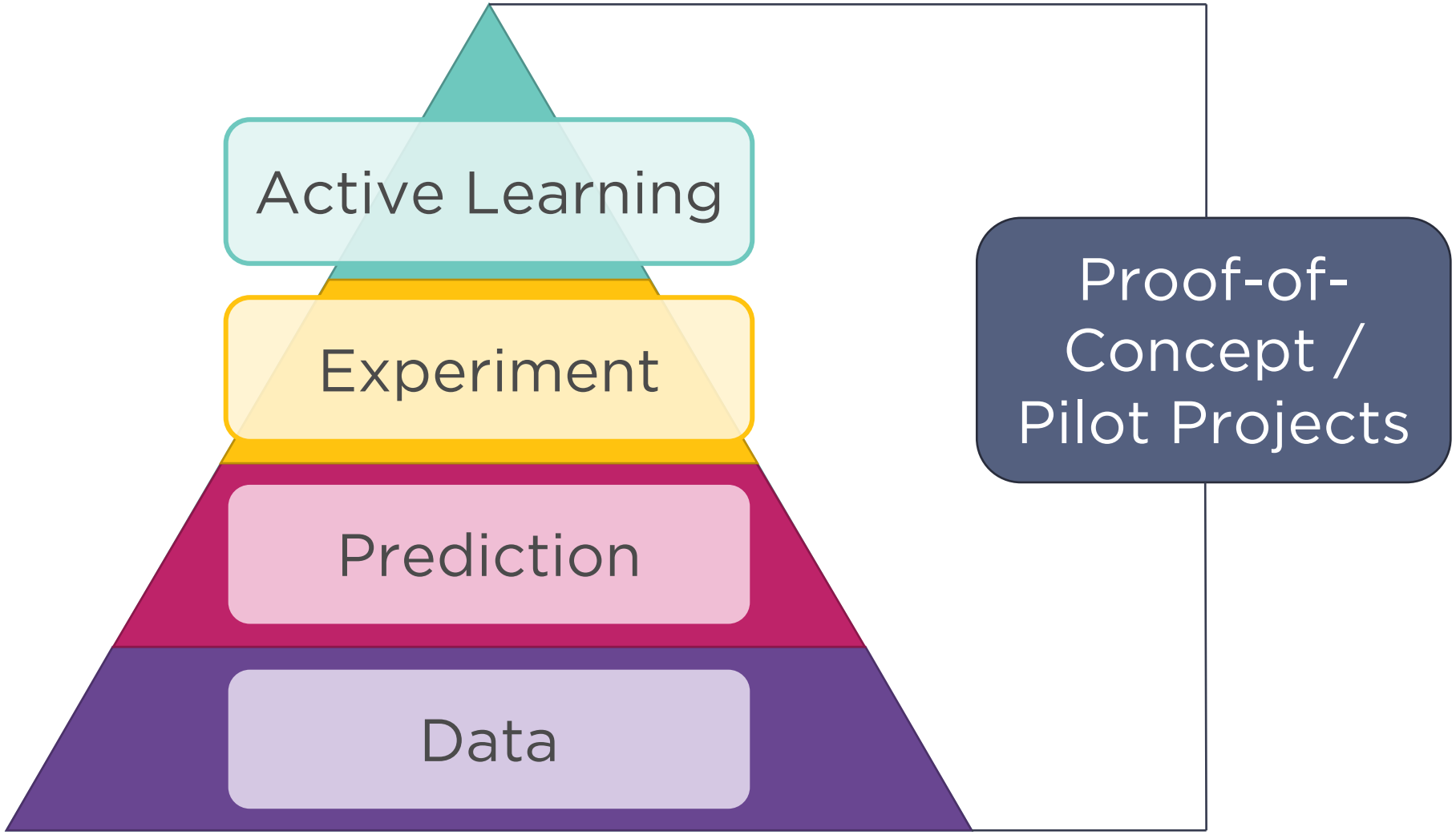


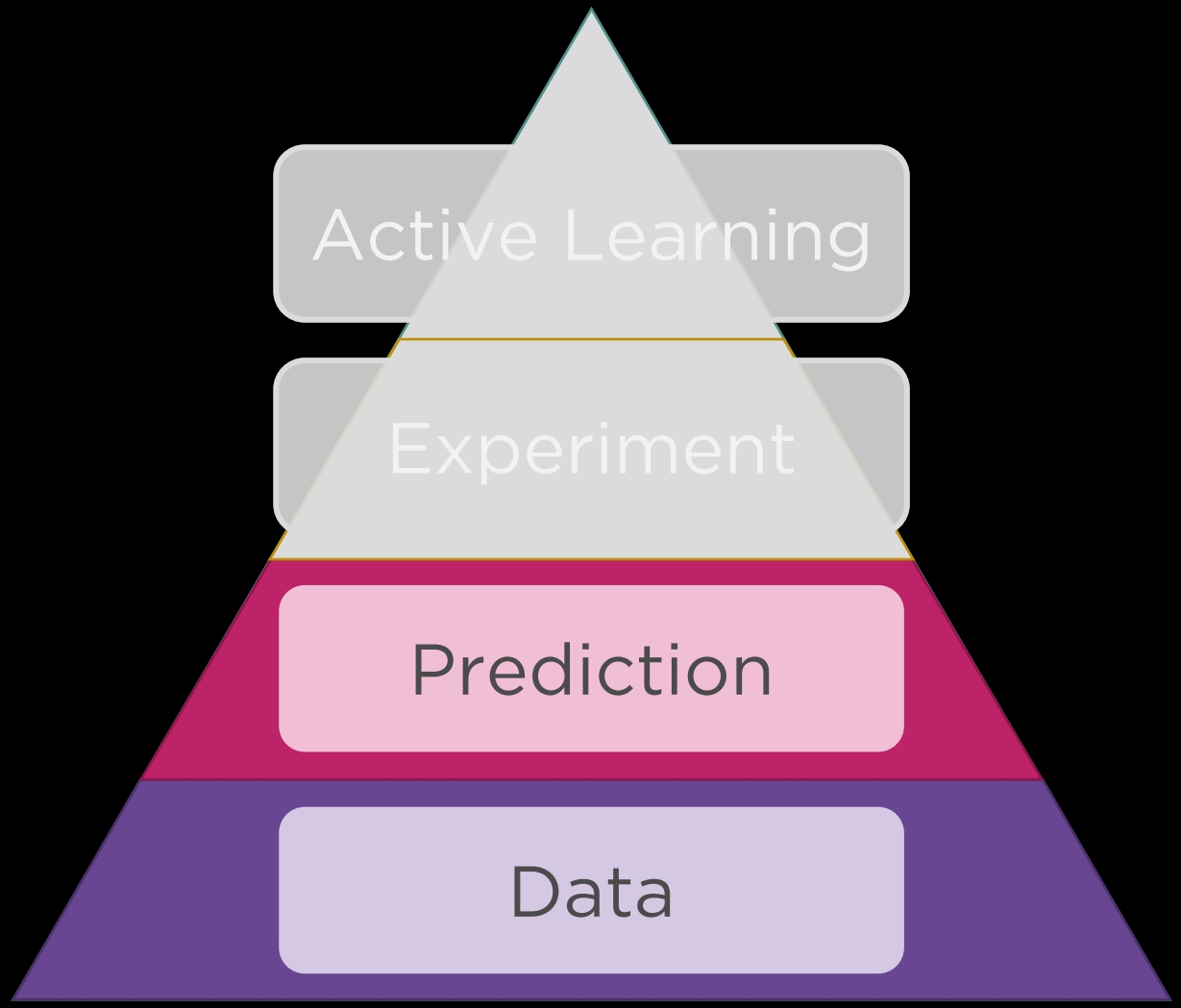
The ATOM Platform

Active Learning Drug Discovery Framework



Components of the ATOM platform




Active Learning
Experiment
Prediction
Data

Example data from 2 million GSK compounds in ATOM

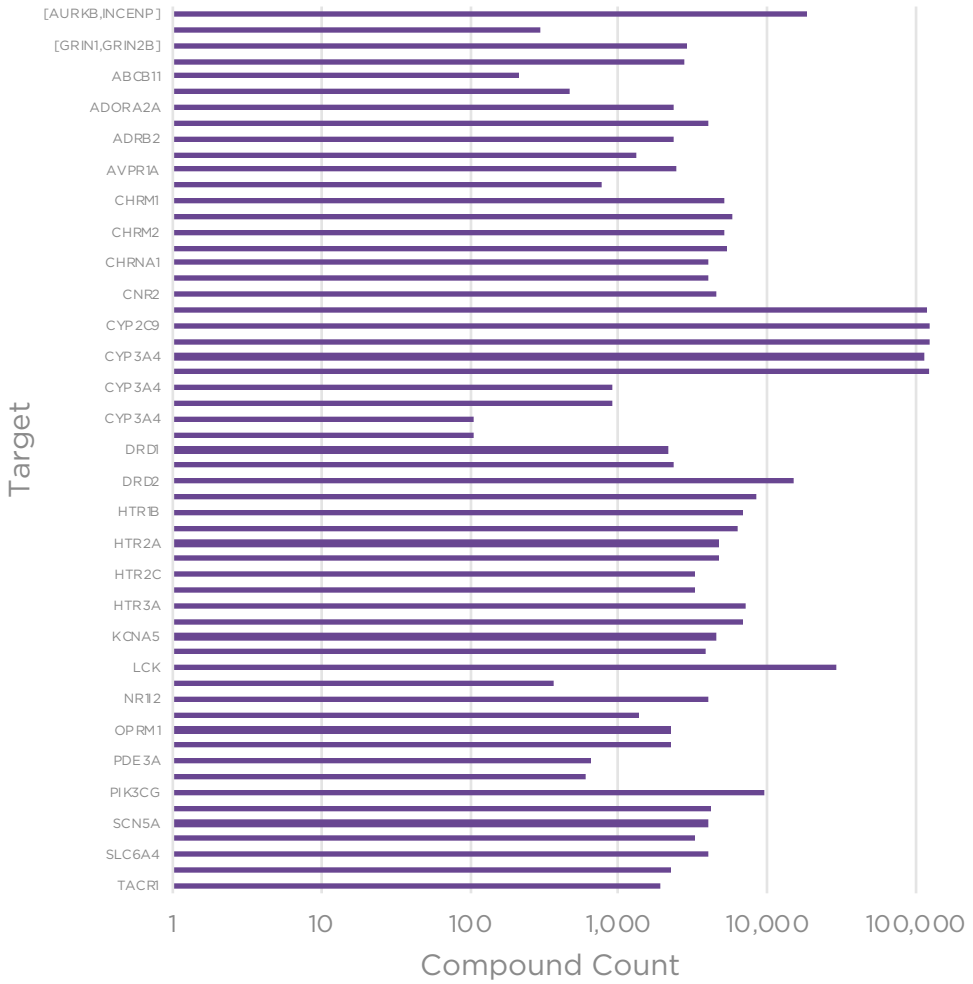
Source	# Compounds	Specific Data Insights
Post-candidate selection programs no longer of interest to GSK	500	• In vitro and in vivo results (see below) • ~100 compounds with anonymized human clinical data
Unique compounds synthesized in lead optimization over last 17 years	515 k	• Structure-activity relationships with learnings on protein target pocket
Retired High Throughput Screening (HTS) compounds	1 M	• in vitro assays against diverse protein targets and physicochemical properties
Commercially available compounds in current HTS collection	420K	• in vitro assay data gained over the past ten years for a diversity of protein targets

Above data sets include, as available:

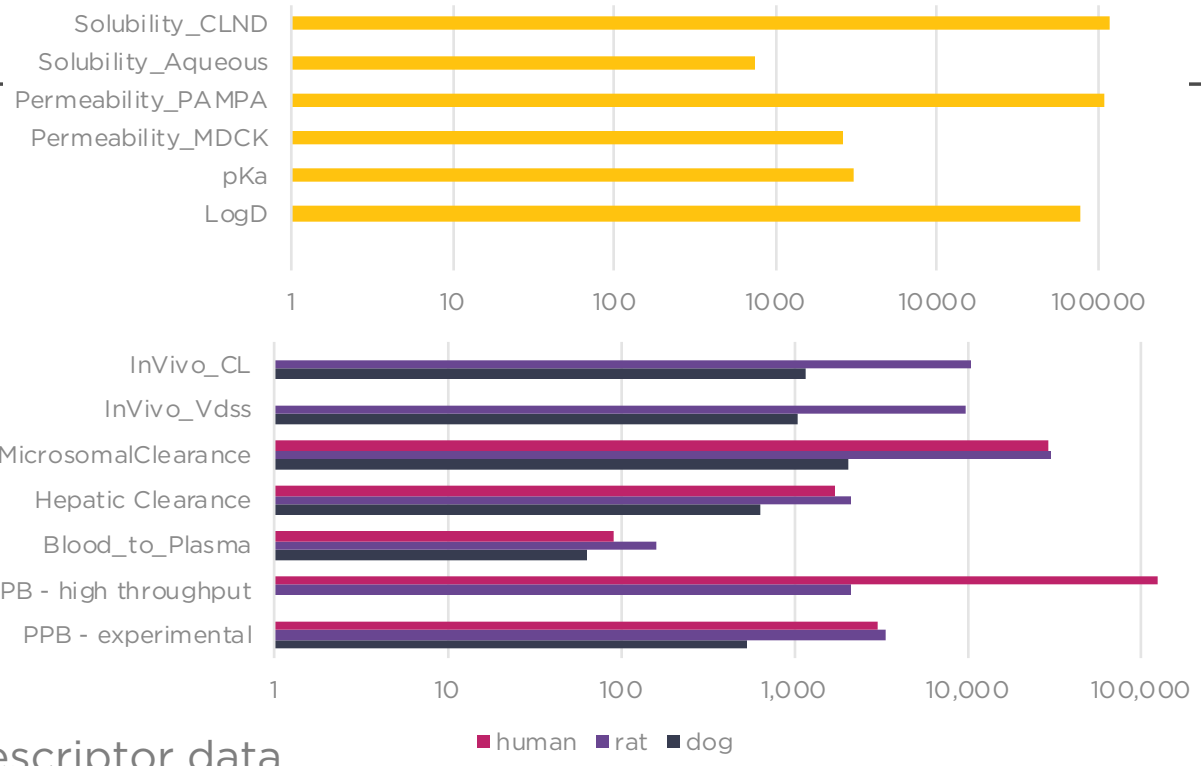
- in vitro enzyme and cell-based assay screening data against
- physicochemical experimental and calculated properties
- ex vivo ADME data
- in vivo pharmacokinetic, toxicokinetic, and animal safety data
- protein ligand crystal structures

ATOM has curated ~150 model-ready data sets

GSK Safety Datasets



GSK Pharmacokinetic Datasets



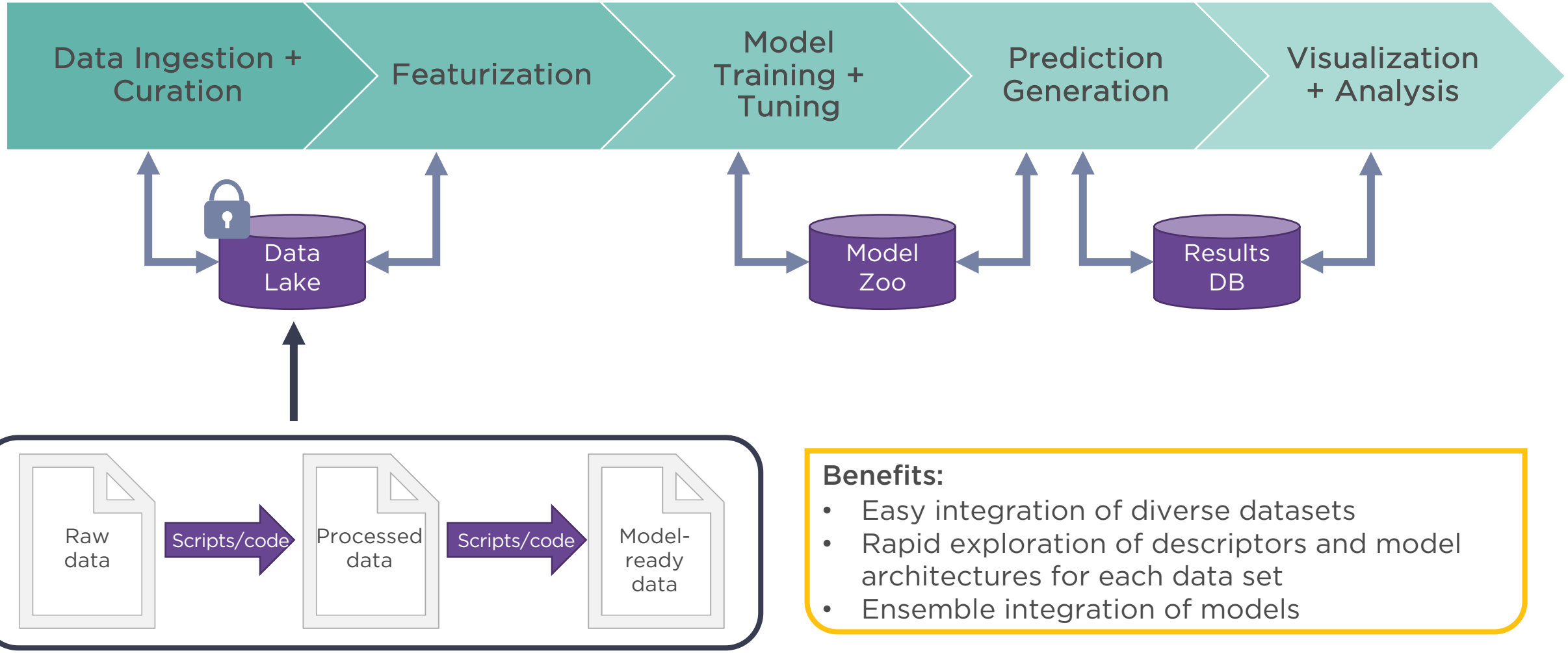
Descriptor data

Data Set	MOE	3D Descriptors	Compounds
GSK	√	√	1.86M
ChEMBL	√	√	1.6M
Enamine	√	√	680M

ATOM

End-to-End Data-Driven Modeling Pipeline

Prediction

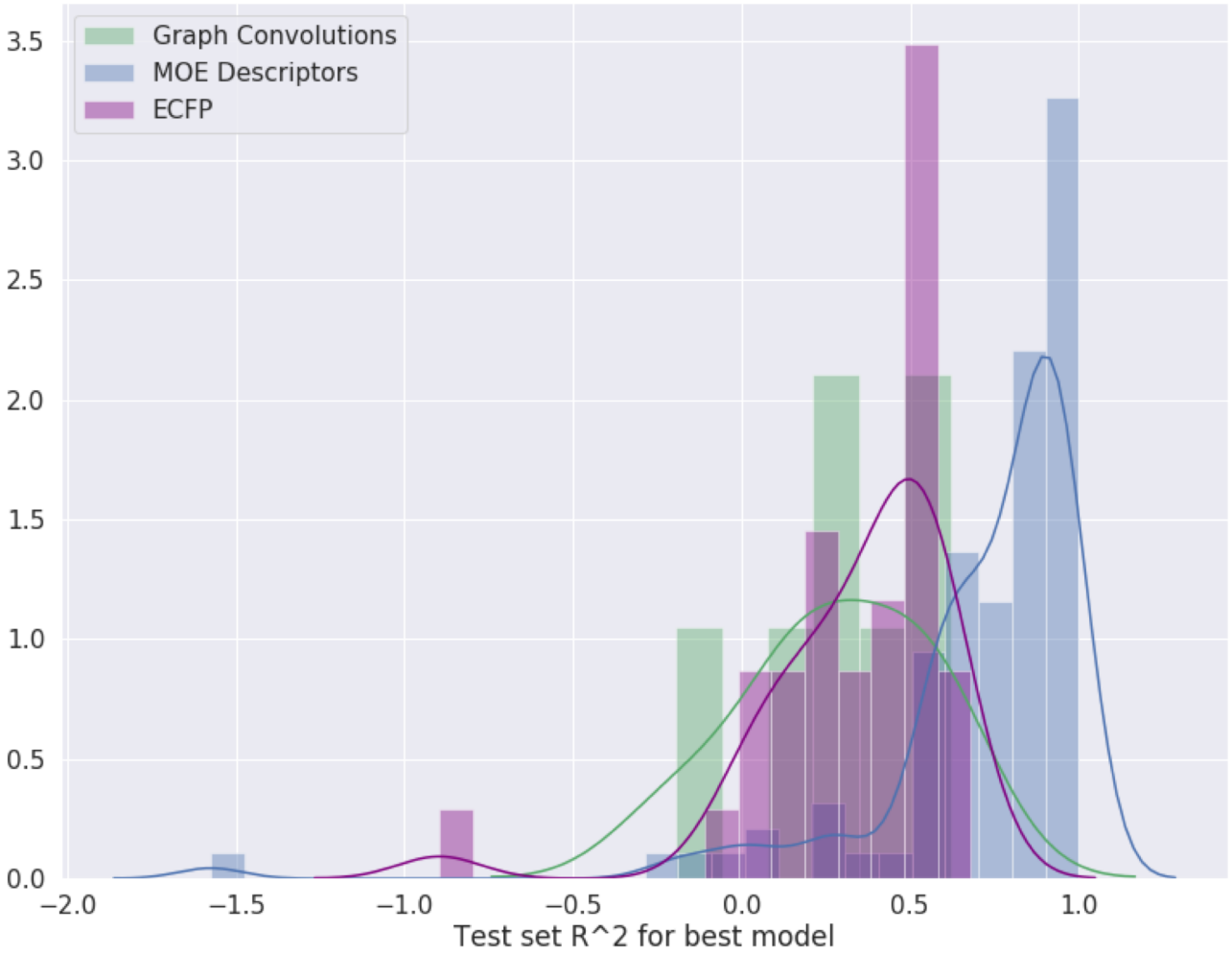


Benefits:

- Easy integration of diverse datasets
- Rapid exploration of descriptors and model architectures for each data set
- Ensemble integration of models

Featurization is key

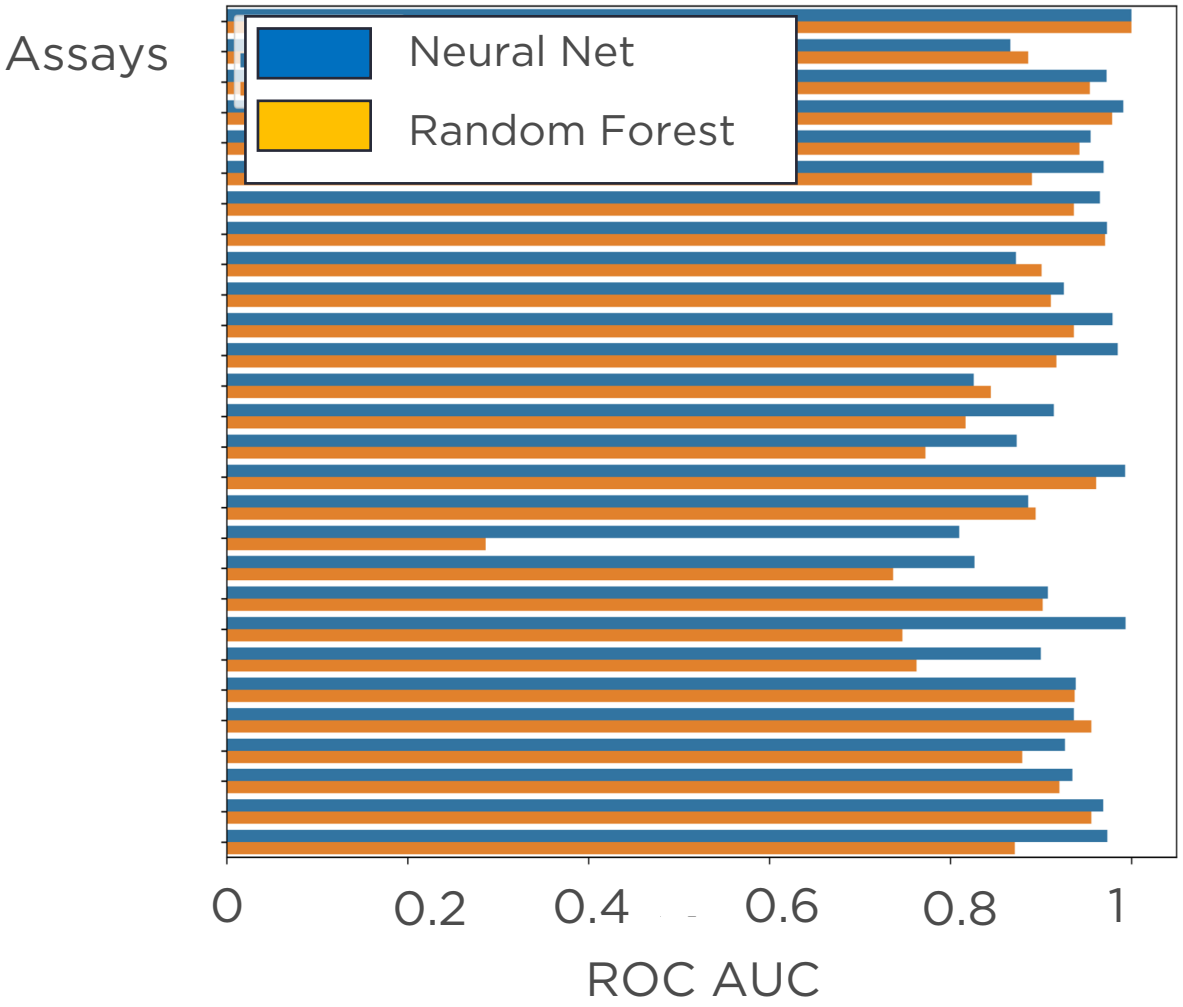
Prediction



- We have found that the best-performing feature type varies by dataset
- In general chemical descriptors out-perform other feature types
- Graph Convolutions occasionally outperform others

Classification performance shows high accuracy for selected safety targets

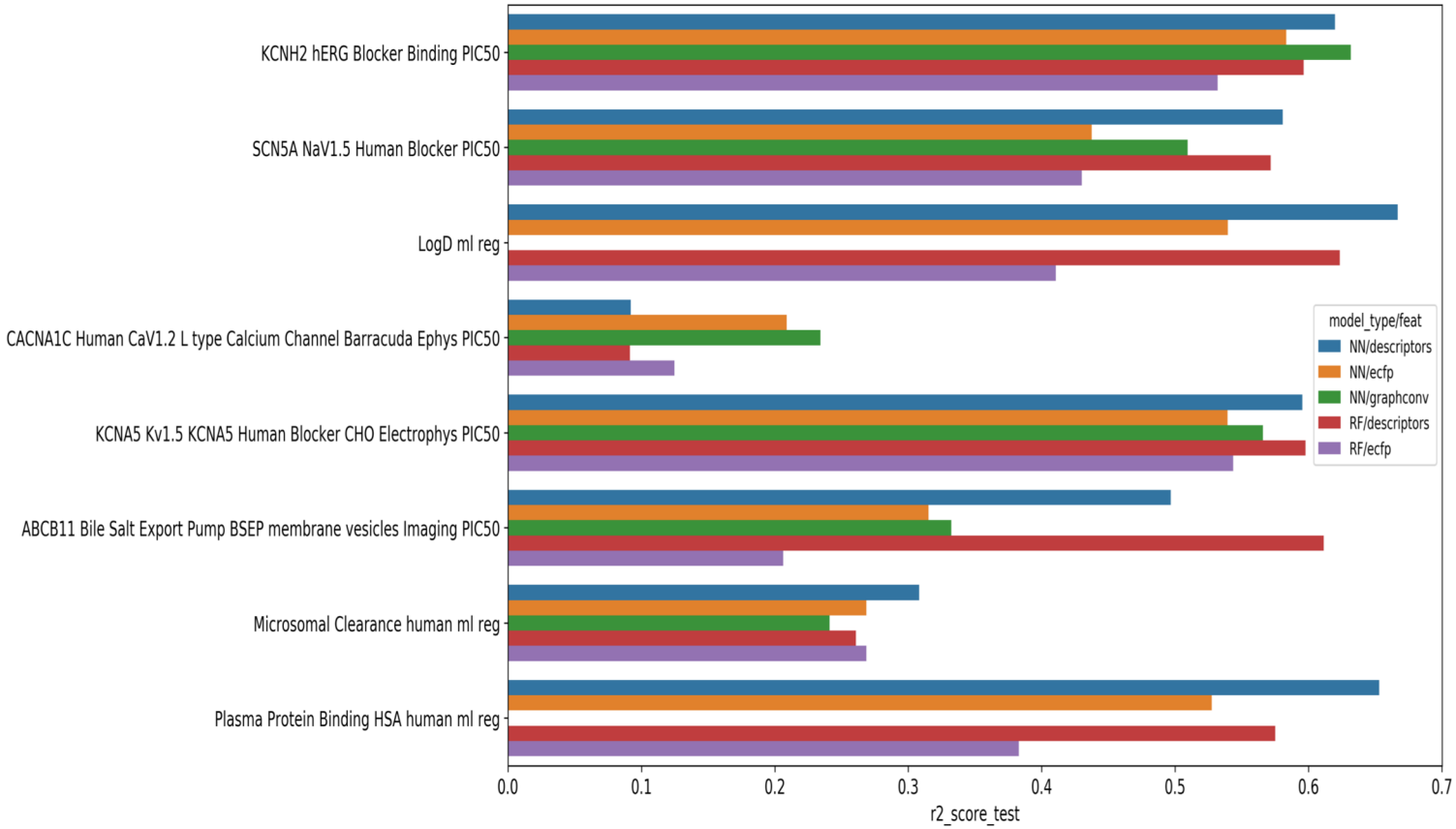
Prediction



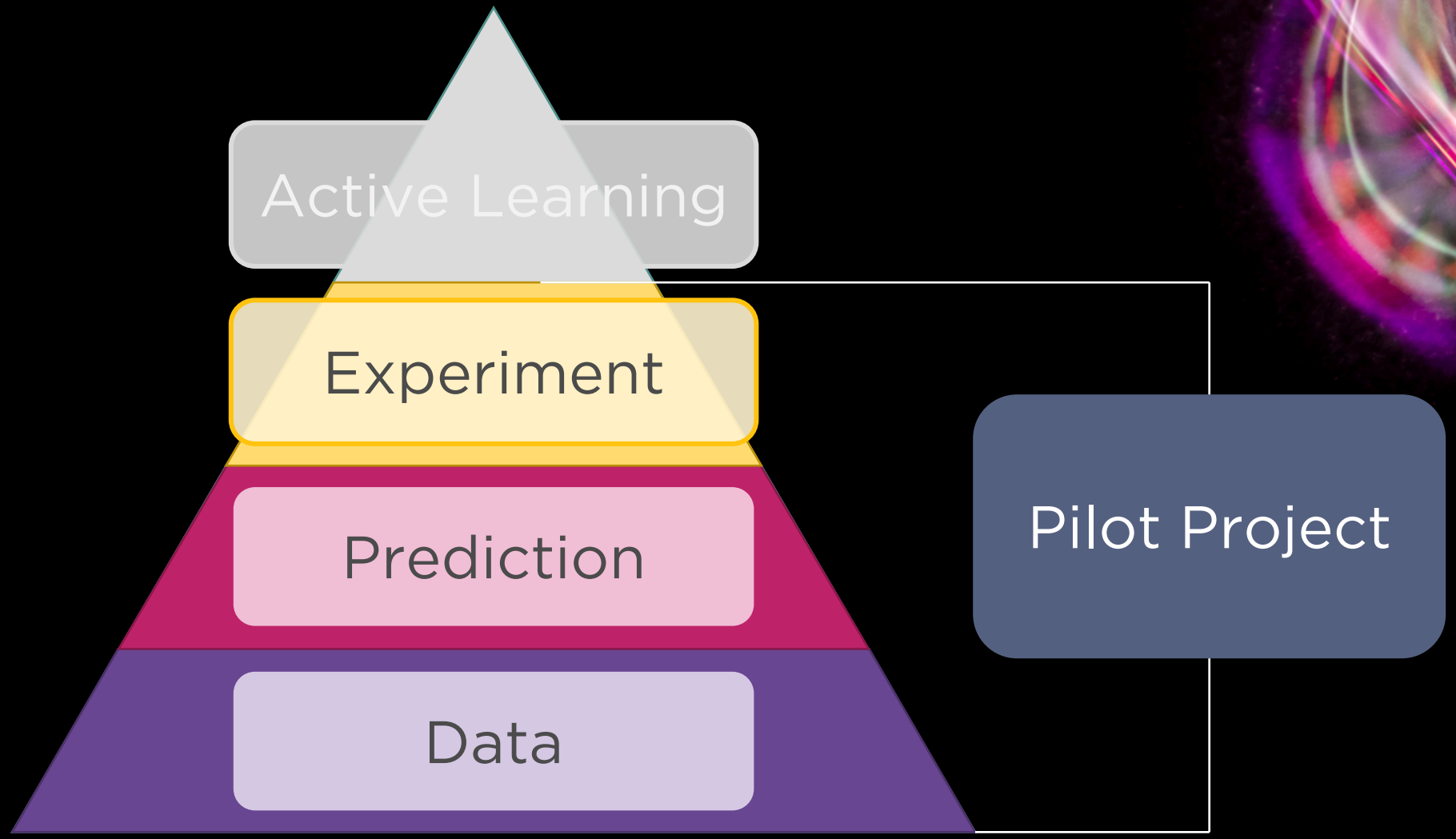
- Assays range in size from 187 to 9173 compounds
- 23 of 28 of the assays show improvement with NN
- KCNE1 shows largest improvement
- Classification accuracy appears to be relatively high (>0.8 ROC-AUC)

Regression models present a greater challenge

Prediction



- Assays range in size from 101 to 123,759 compounds
- 4 of 8 of the assays show improvement with NN
- Descriptors and Graphconv outperform ECFP
- Test set R^2 ranges from ~0.1 to ~0.7



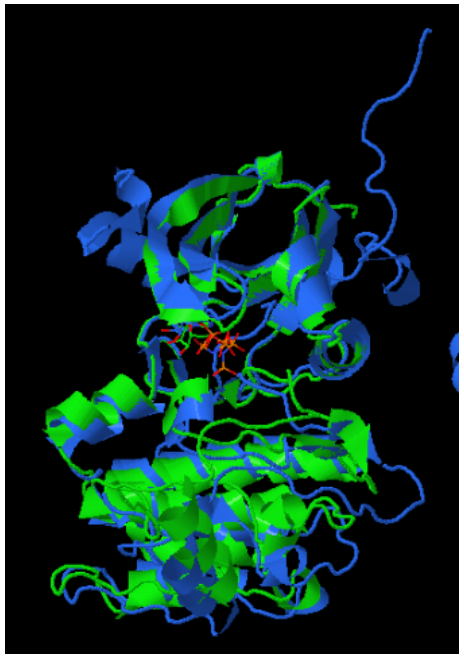
Pilot Project

Computational loop validation

Generative molecular design (GMD) of AURK B inhibitors

Pilot Project

PILOT 1 | Starting point: Early program data | Lead Optimization | End point: Experimental validation



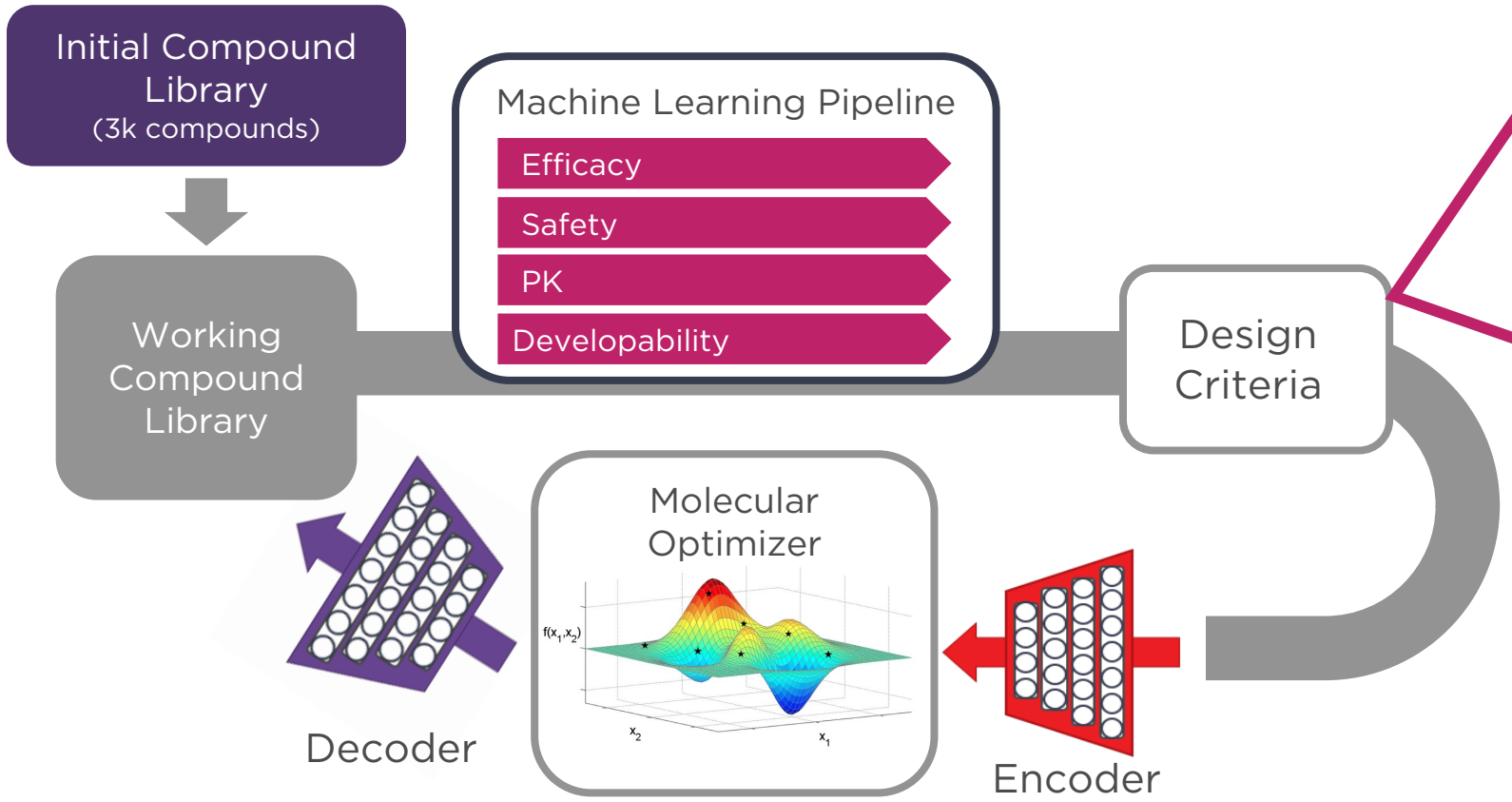
Structure overlay of AURK A and AURK B

Why Aurora Kinase?

- **Cancer relevant**: >30 clinical trials are ongoing or completed for AURKA selective, AURKB selective, and AURKA/B dual inhibitors
- **Data available at ATOM:** Potency data on ~24k compound available for AURK B and/or AURK A
- **Pharmaceutical discovery relevant problem:** Selectivity between kinases is an important and common pharmaceutical discovery problem

Pilot Design Criteria

Pilot Project



Candidate Quality Panel

Efficacy
AURK B (pIC50 > 9)
AURK B/A Selectivity > 1000

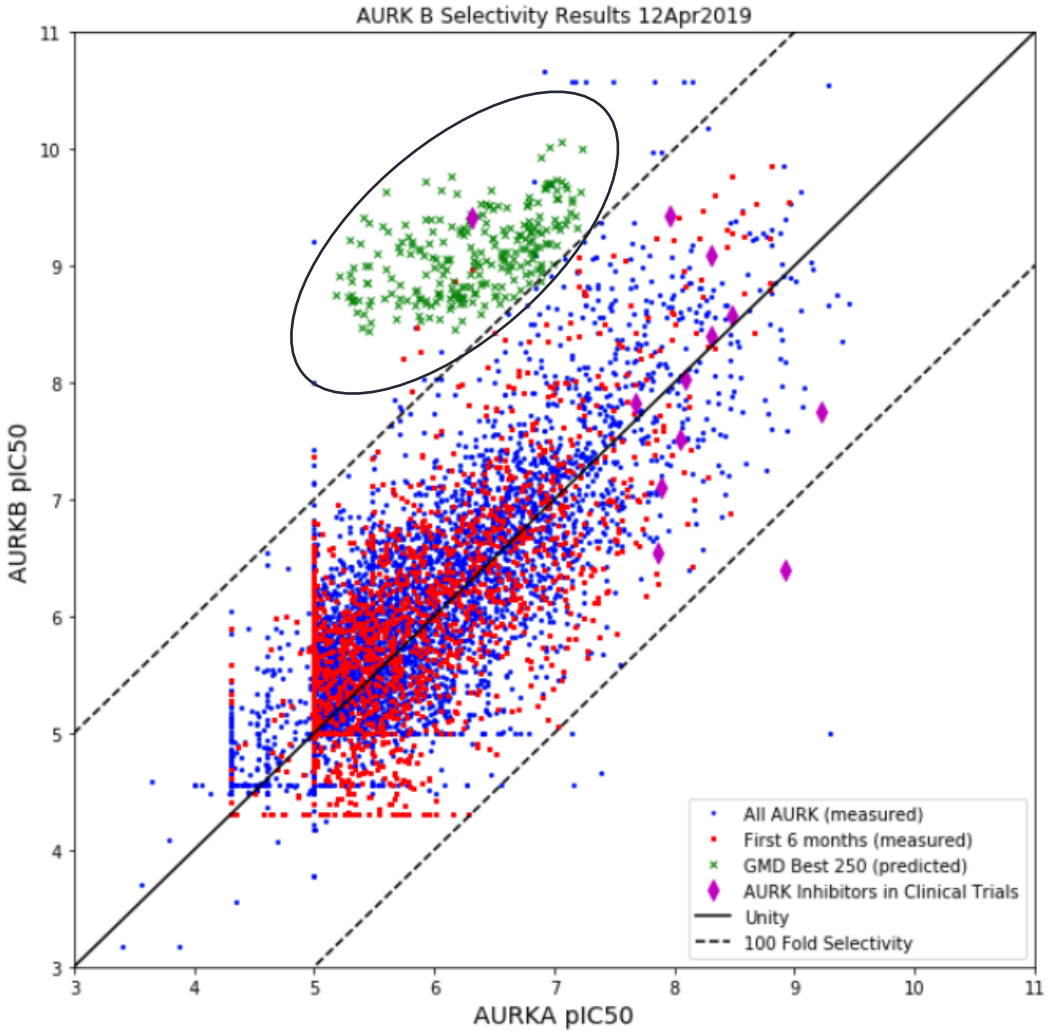
Safety
BSEP (pIC_{50} < 4)
hERG (pIC_{50} < 4)

PK
Solubility (>10uM)
CL_{int} (<3 mL/min/g)

Developability
Solubility (>10uM)
SAS: Synthetic Accessibility Score
QED: Quantitative Estimation of Druglikeness

Results: >200 new potent, selective AURK B compounds with favorable other properties

Pilot Project



Multi-Parameter Optimization:

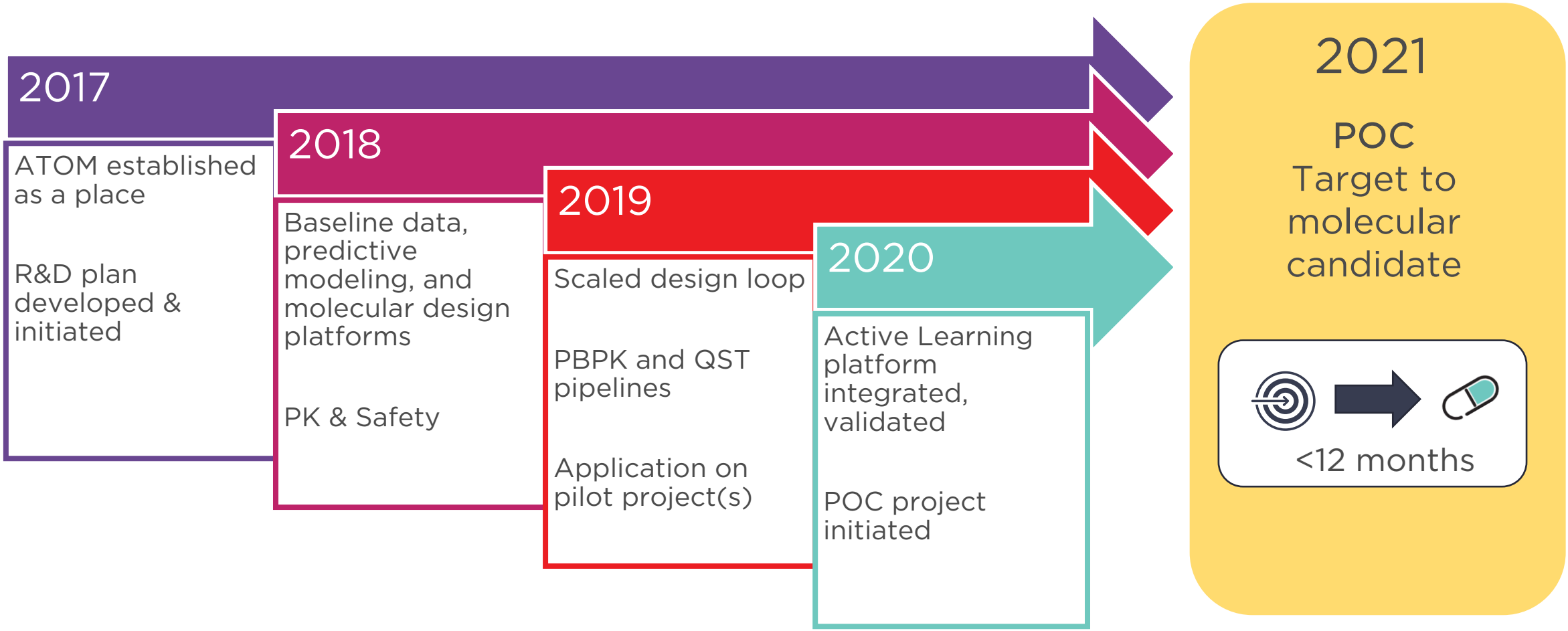
AURKB pIC50	AURKA pIC50	A/B Selectivity (fold)	hERG pIC50	BSEP pIC50	Solubility	hLM CLint	Solubility	SAS
9.627	5.60	10772	3.260	4.010	6.022	1.819	412.492	2.640
9.724	5.92	6381	3.202	4.029	4.241	1.338	69.457	2.632
9.762	6.14	4174	3.197	4.027	4.535	1.322	93.249	2.410
9.298	5.98	2065	3.198	3.969	5.988	1.455	398.809	2.392
9.209	5.73	3024	3.200	4.027	7.000	4.371	1096.282	2.498
9.208	5.81	2477	3.195	4.027	5.413	1.868	224.400	2.397
9.626	6.18	2784	3.868	3.982	5.447	1.434	232.073	2.332
9.407	5.41	9984	3.259	4.018	3.704	1.252	40.620	2.784
9.353	5.75	4028	3.199	4.018	4.470	1.835	87.357	2.339
9.517	6.45	1160	3.223	3.976	4.353	2.024	77.733	2.222
9.252	5.79	2922	3.794	3.977	5.207	1.405	182.459	2.441
9.293	5.61	4851	3.197	3.994	4.006	1.479	54.916	2.627
9.334	5.56	5926	3.198	4.043	6.552	0.986	700.482	2.818
9.393	5.93	2911	3.198	4.026	5.343	1.595	209.163	2.624
9.397	6.05	2247	3.199	4.016	4.017	1.421	55.541	2.640
9.399	5.97	2682	3.211	3.993	3.554	1.632	34.955	2.255
9.193	5.96	1720	3.646	3.970	5.044	1.816	155.047	2.472
9.222	5.30	8342	3.215	4.048	5.936	0.888	378.391	2.628
9.327	6.25	1205	3.198	4.055	6.356	1.498	575.970	2.380
9.440	6.39	1116	3.380	3.968	4.635	1.775	103.039	2.361
9.129	5.88	1775	3.657	4.070	7.134	1.553	1254.501	2.278
9.338	6.14	1579	3.198	3.967	3.507	1.269	33.360	2.369
9.516	6.46	1136	3.202	4.067	6.818	0.858	913.920	2.464
9.278	6.21	1171	3.416	4.069	4.565	1.777	96.042	2.330
9.090	5.91	1509	3.210	4.052	6.827	0.880	922.242	2.433
9.365	6.43	869	3.210	4.020	6.059	0.936	427.917	2.686
9.107	5.53	3788	3.545	3.993	6.339	3.112	565.978	2.501
9.375	5.45	8340	3.199	4.022	2.663	1.340	14.338	2.754
9.650	6.05	3951	3.205	3.990	2.503	1.604	12.224	2.426
8.896	5.64	1821	3.209	4.027	6.561	1.374	707.083	2.181
9.648	6.48	1482	3.244	4.021	4.026	1.318	56.041	2.577
9.389	6.28	1284	3.198	4.055	5.250	1.037	190.604	2.754
9.075	5.98	1235	3.199	4.055	6.027	1.711	414.475	2.527
9.422	6.35	1179	3.202	4.014	3.214	1.569	24.878	2.503
9.298	6.27	1063	4.026	4.058	5.720	1.863	304.910	2.474
9.108	5.80	2028	3.198	4.115	5.448	0.973	232.219	2.608
8.969	5.60	2333	3.925	3.987	5.674	3.901	291.332	2.224
9.162	5.70	2884	3.198	4.069	4.387	0.800	80.426	2.520
9.154	5.87	1907	3.198	4.024	3.459	1.411	31.775	2.319
9.294	6.41	767	3.253	4.000	4.356	0.939	77.924	2.522

Next steps :

Experiment

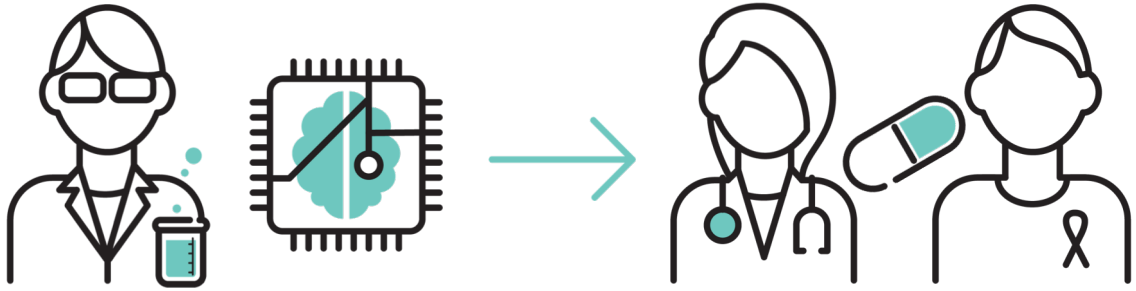
- Comparison to held out, ground truth data
- Experimentally make and test top compounds

High Level ATOM R&D Milestones towards Full Platform POC



Join ATOM

Visit atomscience.org/membership
Contact info@atomscience.org
@ATOM_consortium #ATOMscience



Transform drug discovery, accelerate R&D, and integrate data, AI, and supercomputing to benefit patients

ATOM

Consortium Members:

Frederick National Laboratory for Cancer Research

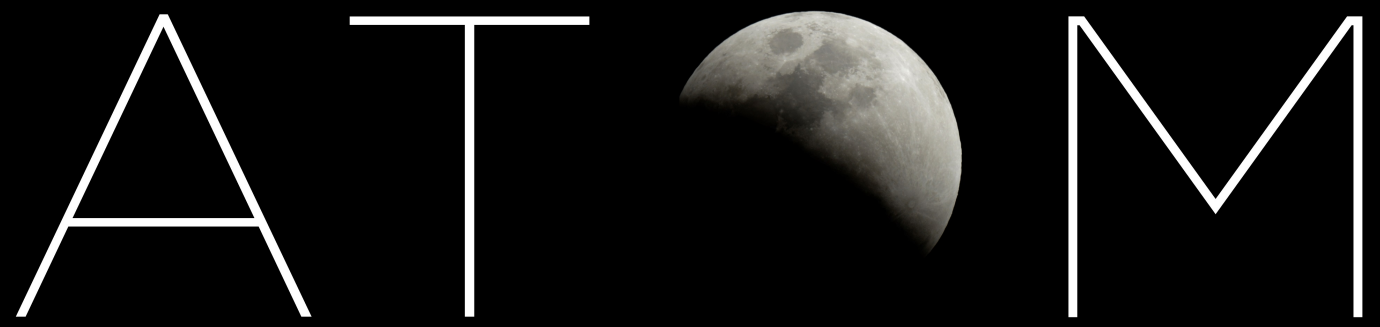
gsk

Lawrence Livermore National Laboratory

UCSF

The ATOM Team

ATOM One Year Anniversary | October 22, 2018
San Francisco, CA | atomscience.org


ATOM