NIH NATIONAL CANCER INSTITUTE

Welcome!

Today's Agenda

- **Welcome and Introduction**

  Ryan Weil, IMPROVE Co-PI, Frederick National Laboratory for Cancer Research

- **Overview of IMPROVE**

  Rick Stevens, IMPROVE Co-PI, Argonne National Laboratory

- **Overview of RFP and Aim 1**

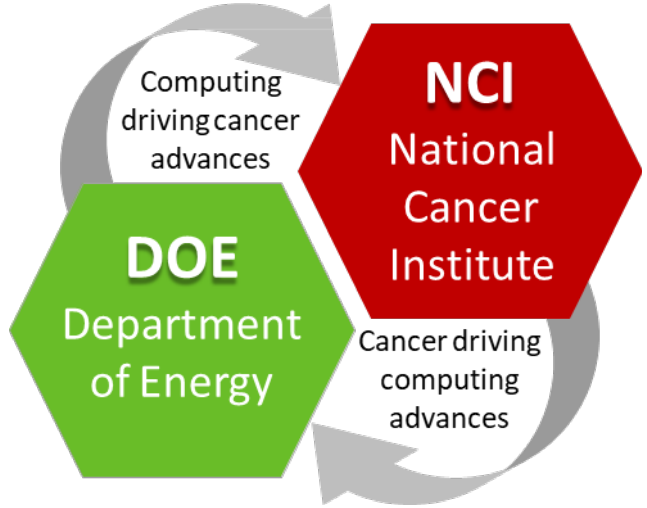
  Ryan Weil, Frederick National Laboratory for Cancer Research

- **Overview from Leidos Biomedical Research Subcontracts Team**

  Josh Wynne and Natalie Fielman, Frederick National Laboratory for Cancer Research

- **Open Discussion and Wrap Up**

  Ryan Weil, Frederick National Laboratory for Cancer Research

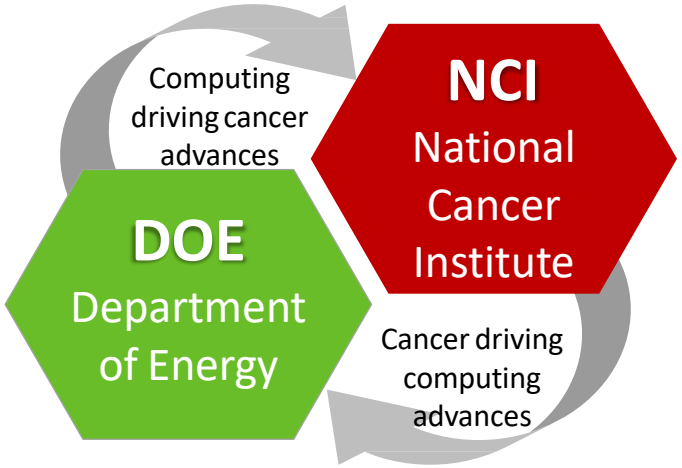


IMPROVE: Innovative Methodologies and New Data for Predictive Oncology Model Evaluation

Rick Stevens
Argonne National Laboratory
University of Chicago

Jeff Hildesheim
Division of Cancer Biology
National Cancer Institute

Ryan Weil
Frederick National Laboratory for Cancer Research



The IMPROVE Project

- IMPROVE is a new project that builds on lessons learned from the NCI-DOE Pilot 1 and uses a new engagement model based on extensive collaboration with the cancer research community
- Two related goals aimed at IMPROVING deep learning models for predicting Drug Response in Tumors:
  - **Aim 1: IMPROVE Models :** Development of semi-automatic protocols for comparing deep learning model and identifying model attributes that contribute to prediction performance with the goal of IMPROVING predictive models of drug response
  - **Aim 2: IMPROVE Data:** Development of protocols for specifying drug screening experiments and to generate new data explicitly aimed at IMPROVING predictive models of drug response

Anticipated Impact of IMPROVE

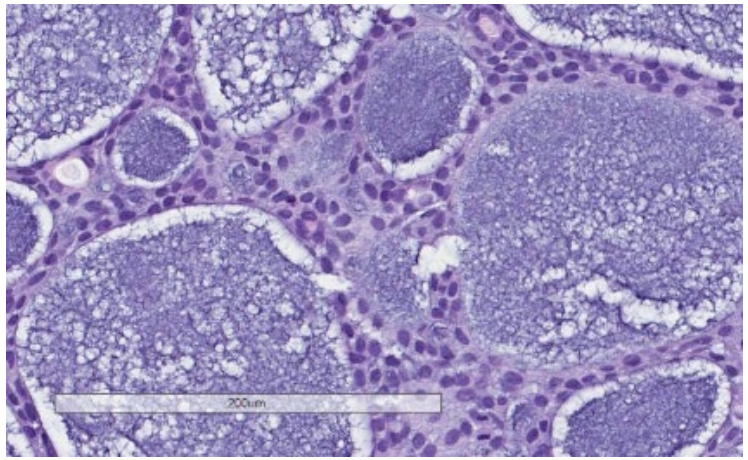
Closing gaps in the development and application of deep learning models for predictive modeling of therapeutic response, including:

- Generating well-curated, clinically relevant, standardized training and testing datasets
- Developing standardized, easily-applicable workflow (including software pipeline, performance metrics, data, etc.) for evaluating and comparing prediction models to drive model improvement and new model development where possible, hastening translation to the clinic
- Understanding the model attributes related to predictive power, interpretability, and uncertainty quantification (including errors and failure to predict and how this is handled) for guidance on future model design
- Engaging the community for expert opinions and collaborations on developing model evaluation framework and generating benchmark data

Identifying approaches for evaluating and improving modeling are intended to be generalizable to deep learning models in other domains in NCI and DOE

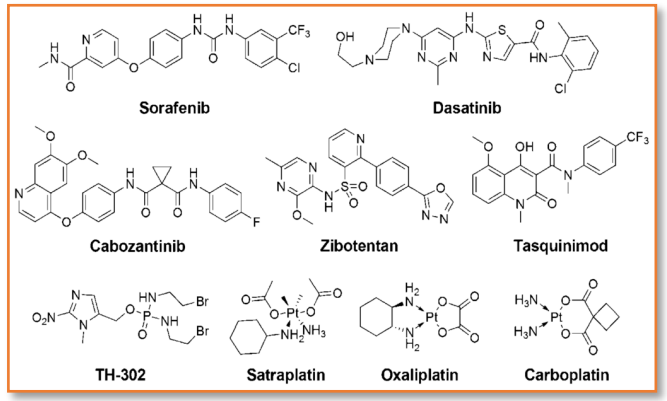
- Materials design, HPC surrogates, etc.
- Have the potential to generate new hypothesis and identify previous hidden cancer types and treatment targets.

Recall: Data Driven Modeling of Cancer Drug Response

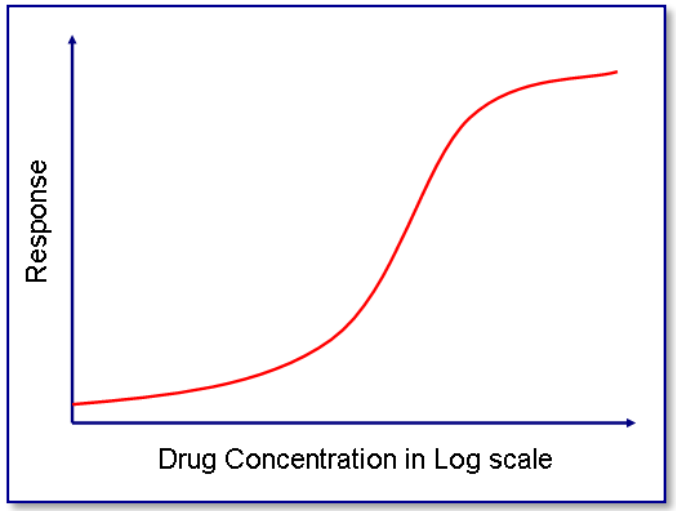


Drug (s)

- descriptors
- fingerprints
- structures
- SMILES
- dose

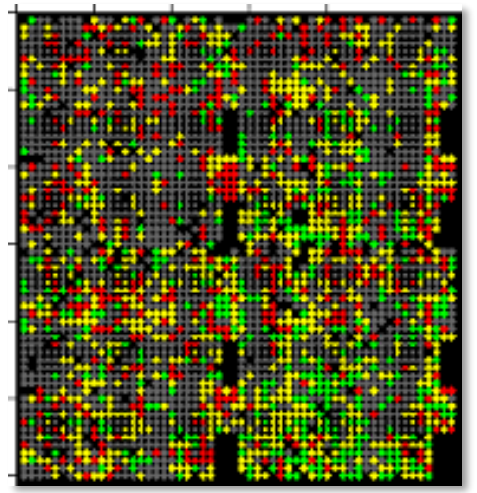


$$R = f(T, D, [P]^*)$$



Response

- IC50
- AUC
- GI50
- % growth
- Z-score



Tumor

- gene expression levels
- SNPs
- protein abundance
- microRNA
- methylation

[P] (patient /treatment history, etc.)*

How much of the predictive power of a given model is due to the structure and nature of the model itself vs. the quality and coverage of the data the model is trained and tested on?

[STRUCTURE] + [PROTOCOL] + [DATA] $\Longrightarrow$ MODEL

If we want to **IMPROVE** the predictive performance of a model, should we:

a) Focus on changing the model structure and tuning hyperparameters, or
b) Improve the datasets (more and better) used for training and testing, or
c) Both?

[PROTOCOL] == hyperparameters, training scheme, etc.

Given what we know and the expanding landscape of public models, how can we make progress? [Models]

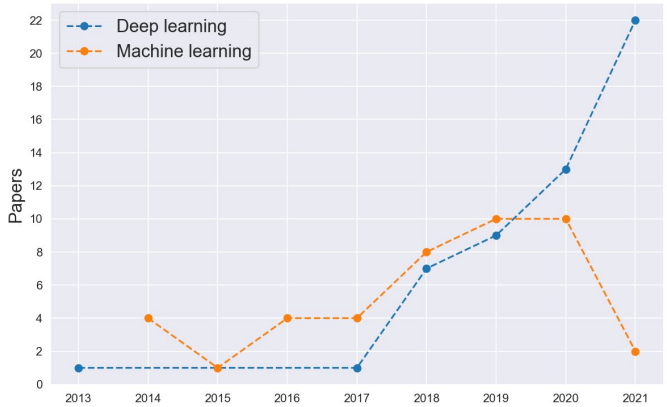
- Our approach focuses on addressing two key bottlenecks for making progress and with broad community engagement
- **Bottleneck 1: Comparing a new model to previous N models (Aim 1)**
  - How to quickly and fairly compare N models and learn which are performing better than others and determine each model's relative strengths and weaknesses
  - Determine what aspects of the model formulation/structure/training protocol, etc. are making a difference in performance while holding training data constant
  - Comparison of training and validation data choices impact on performance
  - Determine the types of errors models are making and why
  - Doing this as automatically as possible
- *Beyond simple validation approaches to more biologically relevant assessment*
- *Work with the community to develop more standard approaches for evaluation*
- ***Goal: an "automated" framework (CANDLE) to make massive cross comparisons feasible***

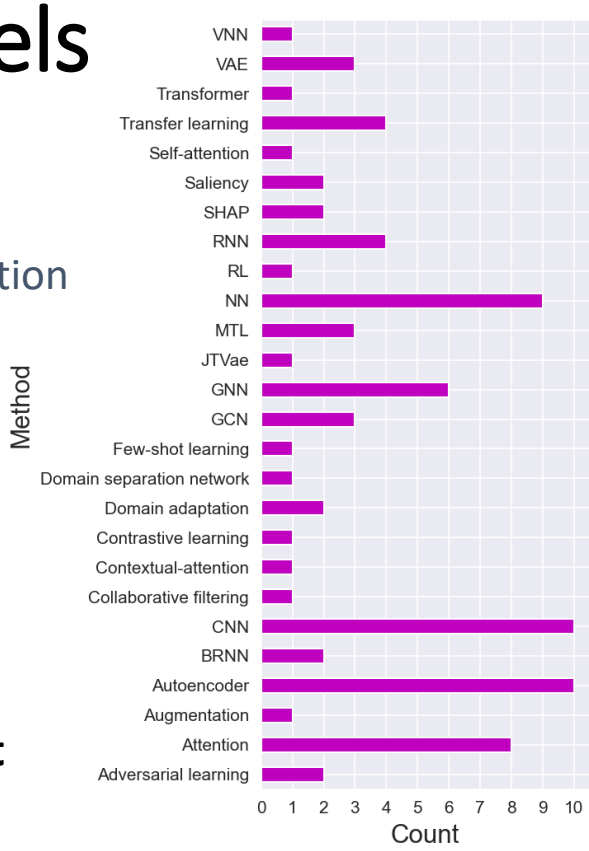
Given what we know and the expanding landscape of public models how to make progress? [Data]

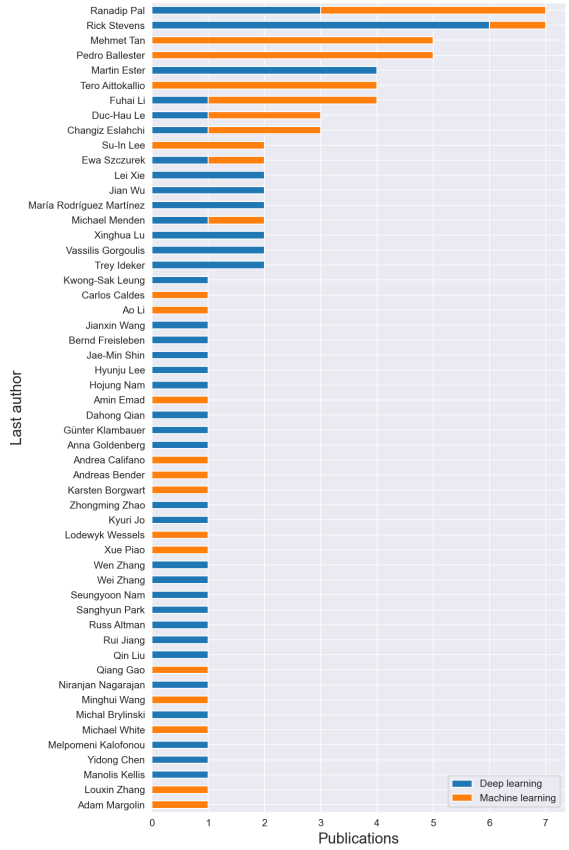
- **Bottleneck 2: What data needs to be generated to improve models (Aim 2)?**
  - Vast majority of the data used to develop current models was **not created for this purpose**
  - By studying **model errors and failures** and how that relates to training and validation datasets *we can determine what new data would be most useful*
  - By understanding **how data quality impacts model performance** *we can determine the standards we need for new training data*
  - By understanding **the learning curve scaling behavior across many models,** *we can determine scale of data needed that would improve models*
  - By understanding **the feature types and modality of training data** *we can determine which assays are needed*
  - By understanding **the impact of data diversity in drug and tumor space** *we can determine the shape (tumor x drugs) of experiments needed to improve performance*
- ***Goal: new datasets explicitly generated to improve models and made widely available***

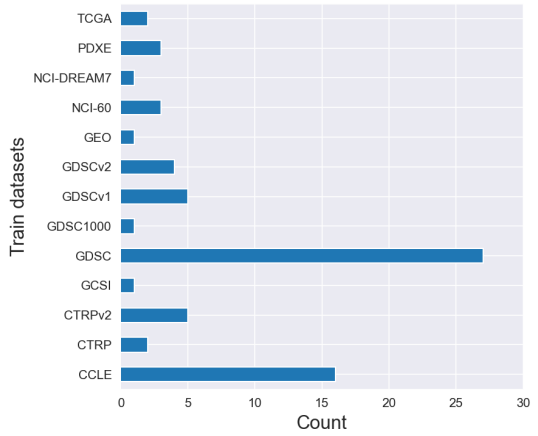
IMPROVE Aim 1: Evaluation and Comparison of State-of-the-art Drug Response Prediction Models

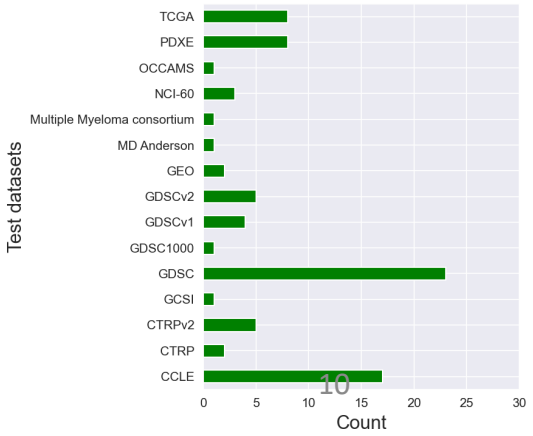
- **Comprehensive literature survey to collect information about research groups and models (ongoing task)**
  - > 100 papers about machine/deep learning drug response prediction
  - Categorize models according multiple criteria to select representative ones for comparison study
  - Model architecture and technique
  - Functionality, e.g. transfer learning, interpretability and uncertainty quantification
  - Code availability and documentation
  - Training and validation data
- **Adapt and modify code to train and test the models, and conduct reproducibility analysis**







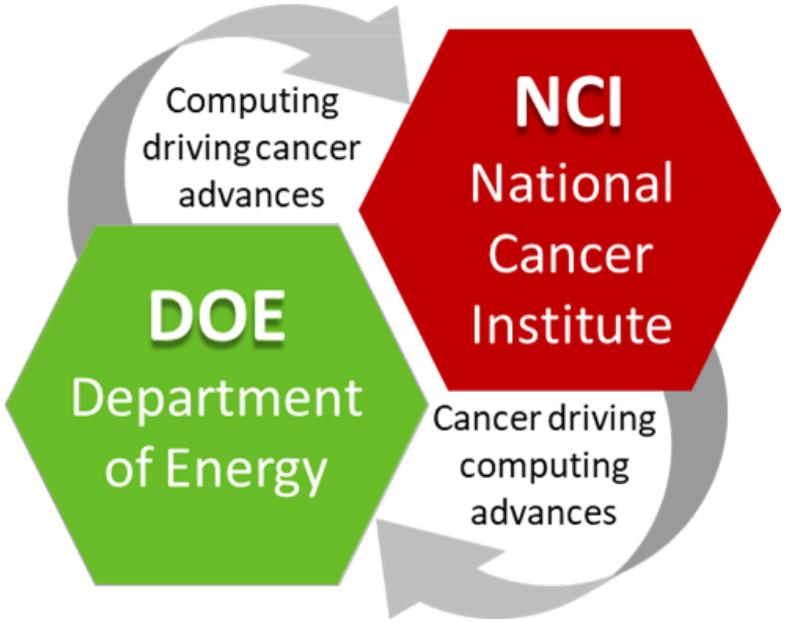




IMPROVE Aim 2: Data Generation to Evaluate and Improve Drug Response Models

- Design and execute high-throughput experiments to generate new data aiming at evaluating and improving drug response prediction models
- Data will include **RNA-seq and DNA-seq** data of cancer models and drug response data with multiple doses and replicates
- Cancer models can be patient-derived organoids (PDOs), xenograft organoids (PDXOs), and primary cell lines (PDCs), which are better representations of patient tumors than immortalized cancer cell lines
- Currently, **most prediction models are built based on drug screening data of immortalized cell lines;** data generated by Aim 2 will be used to:
  - Evaluate the generalizability of prediction models to PDOs, PDXOs, or PDCs
  - Improve prediction models through transfer learning to boost their prediction performance on patient-derived cancer models or patient tumors.
- In addition to data generation, we will continuously curate and standardize new drug screening/response data from public domain

S22-049 IMPROVE Project RFP



S22-049 IMPROVE Project RFP

Goal:

- To award multiple subcontracts to fund extramural research entities with significant experience in AI—especially deep learning research and development—to create the *Collaborative Core Modeling Group (CCMG).*
- The CCMG will work collaboratively with the IMPROVE teams at Argonne National Laboratory (ANL) and Frederick National Laboratory for Cancer Research (FNLCR) to develop semi-automatic protocols for comparing cancer therapeutic response deep learning models and identifying model attributes that contribute to prediction performance with the goal of IMPROVING future models for multiple use cases.

This RFP focuses on AIM 1 IMPROVE Model Comparison:
Development of semi-automatic protocols for comparing cancer therapeutic response deep learning models and identifying model attributes that contribute to prediction performance with the goal of IMPROVING future models.
*A separate RFI has been issued for Aim 2 and can be found **here***

Anticipated Capabilities Developed in IMPROVE

Software

- **A pipeline enables evaluation of new prediction models and comparison with existing state-of-the-art models; standardized evaluation metrics and scenarios will be implemented**
  - GitHub link: https://github.com/JDACS4C-IMPROVE
  - Multiple prediction performance metrics and functional metrics, e.g., interpretability and uncertainty quantification
  - Multiple validation scenarios:
    - Cross-validation within and between benchmark datasets
    - Cross-validation with hard partitions on tumor-drug pairs, tumors, and drugs, simulating different applications
    - Transfer learning between different types of cancer models (e.g., cell lines, PDXs, patients) and different cancer types

Models

- **Existing state-of-the-art drug response prediction models included in the pipeline that can run in batch mode that have been curated/validated and are placed in MODAC for easy adoption by the cancer research community.**
- Improved prediction models through transfer learning on newly generated/curated data

Anticipated Translational Goals of IMPROVE

Benchmark Data

- Newly generated drug screening data on PDOs, PDXOs, or PDCs.
- Newly curated, standardized, and aggregated drug screening/response data on cell lines, PDOs, PDXs, and patients

Advancing the state of the art

- Systematic errors in the ability of AI to predict outcomes/treatments can indicate novel subtypes and highlight previously unappreciated therapeutic targets.
- Potential help move from stage/grade classification to classification based on treatment classes and likelihood of favorable outcome.
- Aiding researchers in knowing, which models are believable and how they can be applied in real world situations.
- Providing a systematic measurement of the value of each type of test/data in relation to cost and patient impact.

Previous Deep Learning Model Curation Efforts

Easy model

- Data and code are available in GitHub
  - Missing instructions for reproducing the environment (only major package names are provided)
  - Missing script of computing saliency maps for feature selection
- Most prediction scores from our re-trained models match the reported ones reasonably well
  - Straightforward to preprocess data from other sources

Medium model

- Dependency versions and full data files are not included in the Github repository
  - A few bugs to fix out of the box to function in the new environment
- Results of re-trained model match published results reasonably well

Difficult model

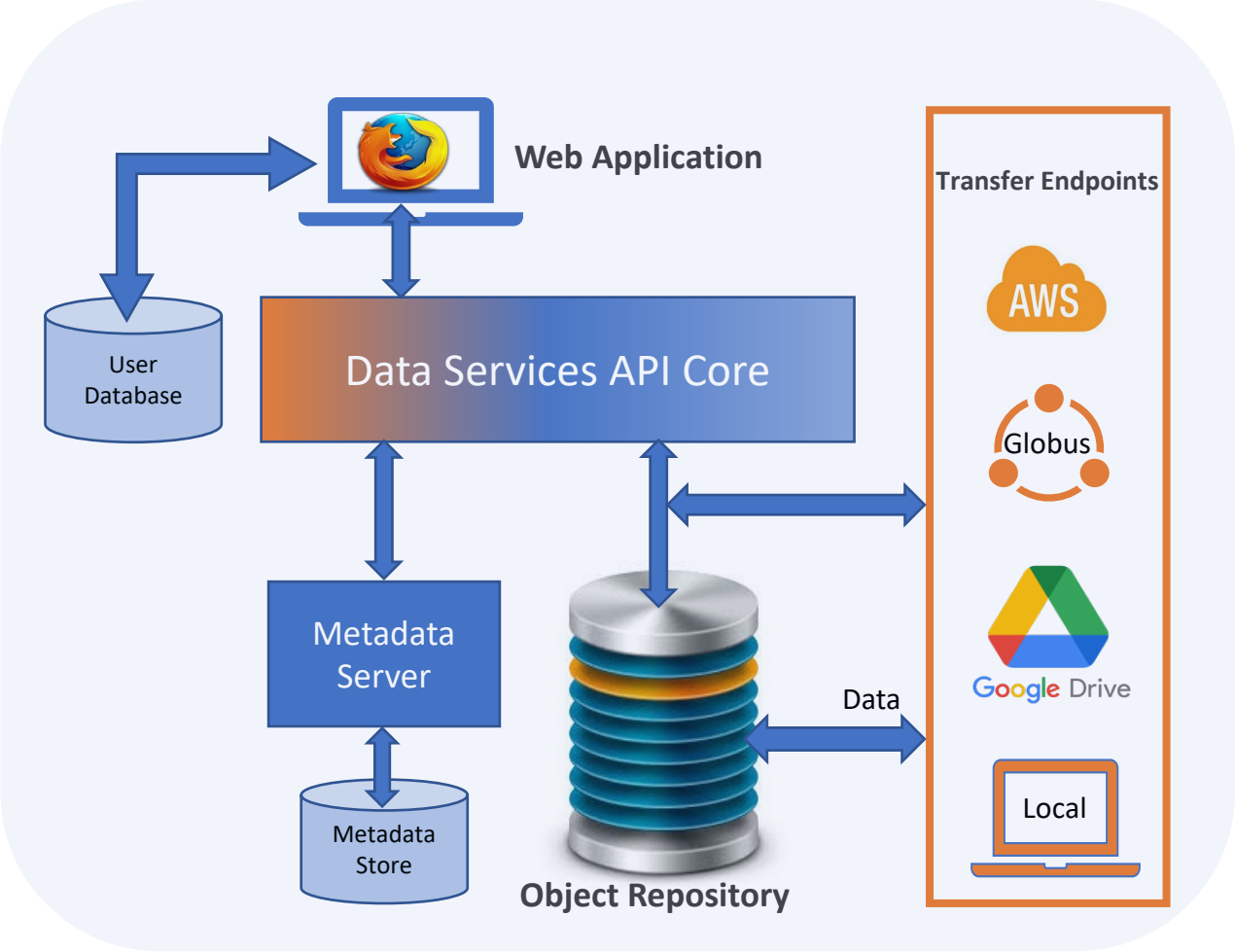
- Pretrained model is accurate to reported results from paper
  - Code, environment and extensibility to new data is easy to do
- The re-trained model using scripts and details provided by publication does not replicate the pretrained model provided by publication.
  - More work is needed to train the model to match the paper

Level of Effort for Previous Model Curation Efforts

- **Easy models** (*code and data are fully available*)
  - 0.5 AI Developer/0.5 Ph.D. level researcher
  - Total time: **one person month**
- **Medium models** (*gaps in available code/data or other issues*)
  - 0.5 AI Developer/0.5 Ph.D. level researcher with fractional efforts from other skillsets
  - Total time: **two person months**
- **Difficult models** (*significant gaps in available resources or results are not reproducible*)
  - 0.5 AI Developer/0.5 Ph.D. Researcher with significant fractional efforts from other skillsets
  - Total time: **four person months**, though this is open and dependent on potential impact
- This is *an interdisciplinary team science approach* utilizing AI (PyTorch, Keras, TensorFlow, etc.), bioinformatics, physics, biology, mathematics, statistics and cancer drug response).
- It is *up to each team to determine the best mix and distribution of skill sets*, but significant gaps should be called out as execution risks.

Year	Publisher	Primary outputs	Code	Framework	Drug set	Cancer features	Drug features	Train bio model	Test bio model	C/W/CW	DL methods	ML methods	Baselines	UQ	XAI
2019	Bioinformatics	AAC	Github	PyTorch	Single	Gene expression	None	CTRPv2	CTRPv2	W	VAE	None	RidgeLR, RF,		
2020	IEEE Journal of Biomedical and	GI50	Github	R		Gene expression	PaDEL fingerprints	NCI-DREAM7			RL	None	IntegratedMR		
2021	BMC Bioinformatics	IC50	Github	NumPy							Autoencoder	HMM, Matrix			
2021	BMC Bioinformatics	IC50	Github	PyTorch	Single	Gene expression,	Morgan	CCLE, GDSC	CCLE, GDSC	W	GNN, Saliency	None	CDRscan,		
2020	IEEE/ACM Trans Comput Biol	AUC, IC50	Github	PyTorch	Single	CNA	Molecular graphs	GDSC	GDSC	W	Attention,	None	tCNNS		Yes
2021	Scientific Reports	AUC, IC50	Github	PyTorch	Single	mRNA expression,	Kinase inhibition	GDSC	GDSC	W	Autoencoder	None	ElasticNet,		Yes
2021	Methods Mol Biol		No	NA	Combo						Autoencoder	None	DeepSynergy,		
2018	Bioinformatics		Github	TF1	Combo	Gene expression	Morgan				NN	None	ElasticNet,		
2021	Journal of Chemical Information	IC50	Github	TF1	Single	Gene expression	Morgan	GDSC1000	CCLE, GDSC1, GDSC2,	CW	Attention,	None	CDRscan,		Yes
2021	BMC Bioinformatics	IC50, binary	Github	PyTorch	Single	Copy number, Gene	None	GDSC	CCLE, CTRP, PDXE, TCGA	CW	Autoencoder	None	AutoBorutaR		
2018	Scientific Reports	IC50	No	TF1	Single	Mutation	PaDEL descriptors	CCLP, GDSC	CCLP, GDSC		CNN	None	RF, SVM		
2021	Bioinformatics	IC50	Github	PyTorch	Single	CNV, Gene	Molecular graphs	GDSCv2	GDSCv2	W	Attention,		CDRscan,		
2019	IEEE/ACM Transaction of	IC50	No	TF1	Single	Gene expression	Morgan	CCLE, GDSC	CCLE, GDSC	W	Autoencoder	None	DNN, KBMF,		
2019	BMC Bioinformatics	IC50	Github	TF1 (w/o Keras)	Single	Mutation	SMILES	GDSC	GDSC	W	CNN	None			
2021	Mathematics	IC50	Github	TF1 (w/o Keras)	Single	Gene expression	Graph structure	GDSC	GDSC	W	GCN	None	Bagging		
2021	bioRxiv		Github	PyTorch	Single						Domain	None	Simpler		
2021	PLoS Computational Biology		Github	PyTorch	Combo	Gene dependency,	Network				Attention,	None	DeepSynergy		Yes
2021	ICLR 2021	IC50	Github	PyTorch	Single	Gene expression	SMILES	CCLE, GDSC	CCLE, GDSC	W	JTVae, VAE	None	MLP, SVR		
2018	ICML	IC50	Yes	TF1/PyTorch	Single	Gene expression	Morgan	GDSC	GDSC	W	Attention,	None	Simplifie		
2019	Molecular Pharmaceutics	IC50	Yes	TF1/PyTorch	Single	Gene expression	Morgan	GDSC	GDSC	W	Attention,	None	Simplifie		
2019	Bioinformatics	IC50, binary	Github	PyTorch	Single	Gene expression	None	GDSCv1	TCGA, PDXE	CW	Autoencoder	None	DNN		
2020	Bioinformatics	IC50, binary	Github	PyTorch	Single	Copy number, Gene	None	GDSCv1	TCGA, PDXE	CW	Adversarial	None	ADDA, MOLI,		
2021	Nature Machine Intelligence	AAC	Github	PyTorch	Single	Gene expression	None	CTRPv2, GDSCv2	CTRPv2, GDSCv2, GEO,	CW	Transfer	None	DeepAll-ERM,		
2021	KDD	AAC	Github	TF2	Single	Gene expression	None	GDSCv2, CTRPv2	GDSCv2, CTRPv2	CW	SHAP	None			Yes
2019	IEEE Biomedical Circuits and	ActArea	No	NA	Single	Gene expression	None	CCLE	CCLE	W	CNN, RNN	None	ElasticNet,		
2013	Plos ONE	IC50	No	Java	Single	Copy number,	PaDEL descriptors	GDSC	GDSC	W	NN	None			
2020	bioRxiv	IC50 to binary	Github	PyTorch	Single			CCLE	CCLE		GNN	None			
2021	Bioinformatics	IC50, binary	Github	PyTorch	Single			GDSC	GDSC, PDXE, TCGA	CW	Autoencoder,	None	ElasticNet,	Yes	Yes
2020	IEEE Annual Computing and	IC50	No	TF	Single	Gene expression		GDSC	GDSC	W	CNN, RNN	None	RF, SVM		
2018	BMC Bioinformatics		No	NA							NN	RF, KNN			
2018	BMC Bioinformatics	AUC	Github	Matlab		Gene expression	None	CCLE, GDSC	CCLE, GDSC	CW	Transfer	None	Different		
2020	Nature Communications	IC50	Github	TF	Single	Gene expression	Descriptors	GDSC, NCI-60	GDSC, NCI-60	W	CNN	None	DNN,		
2018	BMC Bioinformatics														
2020	arXiv	AUC	Github	PyTorch	Single	Gene expression,	Descriptors	CCLE, GDSC, NCI60	CCLE, GDSC, NCI60	W	Attention, NN	None	None		
2020	Scientific Reports	AUC	No	TF1	Single	Gene expression	Descriptors	CCLE, GCSI, CTRP,	CCLE, GCSI	CW	Transfer	LightGBM	LightGBM,		
2021	BMC Bioinformatics	AUC	Github	TF2	Single	Gene expression	Descriptors	CTRPv2, GDSC1,	CTRPv2, GDSC1, GDSC2,	W	NN	None	LightGBM		
2021	Briefings in Bioinformatics				UNO										
2021	Scientific Reports	AUC	Github	TF1	Single	Gene expression	Descriptors	CTRP, GDSC	CTRP, GDSC		CNN	None	LightGBM,		
2020	Bioinformatics	IC50, binary	Github	TF1, DeepChem	Single	DNA methylation,	Molecualr graph	CCLE, GDSC	TCGA	CW	CNN, GCN	None	CDRscan,		
2021	bioArxiv	IC50, ActArea	Github	PyTorch	Single	Gene expression	Pre-trained	GDSC, CCLE, PDXE	GDSC, CCLE, PDXE	CW	TML	None	RF, ST-NN,		
2020	Scientific Reports	IC50 to binary	Github	TF1 (w/o Keras)		Gene expression	Structure	CCLE, GDSC	CCLE, GDSC	W	NN	ElasticNet	CaDRReS,		
2019	International Journal of	IC50	Github	TF		Mutation	Fingerprints	CCLE, GDSC	CCLE, GDSC	W	CNN	None	Ridge		
2020	Cancer Cell	AUC	Github	PyTorch	Single, Combo	Mutations	Morgan	CTRPv2, GDSCv1,	CTRPv2, GDSCv1, PDXE	CW	VNN	None	DNN		Yes
2021	Nature Cancer	AUC	Github	PyTorch	Single	Gene expression,	None	CCLE, GDSCv1, PDXE	CCLE, GDSCv1, PDXE	CW	Few-shot		KNN, LR, RF		
2017	bioRxiv		Github	R		Gene expression	None				NN				
2019	Cell Reports	IC50	Github	R	Single	Gene expression	None	GDSC	OCCAMS, MD Anderson,		NN	None	ElasticNet, RF		
2020	BMC Medical Genomics	AUC, ED50	Github	PyTorch	Single	Gene expression	None	NSCLC	NSCLC		GNN	None	ElasticNet,		
2021	Briefings in Bioinformatics	IC50 to binary	Github	PyTorch		DNA methylation	Molecular graphs	GDSC	CCLE, GDSC	CW	Attention,		DeepCDR,		
2018	Molecular Cancer Research		No	Matlab							Autoencoder	None			
2020	Proceedings of Machine Learning	ActArea to binary	Github	PyTorch	Single	Gene expression	Drug target	CCLE, GDSC	CCLE, GDSC	W	Collaborative	None	Versions of		
2019	BMC Medical Genomics	IC50	No	TF	Single	Gene expression,	None	CCLE, TCGA			Autoencoder	None	Liner		
2021	Nature Communications	IC50, ActArea	Github	TF and R		Gene expression	None	CCLE, GDSC, TCGA	CCLE, GDSC, TCGA	CW	VAE	ElasticNet			
2021	ArXiv	IC50		PyTorch	Single										
2014	Nature Biotechnology	IC20	Yes	DREAM	Pair	Gene expression	None	CCL (DREAM)	CCL (DREAM)		None	DREAM-methods	DREAM-methods		
2016	Bioinformatics		No	Sklearn	Single		Fingerprints (Morg	NCI-60	NCI-60	W	None	RF, SVM			
2020	Frontiers in Genetics										None	Logistic Matrix Factorization			
2020	Scientific Reports	IC50	Github	NumPy	Single						None	Manifold Learning			
2018	Journal of Molecular Biology	IC50	No	NA	Single			CCLE, GDSC	CCLE, GDSC	W	None	Network based	KBMTL		
2020	IEEE International Conference on Knowledge and Syster		No	NA	Single						None	Regression and classification models			
2020	Scientific Reports	AUC	Yes	Sklearn	Single			GDSC	GDSC	W	None				
2017	AMIA Jt Summits Transl Sci Proc				Combo						None	Genomic interaction based network approach			
2018	Pac Symp Biocomput		Upon request		Combo						None	Disease signaling network based approach			
2019	NPJ Syst Biol Appl		Yes	R	Combo						None	genomic interaction based network approach			
2017	BMC Cancer		Github								None				

Predictive Oncology Model and Data Clearinghouse (MoDaC)



- **Clearinghouse for annotated mathematical models and datasets** from NCI collaborations
- **Public facing web interface and RESTful APIs** for submitting data
- **Metadata based search capability** for locating models and datasets. Browsing and filtering support
- **Models and datasets can be staged** in restricted access mode until ready for sharing
- **Multiple endpoint types supported** for data transfer
- **DOI Support**
  - Global identifier per asset
  - Shareable link for citations

https://modac.cancer.gov

MoDaC organization

- Domain agnostic data hierarchy and metadata structure
- Three collection types – Program, Study and Asset, organized hierarchically
  - Models and datasets constitute the lowest level Asset collection.
  - Assets can contain 2 levels of sub-folders.
- Mandatory metadata defined separately for models and datasets
  - Includes attributes to provide information about ML framework, domain and platform.
  - Needs to be submitted along with data.
- Additional user defined metadata can be included during submission or provided separately later.

NIH NATIONAL CANCER INSTITUTE
Predictive Oncology Model & Data Clearinghouse

Everything Needs to be OPEN

- The IMPROVE framework, our model analysis results, any improved models and all the data produced will be open source and available to the whole community
- IMPROVE will hold development hackathons that will be open and an annual meeting that will be open to the community for participation
- IMPROVE will work with agencies, scientific associations and journals to advocate for open models, open data and open source enabling replication of modeling results

Overview from Leidos Biomedical Research, Inc. Subcontracts Team

- Successful Offerors will be awarded a Contract under the principles outlined in the Federal Acquisition Regulation (FAR) Part 15, **Contracting By Negotiation**.

- The resulting agreement will be a ***Contract*** and not a ***Grant***.
- This solicitation is being issued in accordance with FAR Part 6.1, **Full and Open Competition**.
- A Firm-Fixed-Price proposal is requested in response to this solicitation.
- Non-compliant proposals will not be considered. All proposals shall be written and submitted in English.

Overview from Leidos Biomedical Research, Inc. Subcontracts Team (continued)

- **This will be a "Best Value" award, with Technical Factors outweighing Cost/Price proposed**. All factors referenced in the Request for Proposal (RFP) shall be considered when evaluating proposals. **Evaluation factors are:**
  - Technical Approach
  - Team and Key Personnel
  - Experience and Past Performance
  - Project Plan and Work Breakdown Structure
  - Management
  - Cost/Price Reasonableness

- The expectation is that there will be multiple awards issued under this solicitation.
- To be clear: **Notional or aspirational capabilities will be deemed non-compliant.** Demonstrated experience and expertise is required.

Overview from Leidos Biomedical Research, Inc. Subcontracts Team (continued)

- All proposals are requested no later than **12:00 noon Eastern time on Monday, 5/9/22**
- As mentioned in the RFP, Facilities Capital Cost of Money is an unallowable cost under any resulting agreement in accordance with FAR 52.215-17
- Proposals should be submitted to Ms. Natalie Fielman at Natalie.fielman@nih.gov
- All compliant proposals will be reviewed by a panel of FNLCR, NCI, DOE, and ANL personnel against the stated evaluation criteria. While collaborative, this is a FNLCR award.

Questions and Answers

- By popular demand, a Question-and-Answer (Q&A) period will be permitted for this solicitation.
  - The due date for question submission will Close of Business on Friday, 4/15/22.
  - Proposals are due Monday, 5/9/2022 at 12:00 noon (ET).
- This Q&A period will be formalized in an amendment to the RFP, which is forthcoming.
- Please submit questions **in writing** to Natalie Fielman (Natalie.fielman@nih.gov) or Josh Wynne (josh.wynne@nih.gov).
- All questions received will be answered, and a written Q&A document will be provided to all potential offerors who have requested a copy of this Request for Proposal.

Wrap-Up and Open Discussion

- **Key dates to remember:**

4/15 Questions are due

5/9 Proposal submission

- **Questions and submissions should be addressed to:**

Natalie Fielman natalie.fielman@nih.gov **OR**

Josh Wynne josh.wynne@nih.gov