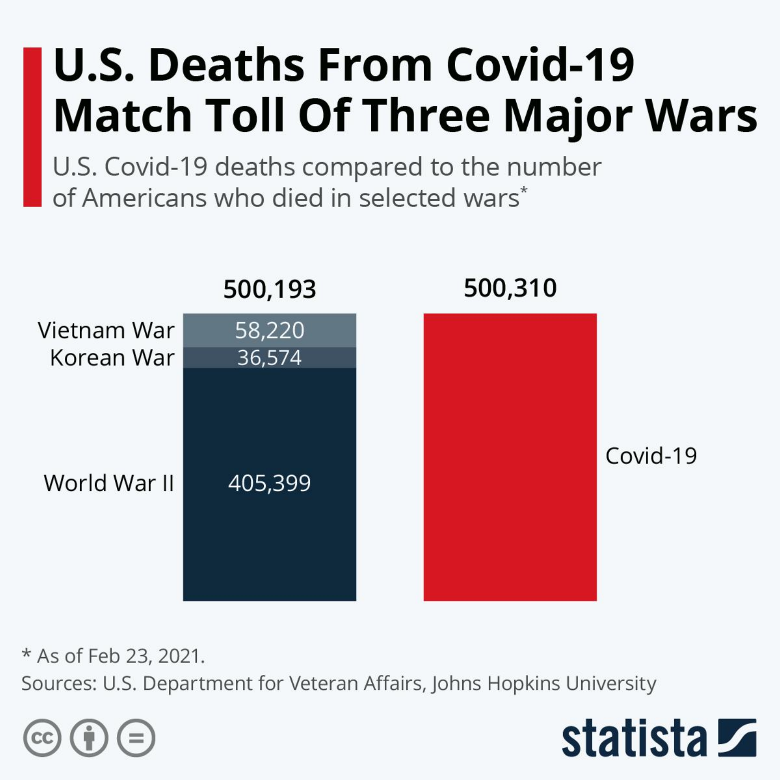
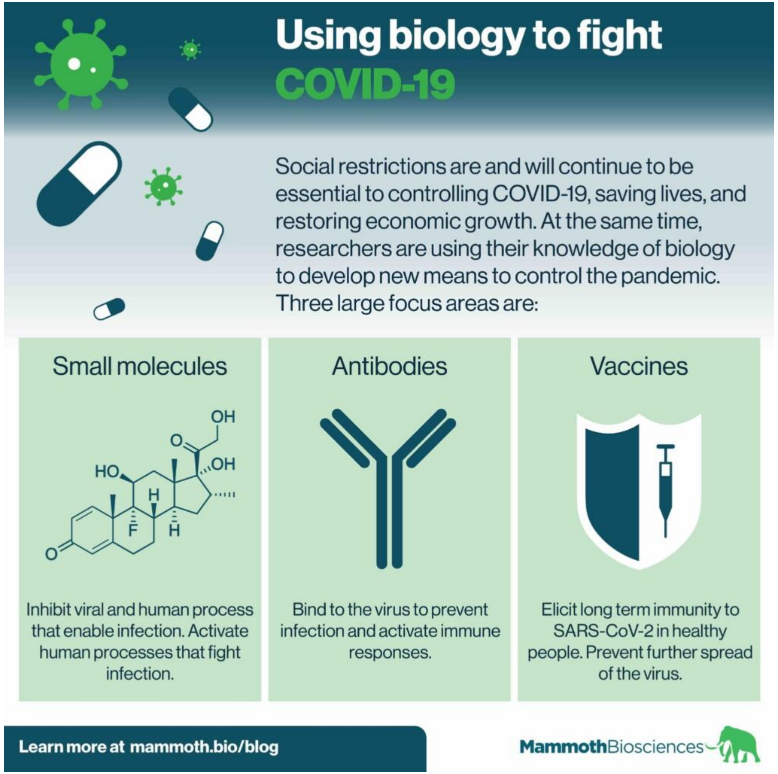
In silico prediction for potential non-phospholipidosis-inducing inhibitors against in vitro replication of SARS-CoV-2

Junqi Lu, Anish Karpurapu

Background

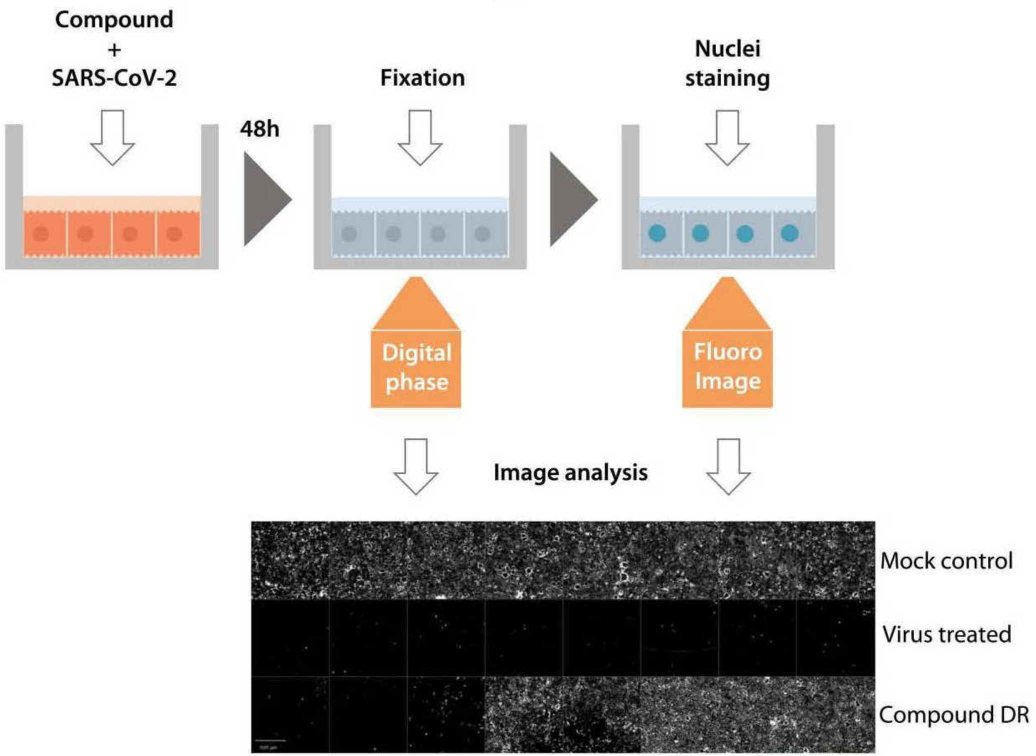
Drug repurposing to mitigate COVID19 emergency





In vitro screening is standard for drug repurposing

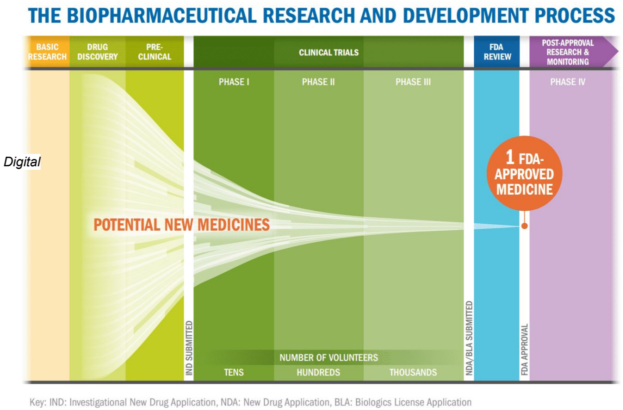
- Ellinger at al screened 5632 compounds for their inhibition of SARS-CoV-2-induced cytotoxicity in Caco-2 cells
- Compounds tested at 10 µM (0.1% in DMSO)
- Inhibition of 75% as the potency cut-off
- 271 hits

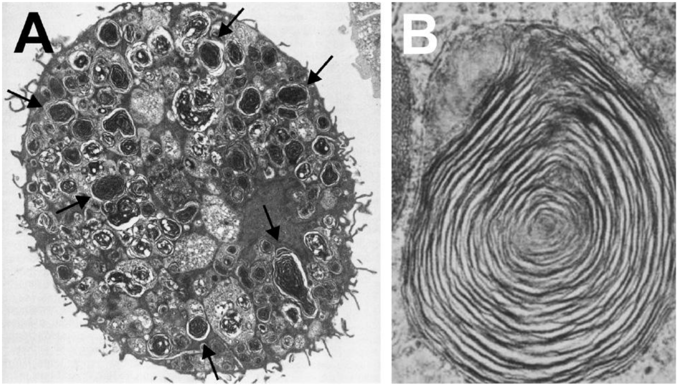


Bernhard Ellinger, Denisa Bojkova, Andrea Zaliani et al. Identification of inhibitors of SARS-CoV-2 in-vitro cellular toxicity in human (Caco-2) cells using a large scale drug repurposing collection, 20 April 2020, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-23951/v1]

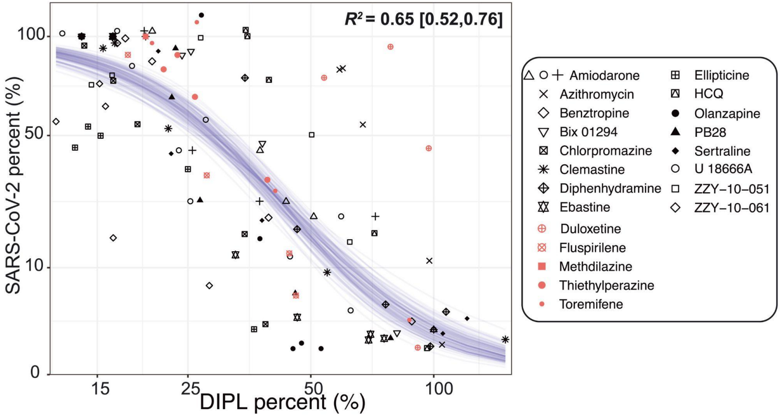
Phospholipidosis: major confounding factor for in vivo translation

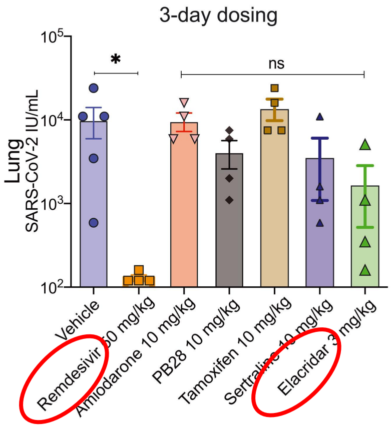
Drug Development and Intellectual Property Theft. *Digital Guardian* 2021.





Breiden, Bernadette and Sandhoff, Konrad. "Emerging mechanisms of drug-induced phospholipidosis" *Biological Chemistry*, vol. 401, no. 1, 2020, pp. 31-46. https://doi.org/10.1515/hsz-2019-0270

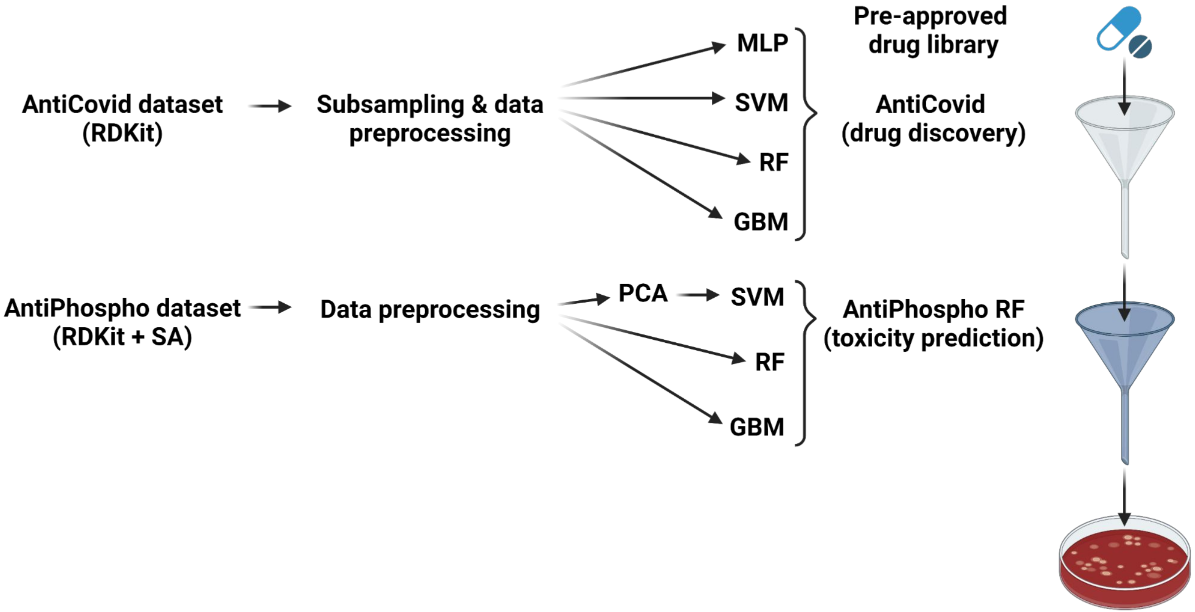




Tummino, T., Rezelj, V., Fischer, B., Fischer, A. et al., Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. *Science* 2021, *373*, 541-547.

Methodology

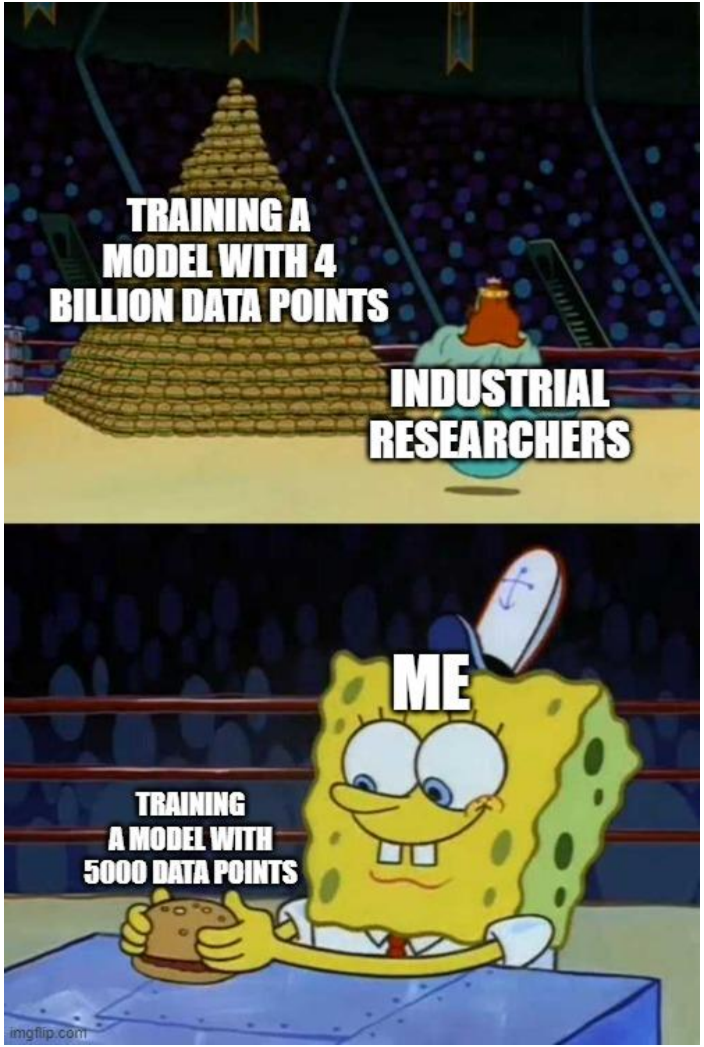
- 20% hold-out testing set
- 10× cross-validation
- Challenges
  - Small datasets
  - Class imbalance
- Baseline success:
  - Hold-out testing set
  - Adversarial controls
  - Compare to random model



COVID-19 Antiviral Model Training

The joys of small datasets...

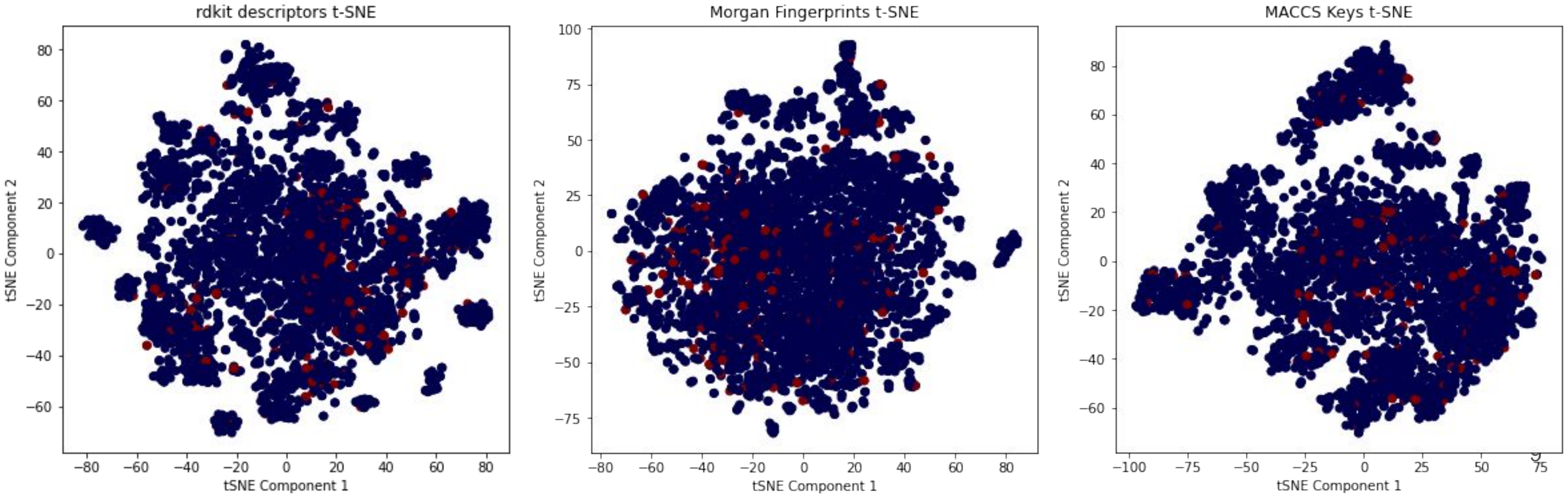
- Typical for wet lab data
- Urgent COVID19 situation



AntiCOVID - exploratory data analysis

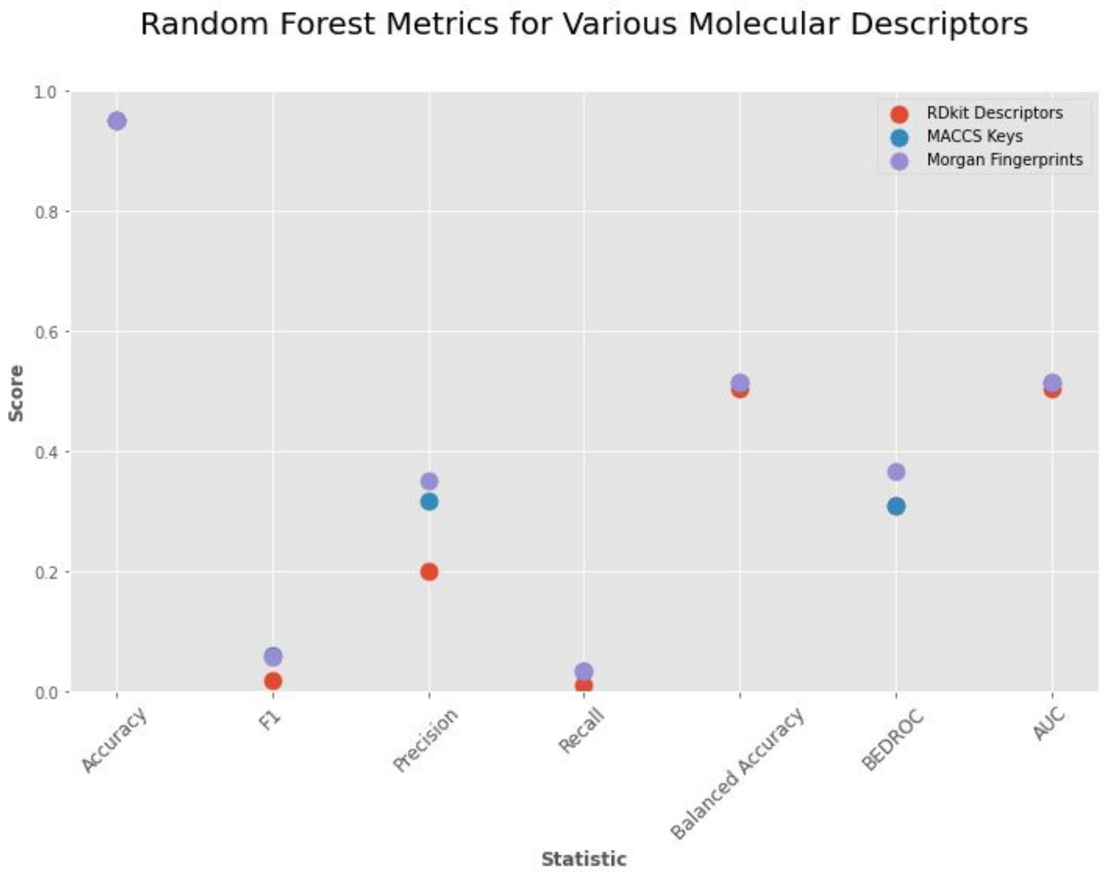
- 5632 compounds with 271 hits

Positives
Negatives

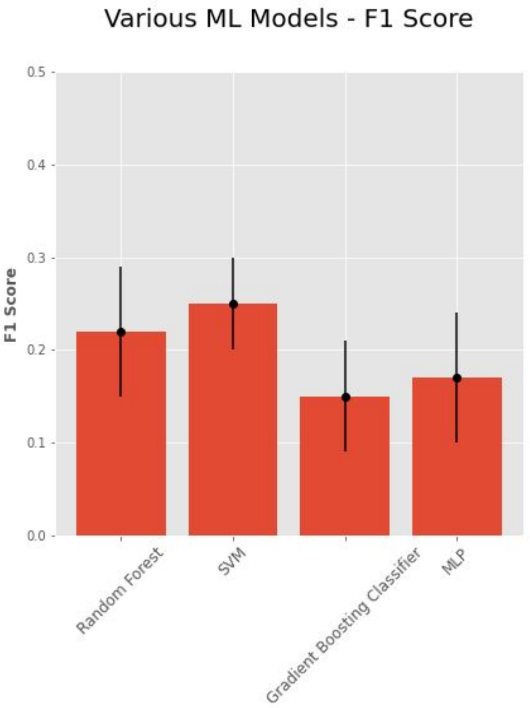


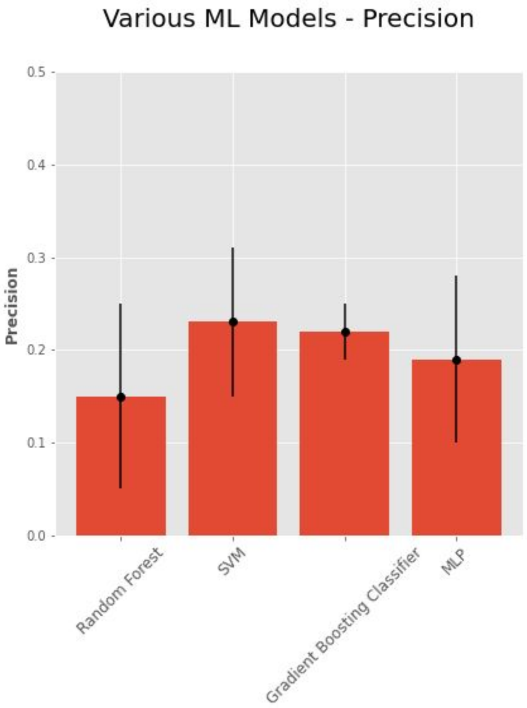
AntiCOVID - descriptors used

- Baseline Random Forest
- Focus on F1 score and Precision



AntiCOVID - models performances with 95% CI



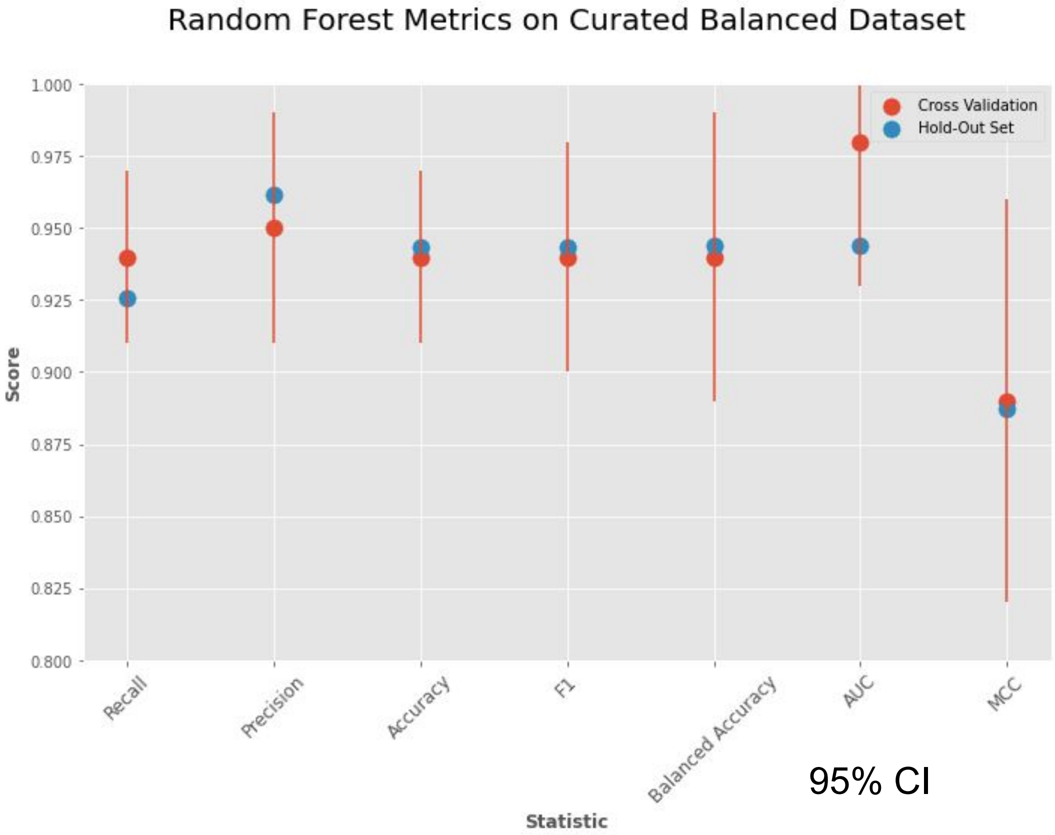


AntiCOVID - balanced subsampled dataset

- Problem: imbalance dataset
- Our Solution: subsample the dataset
    - 271/5632 hits → 271/542 hits
    - 5% → 50%
- Keep hits and “randomly” sample 271 non-hits
    - “Random” → k-medoid on negatives where k = 271
    - Create new dataset

AntiCOVID - balanced subsampled dataset

- RDKit descriptors
- Random Forest Grid Search
- Good results on cross validation and hold-out set

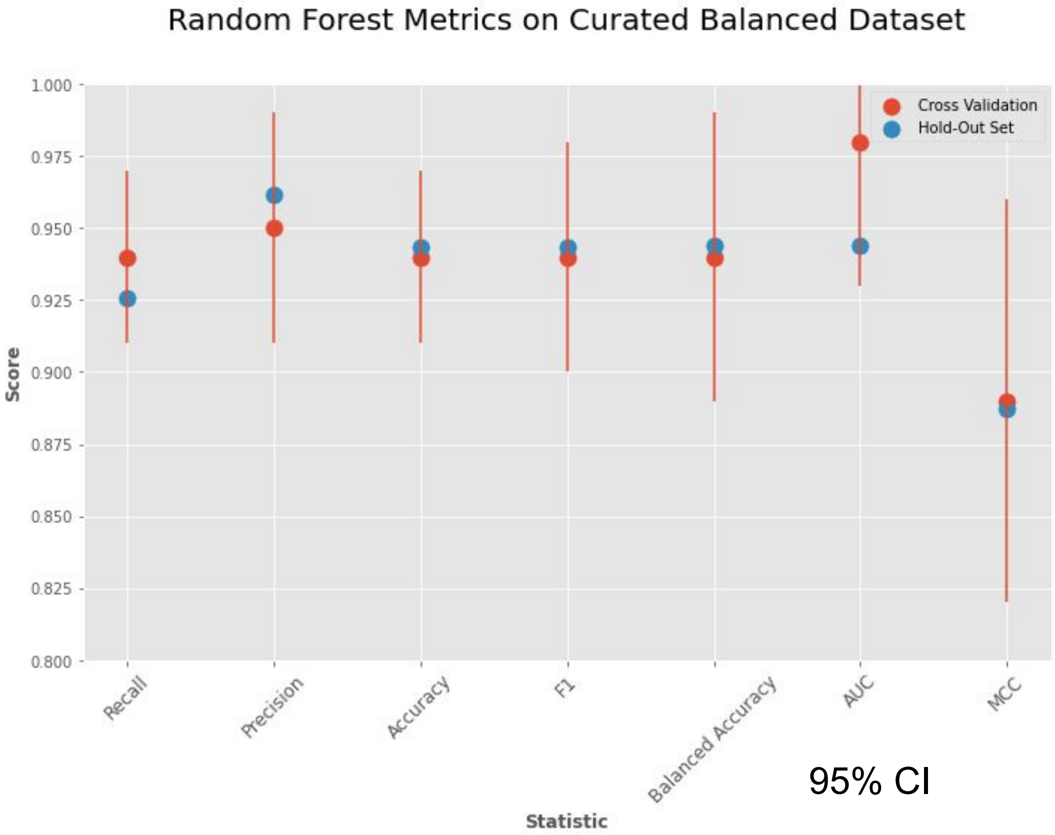


95% CI

AntiCOVID - adversarial control

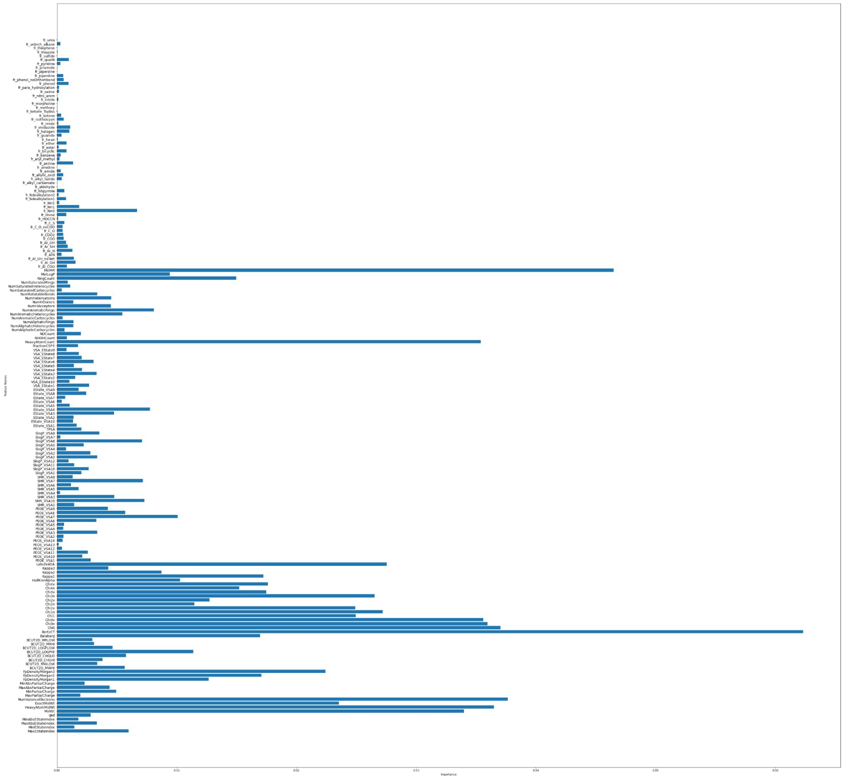
- y-shuffling

```
Accuracy: 0.5471698113207547
F1Score: 0.5471698113207548
Precision: 0.5576923076923077
Recall: 0.5370370370370371
Balanced Accuracy Score: 0.547364
Matthews: 0.0947293447293447З
AUC: 0.5473646723646725
BEDROC: 0.592455000064792 5
```

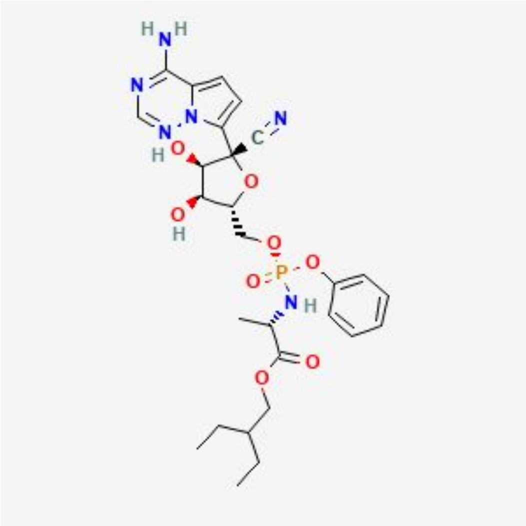


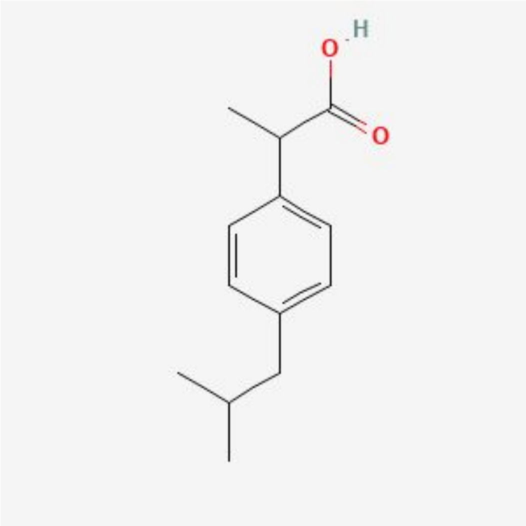
95% CI

AntiCOVID - final RF model’s feature importances



AntiCOVID - test out compounds





Remdesivir
→ FDA-approved anti-COVID-19
→ predicted anti-COVID-19

Ibuprofen
→ exacerbate COVID-19 symptoms
→ predicted NOT anti-COVID-19

AntiCOVID - sanity check

- Run model on whole dataset as well

<u>Results on new balanced hold-out set</u>

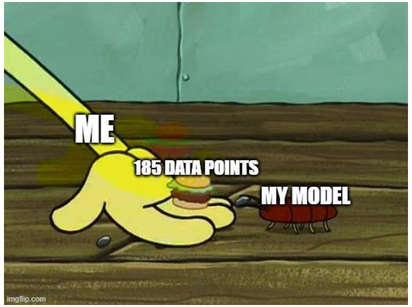
```
Accuracy: 0.9433962264150944
F1Score: 0.9433962264150944
Precision: 0.9615384615384616
Recall: 0.9259259259259259
Balanced Accuracy Score: 0.943732193
Matthews: 0.8874643874643875
AUC: 0.9437321937321939
BEDROC: 0.999903717534156
```

<u>Results on whole dataset</u>

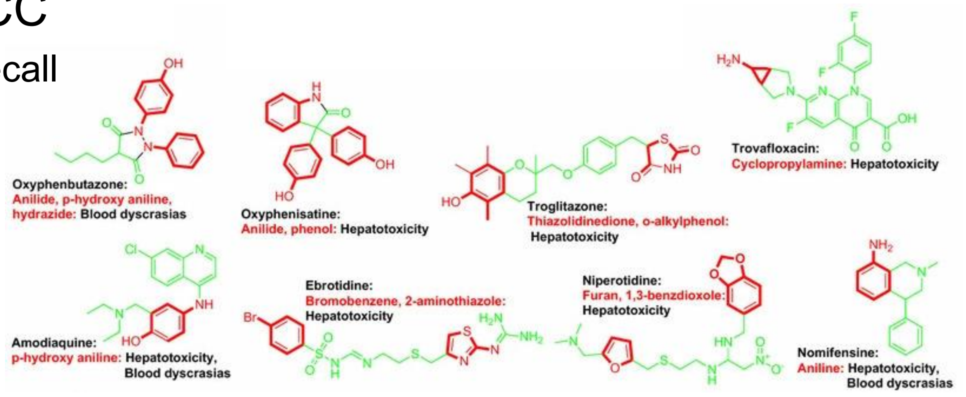
```
Accuracy: 0.27438687073575513
F1Score: 0.1187010078387458
Precision: 0.06315538608198284
Recall: 0.9851301115241635
Balanced Accuracy Score: 0.61121076783
Matthews: 0.11542108722760078
AUC: 0.6112107678303783
BEDROC: 0.252245062588847
```

Phospholipidosis Model Training

AntiPhospho model training



- Challenges: extremely tiny dataset (185 data points)
  - Tree-based models & SVM
  - RF, GBM, SVM (PCs explain 0.95 variances) → RF
- Descriptors: RDKit + Structural Alert (toxicophores)
- Metrics used: recall, ROC-AUC, MCC
  - Toxicity prediction model prefers high recall
- Additional success
  - Cherry picking unknown molecule test
  - Prediction on AntiCovid's dataset



Singh, P.K., Negi, A., Gupta, P.K. *et al.* Toxicophore exploration as a screening technology for drug design and discovery: techniques, scope and limitations. *Arch Toxicol* 90, 1785–1802 (2016). https://doi.org/10.1007/s00204-015-1587-5

AntiPhospho model performances

On 20% hold-out testing set

	Recall	AUC	MCC
svm	0.90	0.832353	0.674702
rf	0.90	0.832353	0.674702
gbm	0.85	0.836765	0.673529
meta rf	0.85	0.866176	0.730208

On shuffled features

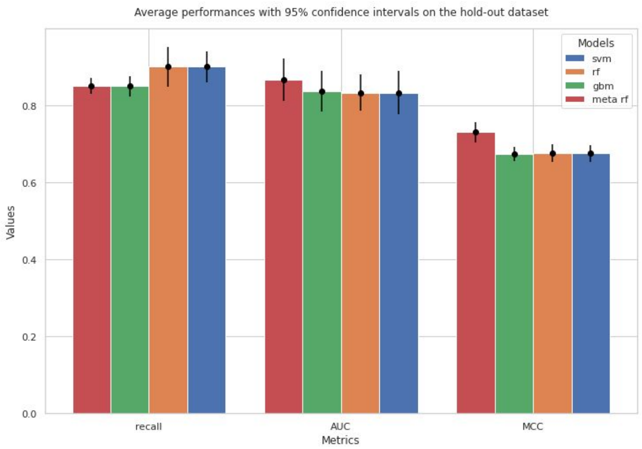
	Recall	AUC	MCC
svm	0.25	0.360294	-0.287115
rf	0.60	0.447059	-0.110531
gbm	0.80	0.547059	0.109323
meta rf	0.45	0.430882	-0.137831

Meta RF is kept as AntiPhospho

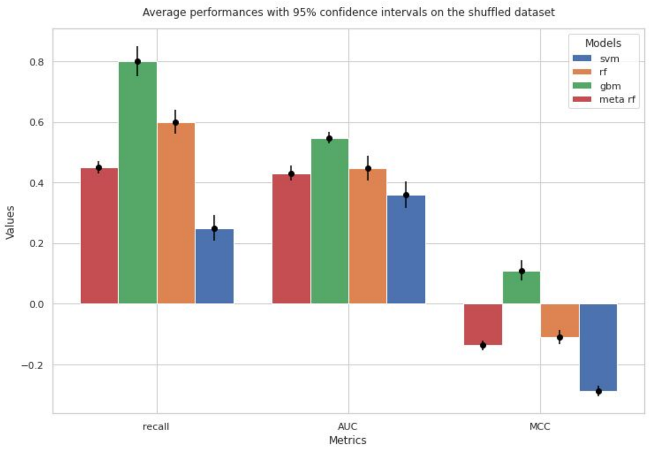
In literature: Meta Learner –RF, Base Learners (Best models +Second best models of GBM, RF, DLNNWOD, DLNNWD) has Recall as 0.86, AUC as 0.89, and MCC as 0.77

AntiPhospho model performances

On 20% hold-out testing set



On shuffled features



Meta RF is kept as AntiPhospho

In literature: Meta Learner –RF, Base Learners (Best models +Second best models of GBM, RF, DLNNWOD, DLNNWD) has Recall as 0.86, AUC as 0.89, and MCC as 0.77

AntiPhospho base RF's feature importances

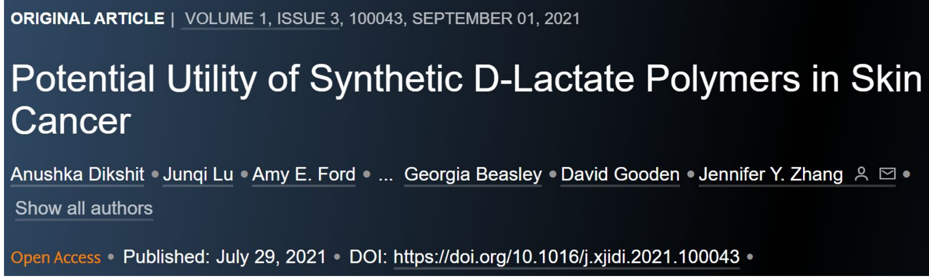
- Chi4n is the most important feature

Cherry picking test set of 3 new compounds

- Remdesivir: the only FDA-approved anti-SARS-CoV-2 drug; non-phospholipidosis-reducing
- Elacridar: an antiviral compound; non-phospholipidosis-reducing
- DLAD: my tested drug (next slide)

	0	1
Remdesivir	0.980198	0.019802
Elacridar	0.722772	0.277228
DLAD	0.831683	0.168317

DIY in vitro drug-test

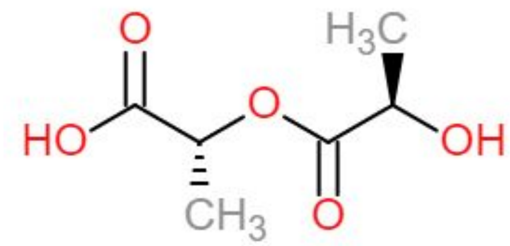

ORIGINAL ARTICLE | VOLUME 1, ISSUE 3, 100043, SEPTEMBER 01, 2021

Potential Utility of Synthetic D-Lactate Polymers in Skin Cancer

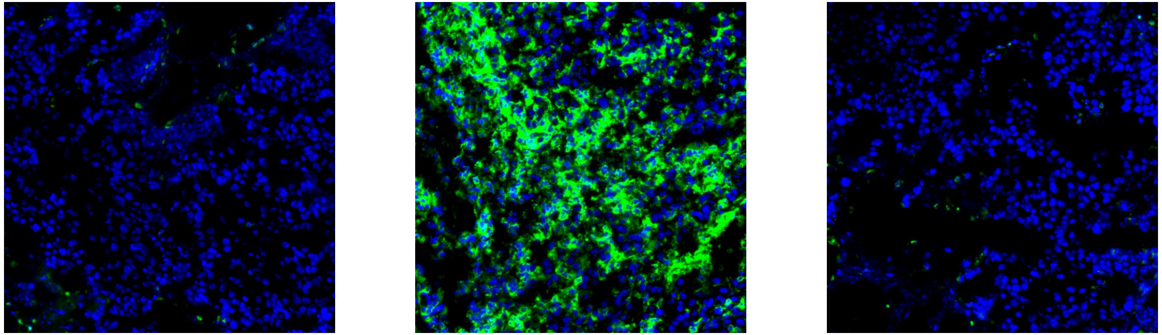
Anushka Dikshit • Junqi Lu • Amy E. Ford • ... Georgia Beasley • David Gooden • Jennifer Y. Zhang •
Show all authors

Open Access • Published: July 29, 2021 • DOI: https://doi.org/10.1016/j.xjidi.2021.100043 •

- Cherry-picking rationale:
  - 2-(2-hydroxy-1-oxopropoxy)propionic acid (DLAD) prohibits skin cancer progression by pushing cellular metabolism profile from anaerobic to aerobic respiration
  - Normal cells prefer aerobic [generate more energy per glucose], while cancer cells and virus-infected cells prefer anaerobic [generate energy faster]
- 83% sure non-phospholipidosis-inducing
- In vitro results: non-phospholipidosis-inducing



DMSO **Sertraline (10 μM)** **DLAD (10 μM)**



DAPI
LAMP-2

34% SARS-CoV-2 positives are phospholipidosis-inducing

- AntiPhospho predicted 34% of the active anti-SARS-CoV-2 molecules from the in vitro screening to be phospholipidosis-inducing

	Classes	Counts
0	0	178
1	1	92


Science
Current Issue First release papers Archive About
HOME > SCIENCE > VOL. 373, NO. 6554 > DRUG-INDUCED PHOSPHOLIPIDOSIS CONFOUNDS DRUG REPURPOSING FOR SARS-COV-2
RESEARCH ARTICLE | CORONAVIRUS
Drug-induced phospholipidosis confounds drug repur-posing for SARS-CoV-2
TIA A. TUMMINO, VERONICA V. REZELJ, BENOIT FISCHER, AUDREY FISCHER, MATTHEW J. O'MEARA, BLANDINE MONEL, THOMAS VALLET, KRIS M. WHITE, ZIYANG ZHANG, [...] BRIAN K. SHOICHET +20 authors Authors Info & Affiliations
SCIENCE • 30 Jul 2021 • Vol 373, Issue 6554 • pp. 541-547 • DOI: 10.1126/science.abi4708

37% AntiCovid's predictions are phospholipidosis-inducing

- AntiPhospho predicted 37% of the AntiCovid's positive predictions to be phospholipidosis-inducing

	Classes	Counts
0	0	34
1	1	20

Science

Current Issue First release papers Archive About

HOME > SCIENCE > VOL. 373, NO. 6554 > DRUG-INDUCED PHOSPHOLIPIDOSIS CONFOUNDS DRUG REPURPOSING FOR SARS-COV-2

RESEARCH ARTICLE | CORONAVIRUS

Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2

TIA A. TUMMINO, VERONICA V. REZELJ, BENOIT FISCHER, AUDREY FISCHER, MATTHEW J. O'MEARA, BLANDINE MONEL, THOMAS VALLET, KRIS M. WHITE, ZIYANG ZHANG, [...] BRIAN K. SHOICHET +20 authors Authors Info & Affiliations

SCIENCE • 30 Jul 2021 • Vol 373, Issue 6554 • pp. 541-547 • DOI: 10.1126/science.abi4708

Summary

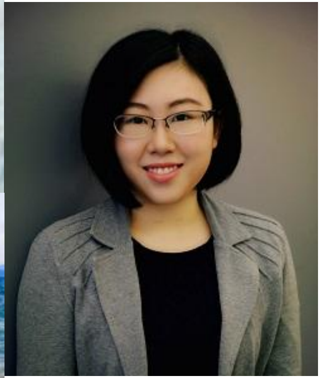
- AntiCOVID has great performance on the subsampled dataset, but not overall
- AntiPhospho has great performance overall with way simpler structure

- K-medoids strategy to undersample imbalanced dataset
- Meta learner with simple base learners to boost performance
- Performance validated by hold-out testing set, adversarial control, orthogonal data, and biological mechanism logics

- Limitation of machine learning: correlation ≠ causation

Thank you!

- Instructor: Daniel Reker, PhD
- TAs: Roujia Wang, Zilu Zhang



Duke | PRATT SCHOOL of ENGINEERING

Duke University
School of Medicine