Current Deep Learning Models for Predicting Kinase Inhibitor Binding Affinities Generalize Poorly

Gabriel Ong | Karanicolas Lab

Presentation for Group Meeting 9/2

An Overview

- Deep Learning, Kinases, and Drug Discovery
- *In silico* methods as an alternative to experimentation?
- Training a CNN model
- Additional experiments
- Implications and Discussion

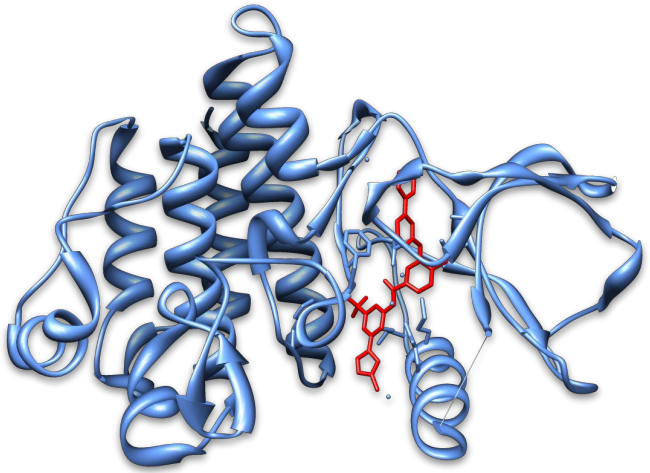
Deep Learning and Drug Discovery

- Deep learning describes training neural networks with more than one hidden layer and is especially suited for discovering patterns in high dimensional data.
- Models are trained by optimizing tunable parameters such as the weights and biases of neurons in the neural network.
- More recent work has highlighted the potential for the use of convolutional neural networks (CNNs) in drug discovery [1].
    - CNNs consider features *invariant to location*, this is makes it suited to recognizing the same chemical feature in different parts of molecules.

[1] Ozturk H, Ozgur A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. *Bioinformatics*. 2018; 34:821-9.

Kinases, Cancer, and Drug Discovery

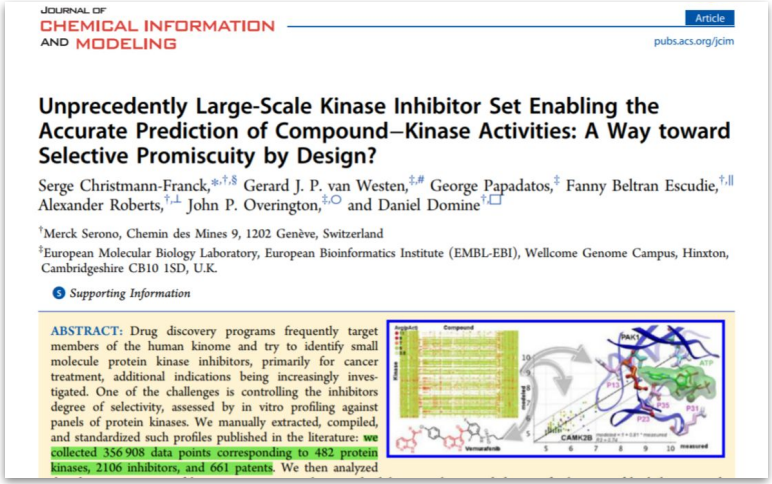
- Kinases play a key role in diseases such as:
  - Cancers
  - Inflammatory diseases
  - Autoimmune disorders
- Kinases already have a strong record as pharmaceuticals, with 48 drugs gaining FDA approval [1].



[Image] "Crystal Structure of nilotnib and Abl kinase" *Wikipedia*. https://en.wikipedia.org/wiki/Tyrosine_kinase_inhibitor
[1] Roskoski R, Jr. Properties of FDA-approved small molecule protein kinase inhibitors: A 2020 update. *Pharmacol Res*. 2020; 152:104609.

The Challenges of Kinase Inhibitor Discovery

- Kinase inhibitor libraries are large due to diverse chemical space and large number of kinases.
- Screening *in vitro* with HTS is time consuming and expensive [1].



JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Unprecedently Large-Scale Kinase Inhibitor Set Enabling the Accurate Prediction of Compound−Kinase Activities: A Way toward Selective Promiscuity by Design?

Serge Christmann-Franck,[*,†,§] Gerard J. P. van Westen,[‡,#] George Papadatos,[‡] Fanny Beltran Escudie,[†,∥] Alexander Roberts,[†,⊥] John P. Overington,[‡,○] and Daniel Domine[†,□]

[†]Merck Serono, Chemin des Mines 9, 1202 Genève, Switzerland

[‡]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, U.K.

Supporting Information

ABSTRACT: Drug discovery programs frequently target members of the human kinome and try to identify small molecule protein kinase inhibitors, primarily for cancer treatment, additional indications being increasingly investigated. One of the challenges is controlling the inhibitors degree of selectivity, assessed by in vitro profiling against panels of protein kinases. We manually extracted, compiled, and standardized such profiles published in the literature: we collected 356 908 data points corresponding to 482 protein kinases, 2106 inhibitors, and 661 patents. We then analyzed

[1] Macarron R, Banks MN, Bojanic D, Burns DJ, Cirovic DA, Garyantes T, Green DV, Hertzberg RP, Janzen WP, Paslay JW, Schopfer U, Sittampalam GS. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov.* 2011; 10:188-95.

Virtual Screening: A Possible Solution?

- In light of the challenges posed by experimental screening of kinase libraries, do *in silico* deep learning techniques provide a viable alternative to *in vitro* screens? *Maybe…*

Goal: Train a model on a set of experimentally tested kinase inhibitors then use it to search synthetically accessible chemical space for previously unknown kinase inhibitors.

- Thus our model must perform well under these conditions: where the model is asked to make predictions on previously unseen inhibitors.

Motivating Question: Do current models perform well in settings where they are asked to predict the binding affinities of previously unseen inhibitors?

Methods: Overview

Motivating Question: Do current models perform well in settings where they are asked to predict the binding affinities of previously unseen inhibitors?

- To answer the question...
  - We design a deep learning model comparable to those in the literature
  - We test our model on previously unseen inhibitors

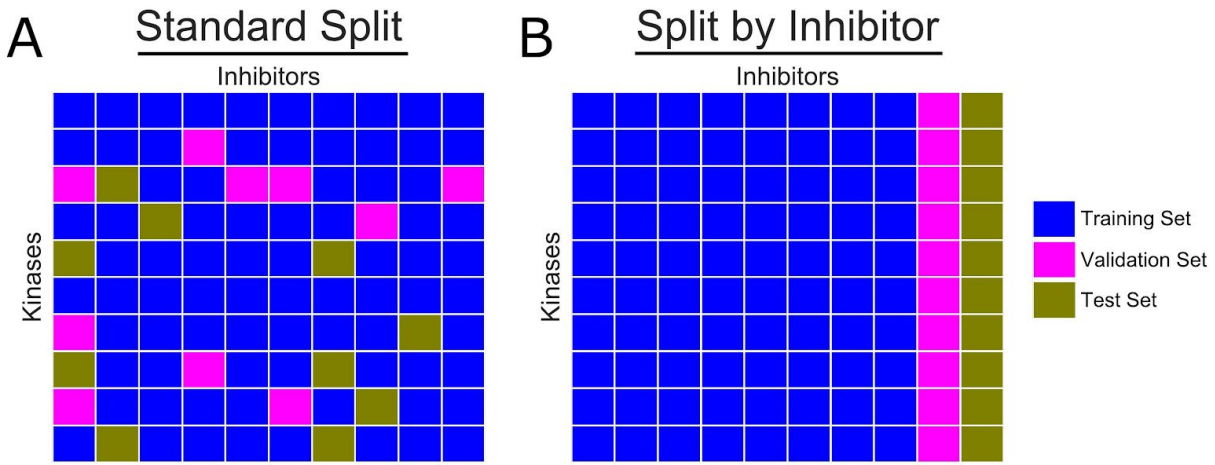
Methods: Datasets and Processing

Methods: Datasets and Processing

- Data points with kinase mutants and SMILES strings >90 characters are discarded, protein kinase information converted to KLIFS sequences [1].
- Anastassiadis Dataset: Convert bioactivity labels to pIC50 using Hill-slope of 1
- Elkins Dataset: Remove data points with Kd with <0 and convert to pKd

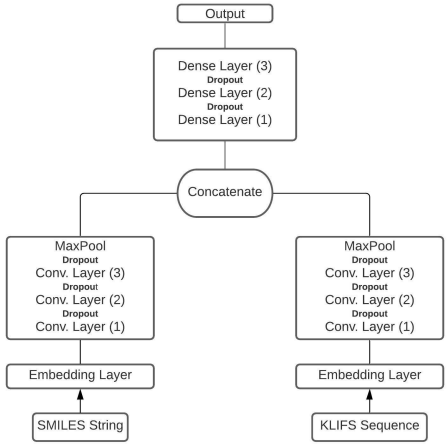
[1] van Linden OP, Kooistra AJ, Leurs R, de Esch IJ, de Graaf C. KLIFS: a knowledge-based structural database to navigate kinase-ligand interaction space. *J Med Chem.* 2014; 57:249-77

Methods: Datasets and Processing

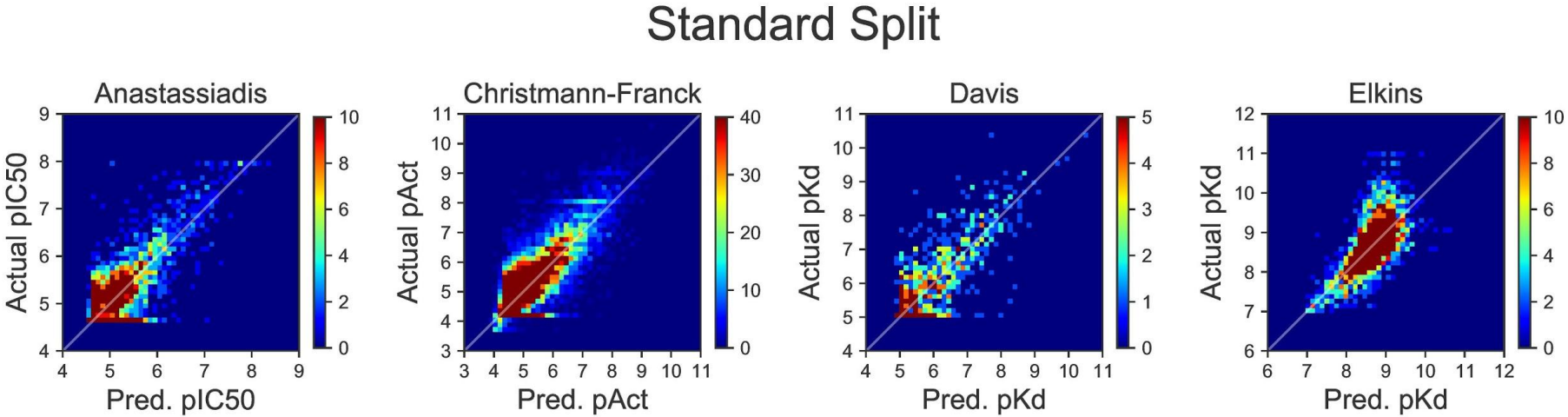


Methods: Neural Net Model

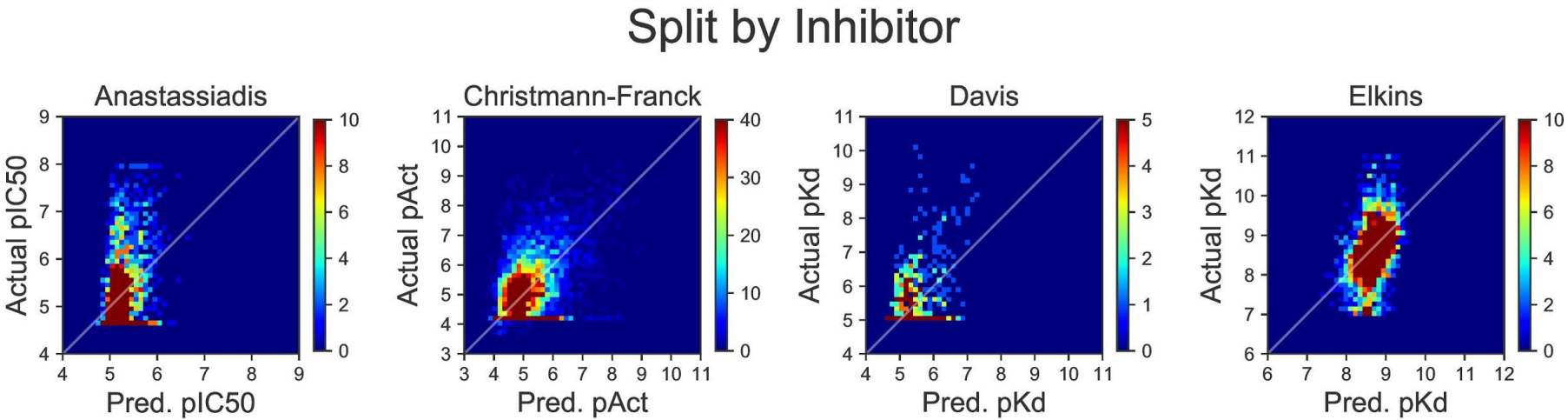
- Embedding layer transforms KLIFS and SMILES information into dense vectors of fixed length
- CNN blocks to extract features from SMILES strings and KLIFS sequences
- Dense network for multivariable regression to predict binding affinity
- Hyperparameters optimized with systematic tuning optimizing for best performance on test-set data



Results: CNN Model



Results: CNN Model



Results: CNN Model

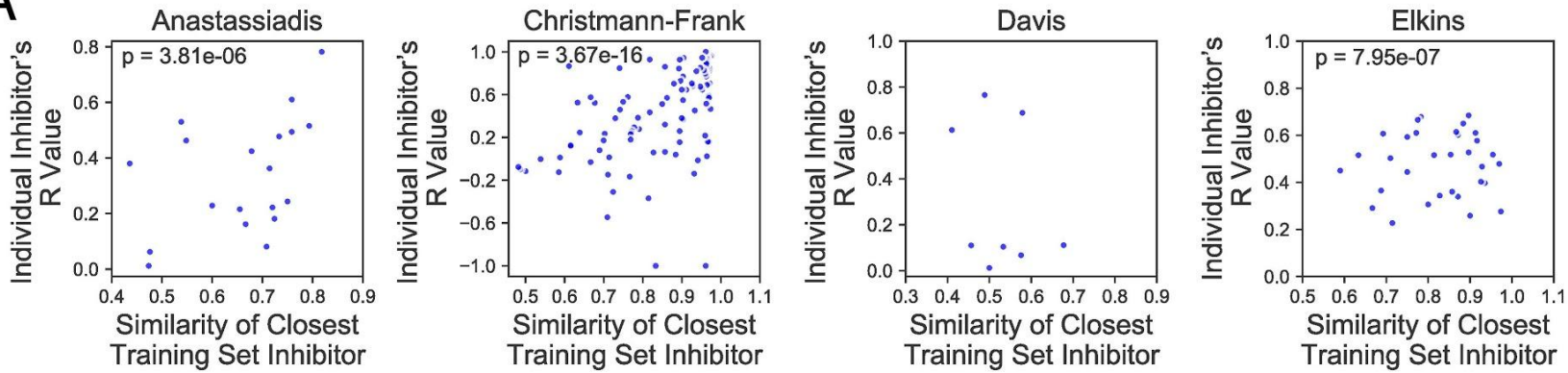
- Our standard split model performs ***at similar levels*** *to* existing models in the literature with a concordance indices and correlation coefficients of about 0.8 and similar mean squared error (MSE).
- Our standard split model ***outperforms*** DeepDTA on the Davis set, the common dataset between our experiments
    - Higher concordance index: **0.896** vs. **0.878**
    - MSE: **0.177** vs. **0.261**
- Performance ***deteriorates significantly*** on the Split by Inhibitor model, exhibiting lower correlation coefficients and concordance indexes and higher MSE on all four datasets, indicating our model is unable to generalize to previously unseen information.

Methods: MCS-Based Experiments

- We use the maximum common substructure (MCS) tool in RDKit to calculate similarity of inhibitors with each other.
- We use this data to run two experiments:
    - One where we test if the presence of closely related inhibitors in the training set affects our model's performance on a given test set compound.
    - Another where we test if our model's predictive capacity arises from it transferring training labels from the most closely related training set inhibitor.

Results: MCS Correlation Experiment

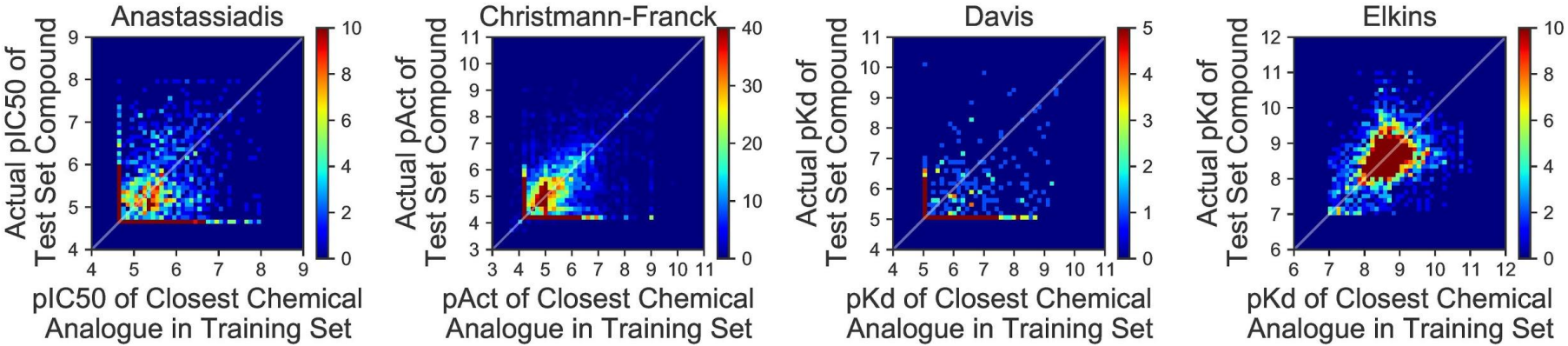
A



Results: MCS Correlation Experiment

- We demonstrate that the ***correlation between an individual inhibitors' R value and the MCS similarity of the closest training set compound is statistically significant*** ($p<0.001$) using the Wilcoxon Rank Sum test.
- These results indicate that our model suffers from information leakage by using training labels of closely related inhibitors to make predictions for test set compounds.

Results: MCS Label Transfer Experiment



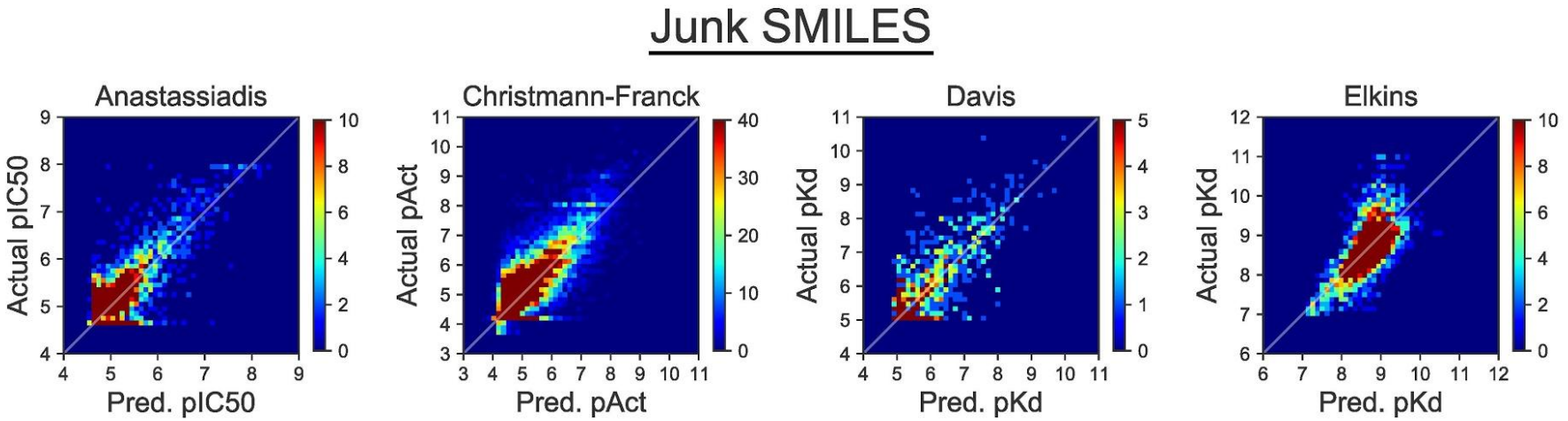
Results: MCS Label Transfer Experiment

- On average, our model that solely transfers training labels from the closest training set inhibitor ***performs similarly to*** our Split by Inhibitor model.
- On the Davis set, our label transfer model ***outperforms*** the Split by Inhibitor model
  - Higher CI: **0.711** vs. **0.694**
  - Higher correlation coefficient: **0.419** vs. **0.409**
- Near-equivalent performance to the Split by Inhibitor model indicates that our machine learning model is ***not doing much more than transferring the training labels of the closest chemical analogue in the training set.***

Methods: Junk SMILES Experiment

- We replace SMILES strings with random character strings that are unique to each inhibitor.
- We use these random strings in place of SMILES strings when training our CNN model.
    - This allows our model to differentiate between different inhibitors without any intuition about chemical structure.

Results: Junk SMILES



Results: Junk SMILES

- Our Junk SMILES model performs ***at similar levels to*** our Standard Split model with a concordance indices and correlation coefficients of about 0.8 and similar MSE.
- In some cases, our Junk SMILES model ***outperforms*** our Standard Split model:
    - Anastassiadis Dataset: Better CI of **0.768** vs. **0.763**
    - Elkins Dataset: Lower MSE of **0.145** vs. **0.147**
- The high performance of our Junk SMILES model where the CNN model is given meaningless strings in place of the inhibitors' information indicates that our model is memorizing kinase phylogeny.

Discussion: Summary

- When trained in a similar way to existing models, our model performs competitively with those in the literature.
- Performance significantly deteriorates when our model is trained on the Split by Inhibitor scheme, the way we would ideally apply such models in the real world.
    - In the case of the Davis set, CI drops from 0.896 to 0.694.
- When unable to use memorized kinase phylogeny to make predictions, our model searches the training set for the most closely related inhibitor and copies the binding affinity labels.

Discussion: Implications

- Most existing models are trained with data randomly sorted into training, test, and validation sets indicating they may not be as predictive as once thought.
- SMILES strings may not be sufficiently meaningful representations of inhibitors under the constraints of current dataset sizes -- 3D representations may be needed.
- Even when trained on the Split by Inhibitor scheme, information leakage still arises.

Further Work

- Learning from this paper -- benchmarking models (eg. 3D CNN project) on Split by Inhibitor scheme
- Examining on other types of similarity measures
- Finding a way to minimize information leakage from closely related inhibitors in the training set.

Thanks and Acknowledgements

Project Collaborators: Kiruba Palani and John Karanicolas

Karanicolas Lab: Grigorii Andrianov, Lei Ke, Jake Khowsathit, Sven Miller, Chris Parry, Daniel Yeggoni

FOX CHASE CANCER CENTER
TEMPLE HEALTH