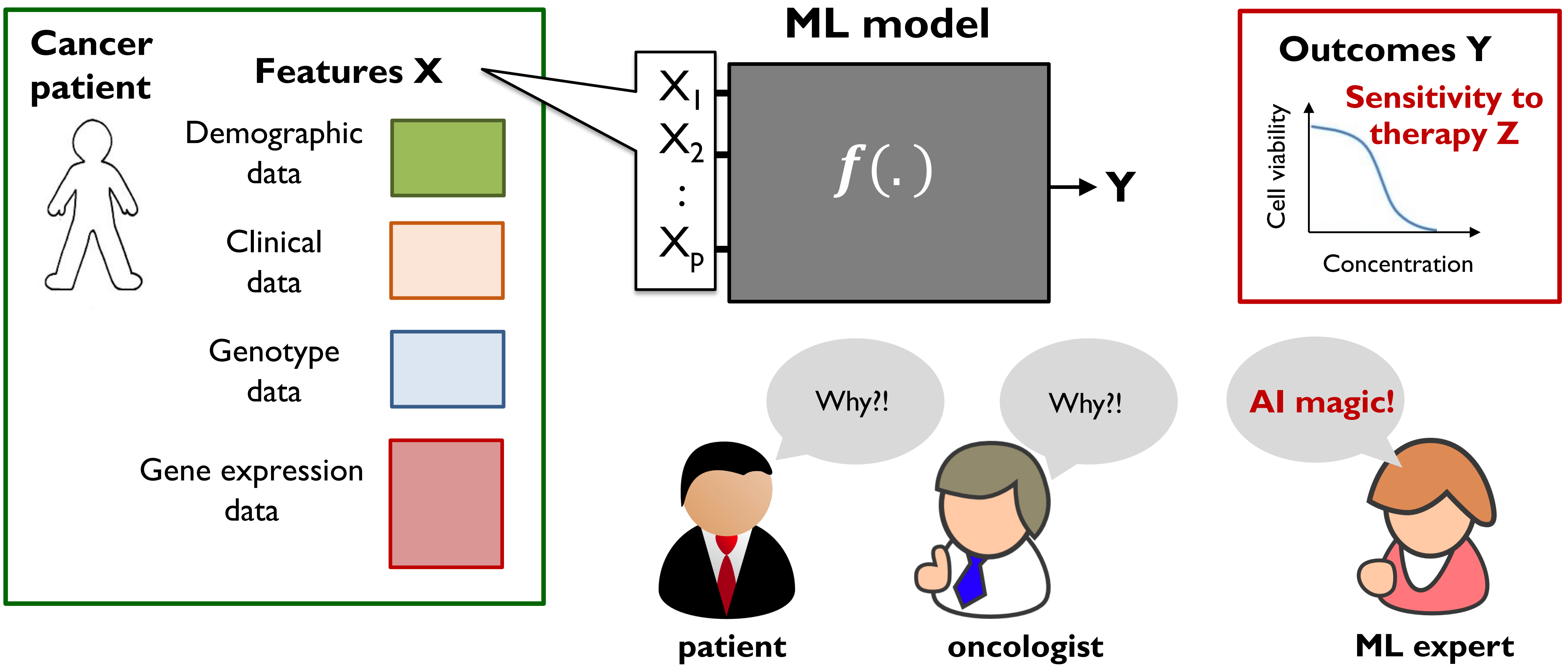
# Explainable AI for cancer precision medicine

Su-In Lee

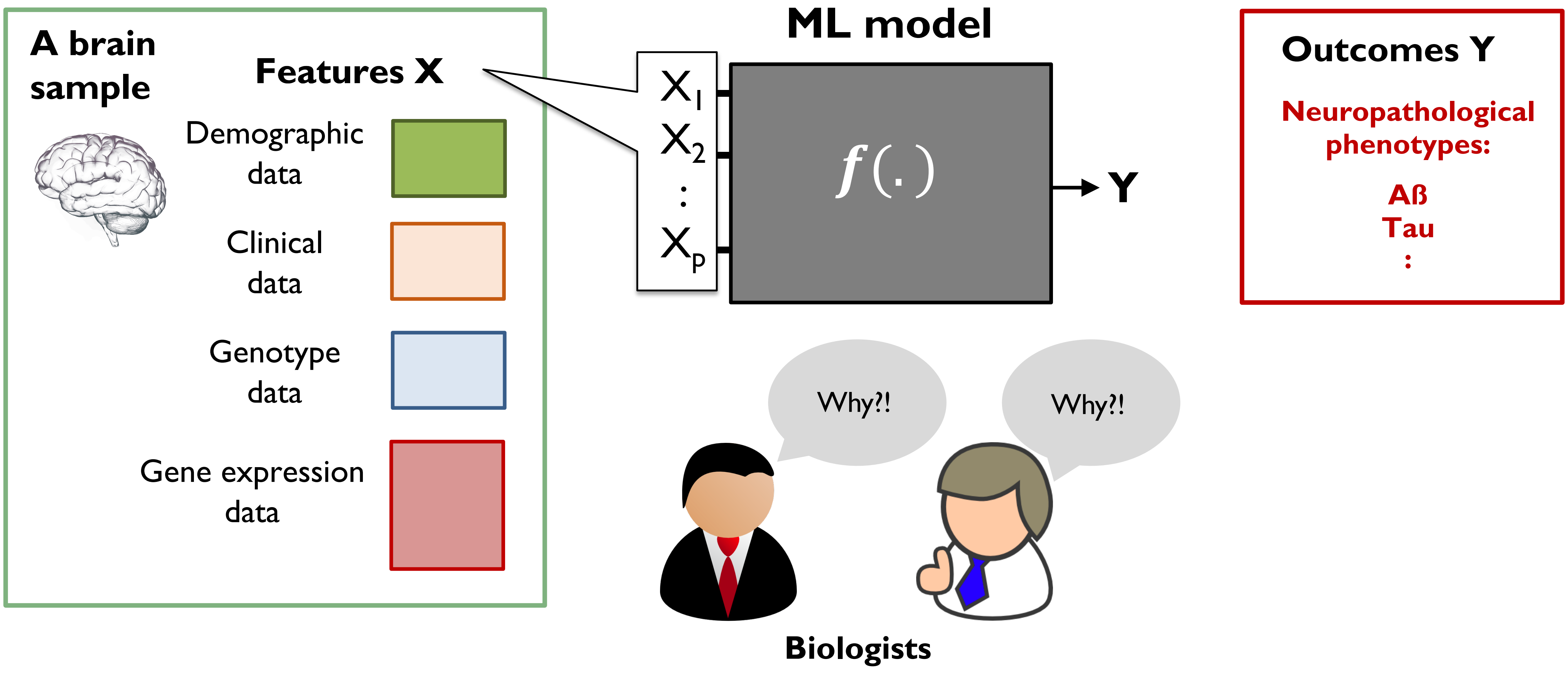
Paul G. Allen School of Computer Science & Engineering

University of Washington, Seattle

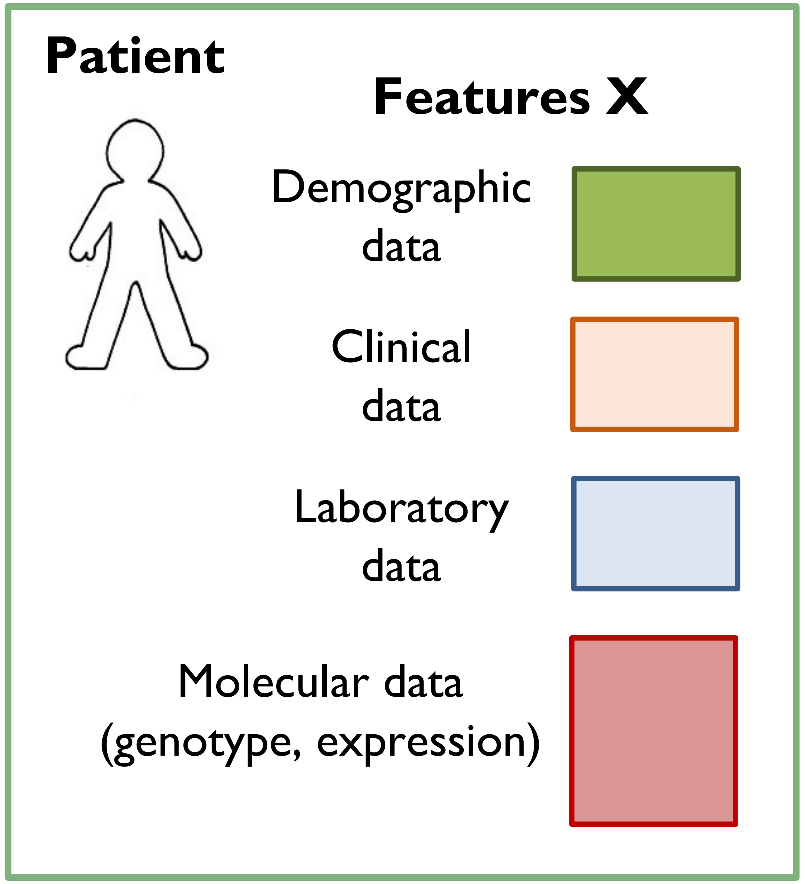
# Accurately predicting a clinical outcome is important but the key question is *why*

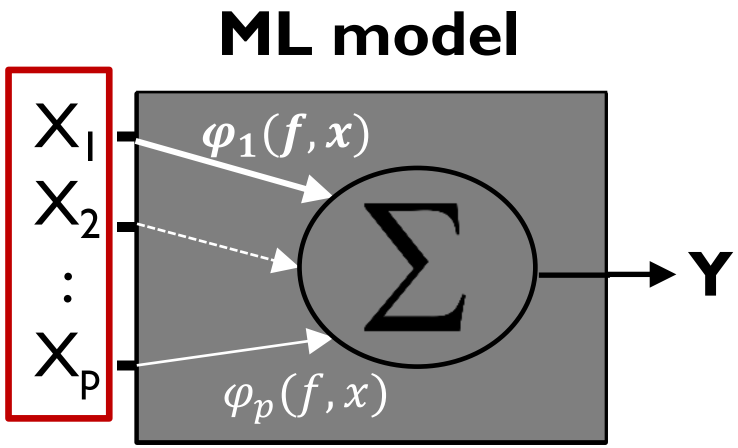


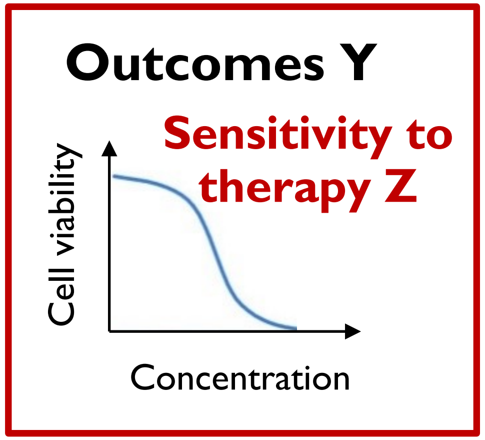
# Identifying predictive markers is important but the key question is *why*



# Explainable AI for clinical genomics

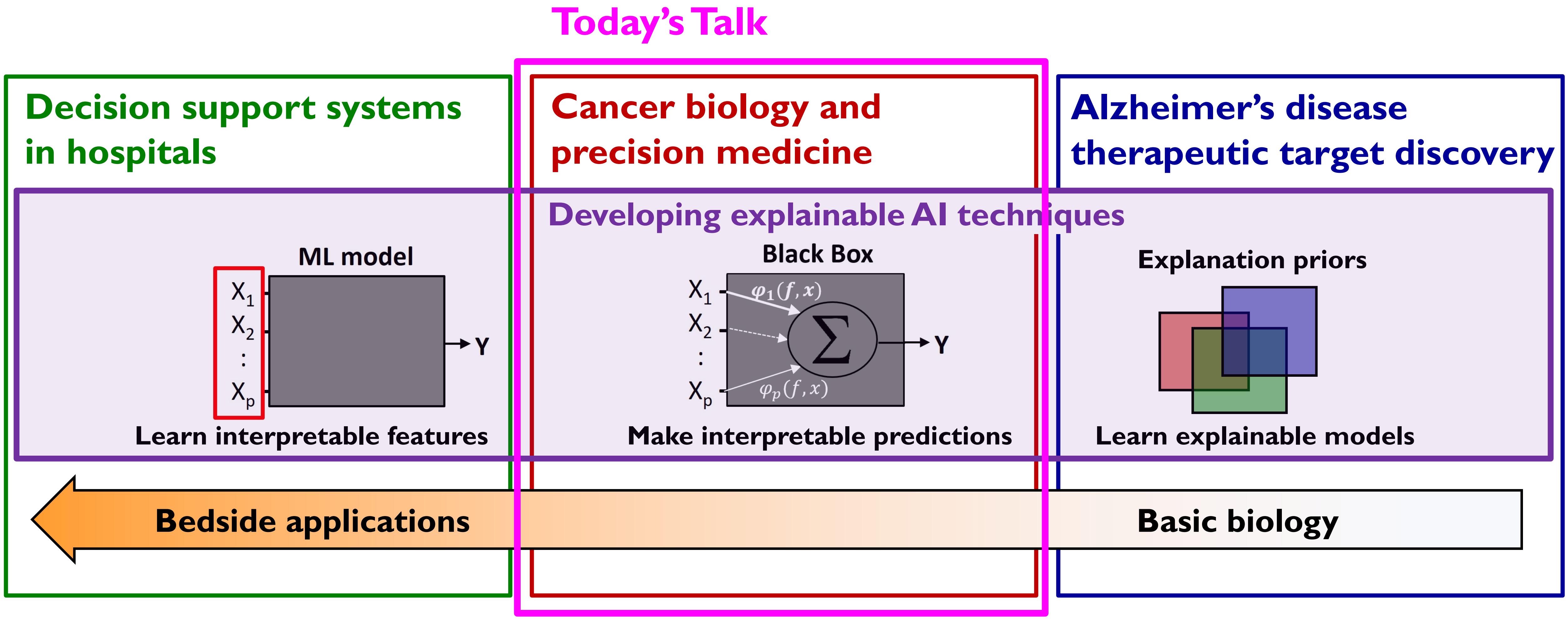


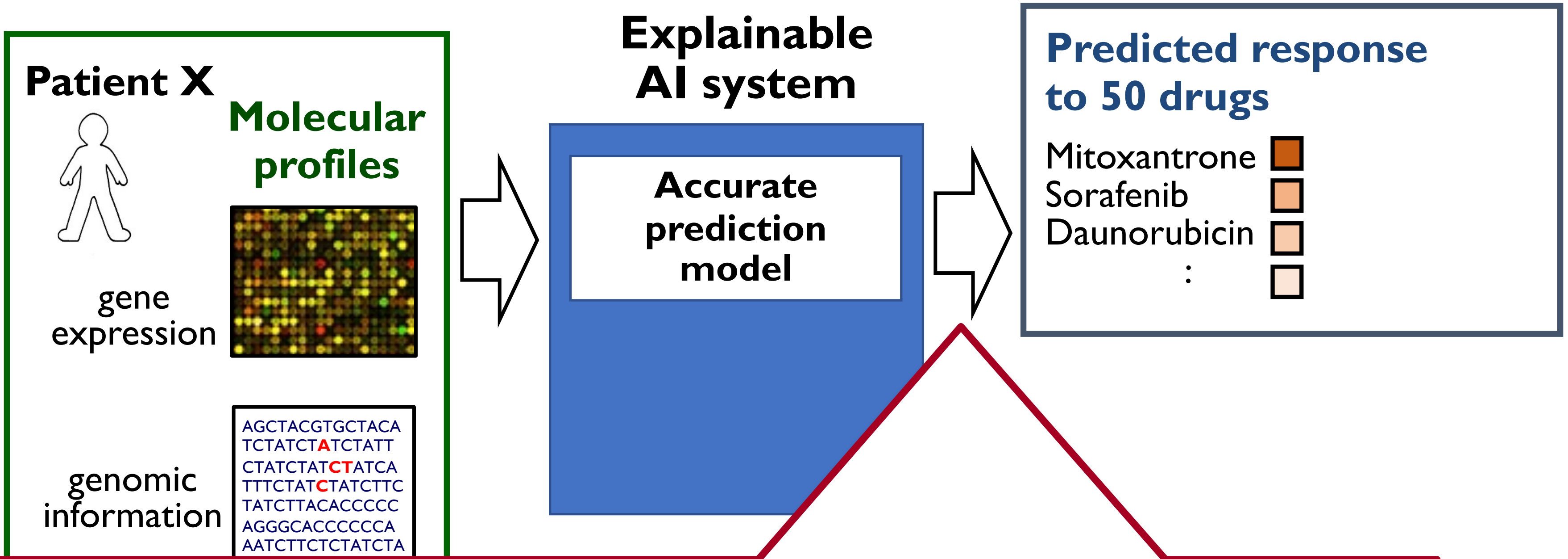

- **Explainability** is more important than accuracy.
  - How to learn or select features that are interpretable?
  - Which features contributed to a certain prediction and how?
  - How to make biological or clinical sense of a black-box model?

# Explainable AI for biology and health



# Explainable AI for cancer precision medicine

- Acute myeloid leukemia (AML)
  - Cancer of the blood and bone marrow cells
  - 5 year survival rate: 26%
- Chemotherapy
  - >100 anti-cancer (62 FDA approve
  - Standard therapy personalized.




- Identify explainable gene expression markers by jointly learning the model with prior knowledge on genes' driver potential (based on genomic, transcriptomic, and functional data).

Lee*, Celik*, et al. (2018) *Nature Communications*

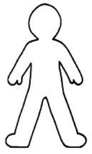
Weinberger et al. (2020) *NeurIPS*

Erion*, Janizek*, Sturmfels* et al. (2021) In Press *Nature Machine Intelligence*

# Explainable AI for cancer precision medicine

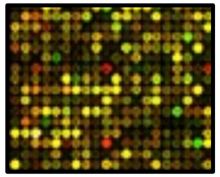
- Acute myeloid leukemia (AML)
  - Cancer of the blood and bone marrow cells
  - 5 year survival rate: 26%
- Chemotherapy
  - >100 anti-cance
    (62 FDA approv
  - Standard therap
    personalized.

**Patient X**



**Molecular profiles**

gene expression



DNA sequence

AGCTACGTGCTACA
TCTATCTATCTATT
CTATCTATCTATCA
TTTCTATCTATCTTC
TATCTTACACCCCC

**Interpretable ML system**

**Accurate prediction model**

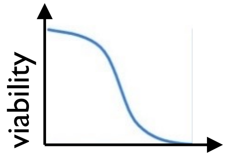
Making **explainable prediction**

**Predicted response to 50 drugs**

Mitoxantrone
Sorafenib
Daunorubicin
:

**Best drug combinations for patient X & Why?**

Mitoxantrone
+ Etoposide
+ Ara-C
:



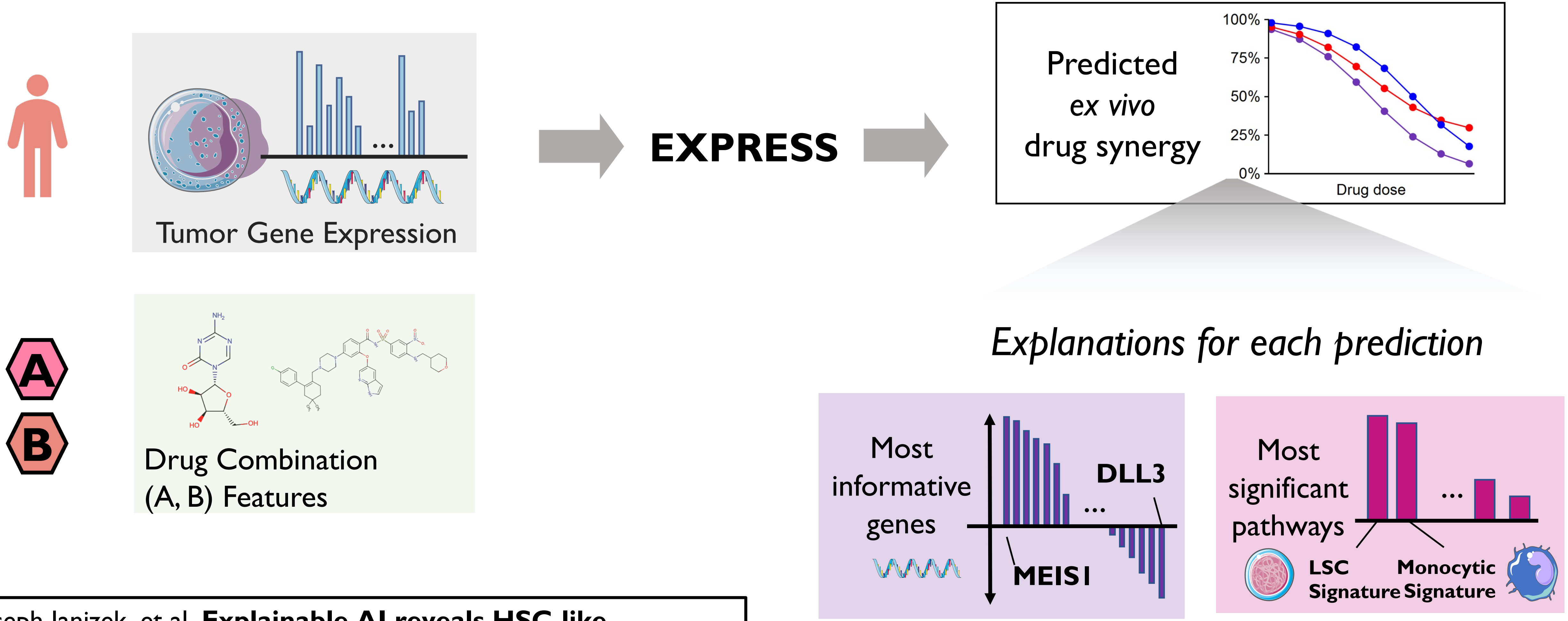

- Cancers can develop drug resistance when being treated with one drug at a time
  - 185 FDA approved drugs → 17,020 pairs

Lee*, Celik*, et al. (2018) *Nature Communications*

Weinberger, et al. (2020) *NeurIPS*

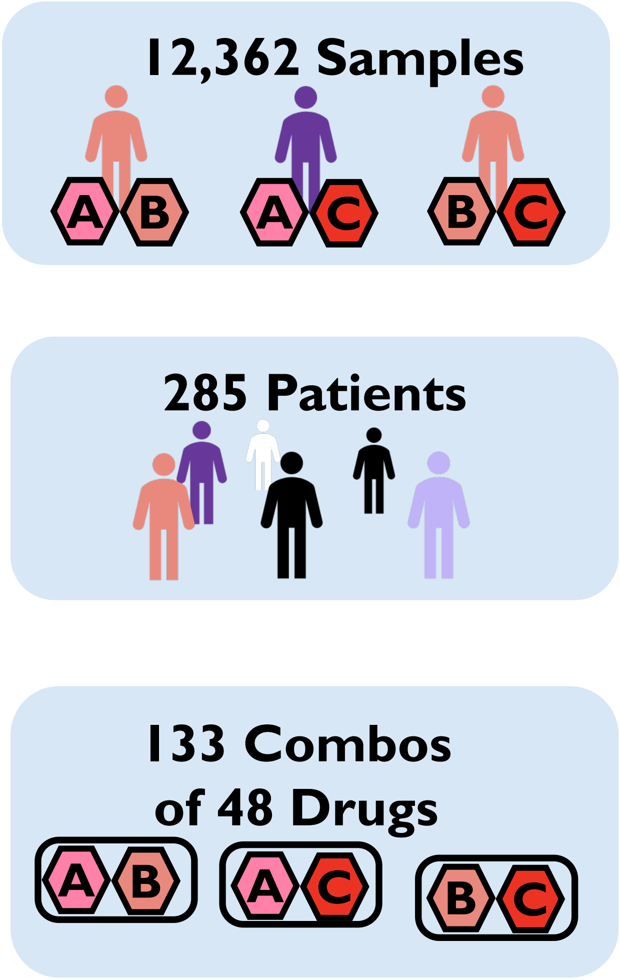
# EXPRESS: Explainable prediction of drug synergy in AML

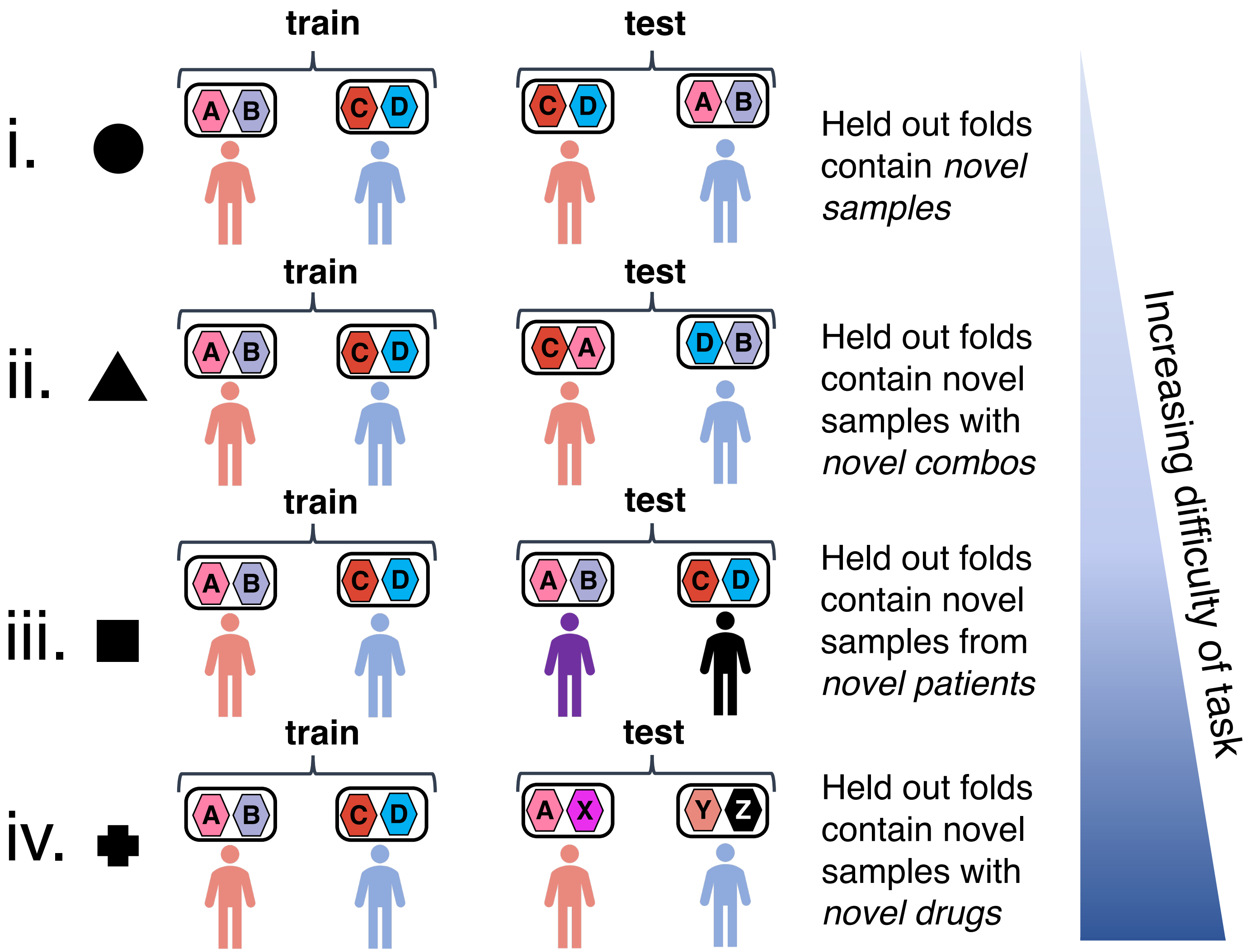
**MGH/Harvard**
Prof. Kamila Naxerova



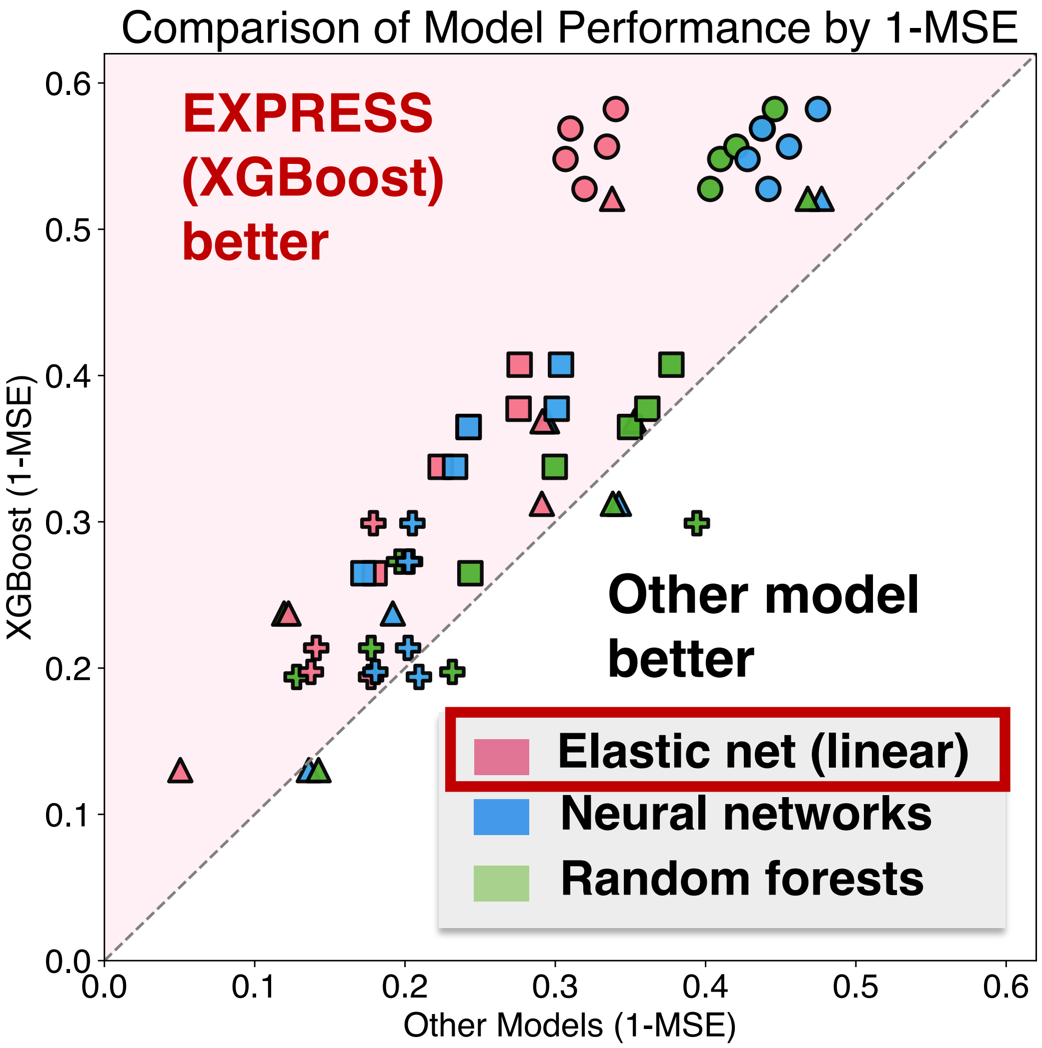

Joseph Janizek, et al. **Explainable AI reveals HSC-like expression signature as relevant to drug synergy in AML.**

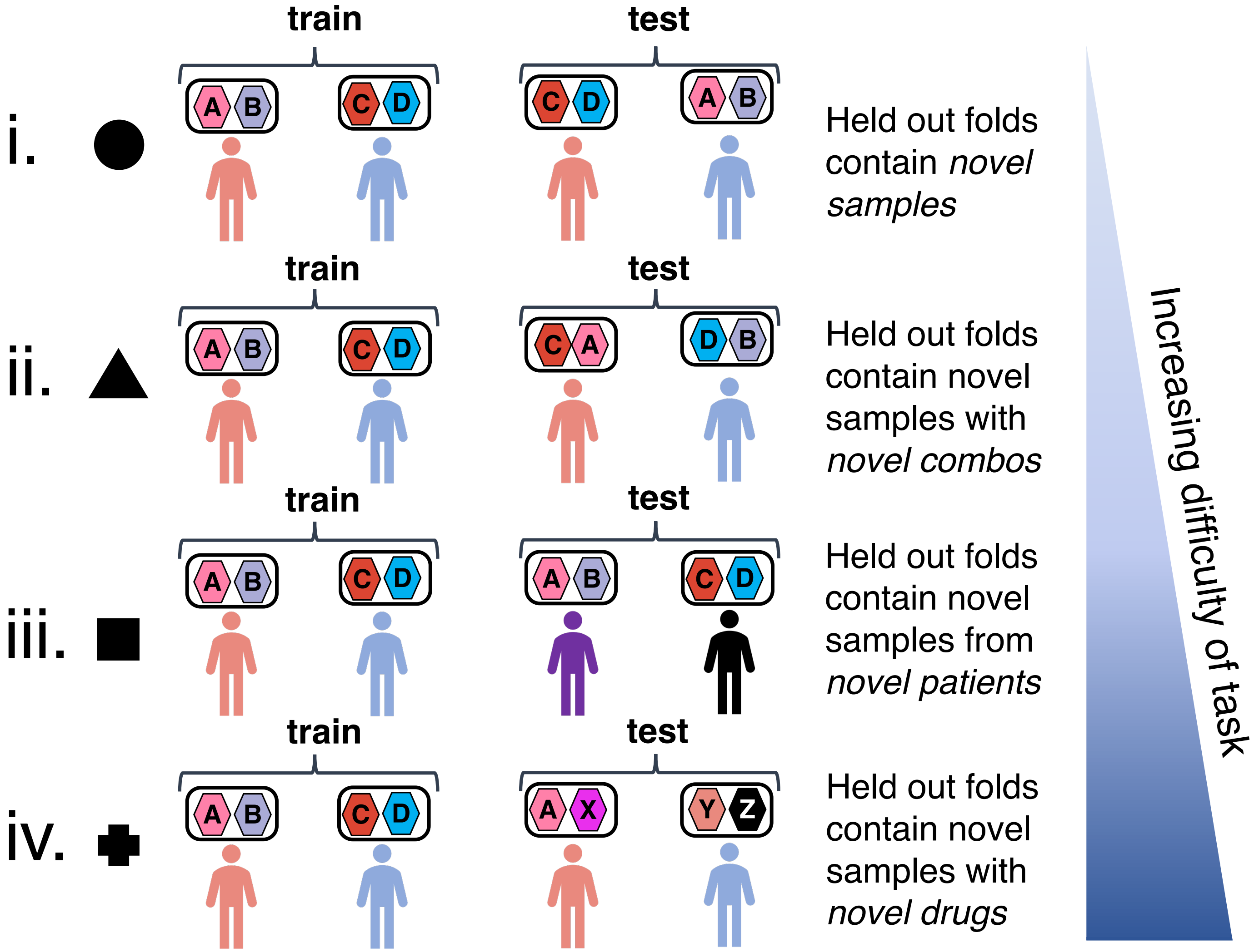
# Training data for EXPRESS: Beat AML data (Tyner et al. Nature 2018)



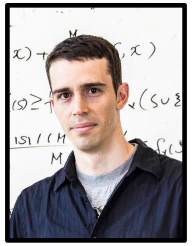



# Complex non-linear models more accurately predicts drug synergy in held-out data



# Our solution is to make a prediction with explanations

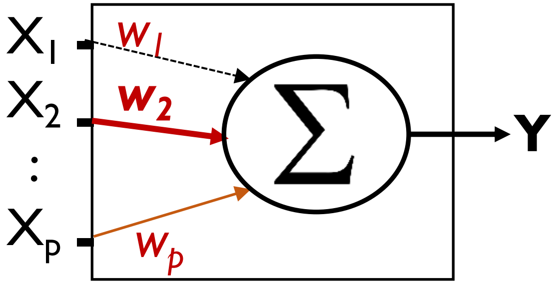

Scott

Eliminating the accuracy vs. interpretability tradeoff
⇨ Broader applicability of ML to biomedicine

- Accuracy vs. interpretability
  - Simple models often lead to lower performance.
  - Complex models are often considered to be a black box.

*Linear model*

**X:** Features **Y:** Outcome

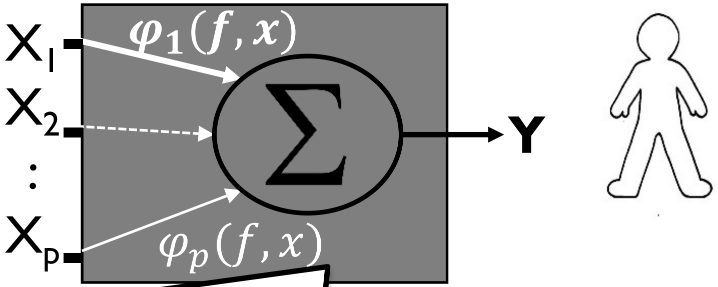



*Complex model f (.)*

**Black Box**


*Our approach, SHAP*

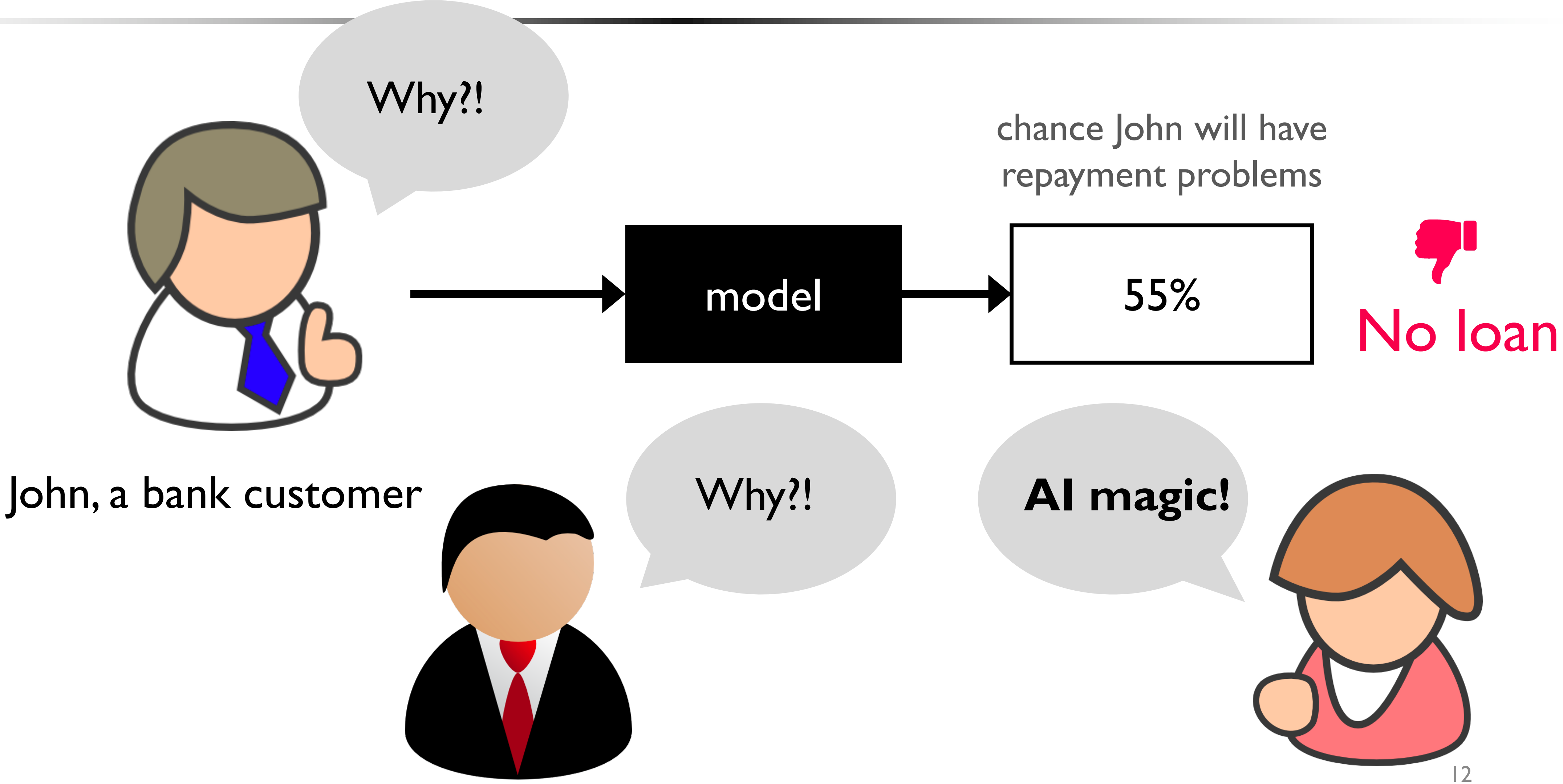
For a particular prediction



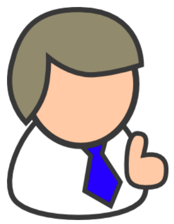

- SHAP can estimate feature importance for a particular prediction for any model.

Scott Lundberg and Su-In Lee. **A Unified Approach to Interpreting Model Predictions.** *NeurIPS (2018)* **Oral presentation (top 1%),** *NeurIPS workshop on Interpretable ML (2016)* – **Best paper award**

# How exactly can we estimate feature importance?
## – SHapley Additive exPlanation (SHAP) values



# How exactly can we estimate feature importance?
## – SHapley Additive exPlanation (SHAP) values



Base rate

20%

$E[f(x)]$

Prediction for John

55%

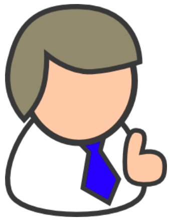
$f(x)$

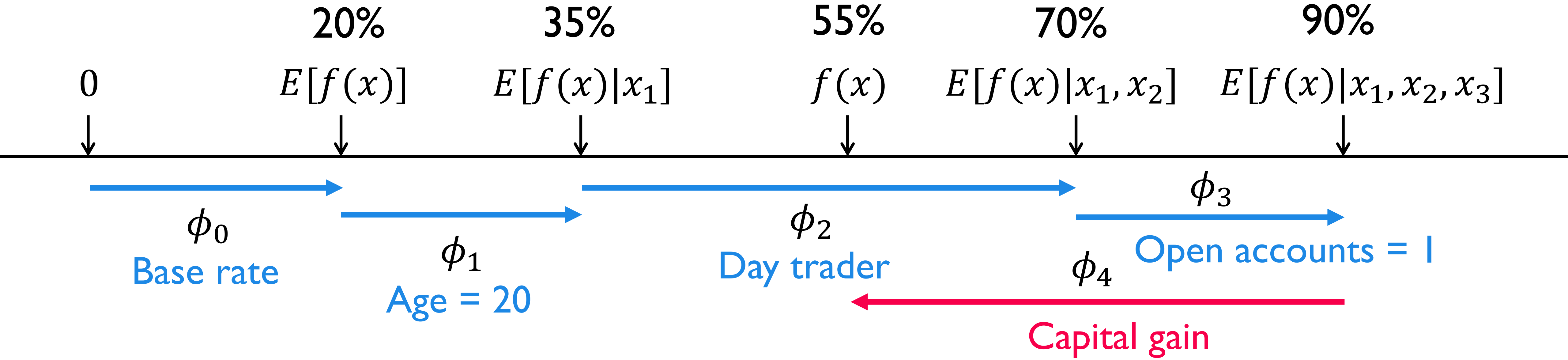
0

How did we get here?

Scott Lundberg and Su-In Lee. **A Unified Approach to Interpreting Model Predictions.** ***NeurIPS (2018)*** **Oral presentation (top 1%),** *NeurIPS workshop on Interpretable ML (2016)* – **Best paper award**

# How exactly can we estimate feature importance? – SHapley Additive exPlanation (SHAP) values






Scott Lundberg and Su-In Lee. **A Unified Approach to Interpreting Model Predictions.** *NeurIPS (2018)* **Oral presentation (top 1%),** *NeurIPS workshop on Interpretable ML (2016)* – **Best paper award**

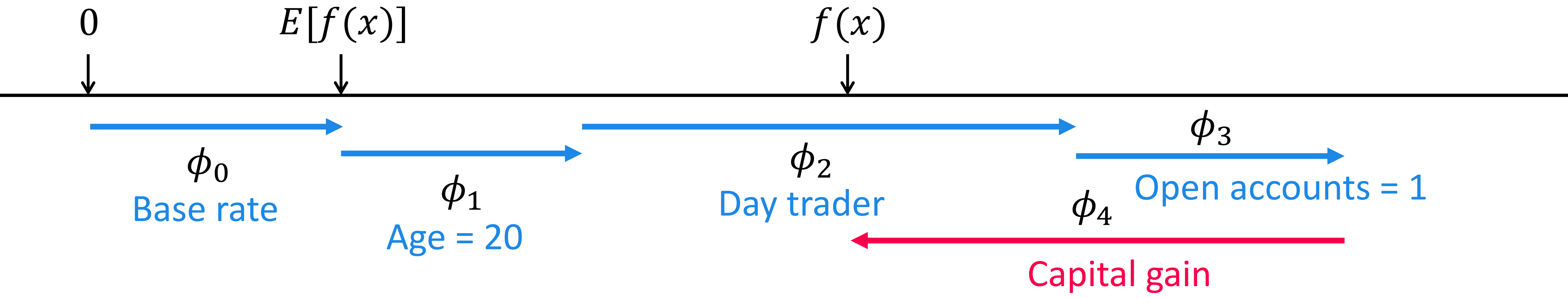
# How exactly can we estimate feature importance?
## – SHapley Additive exPlanation (SHAP) values

**The order matters!**

**SHAP values result from averaging over all N! possible orderings**

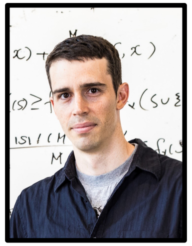
**They are the only solution that satisfies three important properties**

**We need to develop efficient methods to estimate or compute exact SHAP values.**

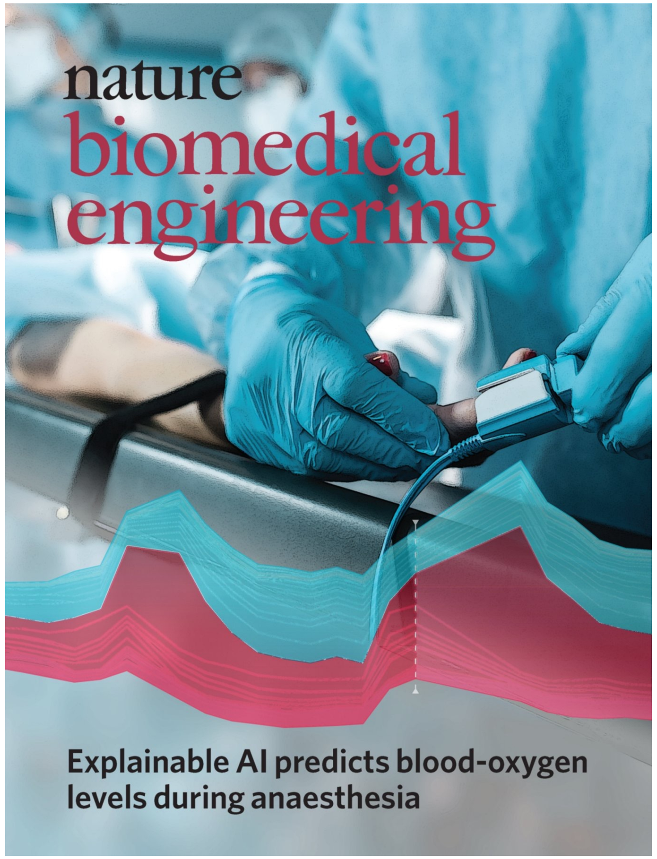



Scott Lundberg and Su-In Lee. **A Unified Approach to Interpreting Model Predictions.** ***NeurIPS (2018)*** **Oral presentation (top 1%),** *NeurIPS workshop on Interpretable ML (2016)* – **Best paper award**

# Providing explainable prediction improves anesthesiologist's ability to predict hypoxemia



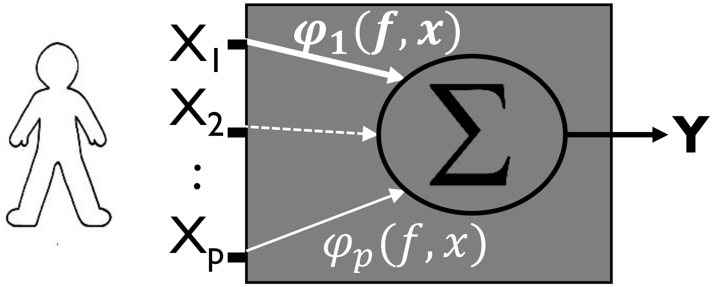
Scott



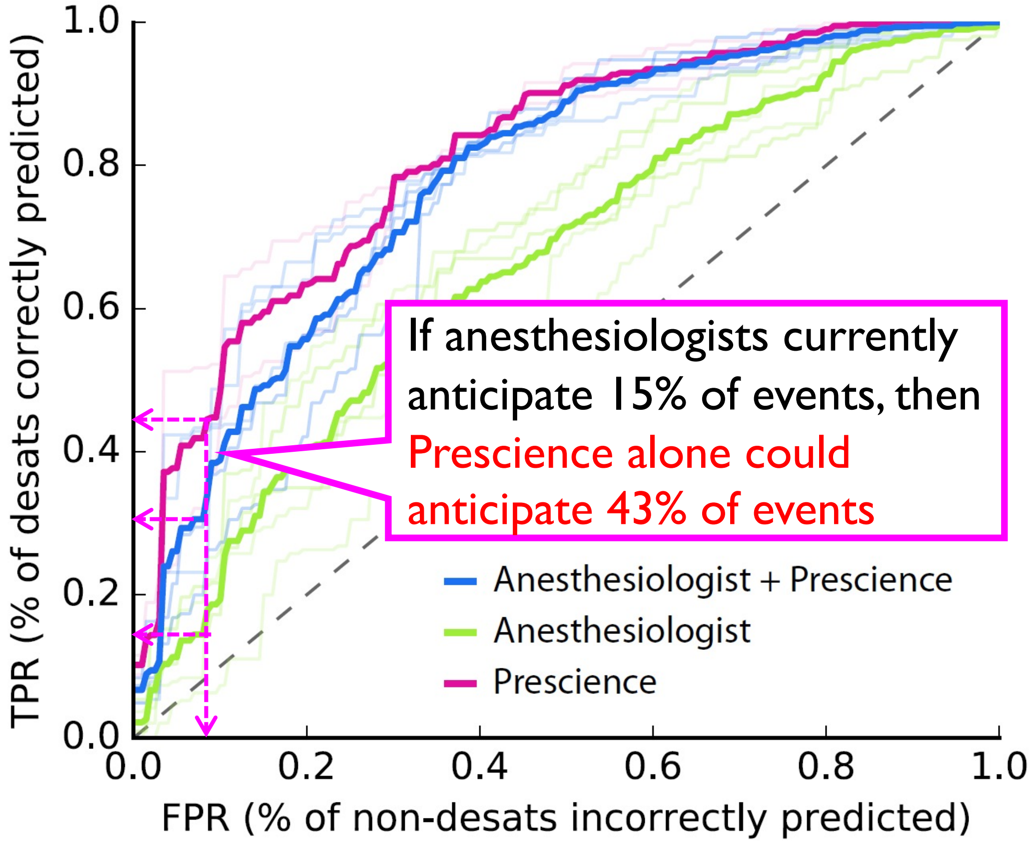

- Our *Prescience* method predicts hypoxemia in the next 5 minutes and provides explanations in real time.

***Our approach, SHAP***

For a particular prediction




Real-time hypoxemia prediction

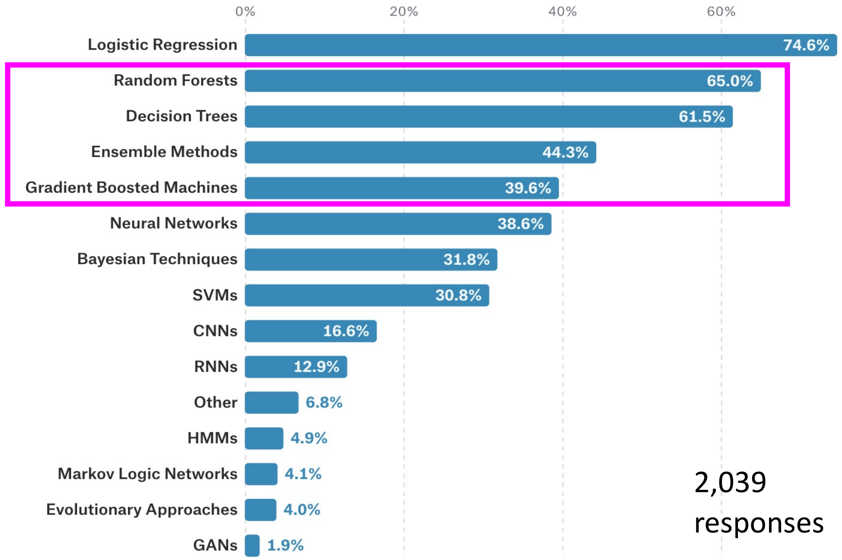



Scott M. Lundberg, Bala Nair, Monica S. Vavilala, Mayumi Horibe, Michael J. Eisses, Trevor Adams, David E. Liston, Daniel King-Wai Low, Shu-Fang Newman, Jerry Kim, and Su-In Lee. **Explainable machine-learning predictions for the prevention of hypoxaemia during surgery.** *Nature BME* 2, 749–760 (Oct 2018) - **Featured on the Cover; cited >150 times over 2 years**

# Making tree ensembles interpretable

- Why tree ensembles? 
  - Gradient Boosted Trees and Random Forests are widely used state-of-the-art models.
  - Over half (17/29) of all Kaggle competition winners in 2015 used XGBoost (Chen and Guestrin).

Tree models




- Tree SHAP reduces the exact computation of SHAP values from exponential to polynomial time.

Direct Solution $O(TL2^M)$ **Exponential**

to

Tree SHAP $O(TLD^2)$ **Polynomial**

Scott Lundberg, Gabe Erion, Hugh Chen, Alex DeGrave, […], and Su-In Lee. **Explainable AI for Trees: From Local Explanations to Global Understanding.** ***Nature Machine Intelligence (2020) as a cover article of the January issue***

# Explainable AI for trees

(https://github.com/slundberg/shap)


Google Cloud

PORTLAND TRAIL BLAZERS

NBA

Microsoft

ROLLS ROYCE

Cleveland Clinic

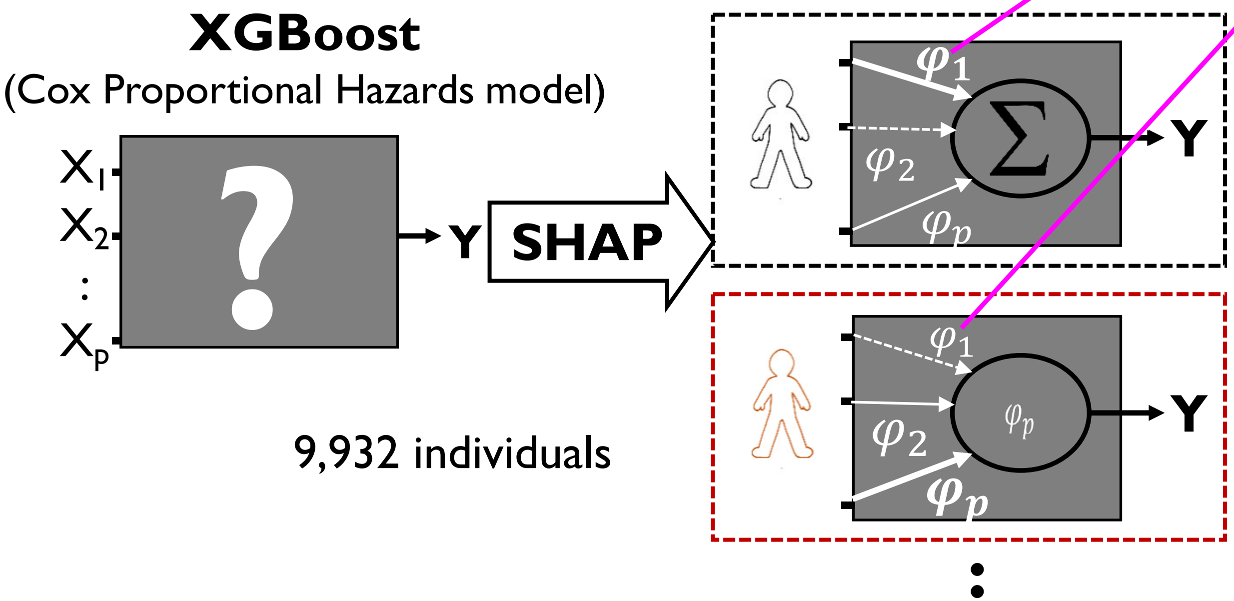
BANK OF ENGLAND

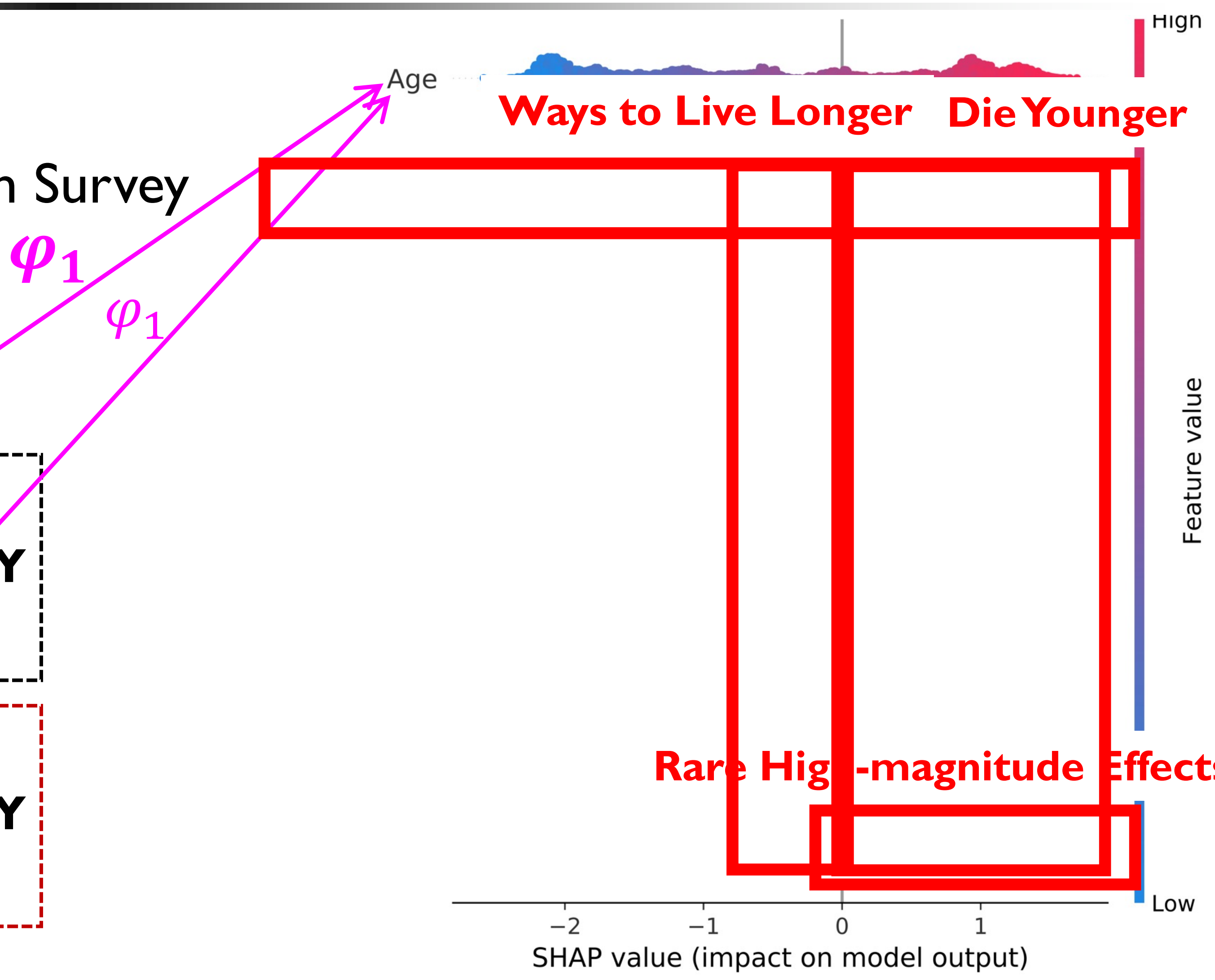
Scott Lundberg and Su-In Lee. **A Unified Approach to Interpreting Model Predictions.** ***NeurIPS (2018)*** **Oral presentation (top 1%),** *NeurIPS workshop on Interpretable ML (2016)* – **Best paper award**

Scott Lundberg, Gabe Erion, Hugh Chen, Alex DeGrave, […], and Su-In Lee. **Explainable AI for Trees: From Local Explanations to Global Understanding.** ***Nature Machine Intelligence (2020) as a cover article of the January issue***

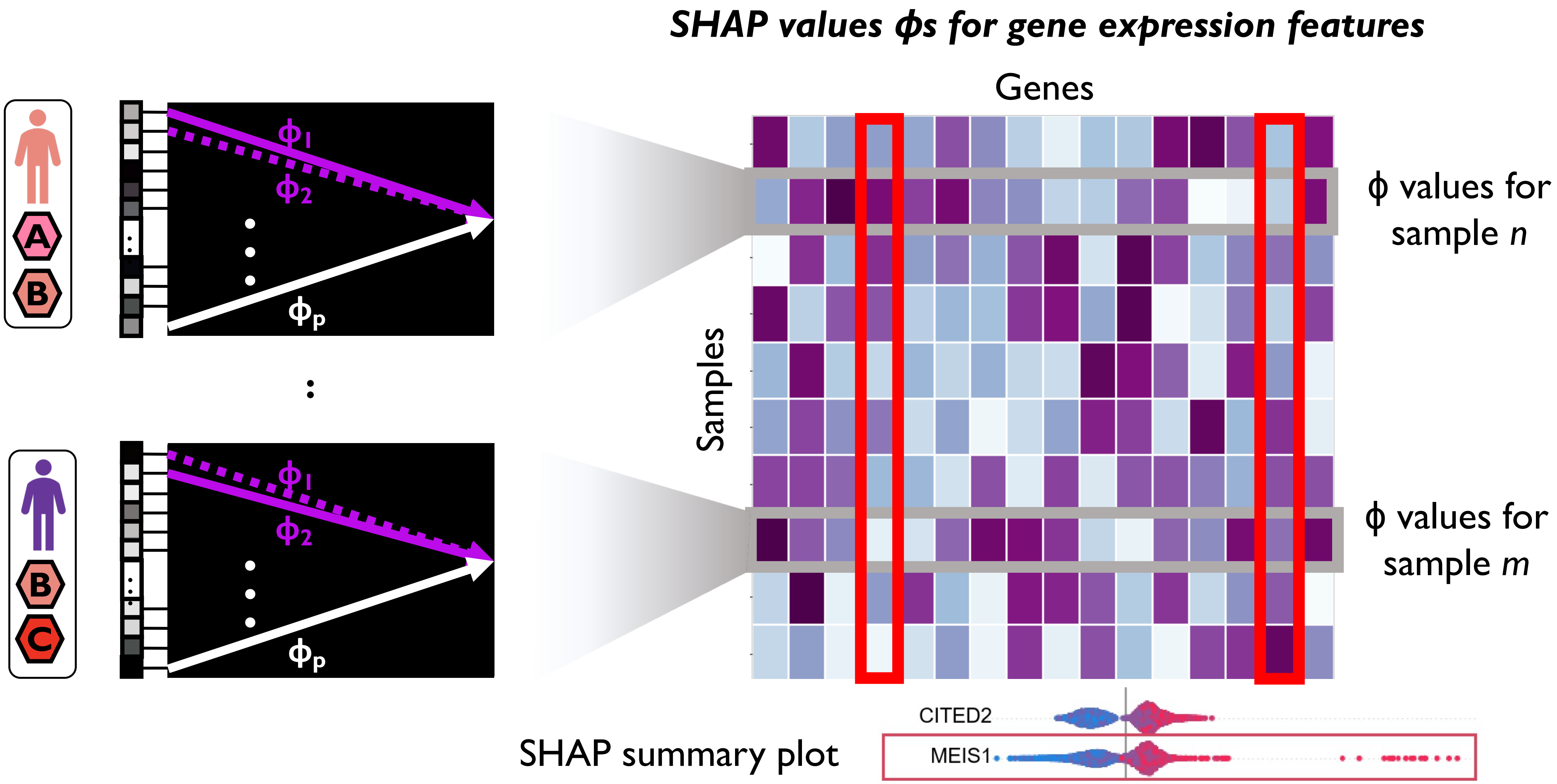
# Using SHAP values as building blocks for interpretable ML – SHAP summary plot

- NHANES I
  National Health & Nutrition Examination Survey
  - X: 59 common measurements
  - Y: Mortality

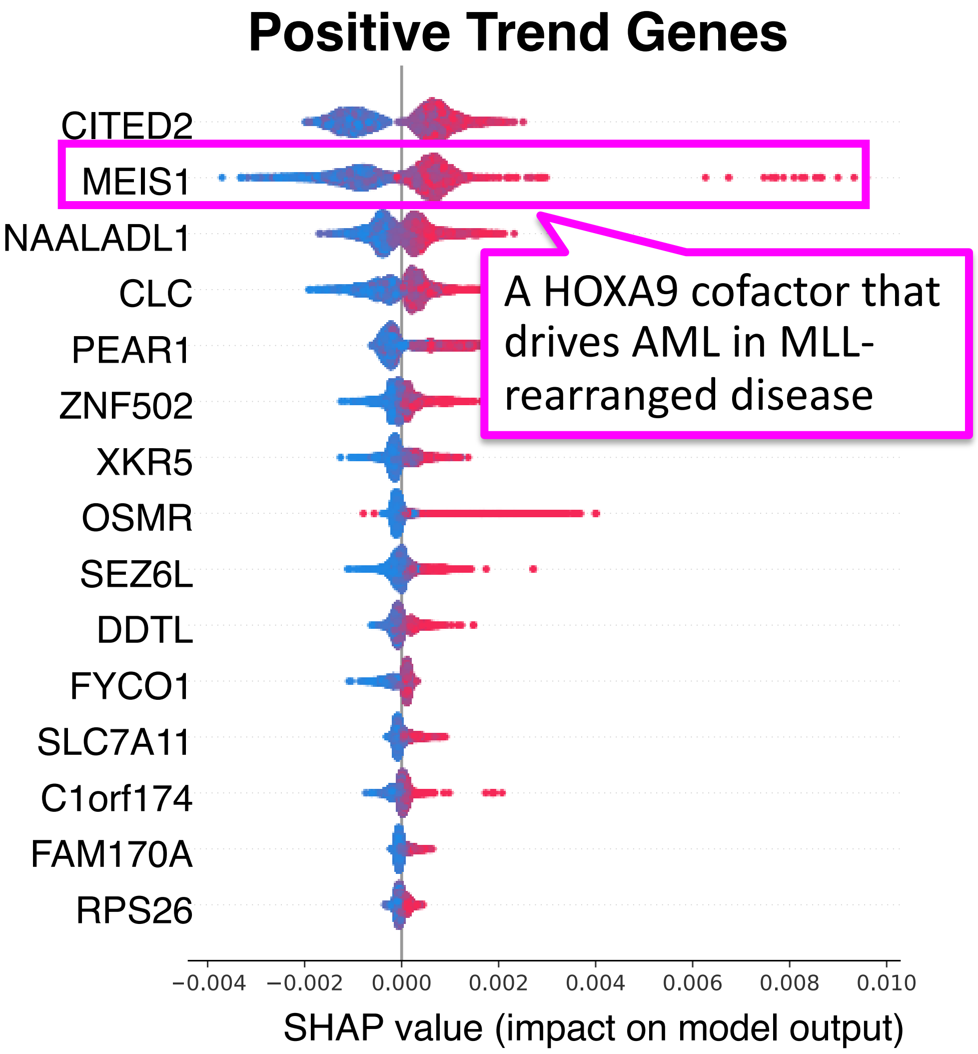


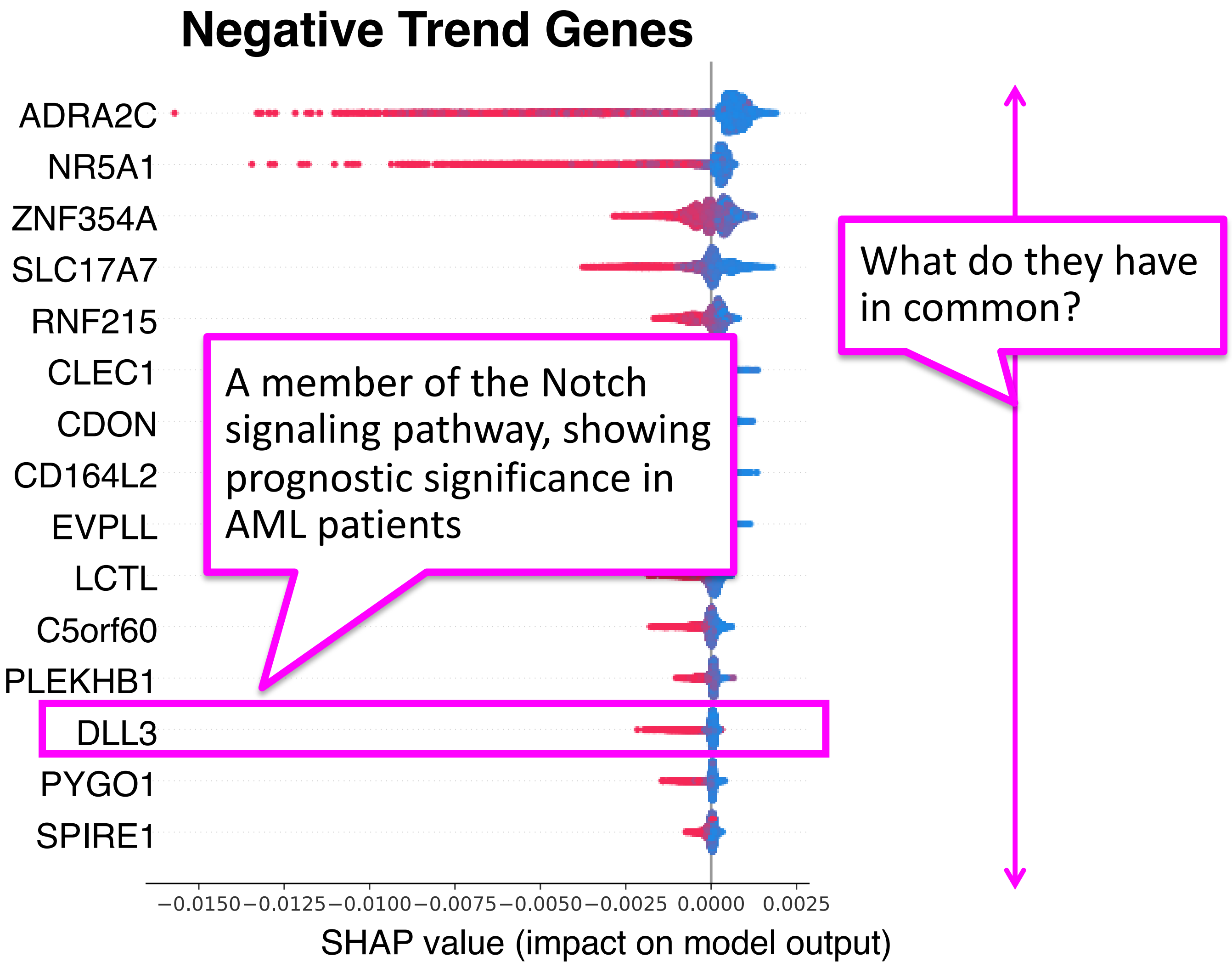



# EXPRESS: Explainable prediction of anti-cancer drug synergy



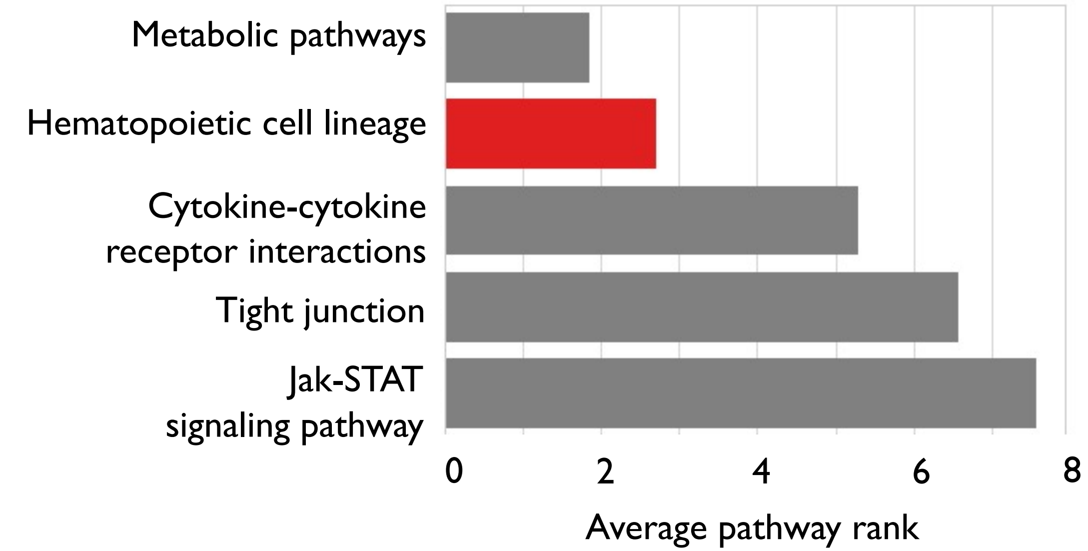
# Genes that are important to drug synergy in AML

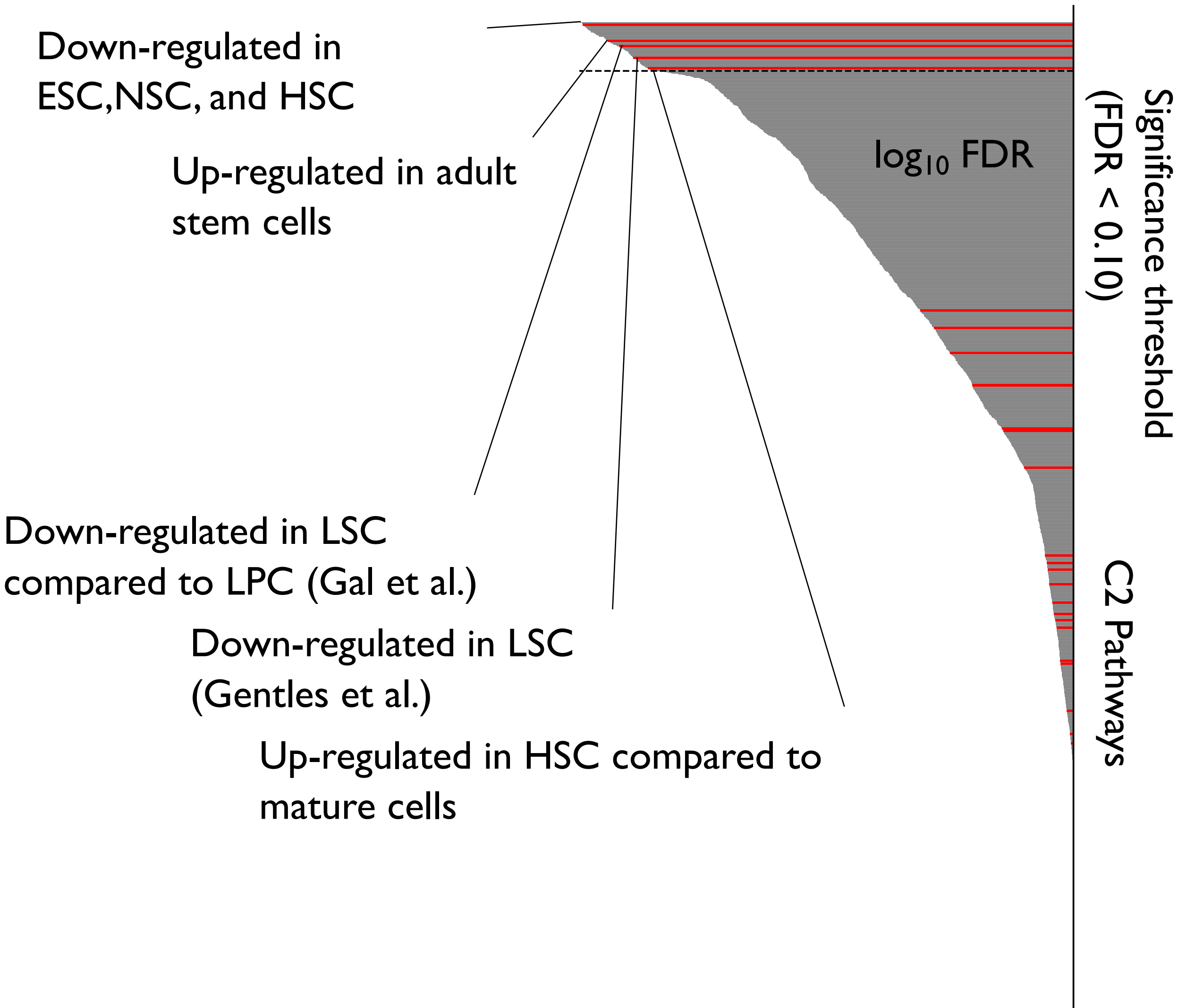


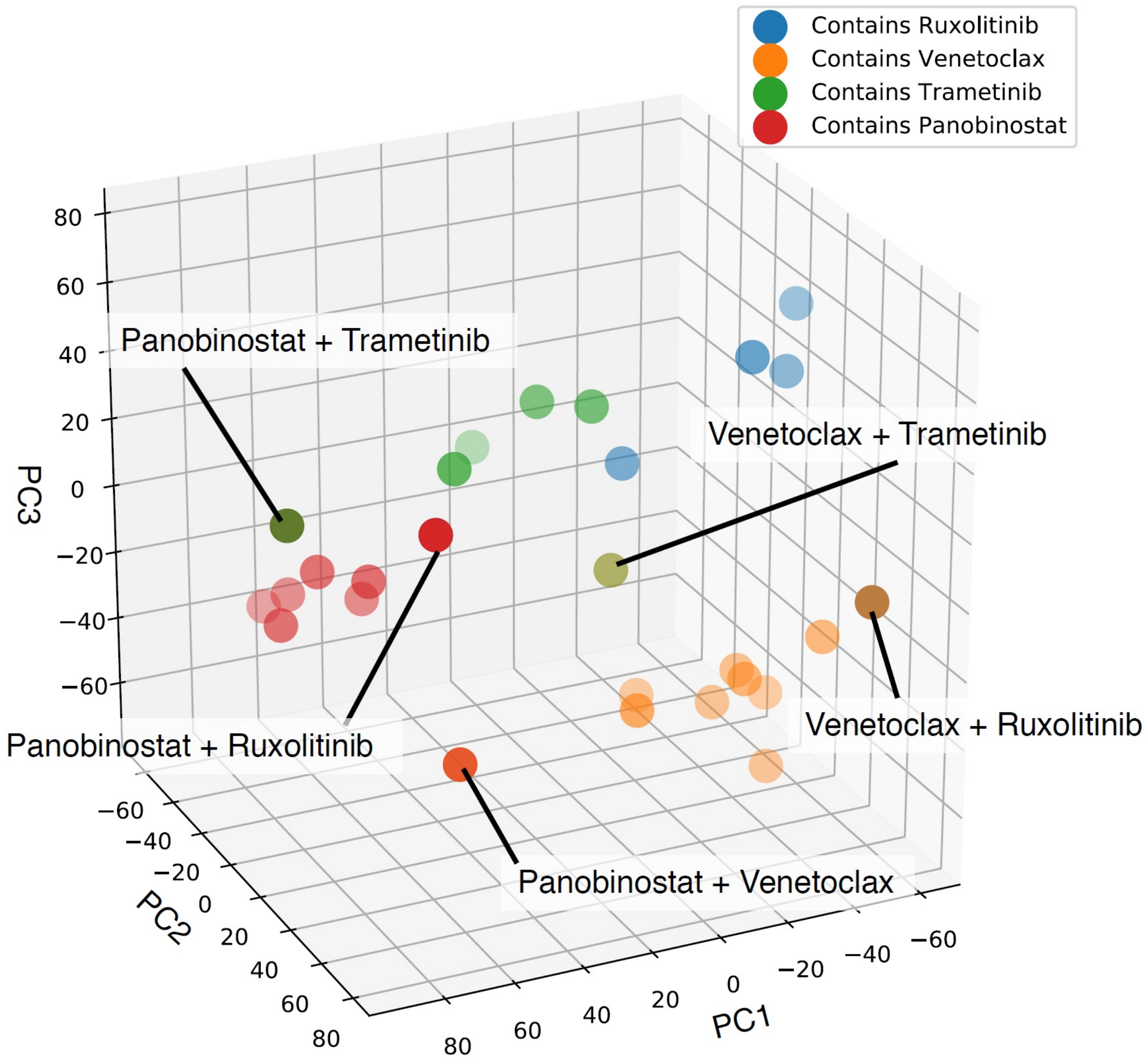
# Hematopoietic stem cell (HSC)-like expression signature is important for drug synergy

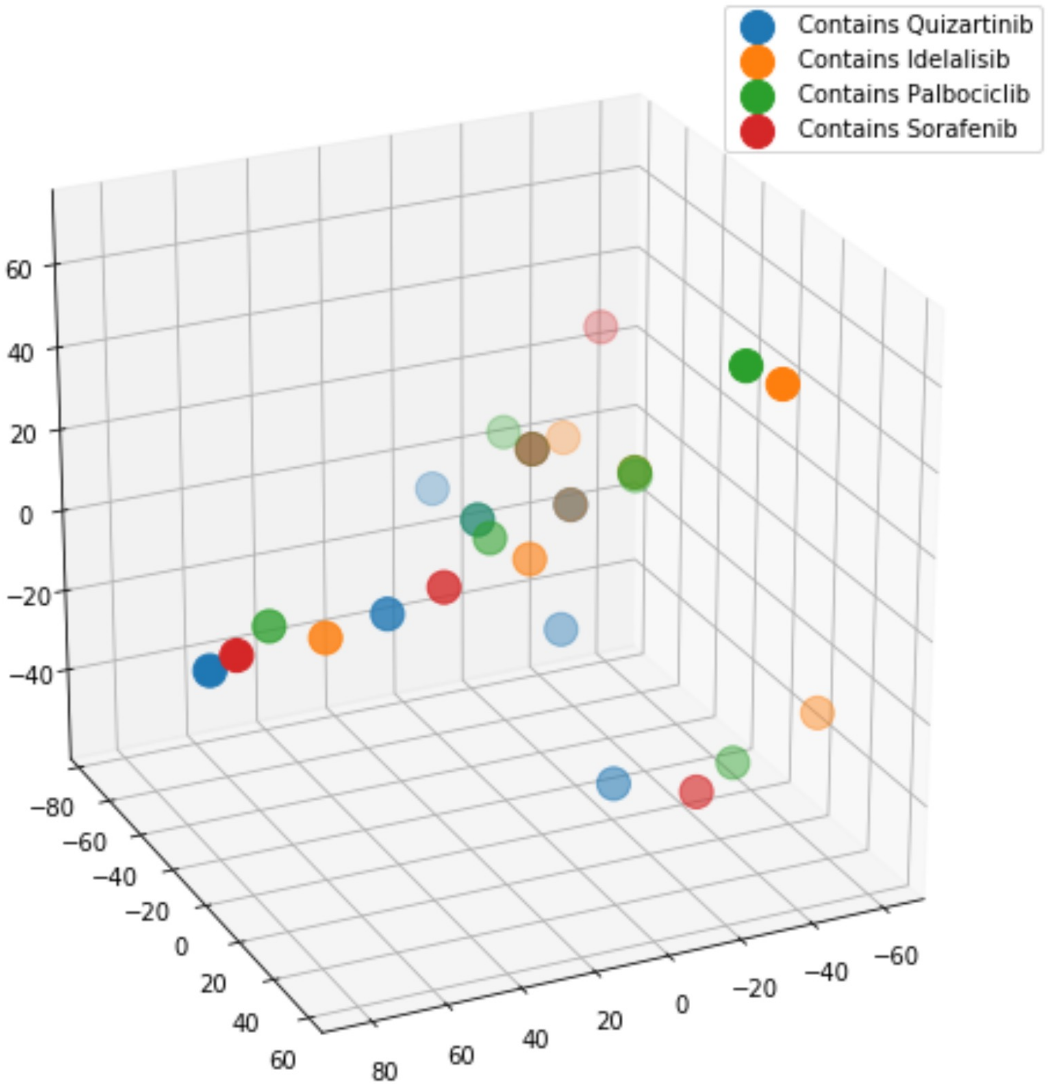
- Pathways enriched in the top-ranked genes
  - Potential mechanistic explanation of anti-cancer drug synergy
- Drug- or combo-specific analysis reveals previously unknown characteristics of drugs.



# Drug-specific analysis



# Explainable AI for biology and health

**Today's talk**

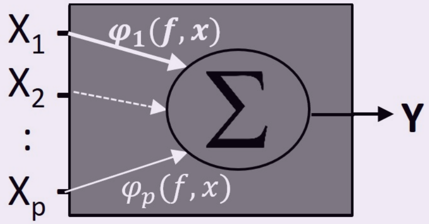
SHAP, Tree SHAP, Deep SHAP, expected gradients, & attribution priors

**Cancer biology and precision medicine**

**Alzheimer's disease therapeutic target discovery**

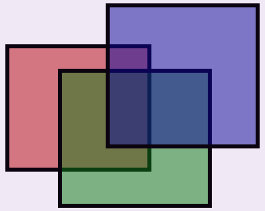
**Developing explainable AI techniques**

**Black Box**




e interpretable predictions

**Explanation priors**

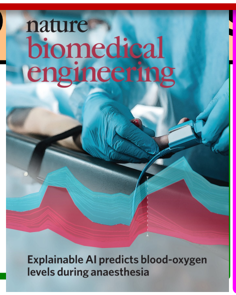


**Learn explainable models**

**Conclusions:**

- Complex models are useful to capture non-linear interactions.
- We need new methods to make biological sense.

Bedside app s




**Basic science**


| | | | |
|---|---|---|---|
| Lee*, Celik*, et al. (2018) *Nature Communications* | Lundberg et al. *Nature Biomedical Engineering* (2018) – **Featured on the cover** | Lundberg and Lee. *NeurIPS (2017)* – **Oral presentation (top 1%)** | Lundberg et al. *Nature Machine Intelligence* – **Featured on the cover** |