



Glasgow Polyomics  
[www.polyomics.gla.ac.uk](http://www.polyomics.gla.ac.uk)

 University  
of Glasgow | School of  
Computing Science

# Omics Design For Basket Trial in Precision Medicine

Joe Wandy, Marina Flores,  
Kevin Bryson, Ronan Daly

# Personalized Medicine: A Paradigm Shift in Healthcare



PERSONALISED  
MEDICINE



GENOMIC REVOLUTION



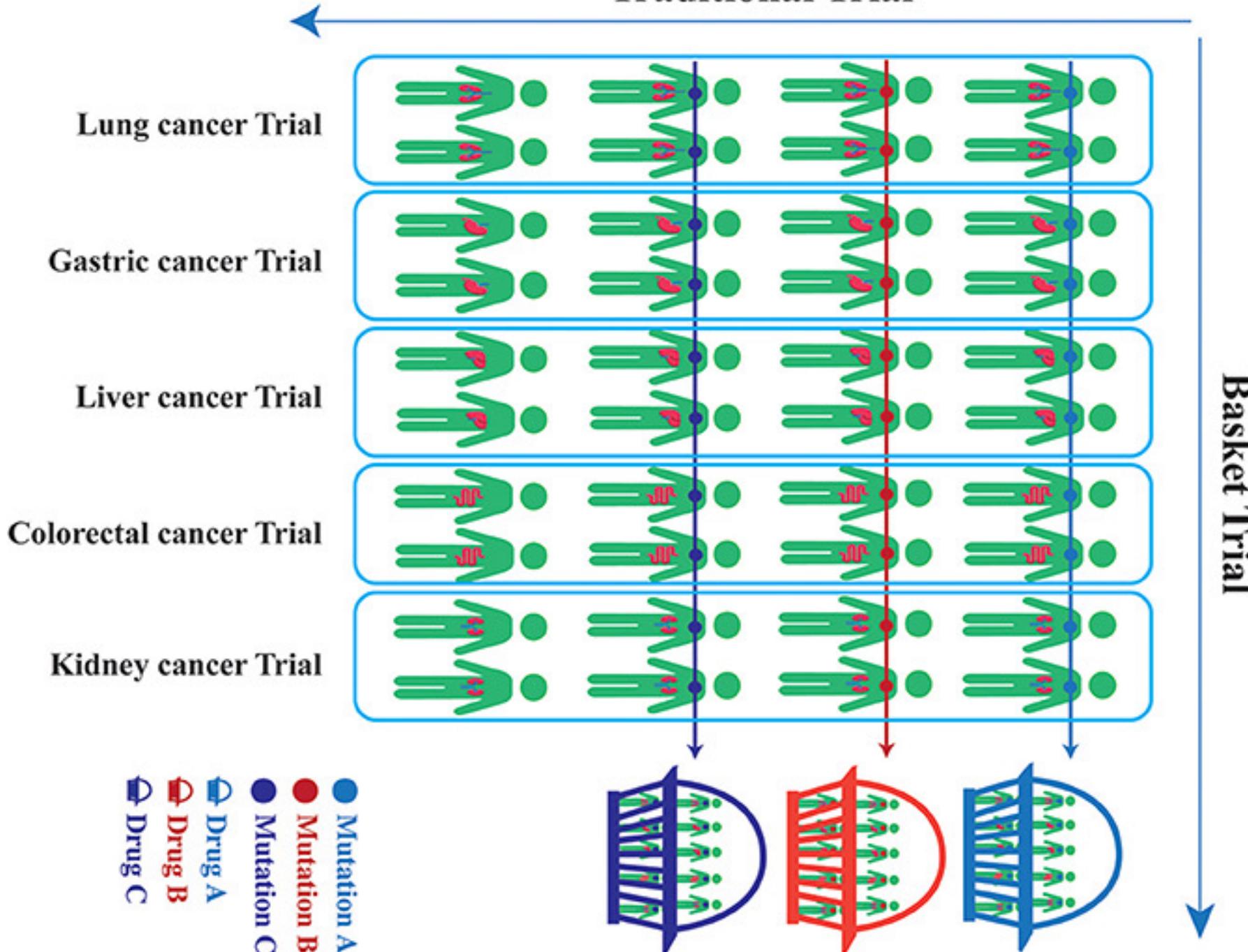
BASKET TRIAL



# Basket trial

- Basket trial enrolls patients with multiple indications (e.g. cancer types) that share the same genomic or molecular biomarkers.
- Assess the efficacy of a treatment (drug) simultaneously on all cancer types.
- Tend to require fewer patients and a shorter trial duration to identify indications that respond to the treatment.
- Allow patients with rare cancer types to participate in clinical trials.

## Traditional Trial



Basket Trial

Qin, Bao-Dong, et al.  
"Basket trials for intractable  
cancer." *Frontiers in  
Oncology* 9 (2019): 229.



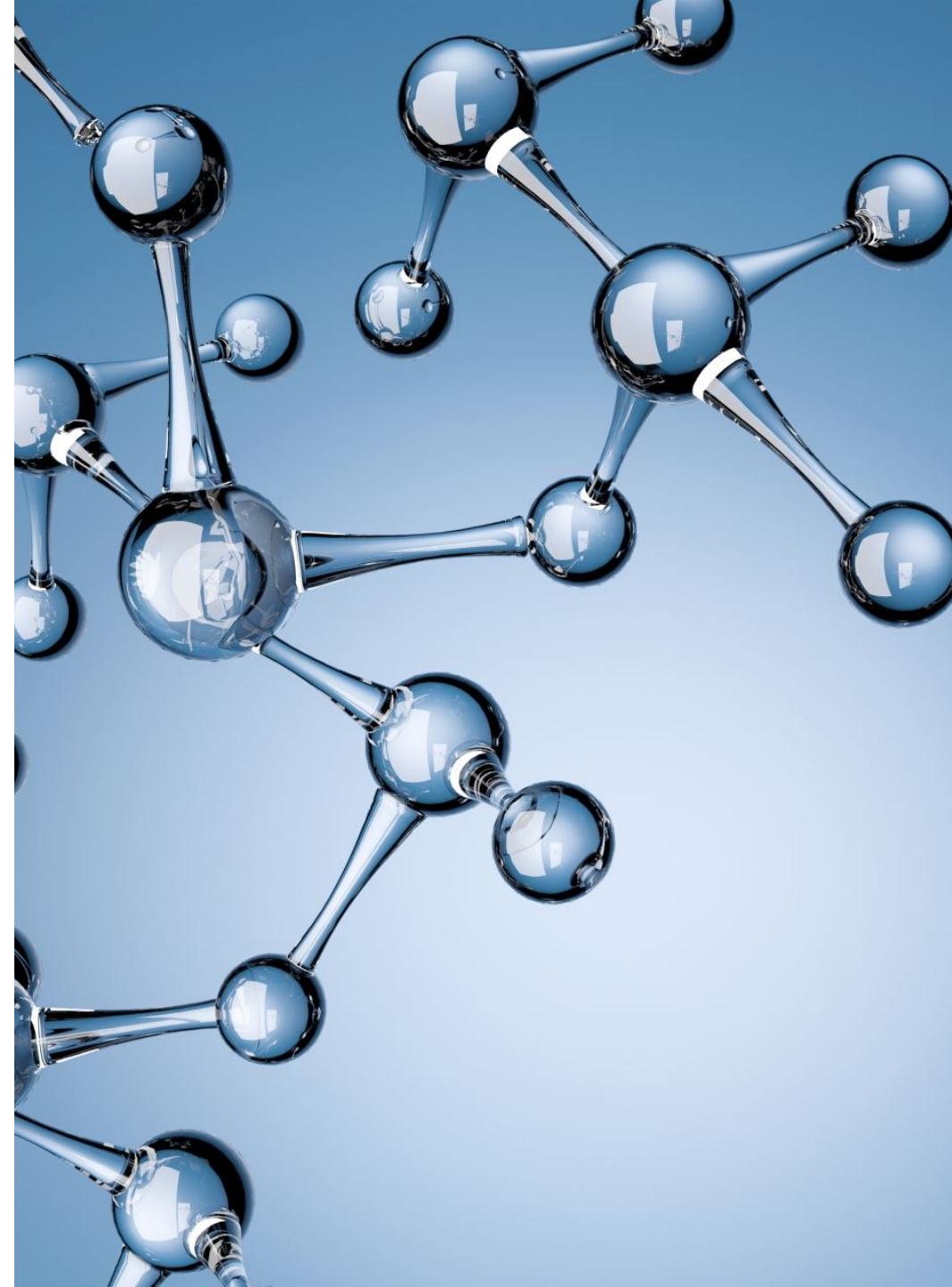
# Basket Trials vs Traditional Trials

---

- Basket trials focus on shared biomarkers, not just the disease type.
- They increase efficiency by evaluating multiple diseases simultaneously.
- Basket trials better align with personalized medicine principles.

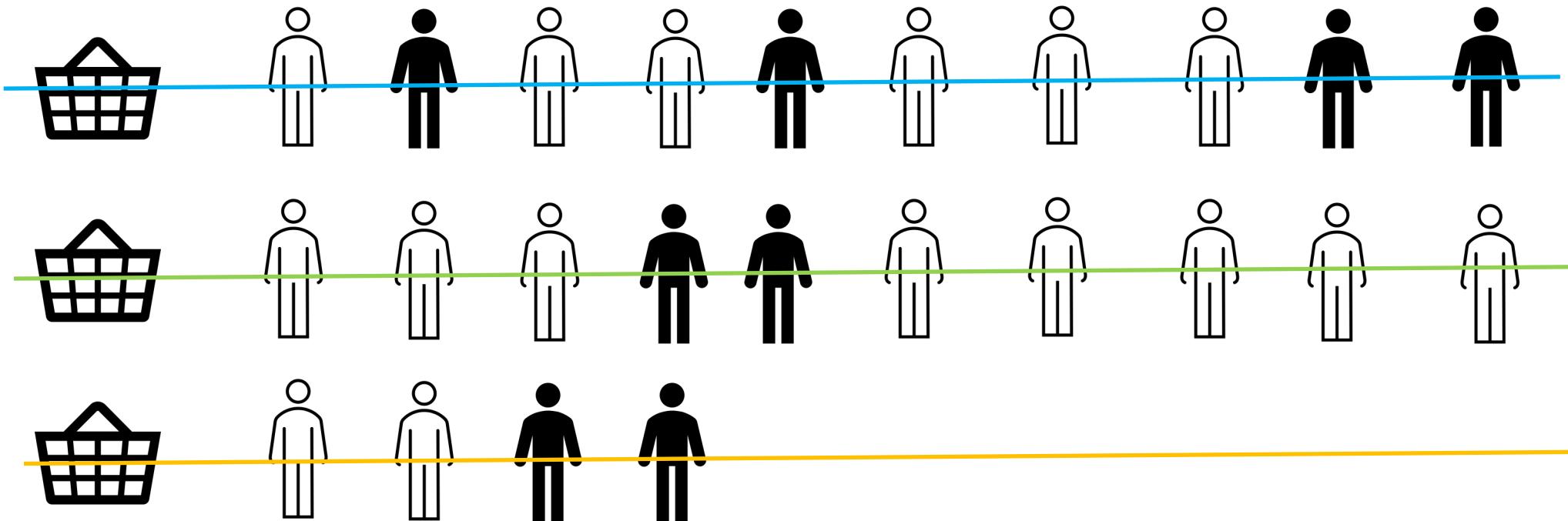
# Omics Data

- Provides extensive molecular information, enhancing patient stratification.
- Can augment basket trials, making them more precise and effective.
- However, the use of omics data in basket trials remains limited (only for screening).



# Basket response inference

- Enrollment in basket trial typically happen in stages.
- Within each stage, for each basket we observe the number of responsive patients.



# Estimating basket response (1)

- Each cancer type (basket  $k$ ) has a true response rate  $p_k$ .
- The standard care response rate,  $p_0$ , is the typical response to currently accepted treatments.
- The target response rate,  $p_1$ , is the desired response rate for the new treatment being tested.
- Hypothesis testing is used to compare the true response rate to the standard and target response rates.

$$H_0: p_k \leq p_0 \text{ vs } H_a: p_k \geq p_1, \text{ for } k = 1, 2, \dots, K$$

# Estimating basket response (2)

- Empirical approach calculates response rate as a proportion ( $\widehat{p}_k = y_k/n_k$ ), where  $y_k$  is positive responses and  $n_k$  is total patients in basket  $k$ .
- Bayesian approach, e.g.  $y_k \sim Bin(n_k, \theta_k)$ ,  $\theta_k \sim Beta(1, 1)$ .
- With a uniform prior ( $Beta(1, 1)$ ), the mode of the Bayesian posterior distribution corresponds to the empirical estimate.
- Each basket is treated as a separate entity. No information sharing across baskets, even when there could be underlying similarities.

# Bayesian Hierarchical Model

- Observed Responses:  $y_k \sim Bin(n_k, p_k)$  for basket  $k = 1 \dots K$
- Success Probabilities:  $p_k = logistic(\theta_k)$
- Log-odds:  $\theta_k \xrightarrow{\text{blue arrow}} \theta = \log\left(\frac{p}{1-p}\right)$
- Basket Prior:  $\theta_k \sim N(\mu_k, \sigma_k^2)$
- Hyperpriors:  $\mu_k \sim N(-1.34, 10^2)$   
 $\sigma_k \sim IG(0.0005, 0.000005)$

Berry (2013) also has an adjustment for the targeted response rate  $p_1$

$$\theta = \log\left(\frac{p}{1-p}\right) - \log\left(\frac{p_1}{1-p_1}\right)$$

Berry, Scott M., et al. "Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials." *Clinical Trials* 10.5 (2013): 720-734.

# Introducing pyBasket

1. Cluster patients using omics.
2. Bayesian inference (logistic regression)
  - Infer patient response to treatment based on its basket, omics cluster and their interactions.
  - Hierarchical structure allows for parameter sharing across baskets, just like BHM.
  - Interaction terms enable the capturing of relationships between baskets and clusters.

# pyBasket Model

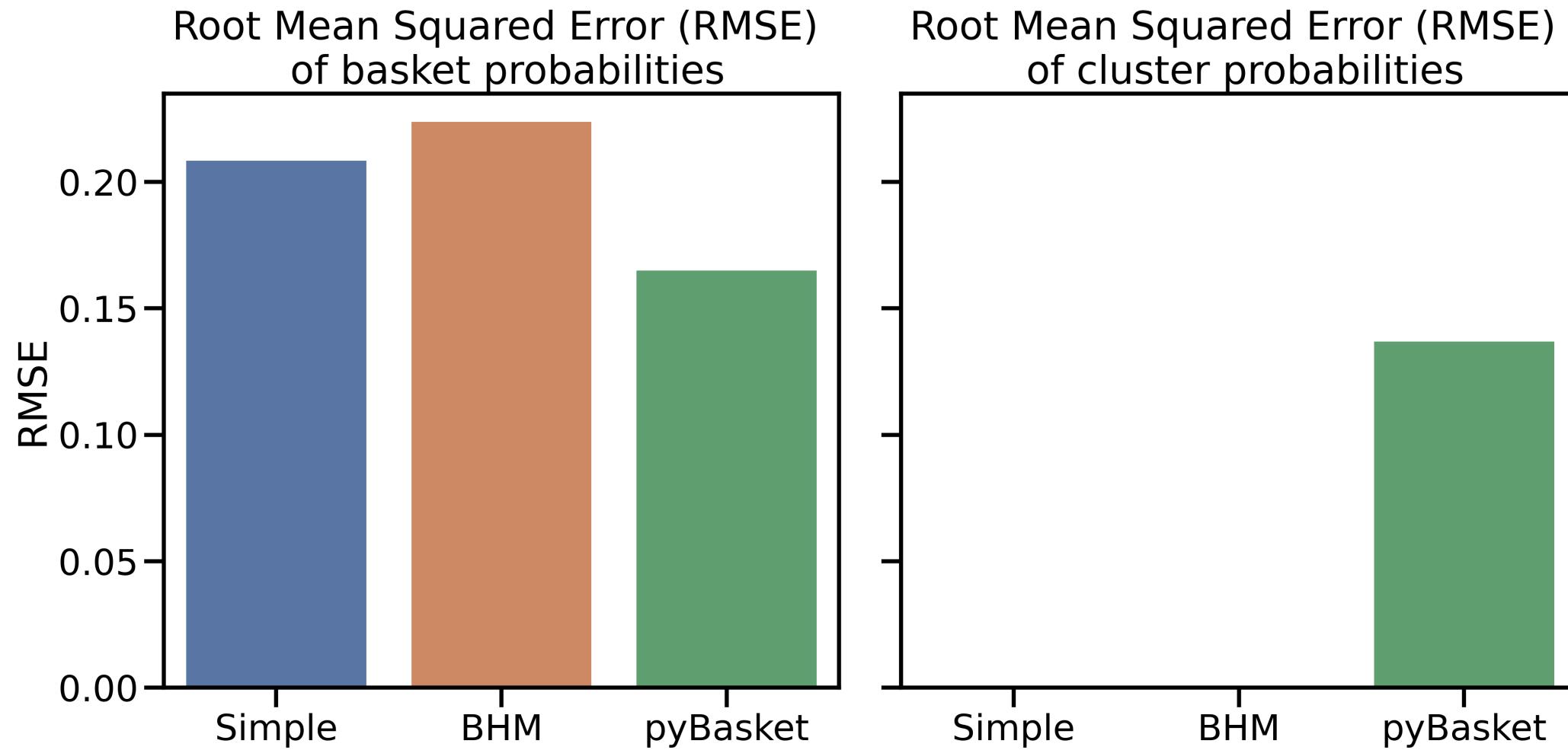
- Observed Responses:  $y_n \sim Bernoulli(p_n)$
- Success Probabilities:  $p_n = logistic(\theta_n)$
- Log-odds:

$$\theta_n = \theta_b^i + \theta_c^j + \theta_{interaction}^{ij}$$

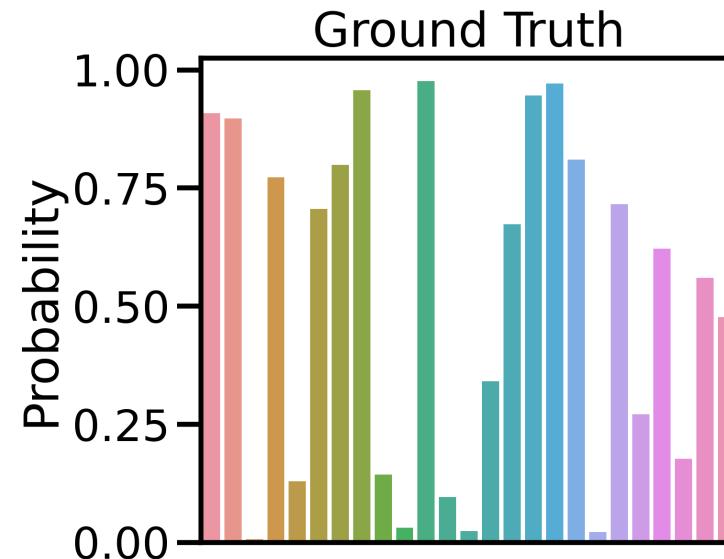
for patient  $n$  in basket  $i$  and cluster  $j$ .

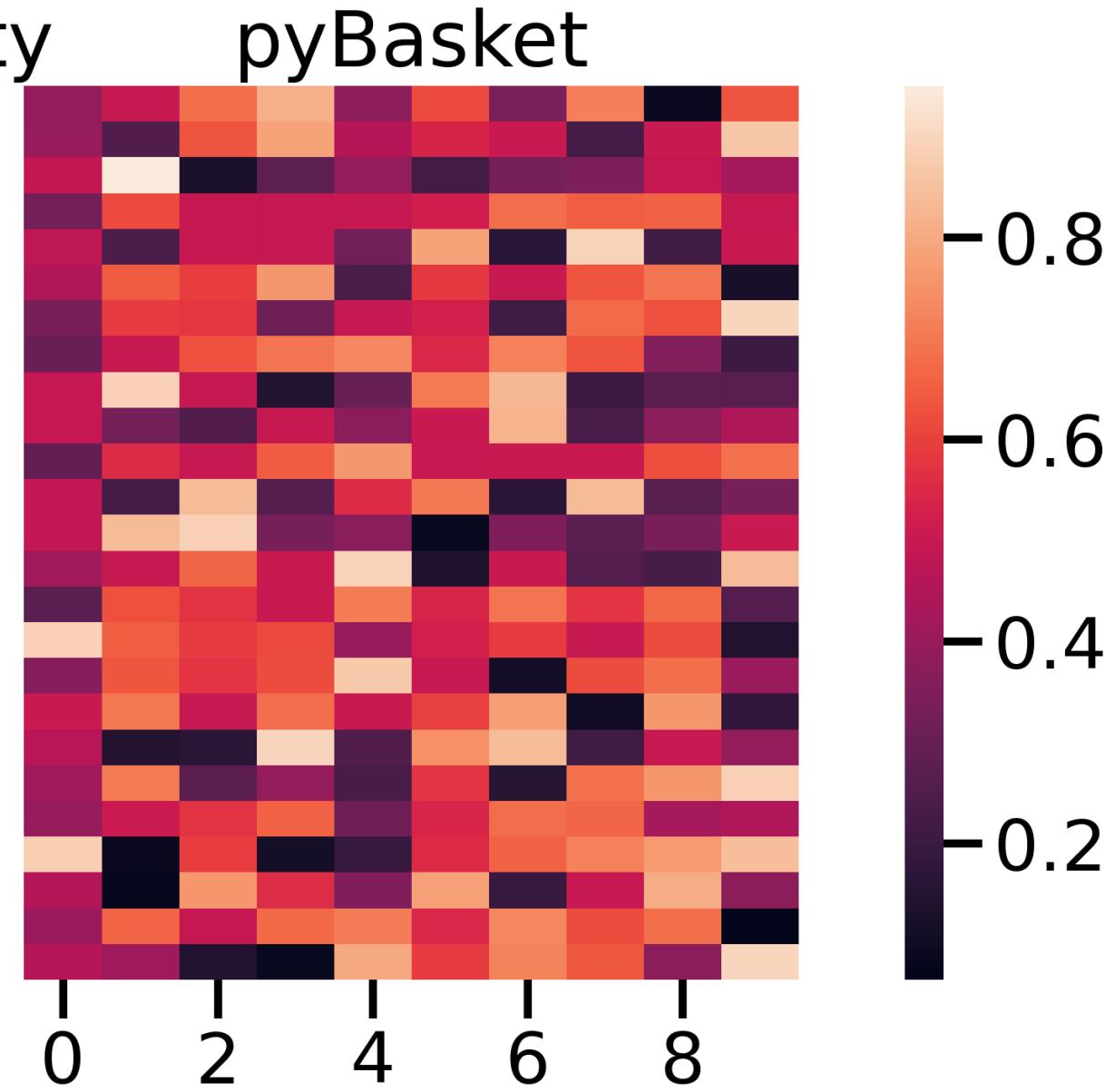
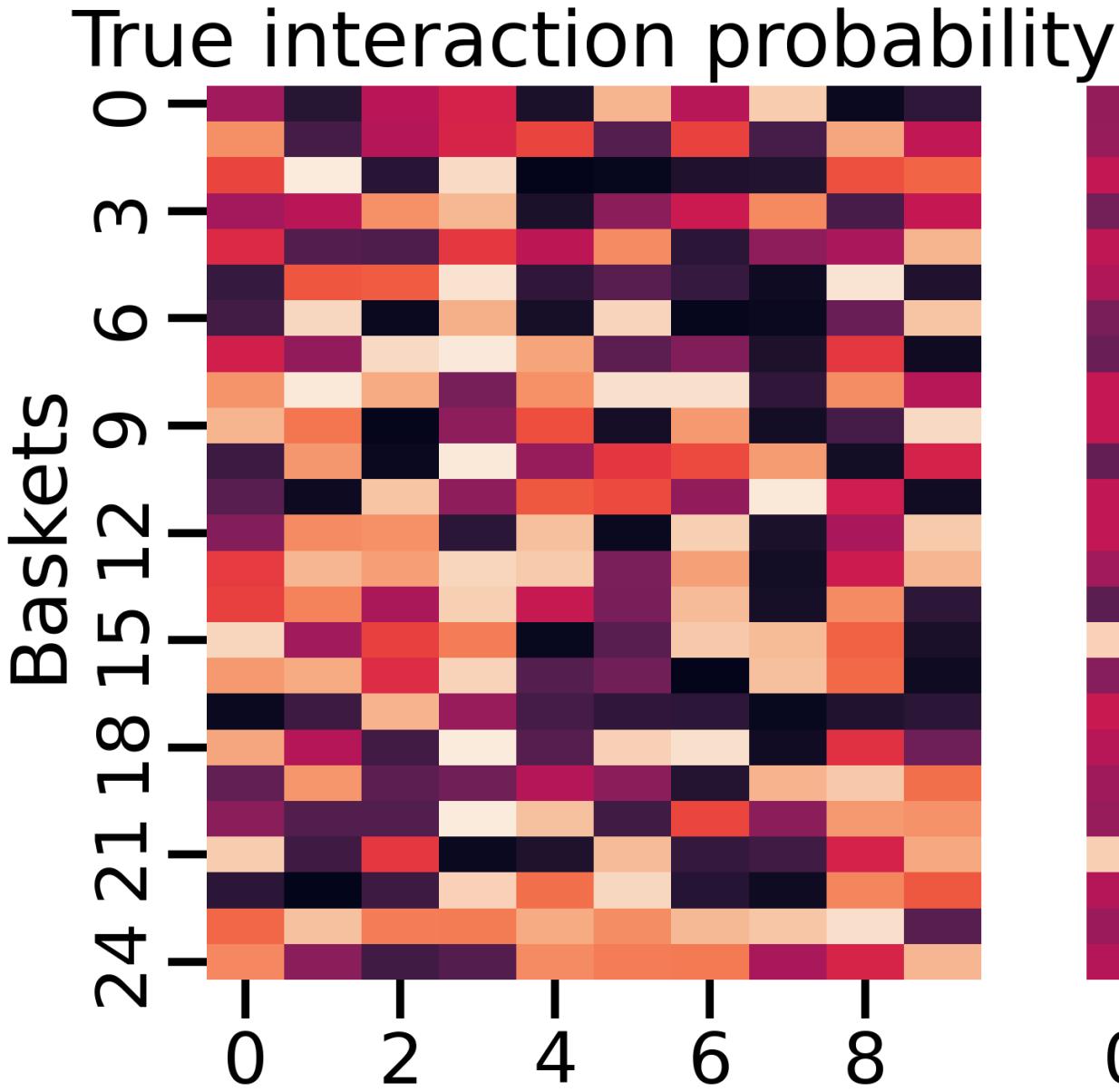
- Basket and Cluster Effects:  $\theta_b \sim N(\mu_b, \sigma_b^2)$ ,  $\theta_c \sim N(\mu_c, \sigma_c^2)$
- Interaction Term:  $\theta_{interaction} \sim N(0, 2^2)$
- Hyperpriors:  $\mu_b, \mu_c \sim N(0, 2^2)$   
 $\sigma_b, \sigma_c \sim HalfNormal(1)$

# Synthetic Results



# Basket Probabilities





# Real-world Application - Genomics of Drug Sensitivity in Cancer (GDSC)

- Due to the rarity of existing basket trials with omics measurements, validation of the pyBasket model is challenging.
- We use the Genomics of Drug Sensitivity in Cancer (GDSC) dataset, where cell lines (**not patients**) are treated with different drugs and have multiple omics measured.
- Can think of this a single-step basket trial.

# Data pre-processing

- Data consists of 303 cell lines (samples) across 24 tissues (baskets)
- Feature selection using random forest to retain the top-500 features more predictive of drug response.
- Transcriptomics log FC were used for K-means clustering (K=10).
- Binary responses to the treatment (drug) and cluster indicators were used as input to the model.

# Case study

- Erlotinib (Tarceva): A drug used to treat non-small cell lung cancer (NSCLC) and pancreatic cancer.
- Target: Erlotinib is specifically effective for NSCLC with mutations in the epidermal growth factor receptor (EGFR).
- So potentially a (hypothetical) basket trial analysis system would come out with this drug being significant for non-small cell lung cancer and also pancreatic cancer.

# pyBasket

Home

Data Exploration

Interactions

Features Importance

About

## Select basket\*cluster interaction

Select a cluster

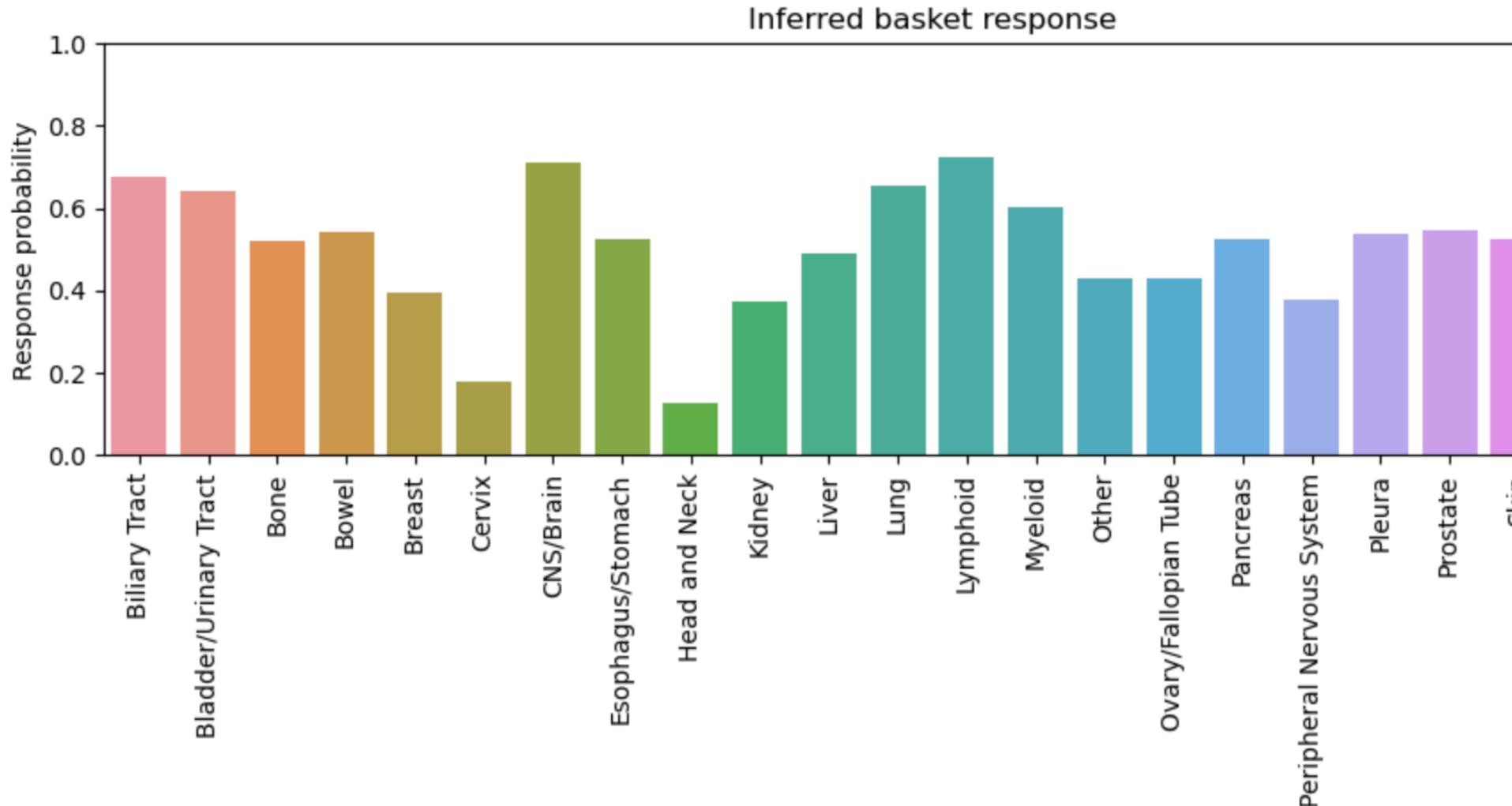
0

Select a basket

Biliary Tract

Samples in cluster 0 & Biliary Tract basket: 1

Interactive viewer (Streamlit) to explore model's ability to capture the effects of different cell lines (basket), clusters and their interactions.



# pyBasket

Home

## Data Exploration

## Interactions

## Select basket\*cluster interaction

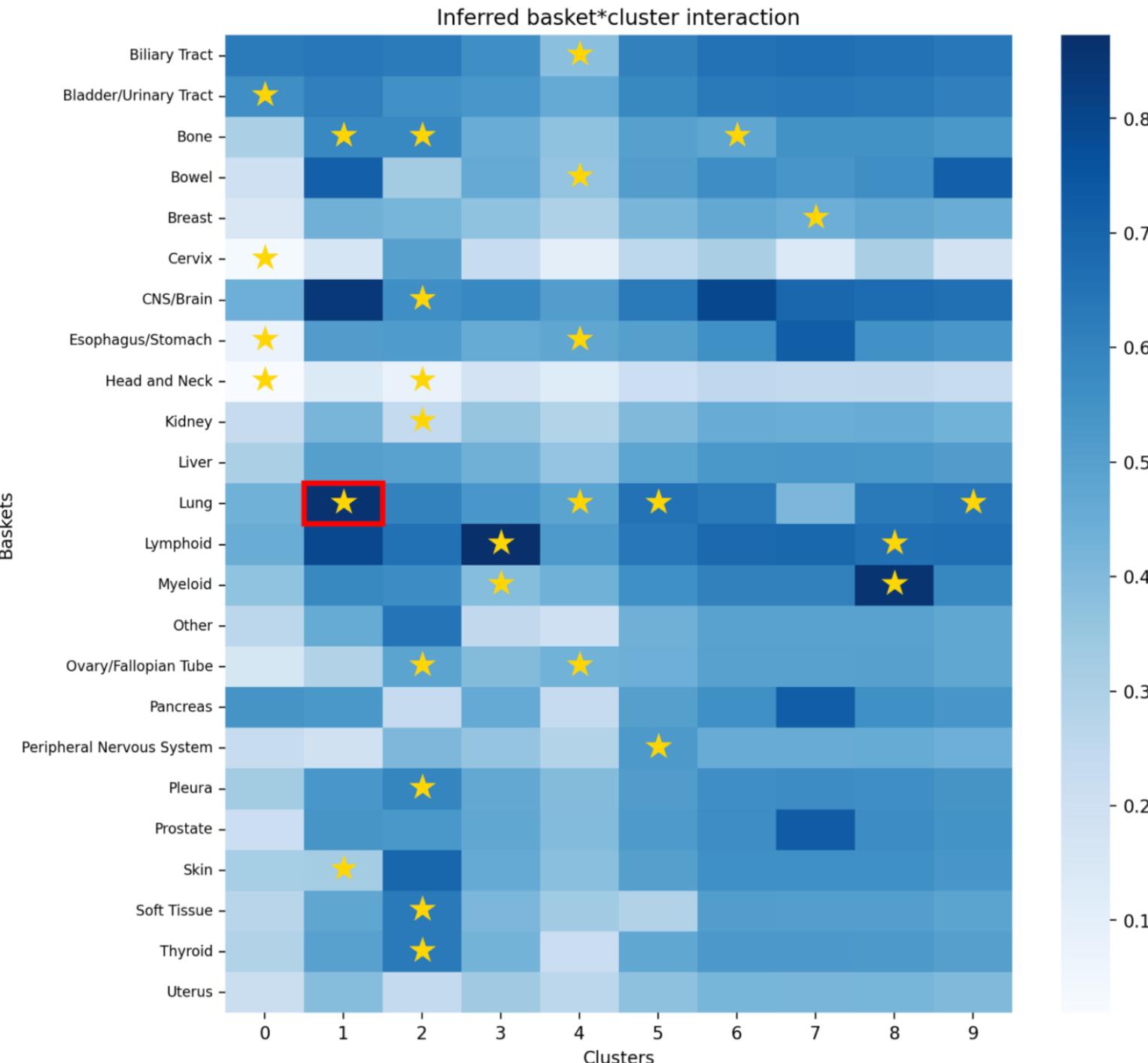
## Select a cluster

1

## Select a basket

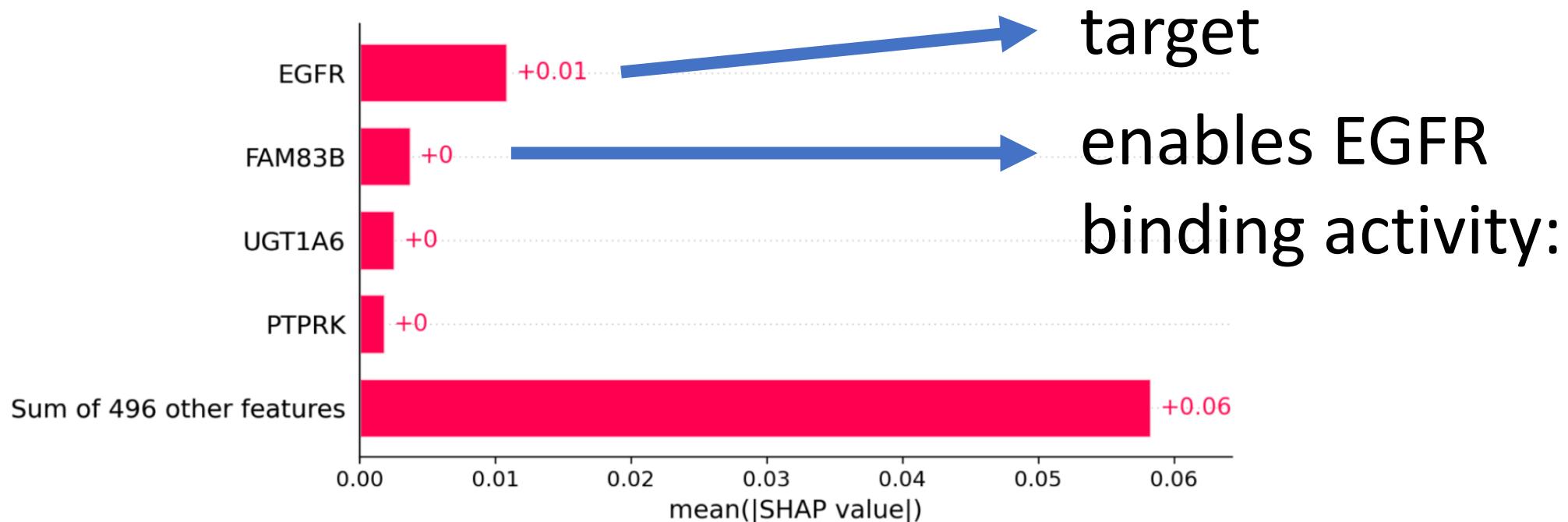
Lung

**Samples in cluster 1 & Lung basket: 3**



# Basket ‘Lung’ & Cluster 1

- SHAP explains the prediction of an observation by computing the contribution of each feature to the prediction.
- Features with highest mean SHAP values for cluster 1 in Lung.



# Future work

- **It's Bayesian, but ...** all the results shown so far are point estimates. Should make full use of the uncertainties we get from the model.
- **Basket Trial Simulations:** assess pyBasket model under a variety of scenarios in multi-stage simulated trials in comparison to other models.
- **Proper Model Evaluation:** use statistical metrics such as the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) to measure the fit of the pyBasket model to the data.
- **Data Integration:** Incorporating additional types of data, such as multi-omics or clinical data to improve predictive accuracy.



Thank you!