

Flexible Predictive Biomarker Discovery

2022 Center for Computational Biology Retreat

Philippe Boileau

Collaborators

Co-authors:

- Nina Qi, Genentech Inc.
- Mark van der Laan, UC Berkeley
- Sandrine Dudoit, UC Berkeley
- Ning Leng, Genentech Inc.

Additional support:

- Zoe June Assaf, Genentech Inc.
- Romain Banchereau, Genentech Inc.

Motivation

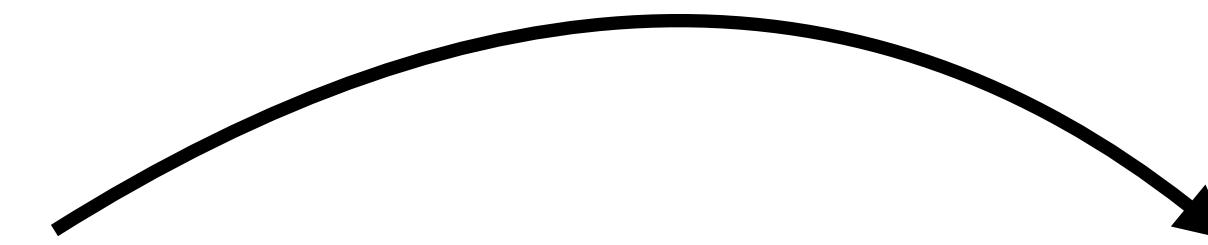
Metastatic Renal Cell Carcinoma

Finding biomarkers predictive of clinical benefit

(Previous) standard of care, tyrosine kinase inhibitors, are ineffective

Metastatic Renal Cell Carcinoma

Finding biomarkers predictive of clinical benefit

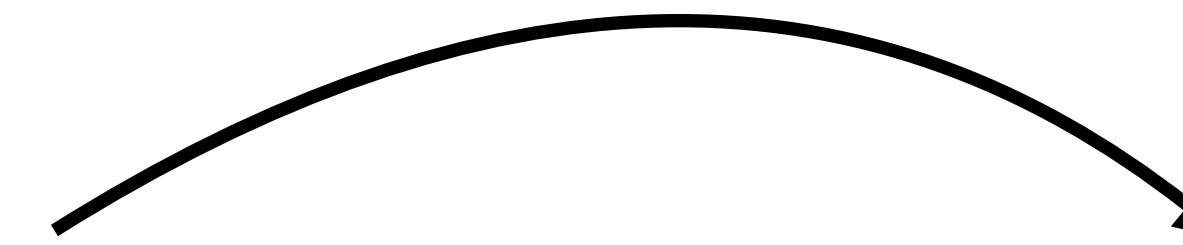


(Previous) standard of care, tyrosine kinase inhibitors, are ineffective

Metastatic Renal Cell Carcinoma

Finding biomarkers predictive of clinical benefit

(Previous) standard of care, tyrosine kinase inhibitors, are ineffective

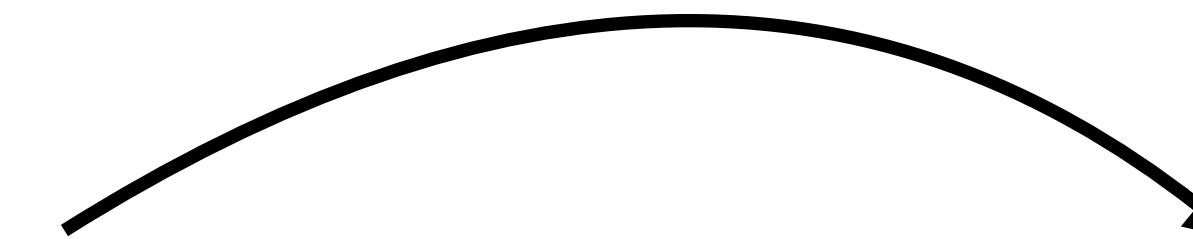


IMmotion 150
Phase 2

Metastatic Renal Cell Carcinoma

Finding biomarkers predictive of clinical benefit

(Previous) standard of care, tyrosine kinase inhibitors, are ineffective



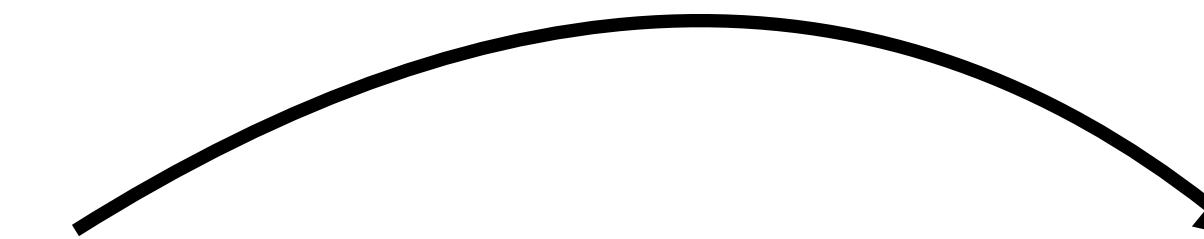
IMmotion 150
Phase 2

Immune checkpoint inhibitor
vs
Immune checkpoint inhibitor
+ tyrosine kinase inhibitor
vs
Tyrosine kinase inhibitor

Metastatic Renal Cell Carcinoma

Finding biomarkers predictive of clinical benefit

(Previous) standard of care, tyrosine kinase inhibitors, are ineffective



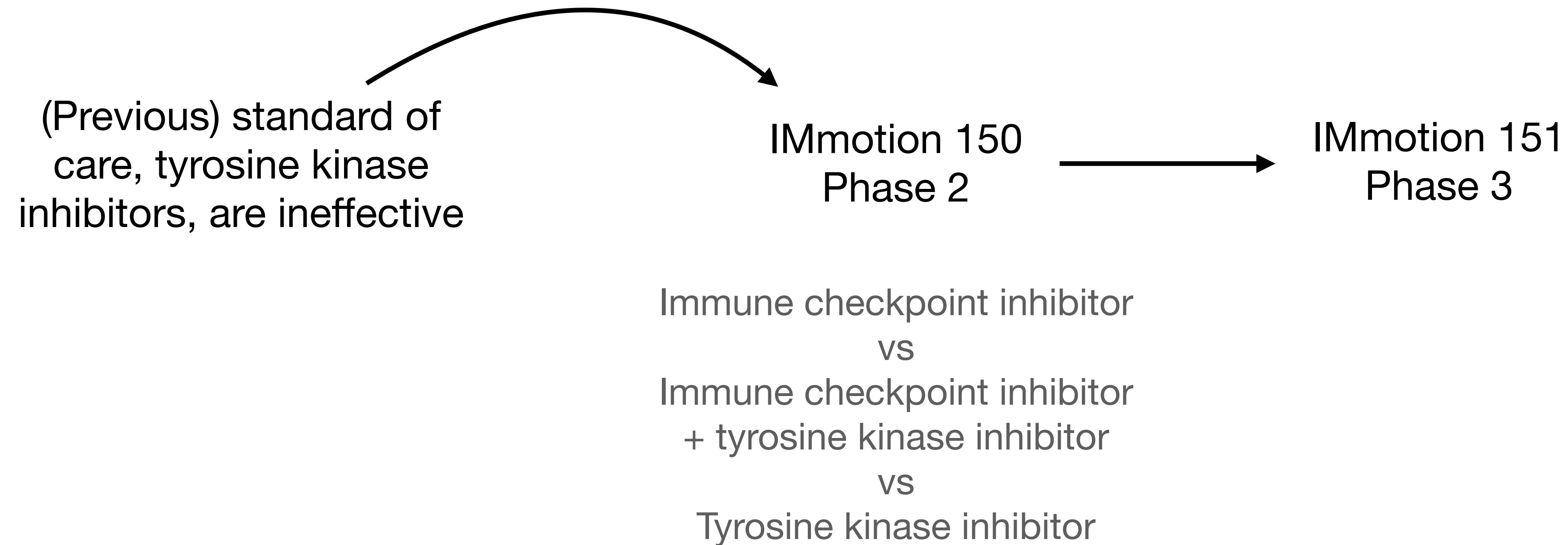
IMmotion 150
Phase 2

IMmotion 151
Phase 3

Immune checkpoint inhibitor
vs
Immune checkpoint inhibitor
+ tyrosine kinase inhibitor
vs
Tyrosine kinase inhibitor

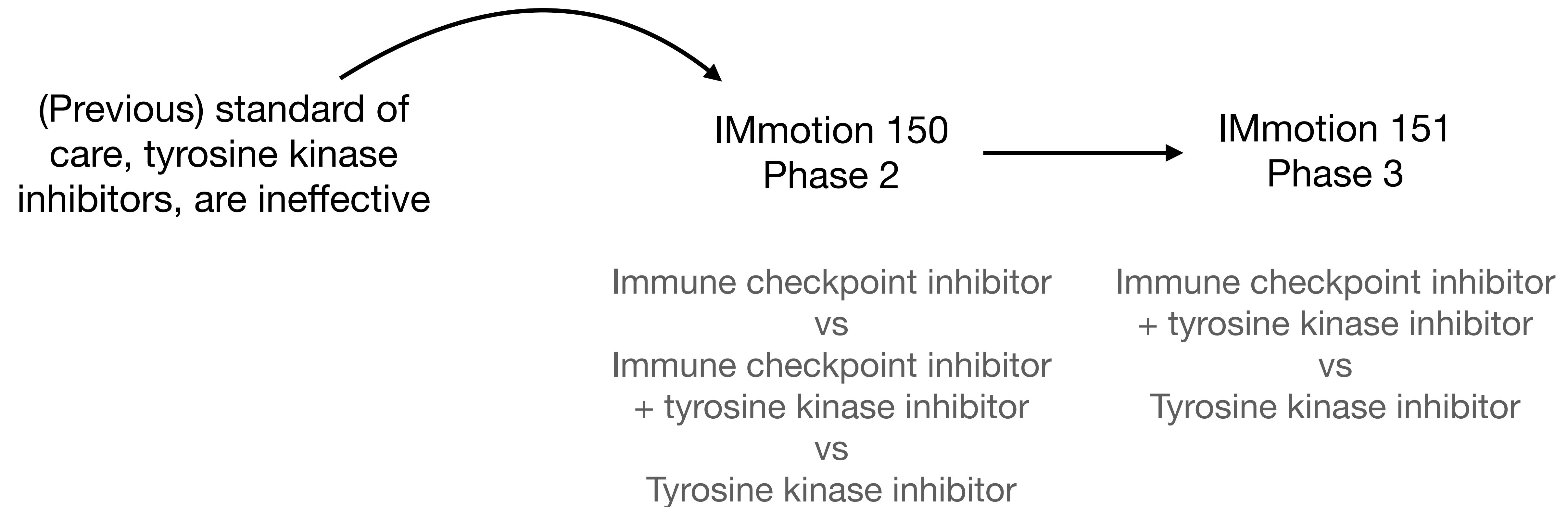
Metastatic Renal Cell Carcinoma

Finding biomarkers predictive of clinical benefit



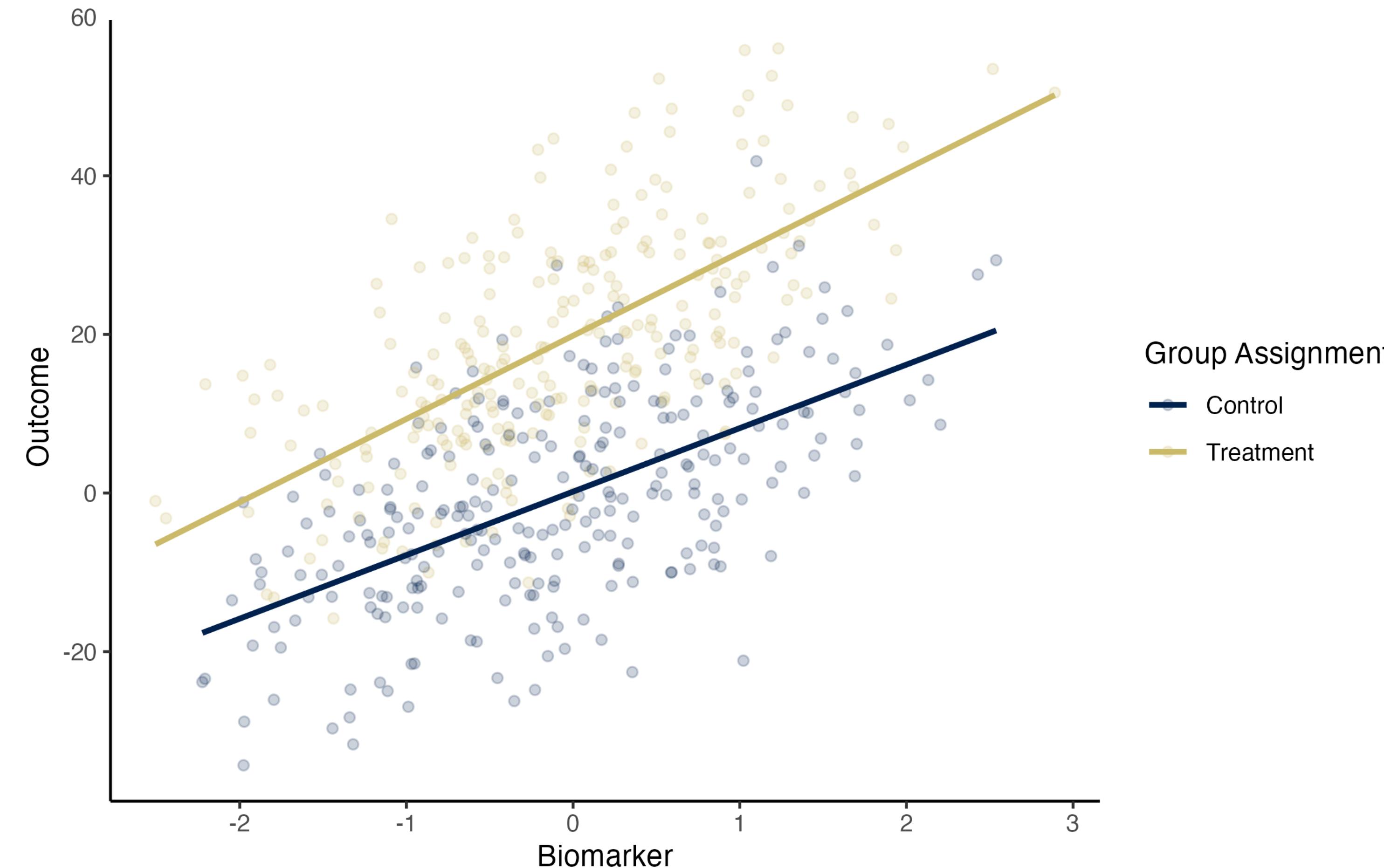
Metastatic Renal Cell Carcinoma

Finding biomarkers predictive of clinical benefit



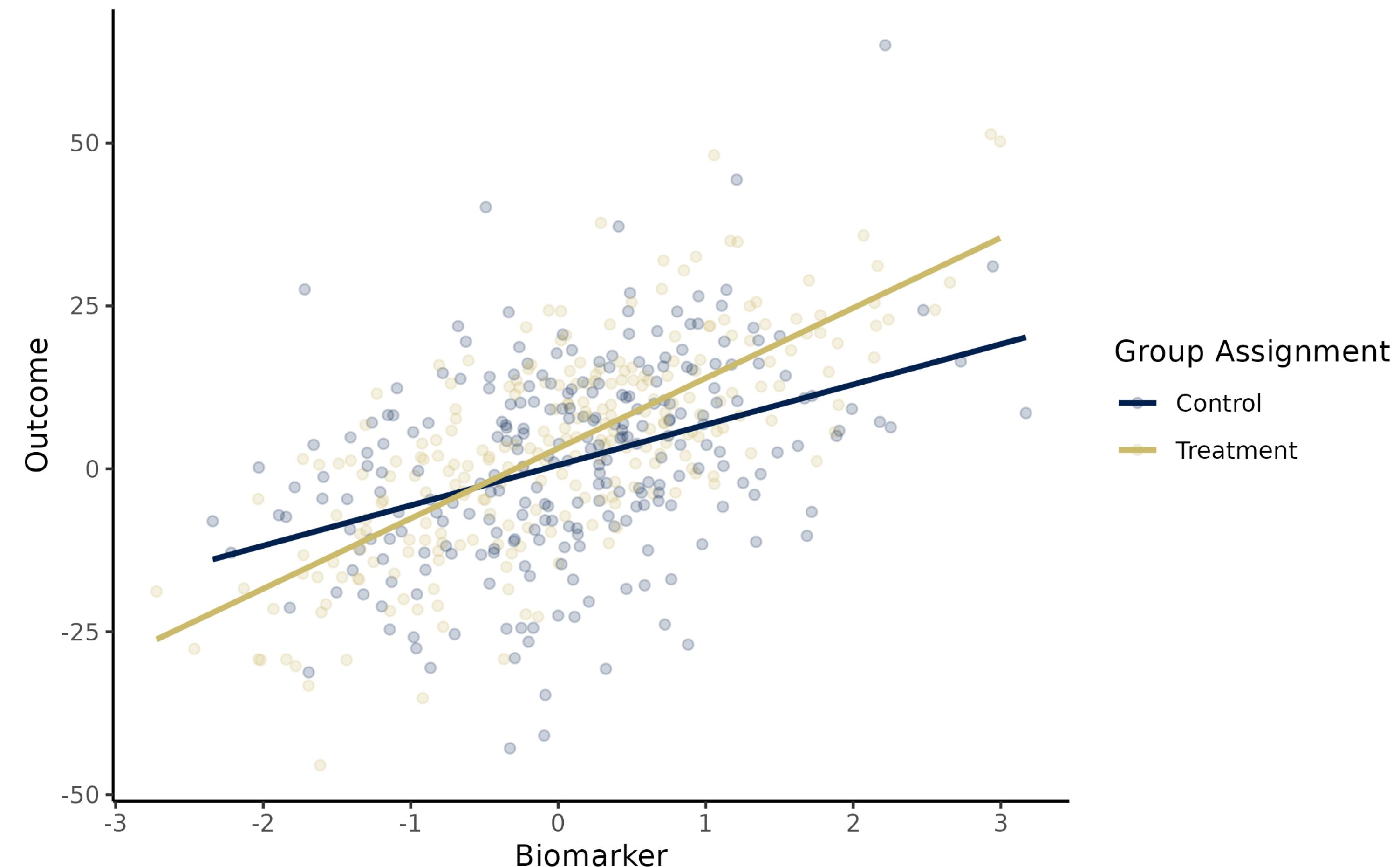
Prognostic Biomarkers

Indicators of outcome, regardless of therapy



Predictive Biomarkers

Treatment effect modifiers



Predictive Biomarker Applications

Predictive biomarkers drive personalized medicine

Predictive Biomarker Applications

Predictive biomarkers drive personalized medicine

- **Diagnostic assay development:** Who benefits most from a therapy?

Predictive Biomarker Applications

Predictive biomarkers drive personalized medicine

- **Diagnostic assay development:** Who benefits most from a therapy?
- **Targeted drug discovery:** What is the biological mechanism of a therapy?

Predictive Biomarker Applications

Predictive biomarkers drive personalized medicine

- **Diagnostic assay development:** Who benefits most from a therapy?
- **Targeted drug discovery:** What is the biological mechanism of a therapy?
- **Refined clinical trials:** Establish a subset of the patient population for which therapy is more efficacious?

Discovering Predictive Biomarkers

Uncovering Predictive Biomarkers

A variable selection problem

Uncovering Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:

Uncovering Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:
 - Linear models with treatment-biomarker interaction terms

Uncovering Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:
 - Linear models with treatment-biomarker interaction terms
 - Conditional average treatment effect (CATE) estimation

Uncovering Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:
 - Linear models with treatment-biomarker interaction terms
 - Conditional average treatment effect (CATE) estimation
- Harder when there are a large number of biomarkers: Penalized versions of the above methods are used.

Uncovering Predictive Biomarkers

A variable selection problem

- Easy when there are few biomarkers to consider:
 - Linear models with treatment-biomarker interaction terms
 - Conditional average treatment effect (CATE) estimation
- Harder when there are a large number of biomarkers: Penalized versions of the above methods are used.
- **Bottom line:** Discovery of predictive biomarkers is the byproduct of another inference procedure.

Example: Modified Covariates Approach

A method for modeling treatment-biomarker interactions directly

Example: Modified Covariates Approach

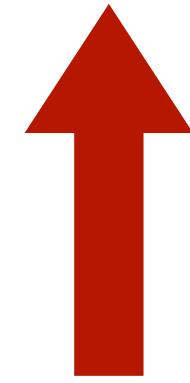
A method for modeling treatment-biomarker interactions directly

$$\text{transformed outcome} = \text{treatment} + \text{treatment} \times \text{biomarker 1} + \dots + \text{treatment} \times \text{biomarker } p$$

Example: Modified Covariates Approach

A method for modeling treatment-biomarker interactions directly

$$\text{transformed outcome} = \text{treatment} + \text{treatment} \times \text{biomarker 1} + \dots + \text{treatment} \times \text{biomarker } p$$



Predictive biomarkers have non-zero coefficients

Issues with Penalized Regression Methods

Unreliable biomarker selection

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

Violations produce to high false positive rates, leading to:

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

Violations produce to high false positive rates, leading to:

- Resources wasted on follow-up experiments and trials

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

Violations produce to high false positive rates, leading to:

- Resources wasted on follow-up experiments and trials
- Decreased signal to noise ratio in diagnostic assays

Issues with Penalized Regression Methods

Unreliable biomarker selection

Strong assumptions: sparsity and correlation structure.

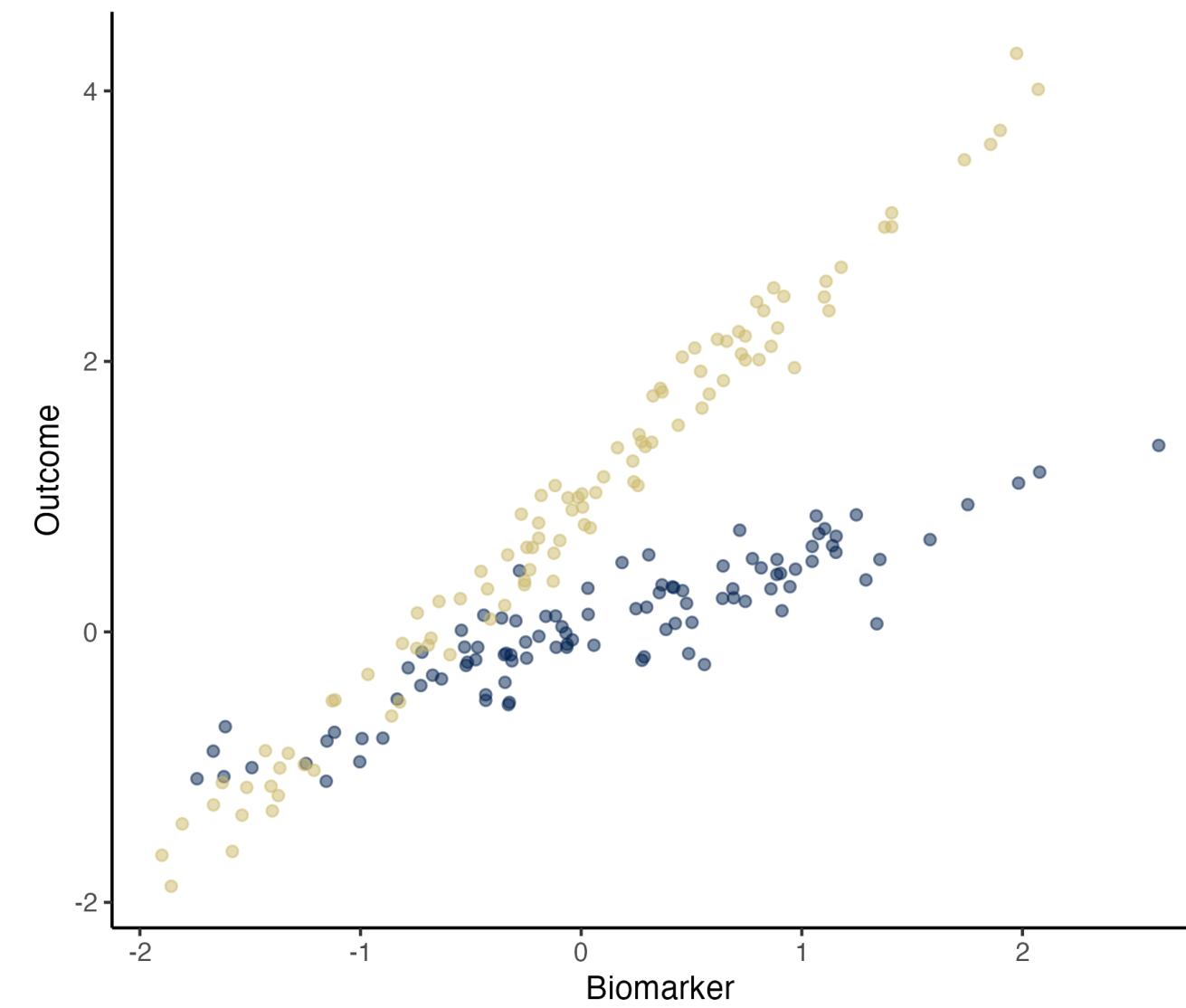
Violations produce to high false positive rates, leading to:

- Resources wasted on follow-up experiments and trials
- Decreased signal to noise ratio in diagnostic assays

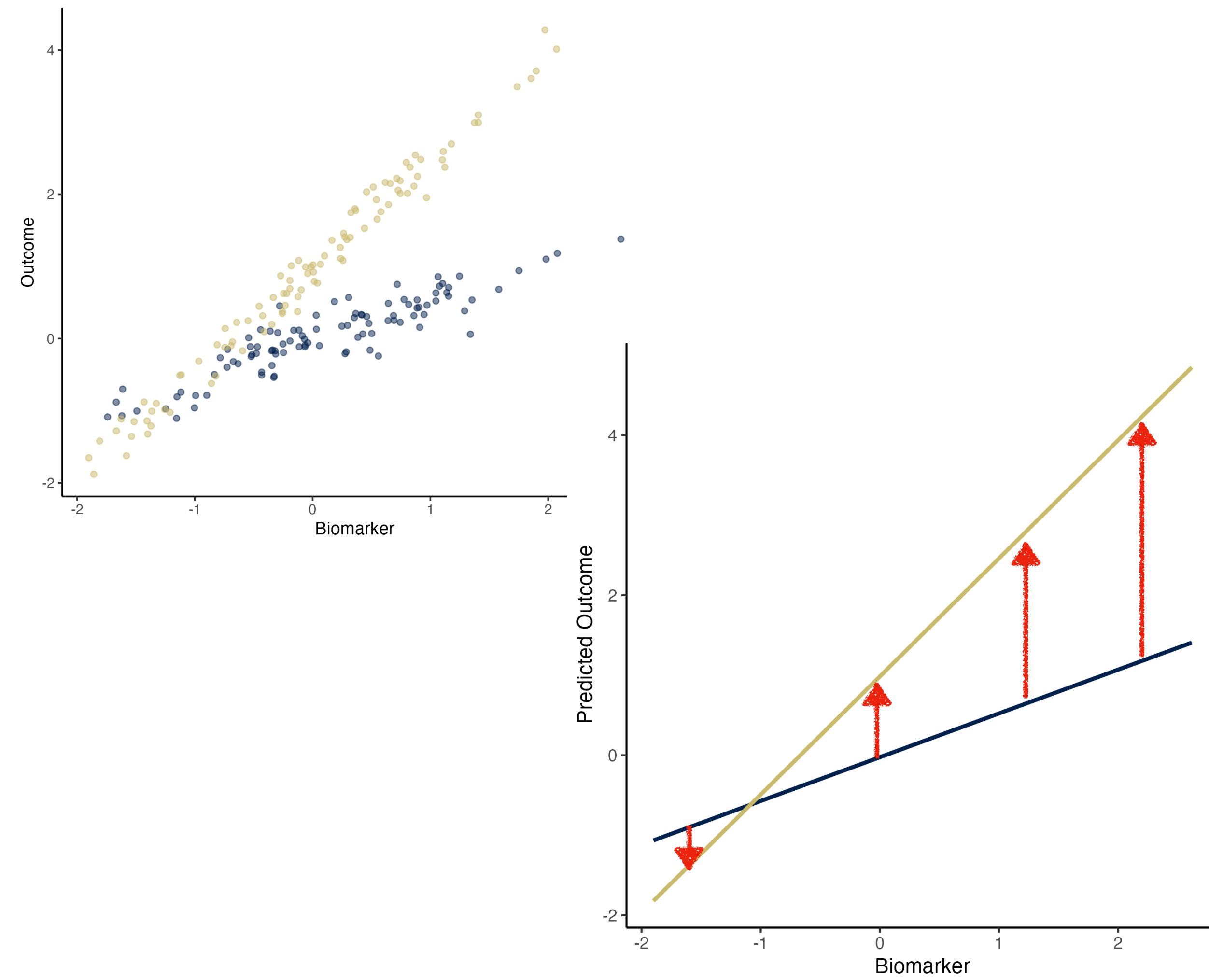
We need to consider alternative problem formulations.

A Dedicated Variable Importance Parameter

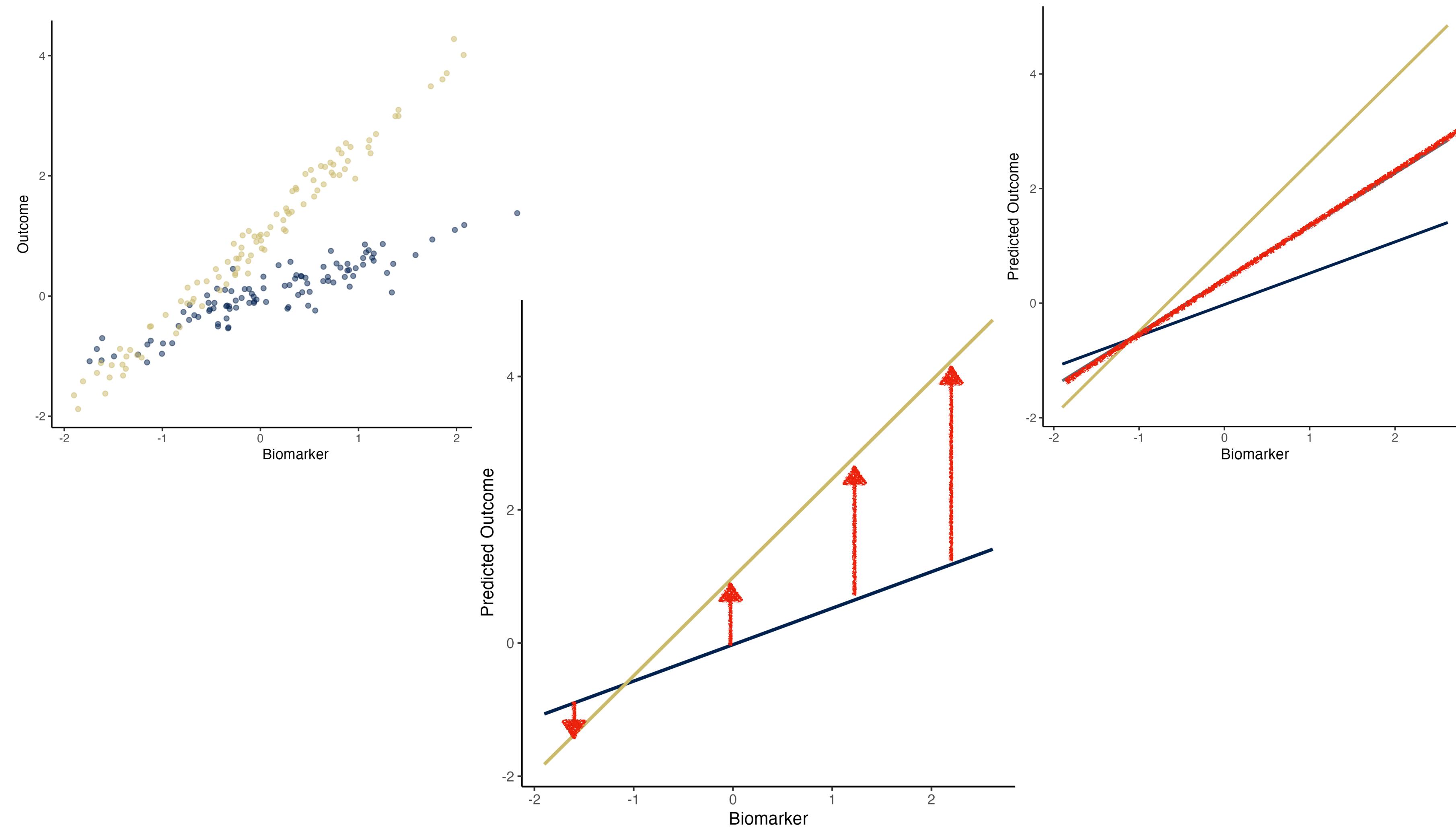
A Dedicated Variable Importance Parameter



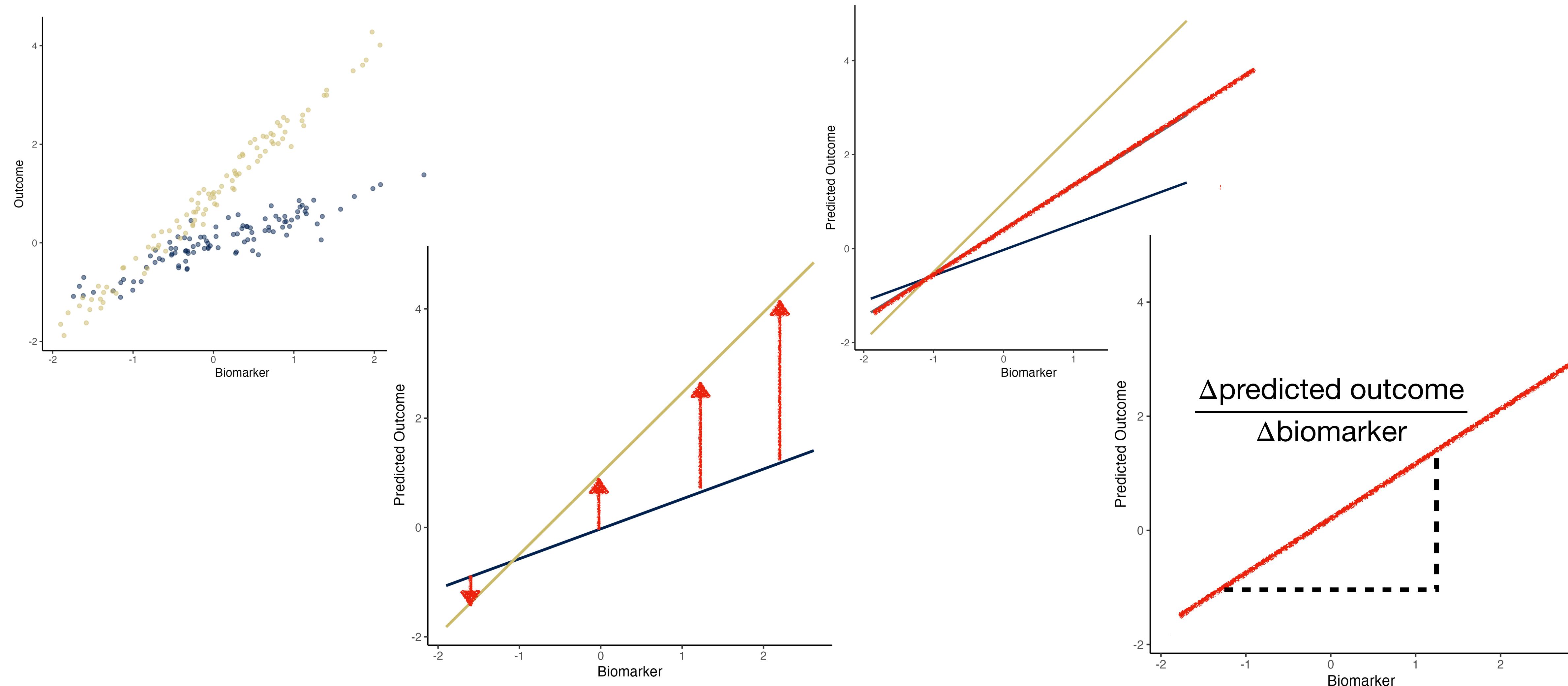
A Dedicated Variable Importance Parameter



A Dedicated Variable Importance Parameter



A Dedicated Variable Importance Parameter



uniCATE

Assumption-lean estimator of biomarker predictiveness

uniCATE

Assumption-lean estimator of biomarker predictiveness

Estimating this parameter for a **centered** biomarker is easy in a nonparametric model!

uniCATE

Assumption-lean estimator of biomarker predictiveness

Estimating this parameter for a **centered** biomarker is easy in a nonparametric model!

$$\frac{\frac{1}{n} \sum_{i=1}^n (\text{predicted outcome difference}^\star)_i (\text{biomarker})_i}{\frac{1}{n} \sum_{i=1}^n (\text{biomarker})_i^2}$$

uniCATE

Assumption-lean estimator of biomarker predictiveness

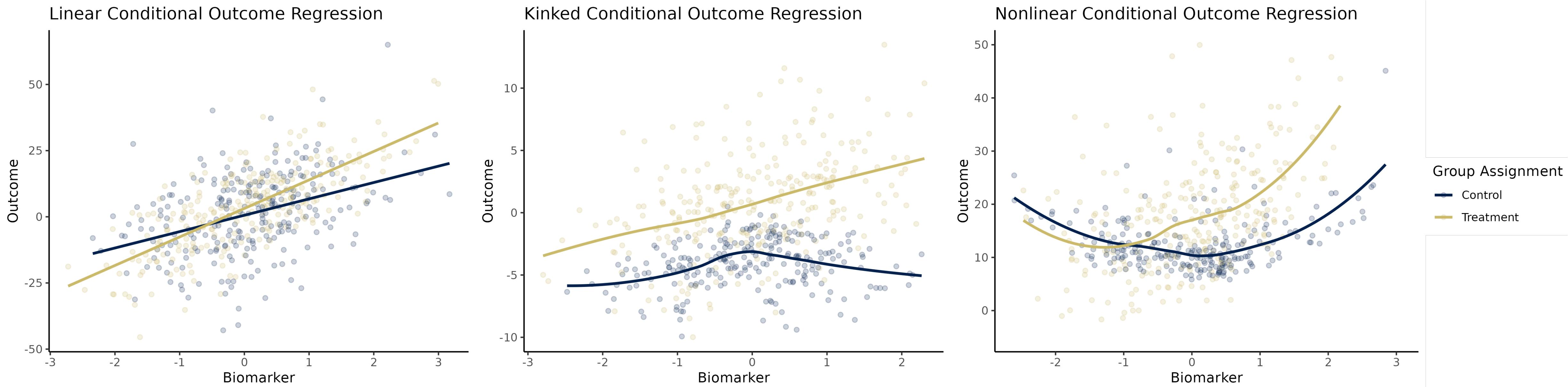
Estimating this parameter for a **centered** biomarker is easy in a nonparametric model!

$$\frac{\frac{1}{n} \sum_{i=1}^n (\text{predicted outcome difference}^\star)_i (\text{biomarker})_i}{\frac{1}{n} \sum_{i=1}^n (\text{biomarker})_i^2}$$

This estimator is **asymptotically Normal**. The only assumption in an RCT: the biomarker has non-zero variance.

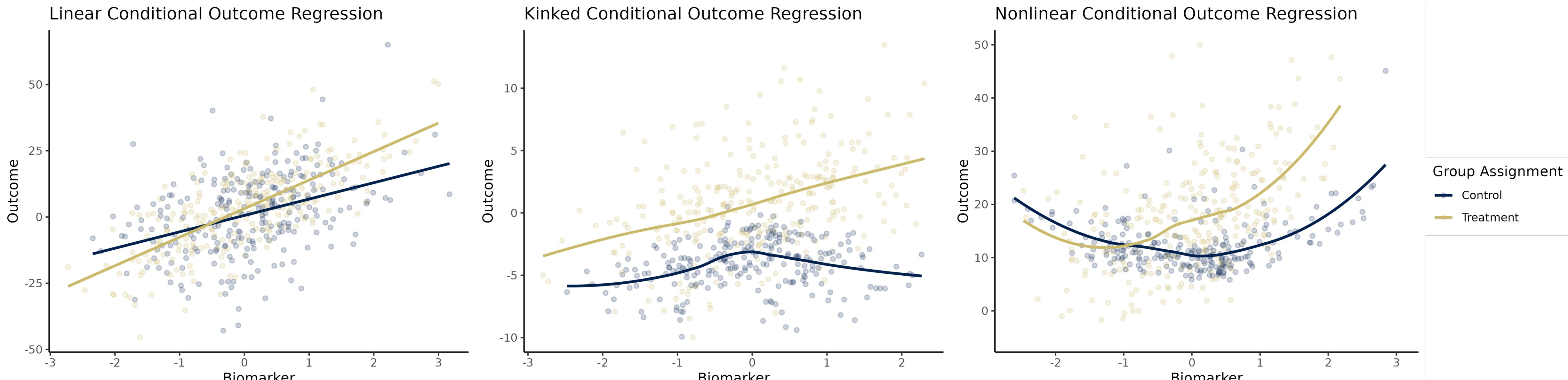
Simulation Studies

Considered Biomarker-Outcome Relationships



Difficulty

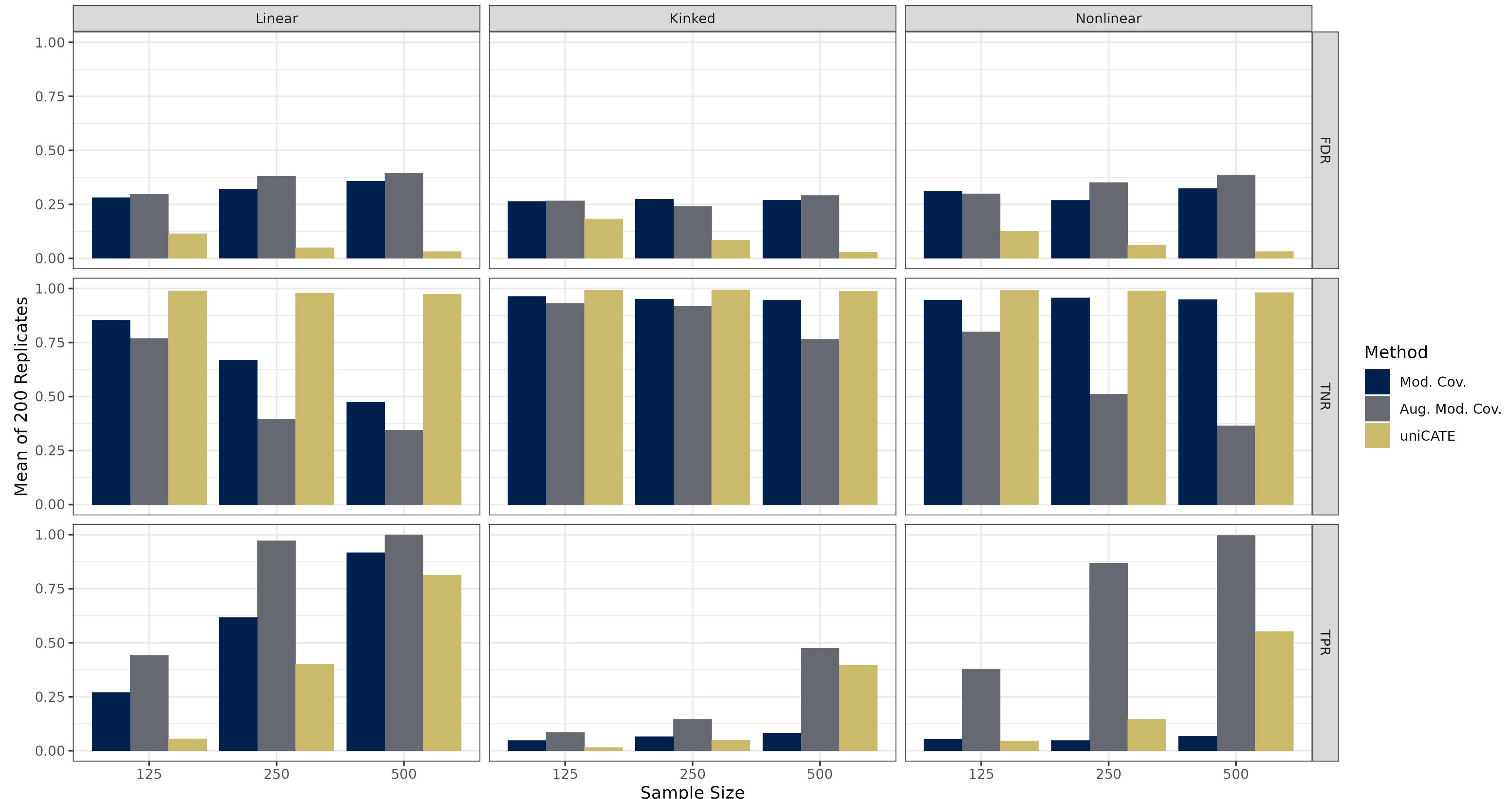
Considered Biomarker-Outcome Relationships



Difficulty

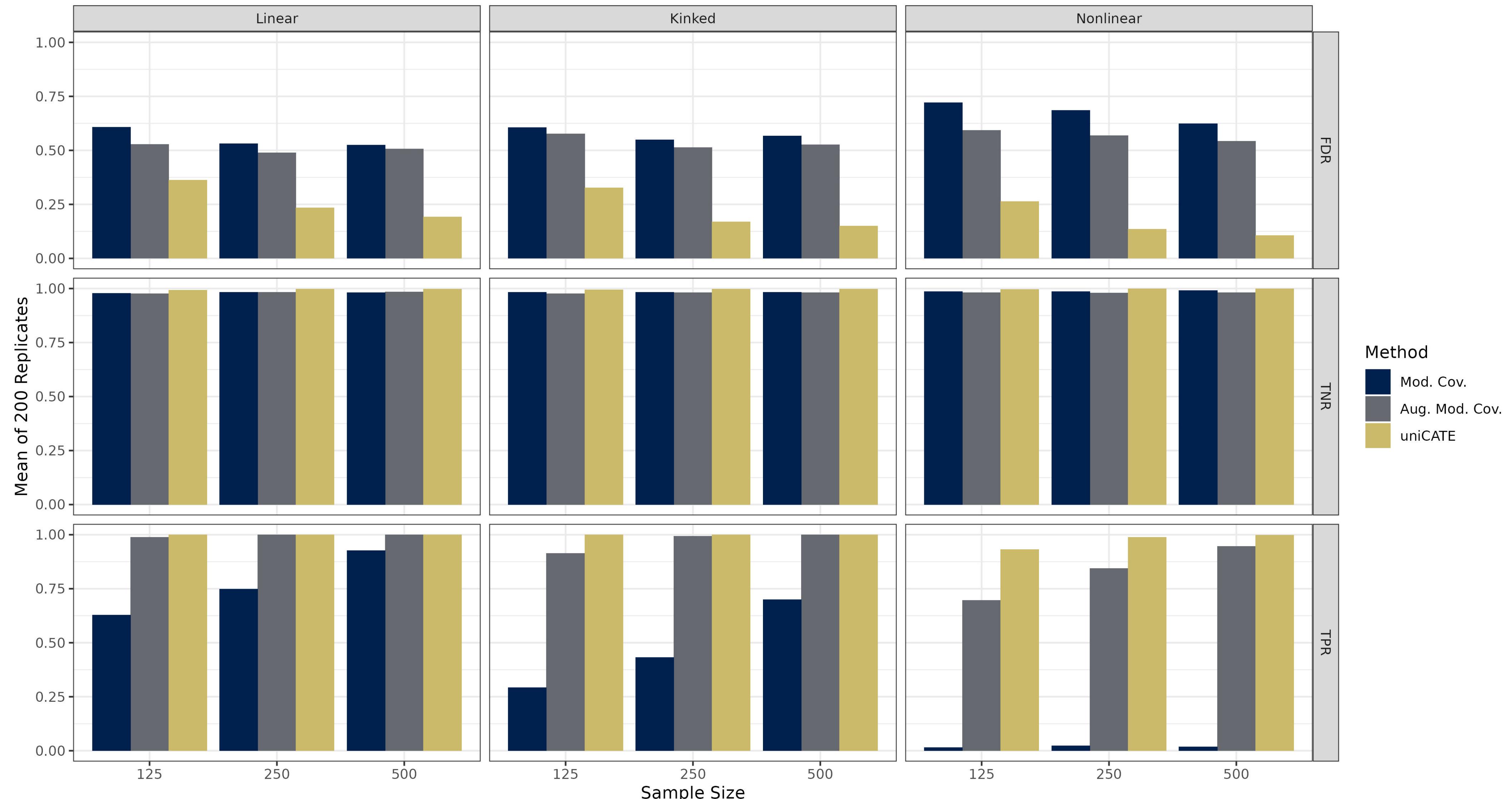
uniCATE Controls False Positive Rates

Classification of Non-Sparse, Moderate-Dimensional, and Uncorrelated Predictive Biomarkers



uniCATE Still Controls False Positive Rates

Classification of Sparse, High-Dimensional, and Correlated Predictive Biomarkers



Application to IMmotion 150/151

Application to IMmotion 150

uniCATE's results align with recent findings in nivolumab

Application to IMmotion 150

uniCATE's results align with recent findings in nivolumab

1. Only patients with tumor RNA-seq data in the sunitinib (n=71) and atezolizumab + bevacizumab (n=77) arms were considered.

Application to IMmotion 150

uniCATE's results align with recent findings in nivolumab

1. Only patients with tumor RNA-seq data in the sunitinib (n=71) and atezolizumab + bevacizumab (n=77) arms were considered.
2. Selected the 500 most variable, log-transformed genes as biomarkers.

Application to IMmotion 150

uniCATE's results align with recent findings in nivolumab

1. Only patients with tumor RNA-seq data in the sunitinib (n=71) and atezolizumab + bevacizumab (n=77) arms were considered.
2. Selected the 500 most variable, log-transformed genes as biomarkers.
3. Objective response was used as the response variable.

Application to IMmotion 150

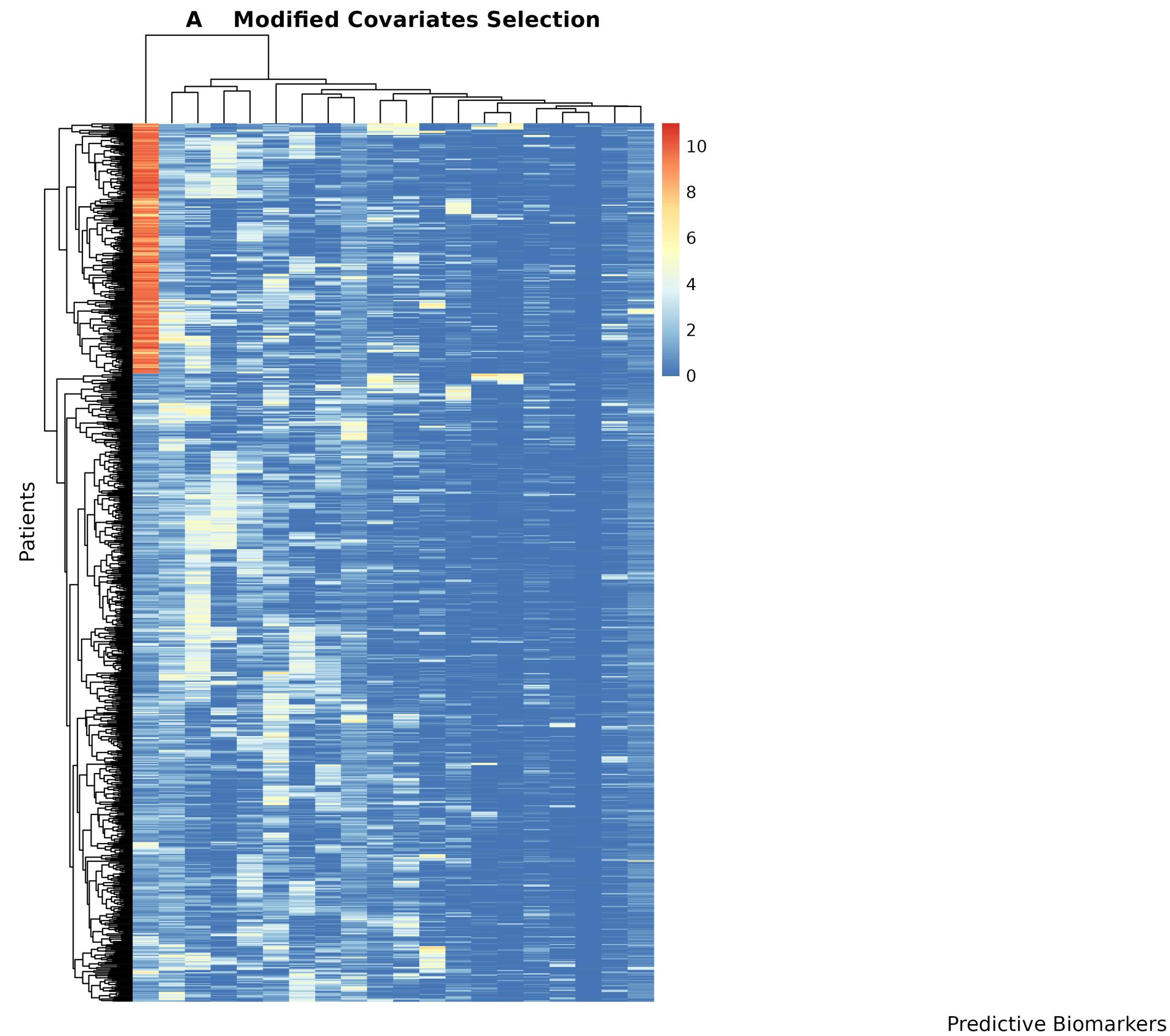
uniCATE's results align with recent findings in nivolumab

1. Only patients with tumor RNA-seq data in the sunitinib (n=71) and atezolizumab + bevacizumab (n=77) arms were considered.
2. Selected the 500 most variable, log-transformed genes as biomarkers.
3. Objective response was used as the response variable.

92 genes were identified as predictive using a 5% FDR cutoff. They are associated with immune responses, including those mediated by B cells and lymphocytes.

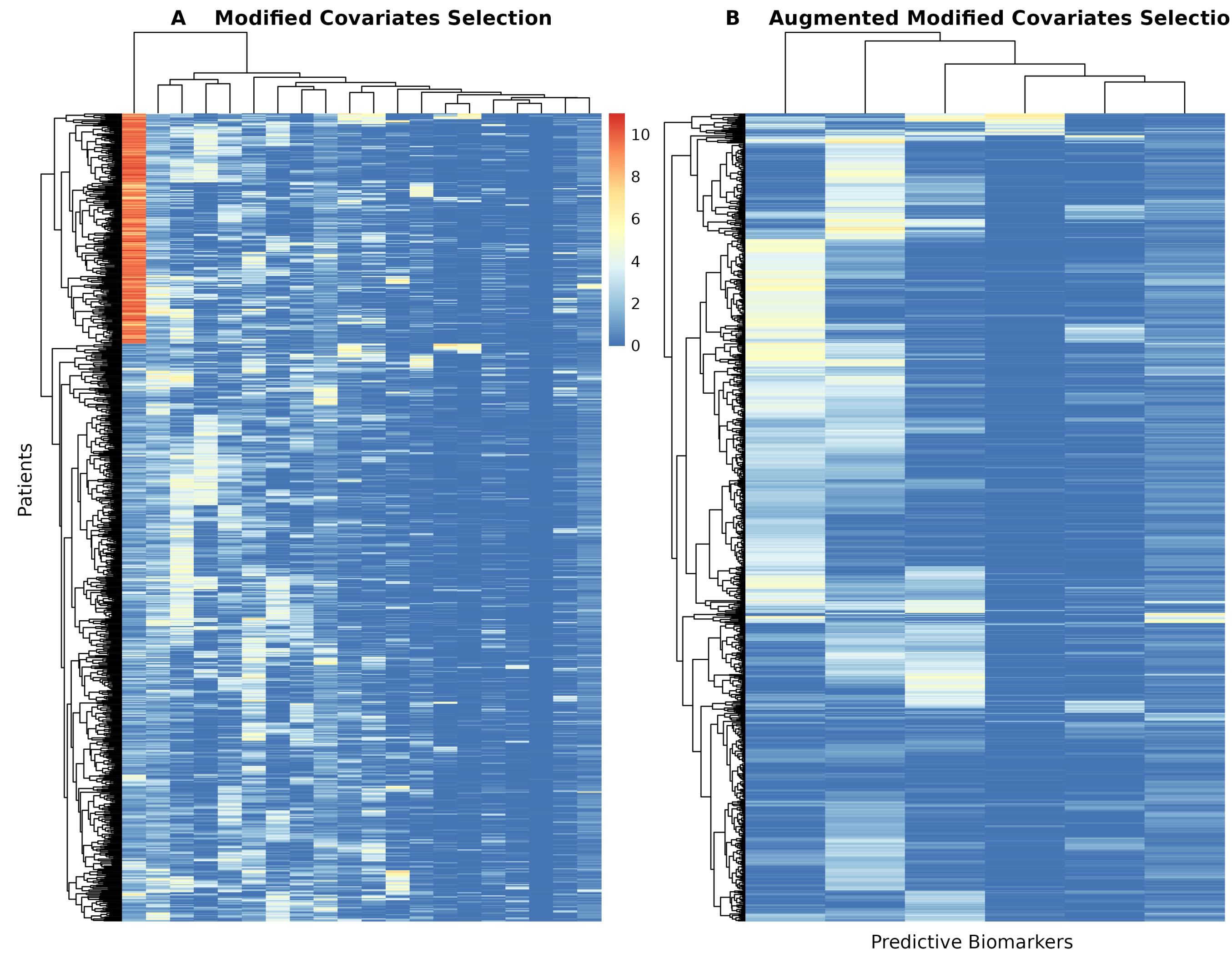
Validation on IMmotion 151

uniCATE identifies meaningful predictive biomarkers



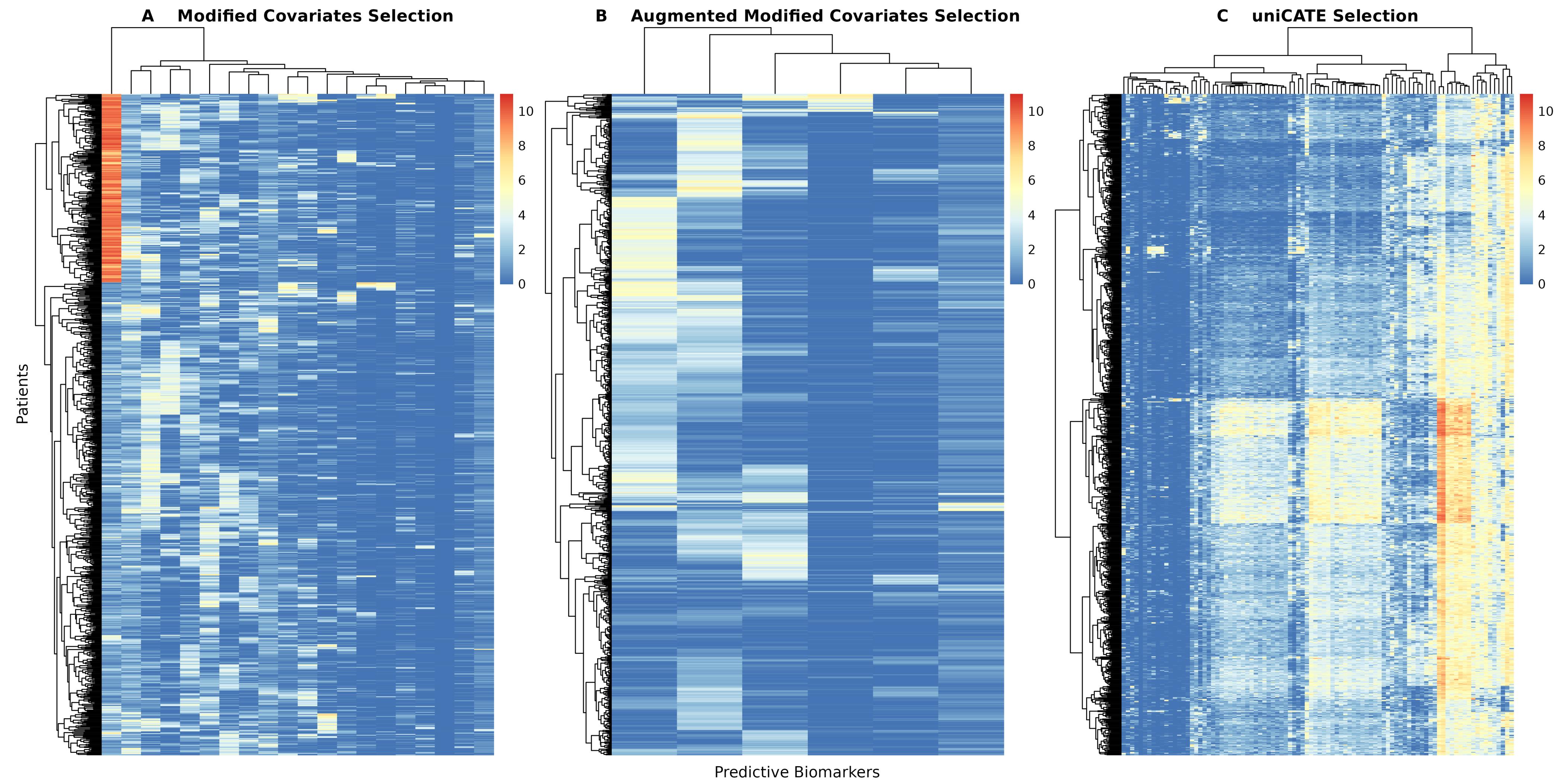
Validation on IMmotion 151

uniCATE identifies meaningful predictive biomarkers



Validation on IMmotion 151

uniCATE identifies meaningful predictive biomarkers



Summary

- uniCATE is an assumption-lean inference procedure that controls the rate of false positive predictive biomarkers in high dimensional RCTs.
- Check out uniCATE's implementation in the uniCATE R package, available at github.com/insightsengineering/uniCATE
- Article published in *Biostatistics: A Flexible Approach for Predictive Biomarker Discovery*