uniCATE: Flexible Predictive Biomarker Discovery

High-Dimensional Statistics Session, SDSS 2022

Philippe Boileau, UC Berkeley — June 2022

Collaborators

Co-authors:

- Nina Qi
- Mark van der Laan
- Sandrine Dudoit
- Ning Leng

Additional support:

- Zoe June Assaf
- Romain Banchereau

Motivation

Metastatic Renal Cell Carcinoma

Finding biomarkers predictive of clinical benefit

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

IMmotion 150
Phase 2

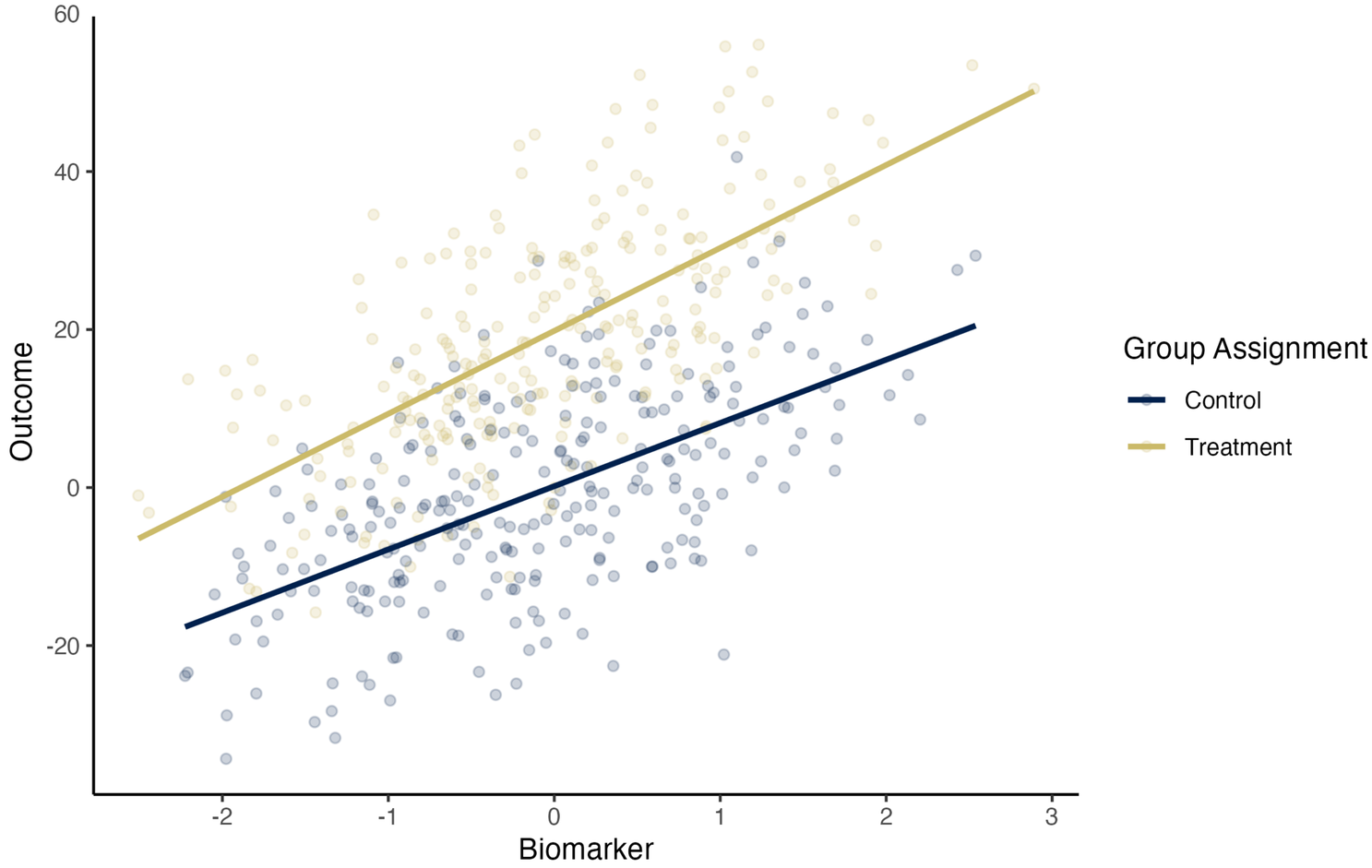
Immune checkpoint inhibitor
vs
Immune checkpoint inhibitor
+ tyrosine kinase inhibitor
vs
Tyrosine kinase inhibitor

IMmotion 151
Phase 3

Immune checkpoint inhibitor
+ tyrosine kinase inhibitor
vs
Tyrosine kinase inhibitor

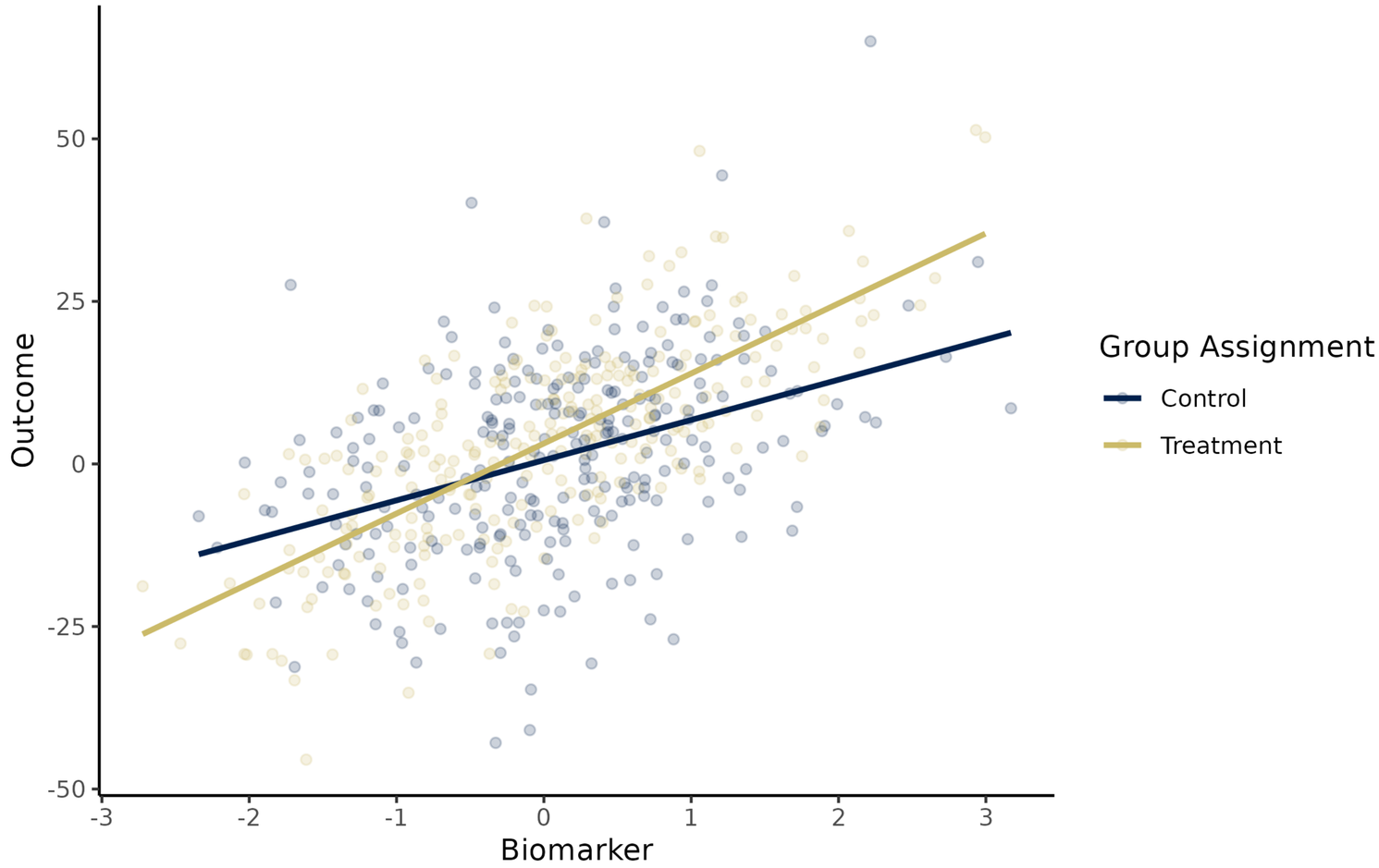
Prognostic Biomarkers

Indicators of outcome, regardless of therapy



Predictive Biomarkers

Treatment effect modifiers



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

- 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

- Tian et al (2014) demonstrated that the treatment-biomarker interactions can be modeled directly through a minor transformation of the outcome.
- In high-dimensions, the interaction coefficients of a linear model are estimated using penalized regression methods, like the LASSO.
- An “augmented” version of the methodology was developed, accounting for prognostic effects. Equivalent to LASSO regression with treatment-biomarker interactions.

Issues with Penalized Regression Methods

Unreliable biomarker selection

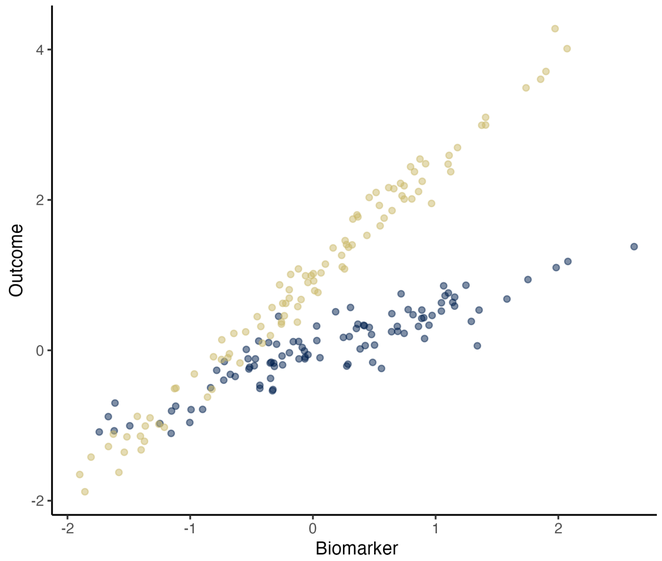
Strong assumptions: sparsity and correlation structure.

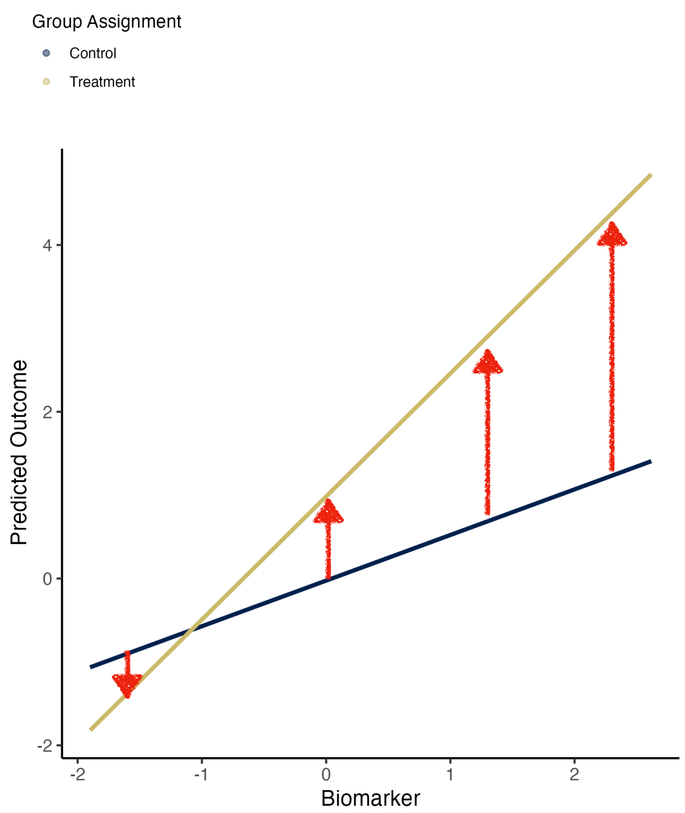
Violations produce to high false positive rates, leading to:

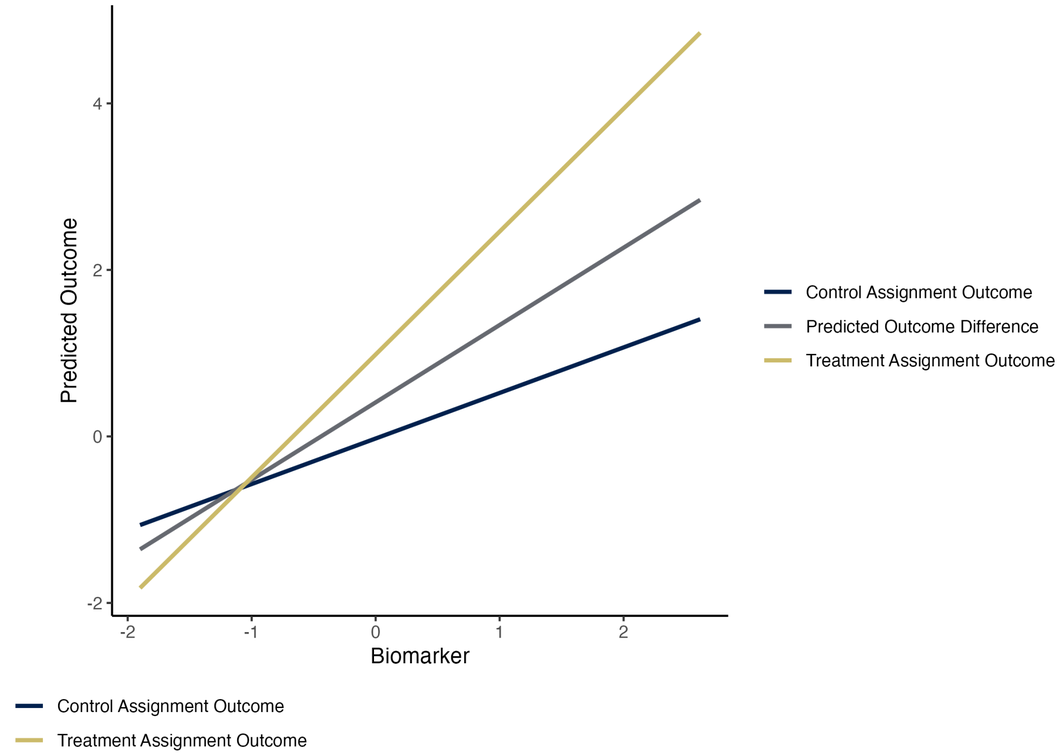
- Resources wasted on follow-up experiments and trials
- Decreased signal to noise ratio in diagnostic assays

We need to consider alternative problem formulations.

A Dedicated Variable Importance Parameter







uniCATE

Assumption-lean estimator of biomarker predictiveness

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

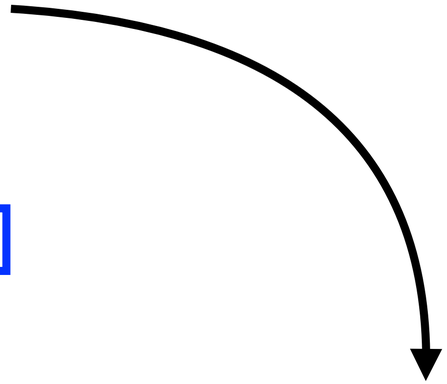
$$\frac{\frac{1}{n}\sum_{i=1}^{n}(\text{predicted outcome difference}^{\star})_i(\text{biomarker})_i}{\frac{1}{n}\sum_{i=1}^{n}(\text{biomarker})_i^2}$$

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

uniCATE in Action

uniCATE ranks biomarkers based on predictiveness

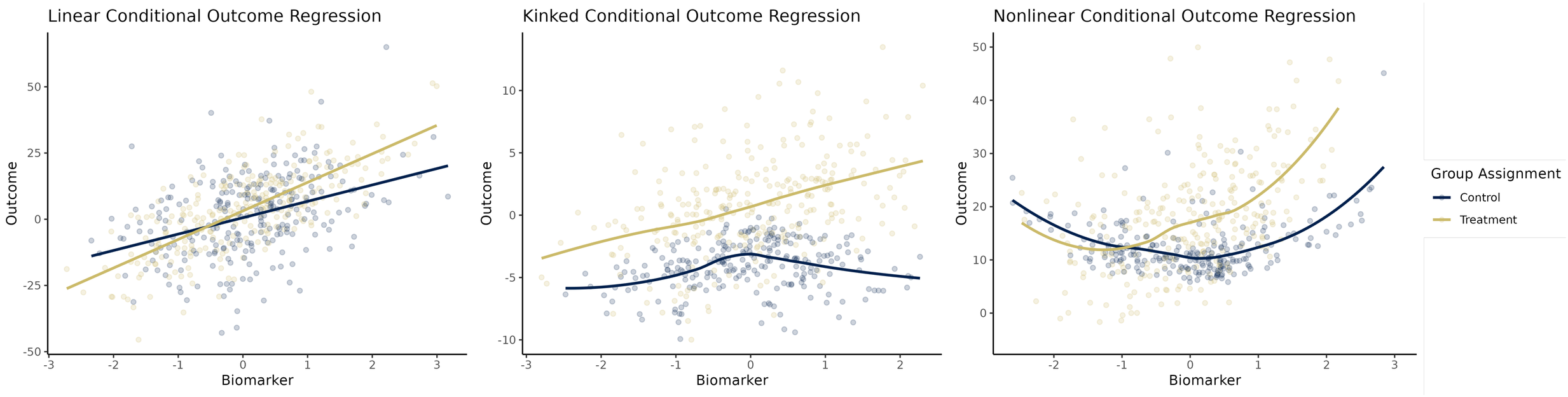
```
n ← 200
biomarker_1 ← rnorm(n, mean = 0, sd = 1)
biomarker_2 ← rnorm(n, mean = 0, sd = 1)
biomarker_3 ← rnorm(n, mean = 0, sd = 1)
biomarker_4 ← rnorm(n, mean = 0, sd = 1)
covariate ← rbinom(n, 1, 0.4)
treatment ← rbinom(n, 1, 0.5)
response ← covar + 1 * biomarker_1 * treatment
              + 2 * biomarker_2 * treatment
```



	Est.	SE	Z-score	P	P (BH)
biomarker_2	1.90	0.0768	24.7	3.58E-13	1.43E-13
biomarker_1	0.820	0.129	6.38	1.75E-10	3.51E-10
biomarker_3	0.0482	0.145	0.333	7.39E-01	9.03E-01
biomarker_4	-0.0202	0.166	-0.122	9.03E-01	9.03E-01

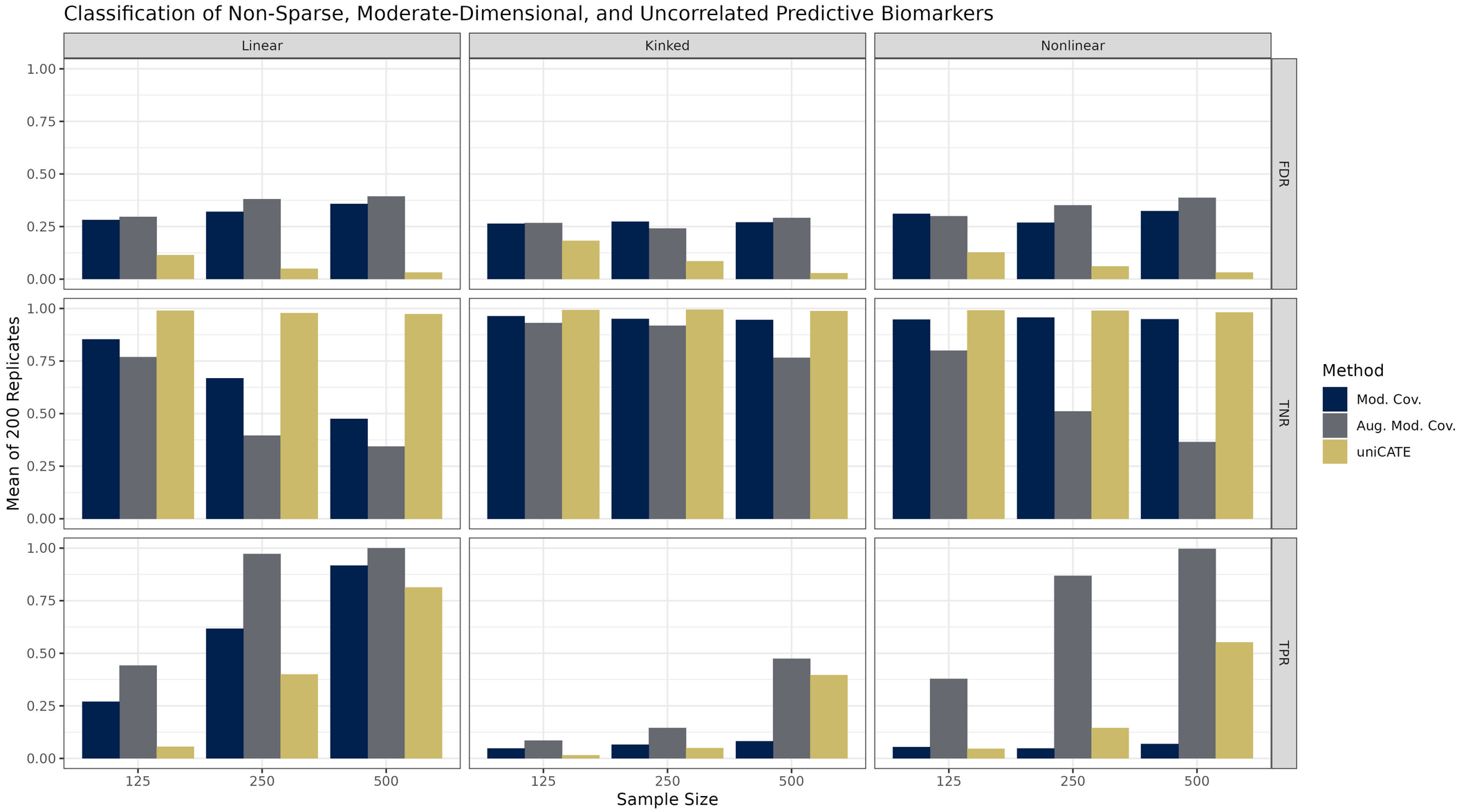
Simulation Studies

Considered Biomarker-Outcome Relationships

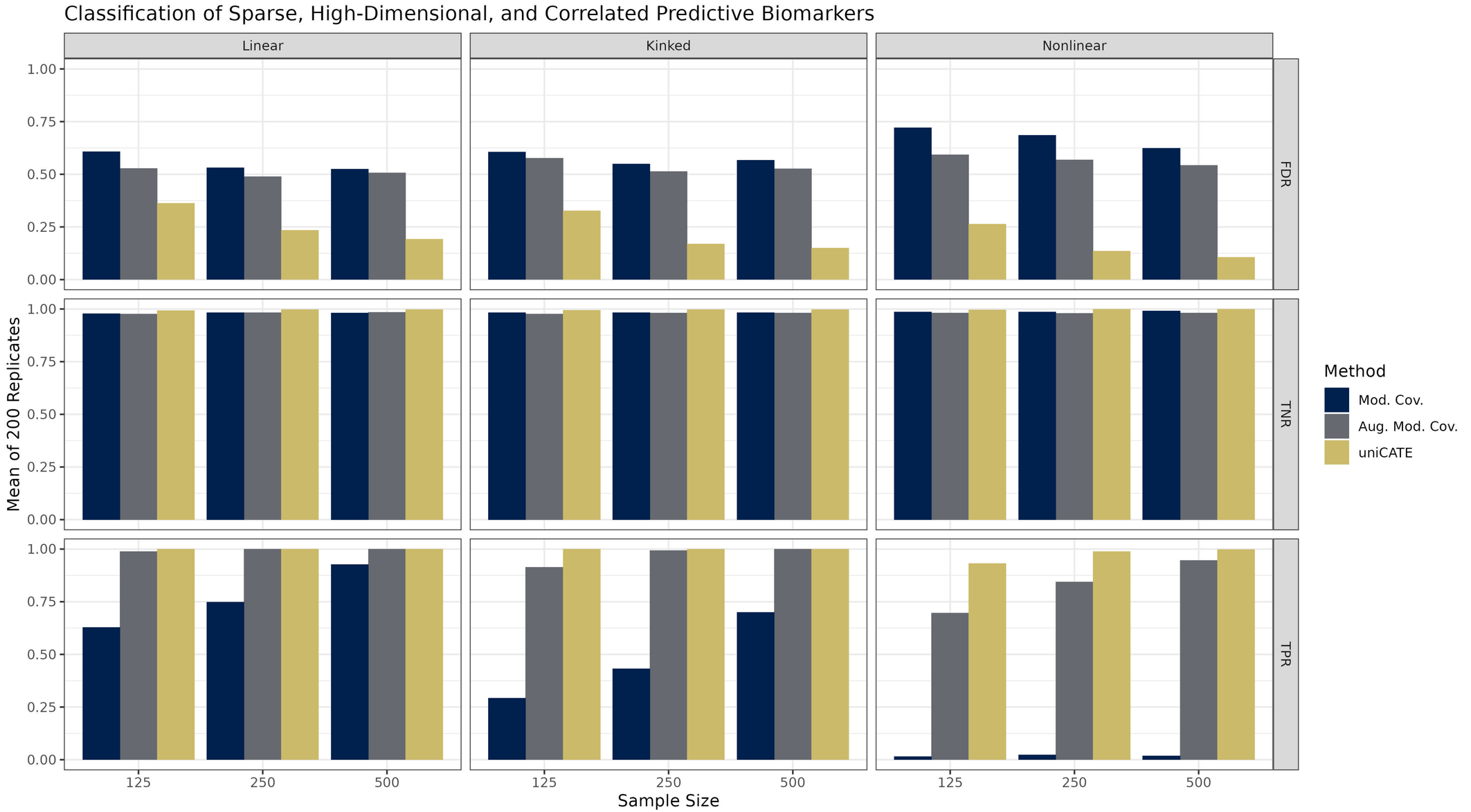


Difficulty

uniCATE Controls False Positive Rates



uniCATE *Still* Controls False Positive Rates



Application to IMmotion 150/151

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.

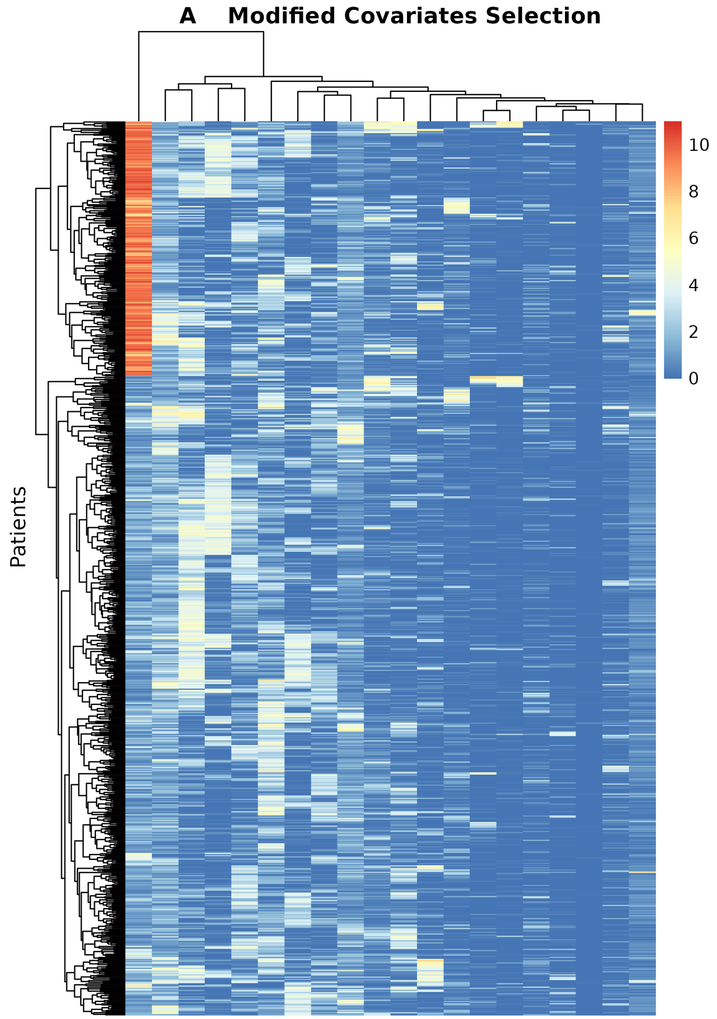
2. Selected the 500 most variable, log-transformed genes as biomarkers.

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



Conclusion

- 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

Questions?

pboileau.ca — philippe_boileau@berkeley.edu