

A benchmark of machine learning feature selection methods for stable identification of biomarkers of resistance to immunotherapy in lung cancer

Anastasiia BAKHMACH^{1,4,6}, Florence MONVILLE², Amélie POUCHIN^{3,5,6}, Marie ROUMIEUX⁴, Maryannick LE RAY⁵, Frédéric VÉLY^{5,6}, Jean Philippe DALES^{5,6}, Florence SABATIER^{5,6}, Mohamed BOUSSENA^{1,4,6}, Mélanie KARLSEN^{1,4,6,*}, Andrea VAGLIO^{1,4,6}, Vanina LECA², Richard MALKOUN⁵, Jacques FIESCHI², Éric VIVIER⁷, Joseph CICCOLINI^{1,4,6}, Laurent GREILLIER^{1,4,5,6}, Fabrice BARLESI⁸, Sébastien BENZEKRY^{1,4,6}

¹ COMPO (COMputational pharmacology and clinical Oncology), Centre Inria d'Université Côte d'Azur, Marseille, France
² Veracyte SAS, Marseille, France
³ Centre d'Immunologie de Marseille-Luminy (CIML), Marseille, France
⁴ Centre de Recherches en Cancerologie de Marseille (CRCM), Inserm U1068, CNRS UMR7258, Institut Paoli-Calmettes, Marseille, France
⁵ Assistance Publique-Hôpitaux de Marseille (APHM), Marseille, France
⁶ Aix Marseille Université, Marseille, France
⁷ Innate Pharma, Marseille, France
⁸ Gustave Roussy, Villejuif, France
^{*} now at Sanofi R&D Montpellier, TMU/PMxDev; LIRMM - Université de Montpellier - CNRS, France



BACKGROUND

- Feature selection (FS) is essential for biomarker discovery.
- High-dimensional feature space, limited sample size, multicollinearity, and missing values make FS non-trivial in biomedical data.
- Stability, i.e. the similarity between feature sets obtained under data perturbations, is a key challenge in FS.** Despite being a long-recognized problem, it is rarely evaluated in current biomedical studies.

OBJECTIVES

- Conduct a **comprehensive benchmark of FS methods**, evaluating both stability and predictive performance on multimodal data integrating clinical variables and complex biomarkers.
- Identify an **optimal FS method** that balances stability and performance **tailored to prediction of primary resistance to anti-PD-(L)1 in non-small cell lung cancer (NSCLC).**

METHODS

DATA

- 435 lung cancer patients from the PIONeer biomarkers clinical study (NCT03493581) [1].
- 374 variables, including clinical and demographic data (p = 10), routine blood tests (p = 49), tumor multiplex immunohistochemistry markers (p = 159), and circulating immune and vasculophenotyping markers (p = 156).
- The outcome to predict was primary resistance to immunotherapy defined as disease progression within 6 months of treatment.

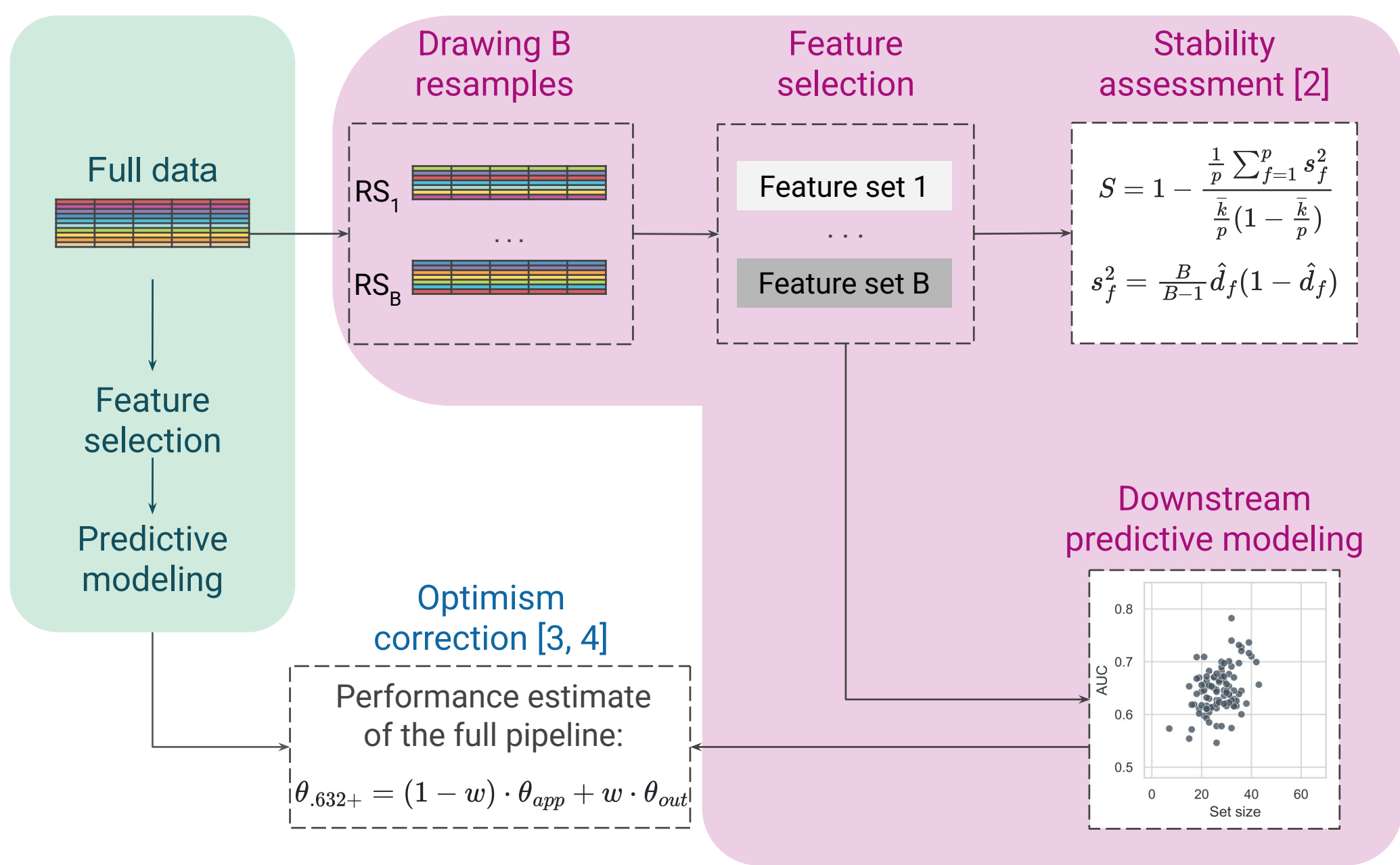
BENCHMARK CONFIGURATION

- 31 FS methods**
 - 16 filters
 - 10 embedded methods
 - 4 wrappers
 - 1 ensemble method
- 11 classifiers**

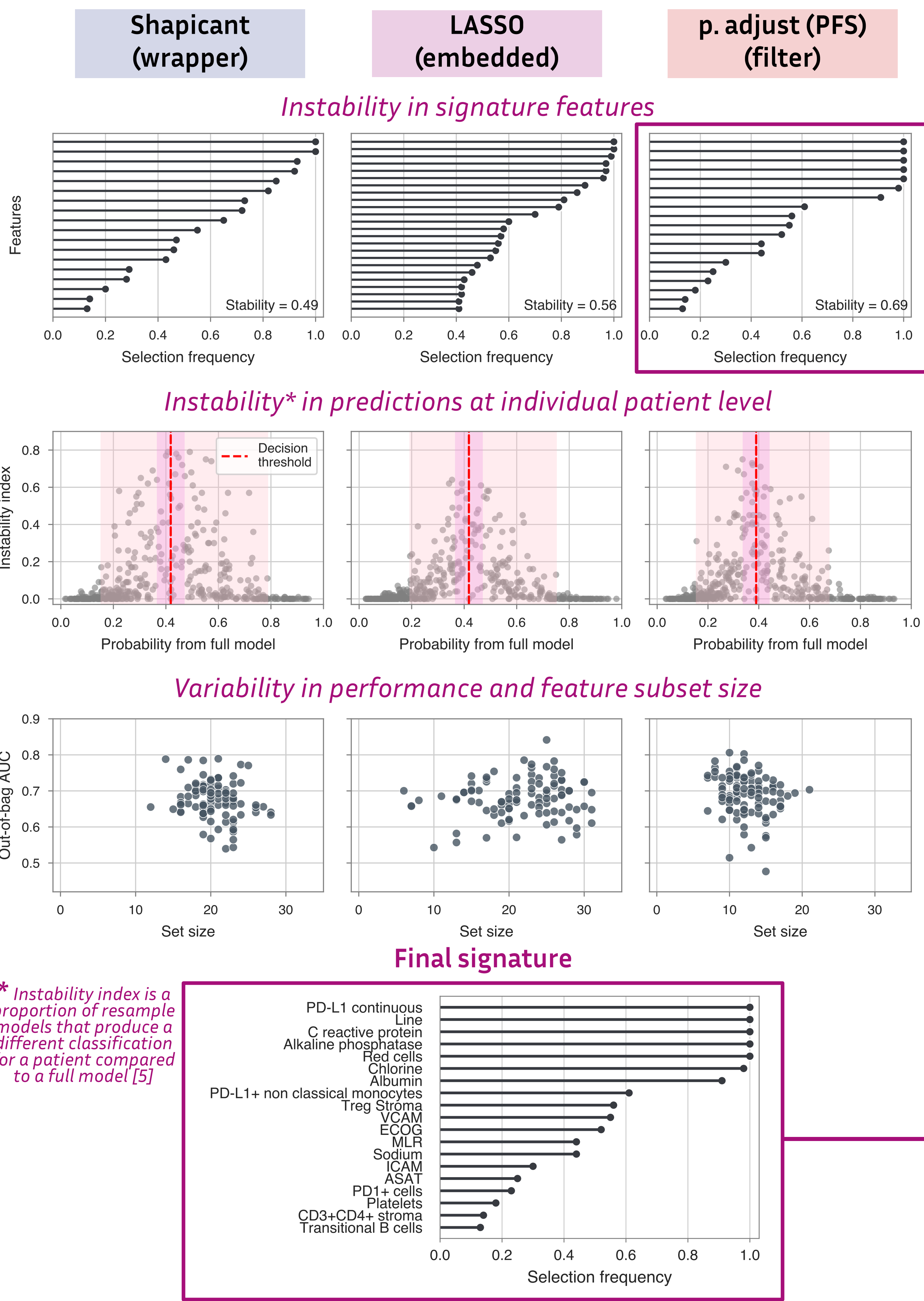
SIMULATION FRAMEWORK

A synthetic target was generated from the original data as a linear combination of the features from the final signature plus random noise.

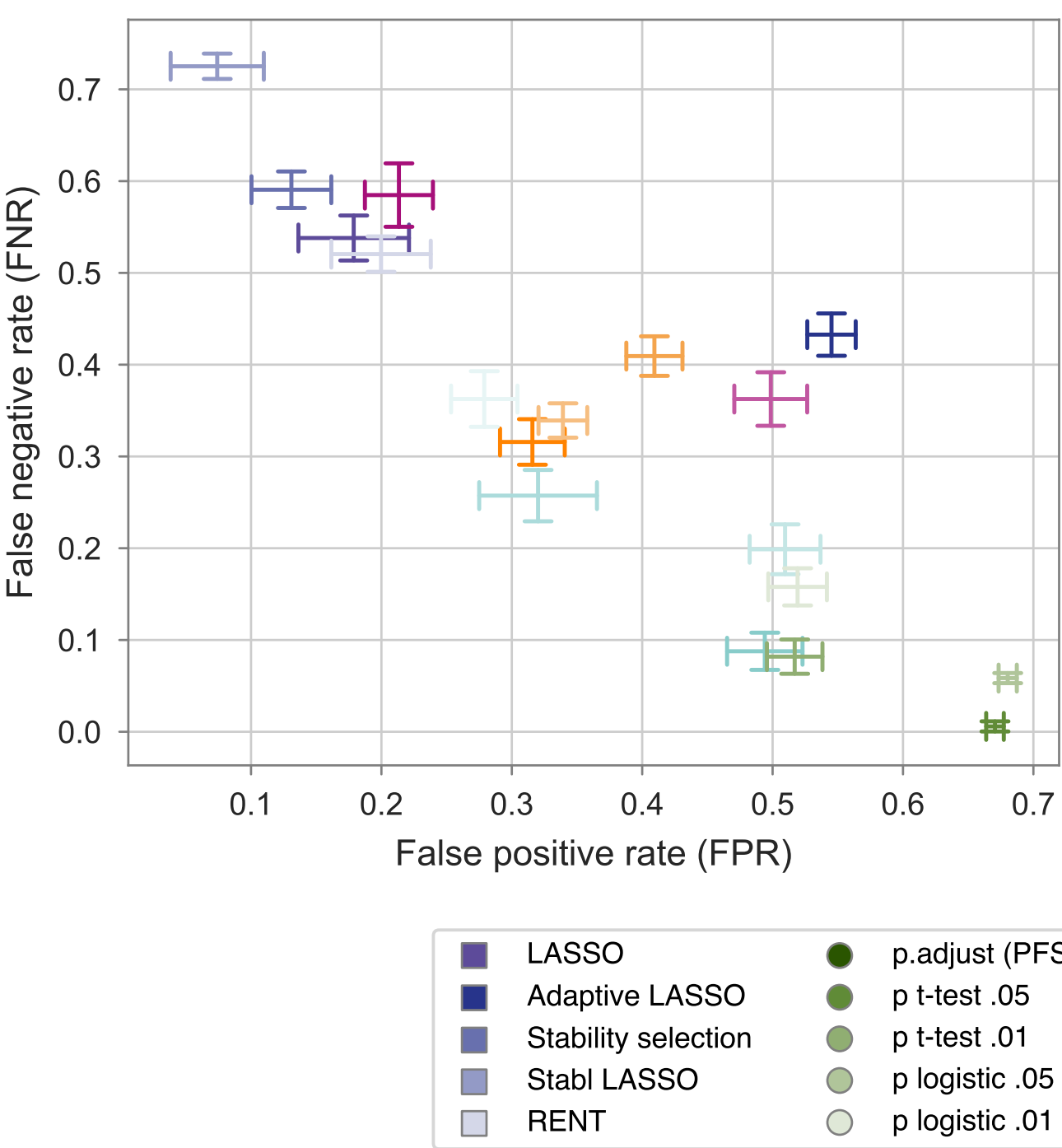
EVALUATION PIPELINE



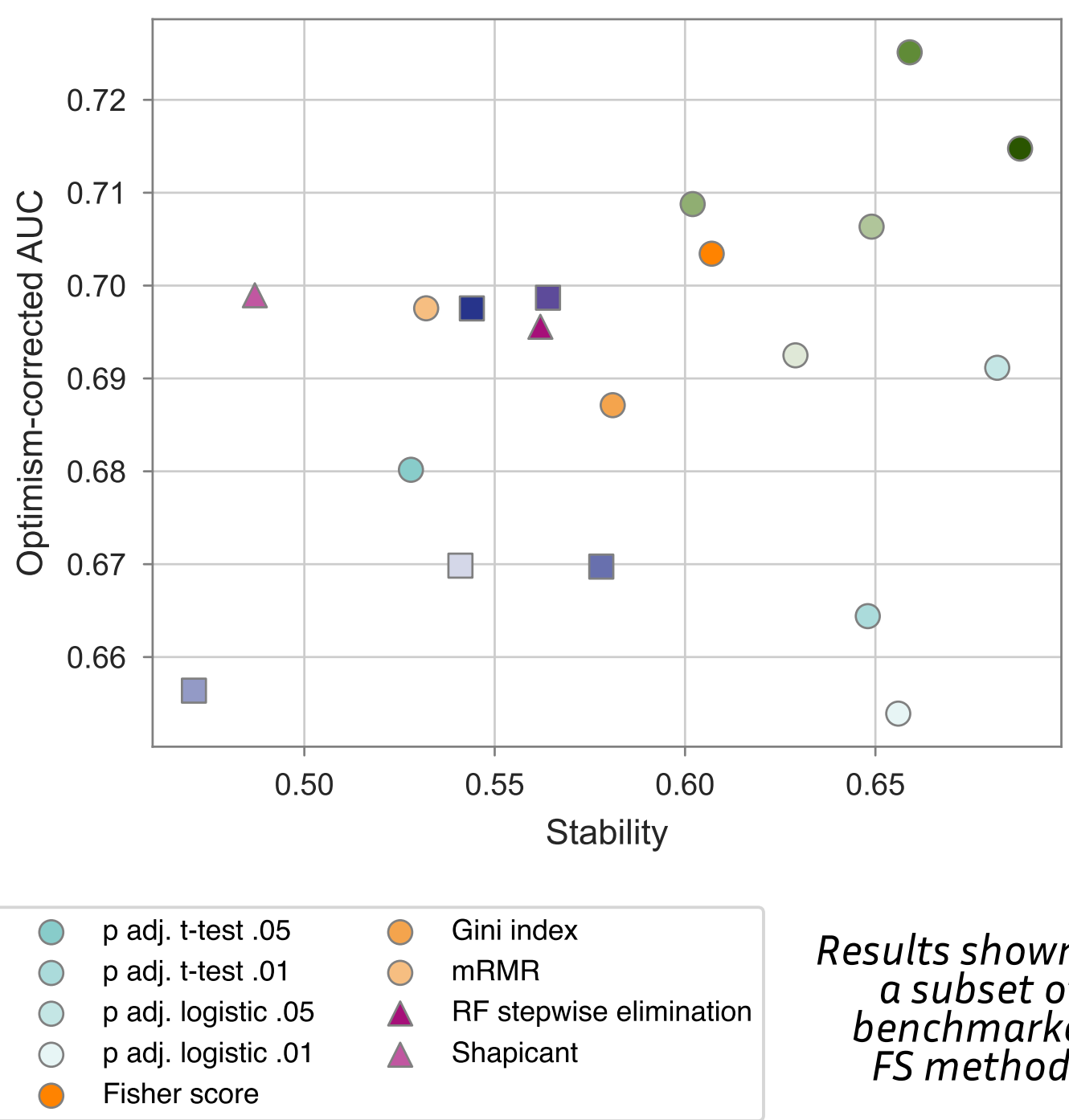
RESULTS



Simulation FPR-FNR trade-off



Performance on real data



Selection frequency of top features

Feature	# methods	p. adjust (PFS)	p. t-test	LASSO	Stability selection	Stabl LASSO	Fisher score	Shapicant
Treg Stroma	21	0.56	1	1	0.98	0.98	1	0.94
CRP	18	1	1	0.96	0.51	0.34	1	0.64
Albumin	17	0.91	1	0.93	0.46	0.43	0.9	0.69
Line	15	1	1	0	0	0.06	1	1
Chlorine	15	0.98	1	0.81	0.18	0.22	0.94	0.98
LDH	15	0.01	0.98	0.94	0.39	0.73	0.74	0.43
PD-L1+ cell dens. in tumor	14	0.09	1	0.94	0.42	0.51	0.76	0.43
Red cells	13	1	0.99	0.67	0.05	0.1	0.79	0.78
VCAM	12	0.55	1	0.9	0.26	0.54	0.92	0.23
PD-L1+ non-classical monocytes	10	0.61	1	0.94	0.26	0.44	0.46	0.01

CONCLUSIONS AND PERSPECTIVES

- Evaluating FS stability and comparing multiple FS methods is crucial to improve the reliability of biomarker discovery related to drug response and resistance.
- Even in the p ~ n case reaching FS stability is a challenge.
- Future perspectives include the study of FS stability in:
 - covariate selection in nonlinear mixed-effects modeling;
 - and higher dimensional settings (p >> n) typical in, e.g., omics data.

References

- Barlesi, F., et al., Benzekry, S., *Annals of Oncology*, 16 (suppl. 1), 2022.
- Nogueira, et al. (2018). *Journal of Machine Learning Research*, 18(174), 1–54.
- Efron, B., & Tibshirani, R. (1997). Improvements on Cross-Validation: The .632+ Bootstrap Method. *JASA*, 92(438), 548–560.
- Collins, G. S. et al. (2024). Evaluation of clinical prediction models (part 1): From development to external validation. *BMJ*, 384.
- Riley, R. D., & Collins, G. S. (2023). *Biometrical Journal*, 65(8), 2200302.