



**HAL**  
open science

# Machine Learning to Predict Survival and Resistance to Immune Checkpoint Inhibition

Sébastien Benzekry

► **To cite this version:**

Sébastien Benzekry. Machine Learning to Predict Survival and Resistance to Immune Checkpoint Inhibition. Twelfth International Workshop on Pharmacodynamics of Anticancer Agents, Sep 2024, Ponta Delgada (Açores), Portugal. ⟨hal-04990492⟩

**HAL Id: hal-04990492**

**<https://inria.hal.science/hal-04990492v1>**

Submitted on 13 Mar 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



# Machine Learning to Predict Survival and Resistance to Immune Checkpoint Inhibition

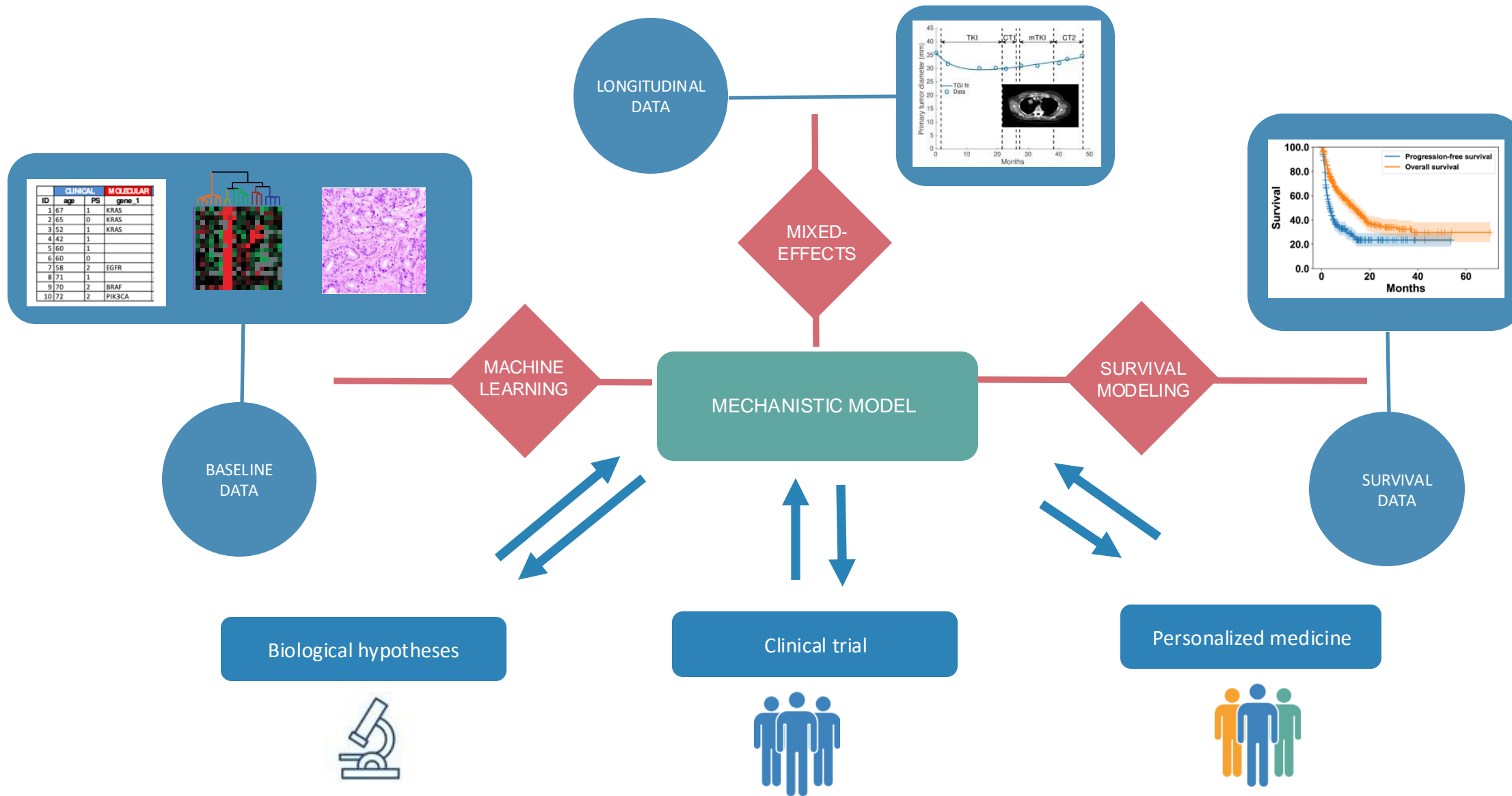
S. Benzekry

*Twelfth International Workshop on Pharmacodynamics of Anticancer Agents*

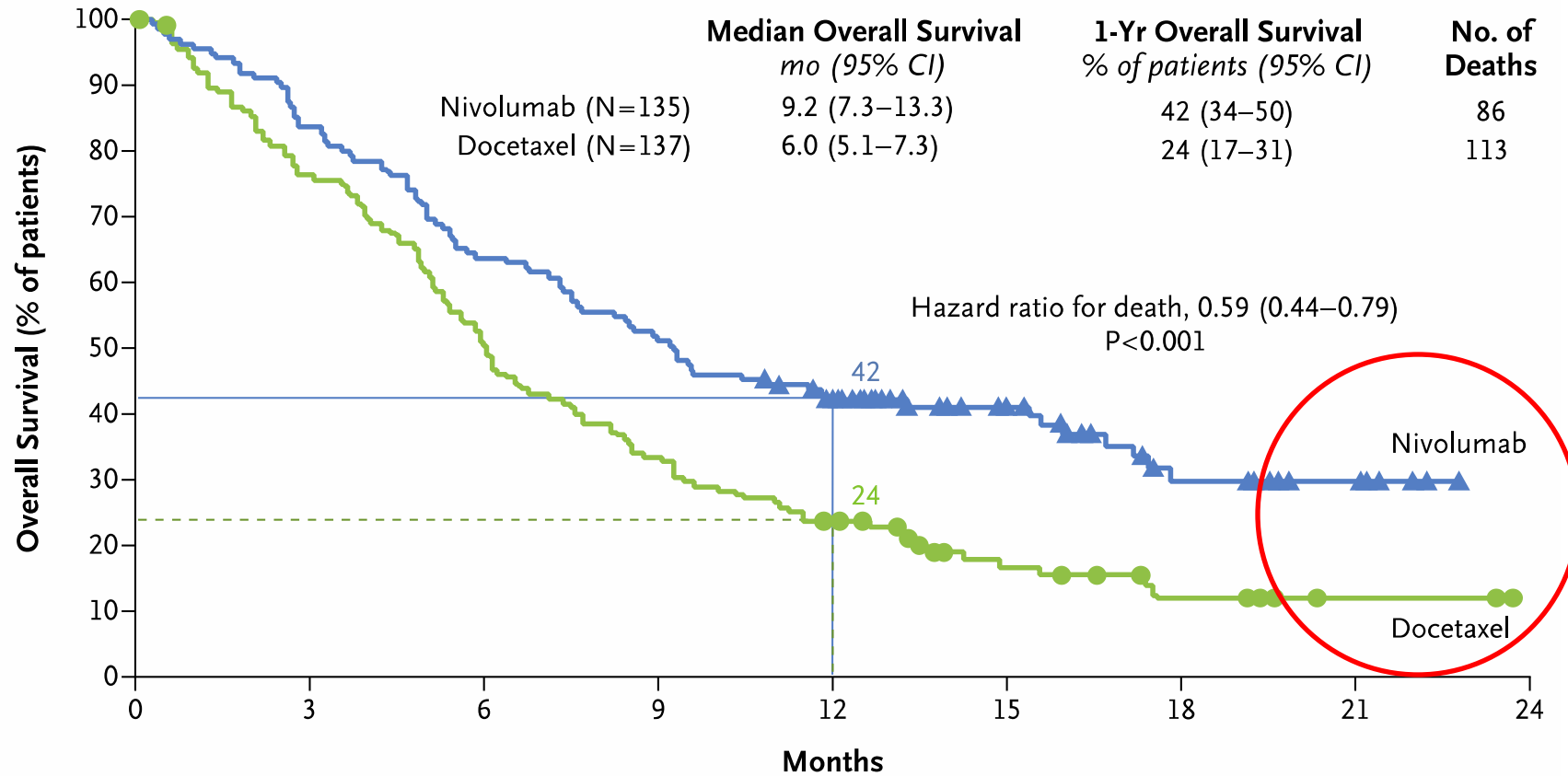
Ponta Delgada, September 2024



# Mechanistic learning



# Long term responders / survivors to anti-PD(L)1 in advanced NSCLC

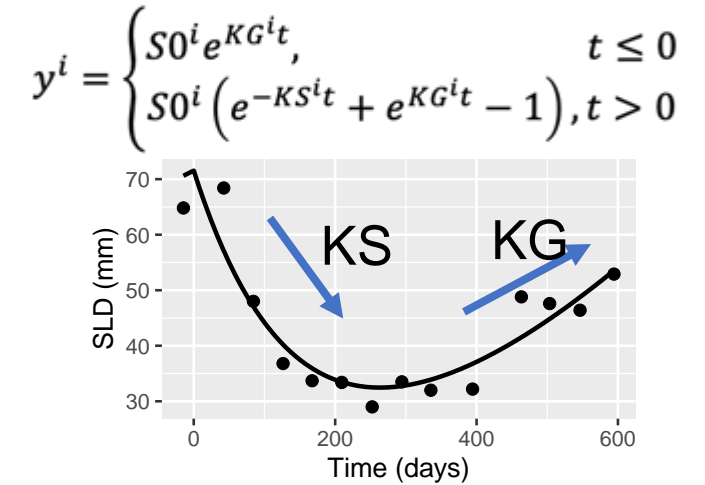


**No. at Risk**

	0	3	6	9	12	15	18	21	24
Nivolumab	135	113	86	69	52	31	15	7	0
Docetaxel	137	103	68	45	30	14	7	2	0

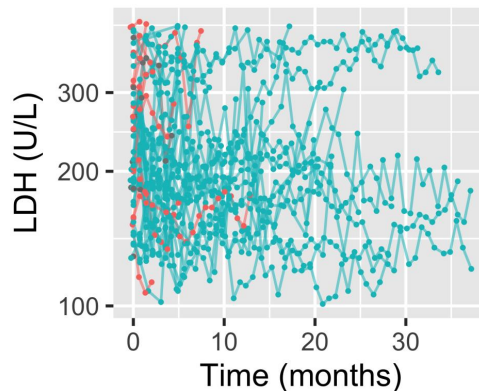
# Current state of the art

- **Baseline**
  - Response: PDL1 (AUC = 0.601<sup>1</sup>), TMB (AUC = 0.646<sup>1</sup>)
  - ROPRO score for survival (27 basic clinical variables) : C-index = 0.69<sup>2</sup>
- **On treatment**
  - **Tumor kinetics** : KG has important predictive power of OS hazard ratio (HR)<sup>3</sup>



## Objectives = predict survival (phase 3 OS) and individual resistance

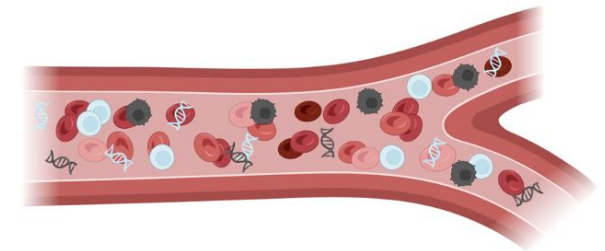
**Blood markers kinetics (BK)**



**Deep phenotyping**



**cfDNA fragmentomics**



<sup>1</sup> Rizvi, H. et al. J. Clin. Oncol. 2018.

<sup>2</sup> Becker, T. et al. An enhanced prognostic score for overall survival of patients with cancer derived from a large real-world cohort. Ann Oncol 31, 1561–1568 (2020).

<sup>3</sup> Claret, L. et al. A Model of Overall Survival Predicts Treatment Outcomes with Atezolizumab versus Chemotherapy in Non-Small Cell Lung Cancer Based on Early Tumor Kinetics. Clin Cancer Res (2018).

## Raw baseline data

Clinical characteristics  
p = 73

Genomic and Transcriptomic  
p = 395 and 58,311

ML Feature selection

## Minimal signature

Clinical characteristics  
p = 11

Genomic and Transcriptomic

## Raw longitudinal data

Tumor Kinetics (TK)  
k = 8,635

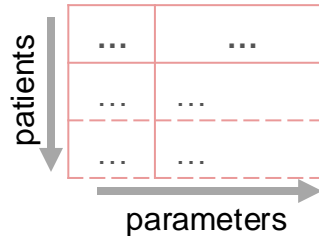
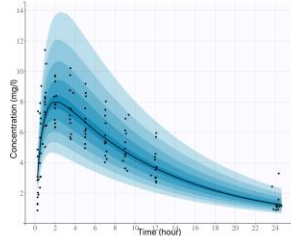
Blood markers Kinetics (BK)  
k = 108,551  
Albumin, CRP, Neutrophils, LDH

Mixed-effects

## Model-based individual parameters

TK  
p = 3

BK  
p = 12



Machine Learning

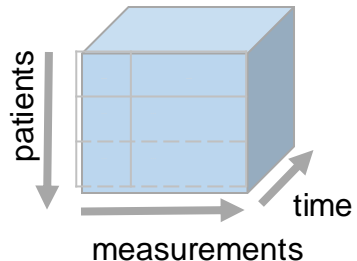
## Survival

- Individual prognosis
- C-index
  - Survival

*Individual digital twin*

- Clinical trial
- Survival curve
  - Hazard ratio

*Trial digital arm*



# Data for model development and external validation

- **Baseline data**

- Patients' and disease characteristics  
**p = 73 parameters**
- Transcriptomic and mutational data  
**p = 58,311 and 395**

- **Longitudinal data**

- Tumor kinetics (TK, SLD)  
**5,570/3,065** observations
- 4 BK markers (Albumin, CRP, LDH and Neutrophils)  
**61,296/47,255** observations

Study	Description	N
<b>FIR GO28625</b>	<b>Phase 2</b> study for the efficacy and safety of ATZ in advanced NSCLC	133
<b>POPLAR GO28753</b>	<b>Phase 2</b> randomised controlled trial of ATZ versus docetaxel in NSCLC	134
<b>BIRCH GO28754</b>	<b>Phase 2</b> study of ATZ in advanced or metastatic NSCLC	595
<b>Train</b>		<b>862</b>
<b>Test - OAK GO28915</b>	<b>Phase 3</b> RCT of ATZ versus docetaxel (DTX) in patients with previously treated NSCLC	<b>553 (ATZ) 521 (DTX)</b>
<b>Train + Test</b>		<b>1936</b>

Train

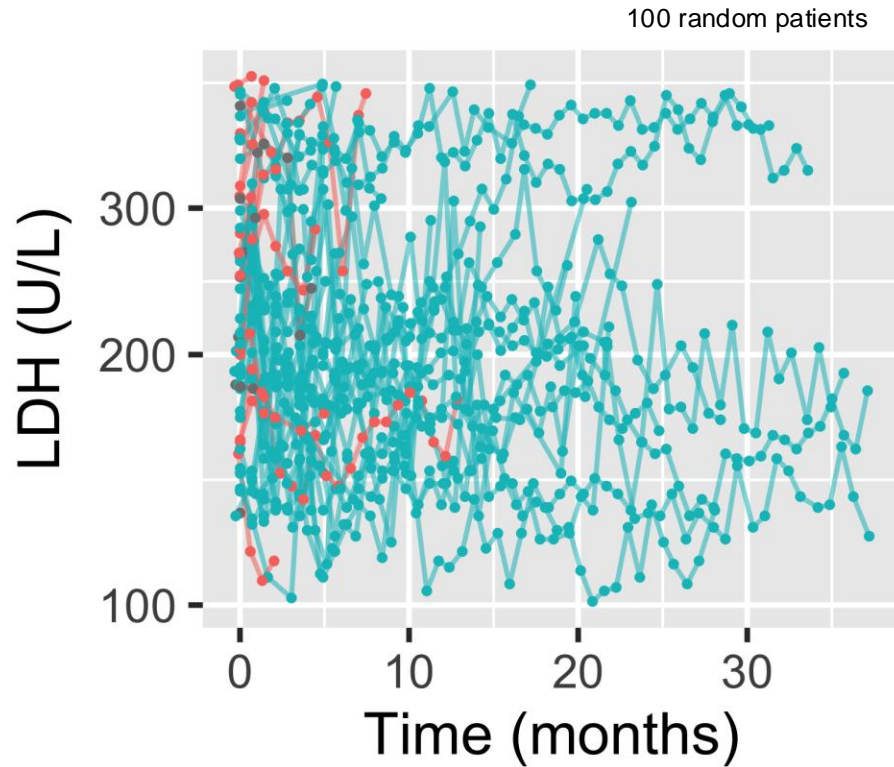
Test

NSCLC: Non-Small Cell Lung Cancer; p = number of parameters, N: number of patients treated with atezolizumab (patients from French centers were excluded for legal reasons (N=118); In total, data from 1074 patients from OAK were used as Test set (553 from the ATZ arm, 521 from the DTX arm); PD: Pharmacodynamic; SLD: Sum of the Largest Diameters. CRP: C Reactive Protein; LDH: Lactate Dehydrogenase.

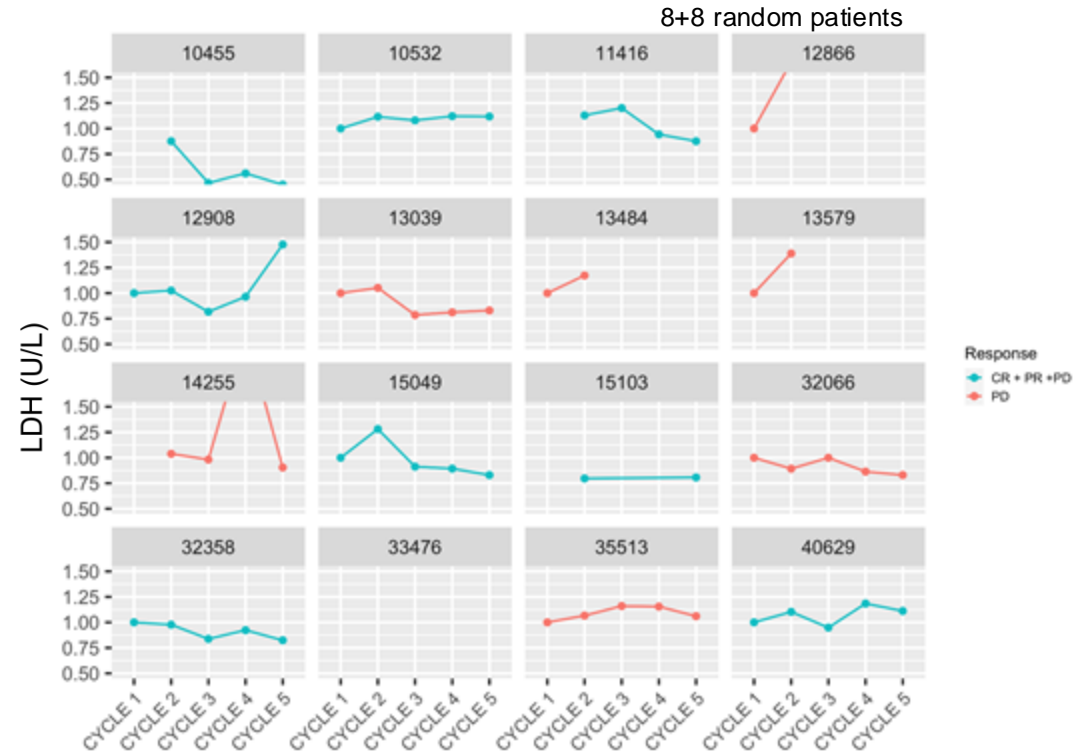
1. Fehrenbacher L *et al.* Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* (2016)
2. Fehrenbacher L *et al.* Updated Efficacy Analysis Including Secondary Population Results for OAK: A Randomized Phase III Study of Atezolizumab versus Docetaxel in Patients with Previously Treated Advanced Non-Small Cell Lung Cancer. *Journal of Thoracic Oncology* (2018)
3. Solange Peters *et al.* Phase II Trial of Atezolizumab As First-Line or Subsequent Therapy for Patients With Programmed Death-Ligand 1-Selected Advanced Non-Small-Cell Lung Cancer (BIRCH). *JCO* (2017)
4. Spigel D.R *et al.* FIR: Efficacy, Safety, and Biomarker Analysis of a Phase II Open-Label Study of Atezolizumab in PD-L1-Selected Patients With NSCLC. *Journal of Thoracic Oncology* (2018)

# Blood markers kinetics (BK)

## Spaghetti plot



## Early cycles (normalized)



**Is there any kinetic pattern ?**

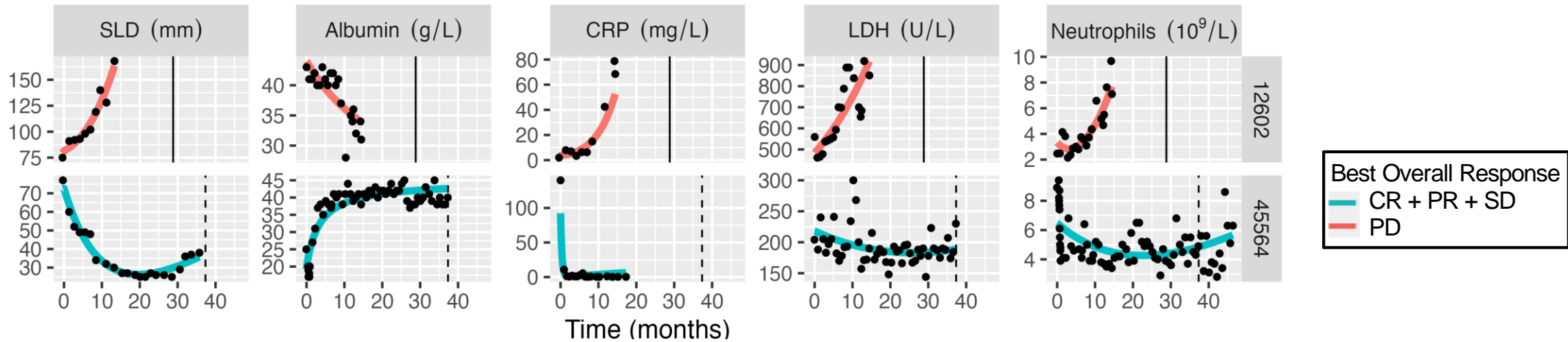
**Does it carry predictive value ?**

**Can we describe dynamics with a mechanistic model ?**

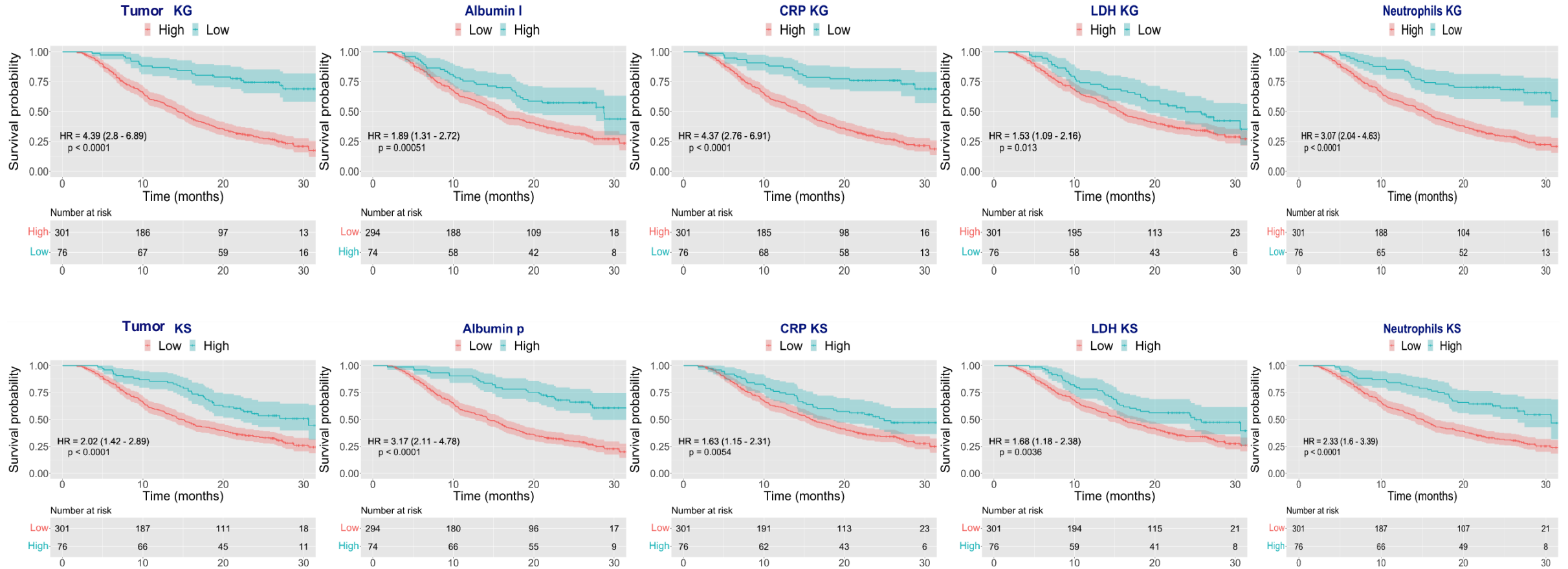
# Is there any kinetic pattern? Yes !

Corrected Bayesian Information Criterion ( $BICc \times 10^3$ ) for four empirical kinetic models of BK.  $b$  : standard deviation of the proportional error model

Model	Albumin		CRP		LDH		Neutrophils	
	BICc	b	BICc	b	BICc	b	BICc	b
<b>Double-exponential</b>	48.4	0.058	<b>28.8</b>	<b>0.21</b>	<b>39.9</b>	<b>0.56</b>	<b>102</b>	<b>0.14</b>
<b>Hyperbolic</b>	<b>48.0</b>	<b>0.056</b>	29.7	0.22	40.9	0.62	103	0.14
Linear	49.4	0.063	30.0	0.23	42.4	0.70	105	0.17
Constant	49.7	0.065	31	0.25	43.0	0.74	106	0.18

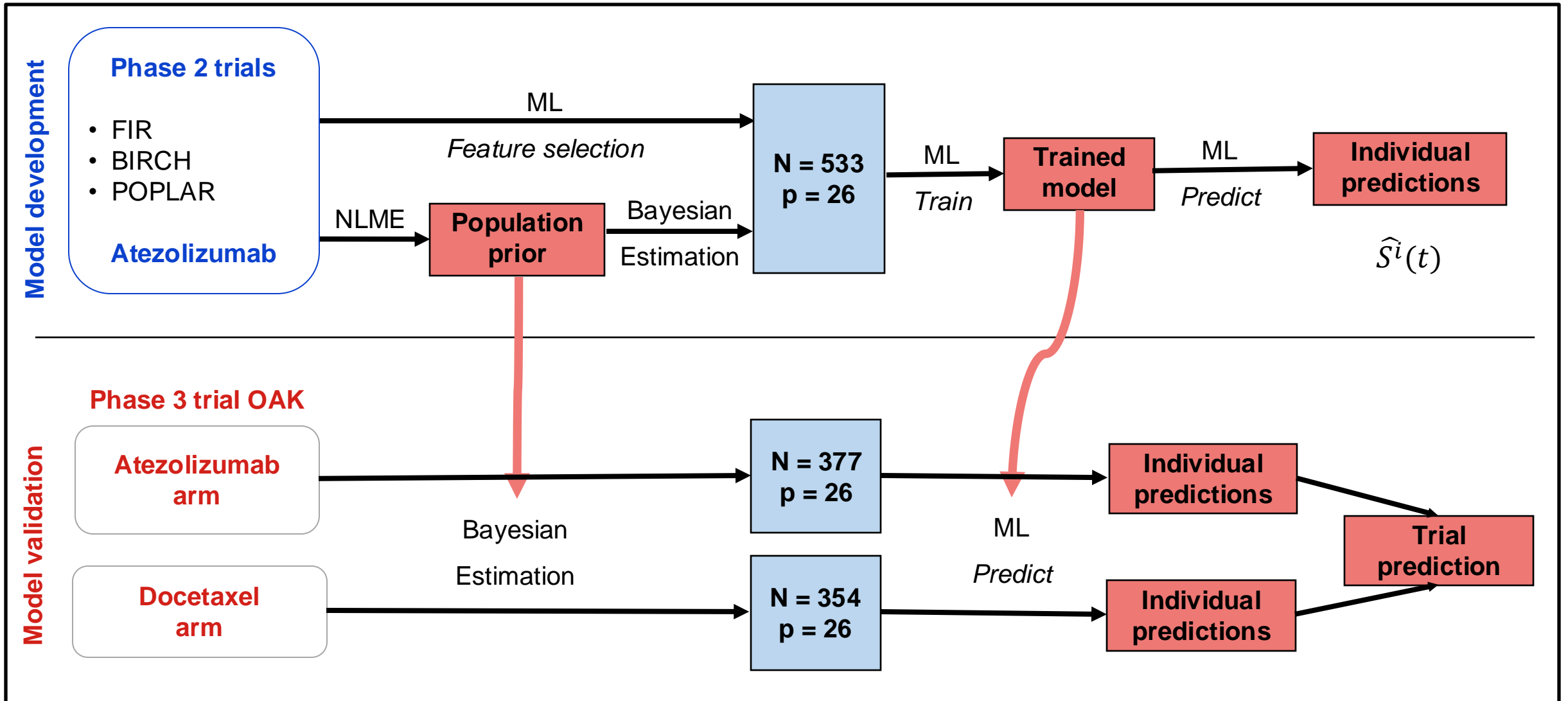


# Does it carry predictive value? Yes !



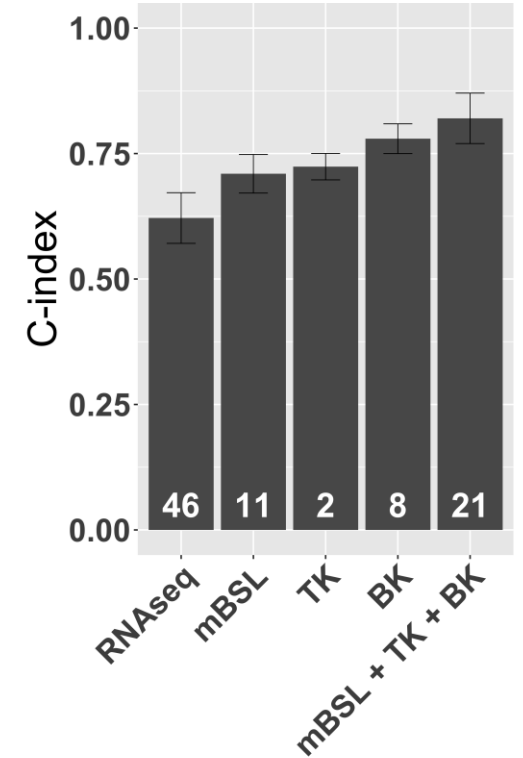
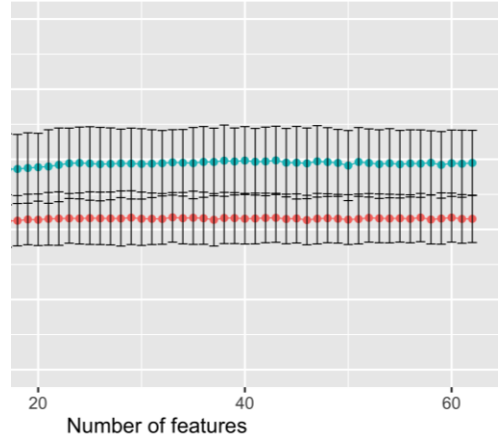
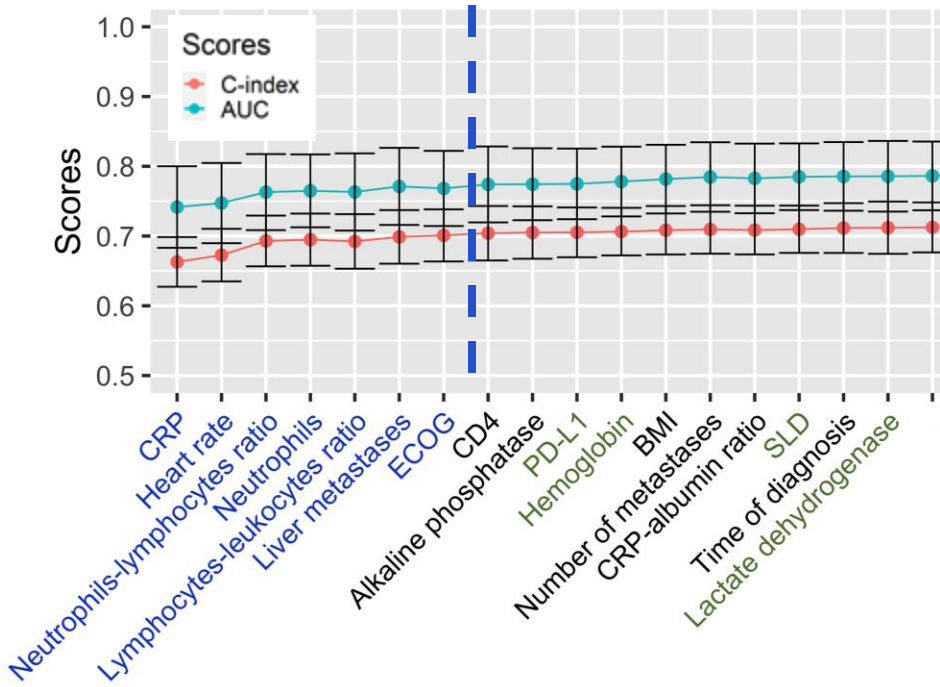
⇒ Individual TK and BK model parameters (without baseline) = **ML inputs**

# kML model development and validation



# Minimal baseline (mBSL) signature kinetics-ML (kML) model

Minimal baseline signature



- All features sorted using LASSO
- Incremental models with increasing number of features
- Minimal set of features that reaches the plateau

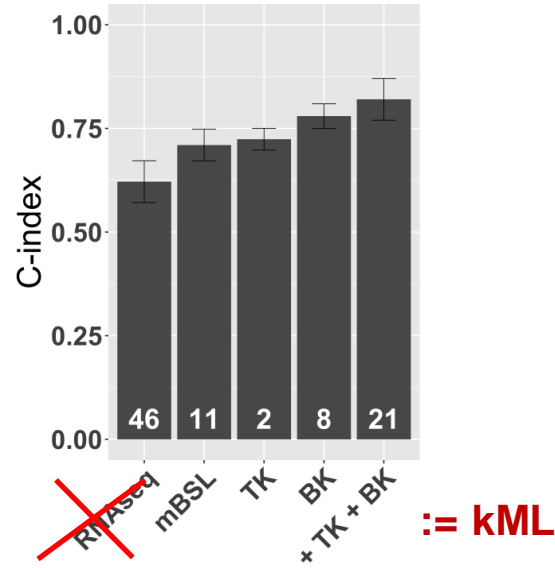
⇒ 11 mBSL variables  
⇒ easily clinically accessible

⇒ RNAseq does not carry predictive value  
⇒ discarded

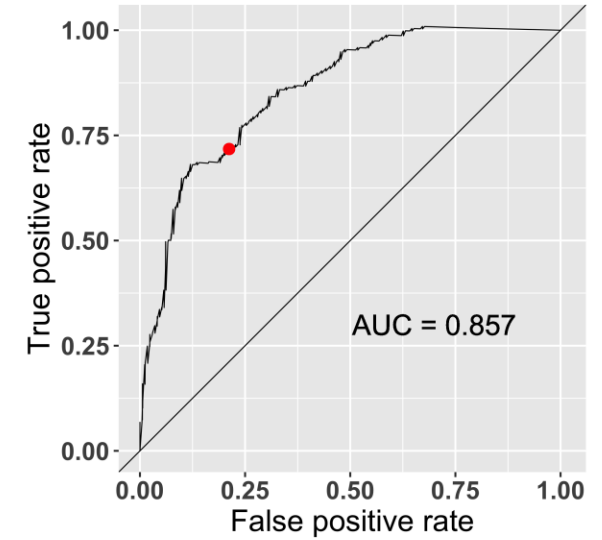
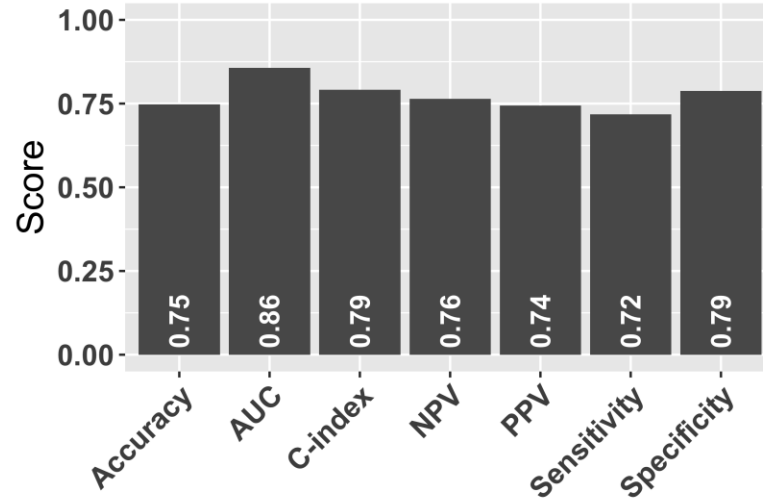
# kML individual predictions

mBSL = minimal baseline  
model = random survival forest

## Cross-val on train



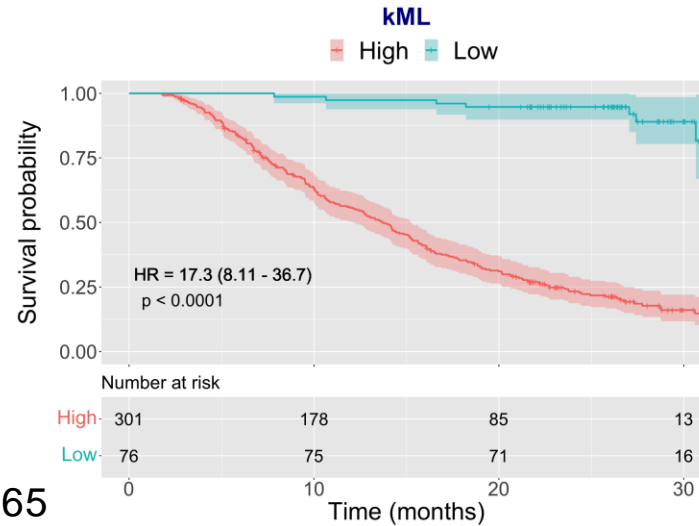
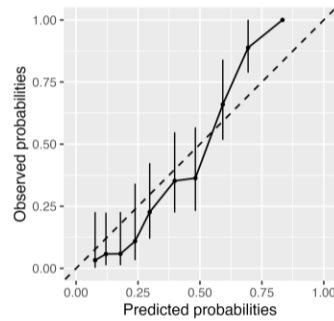
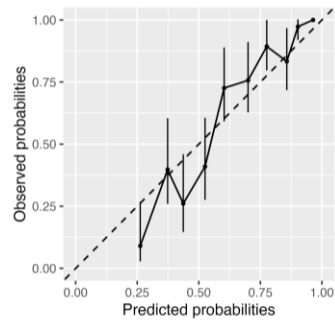
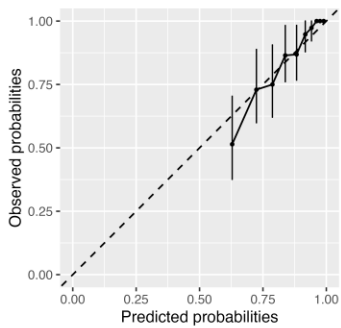
## Test set



6 months

12 months

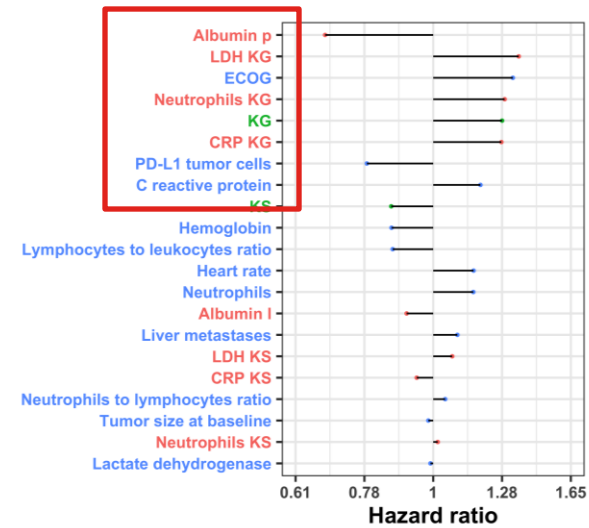
24 months



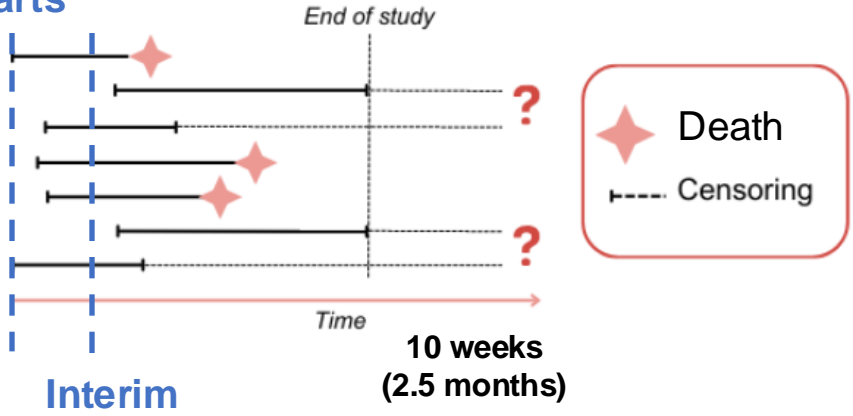
**c-index = 0.79**  
**AUC-12 months = 0.86**

PDL1 : AUC = 0.65

RoPro: c-index = 0.69 AUC-3 months = 0.817



# Study starts



# kML predicts phase 3 trial results from early data

Arm

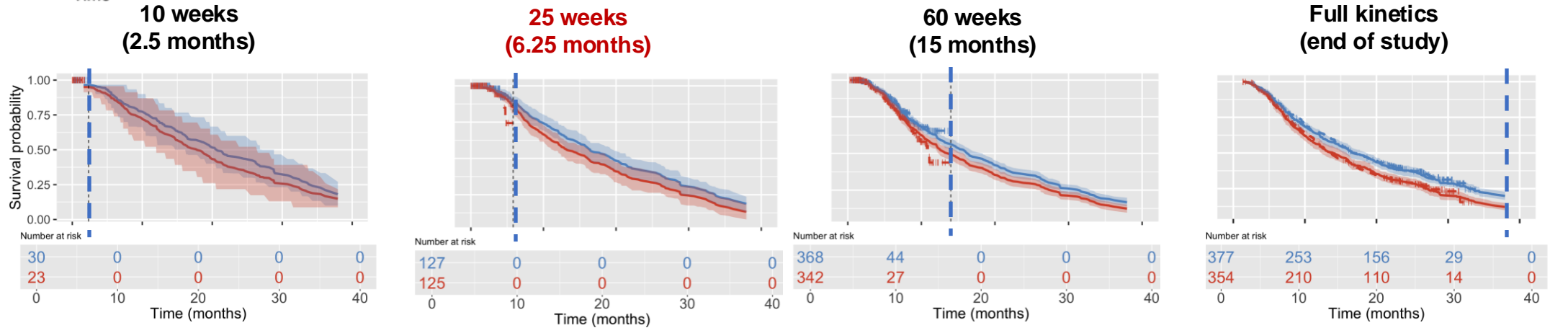
- Atezolizumab
- Docetaxel

Model prediction mean

Data

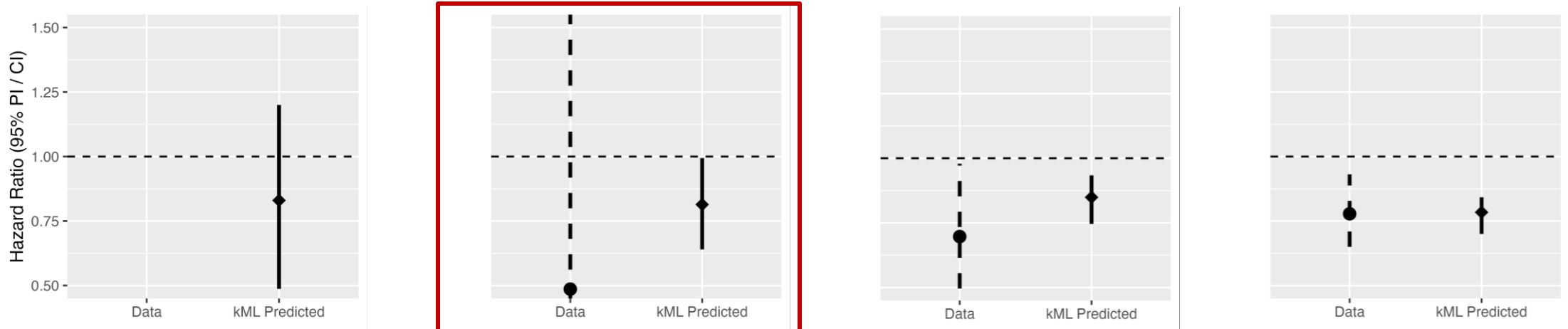
95% PI Atezolizumab

95% PI Docetaxel



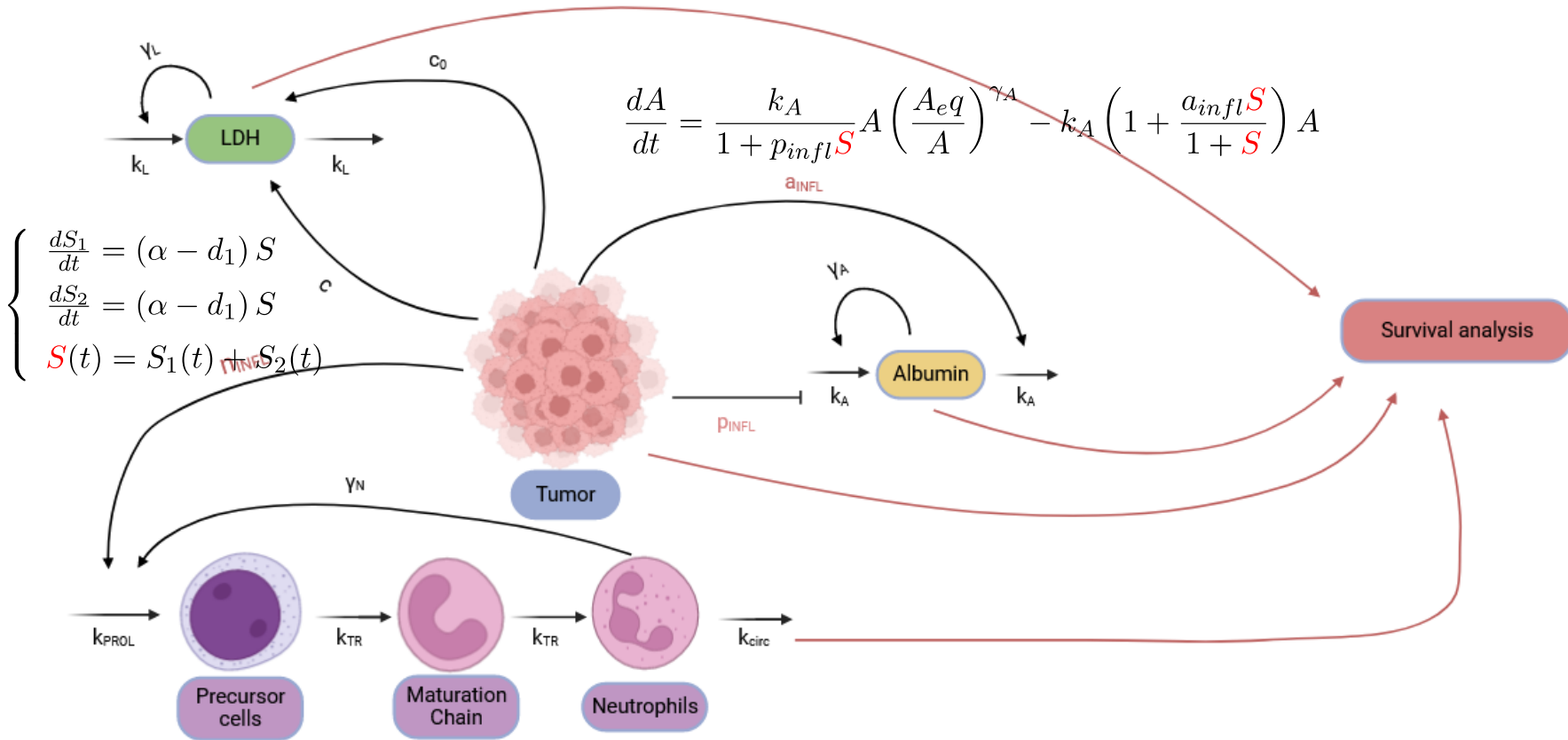
Hazard ratio

- Data
- kML Predicted



# Mechanistic model

$$\frac{dL}{dt} = k_L \left( \frac{L_{eq}}{L} \right)^{\gamma_L} + c(d_1 S_1 + d_2 S_2) + c_0 \mathbb{1}_{S'(t-r) > 0} S'(t) - k_L L$$



$$\frac{dA}{dt} = \frac{k_A}{1 + p_{infl} S} A \left( \frac{A_{eq}}{A} \right)^{\gamma_A} - k_A \left( 1 + \frac{a_{infl} S}{1 + S} \right) A$$

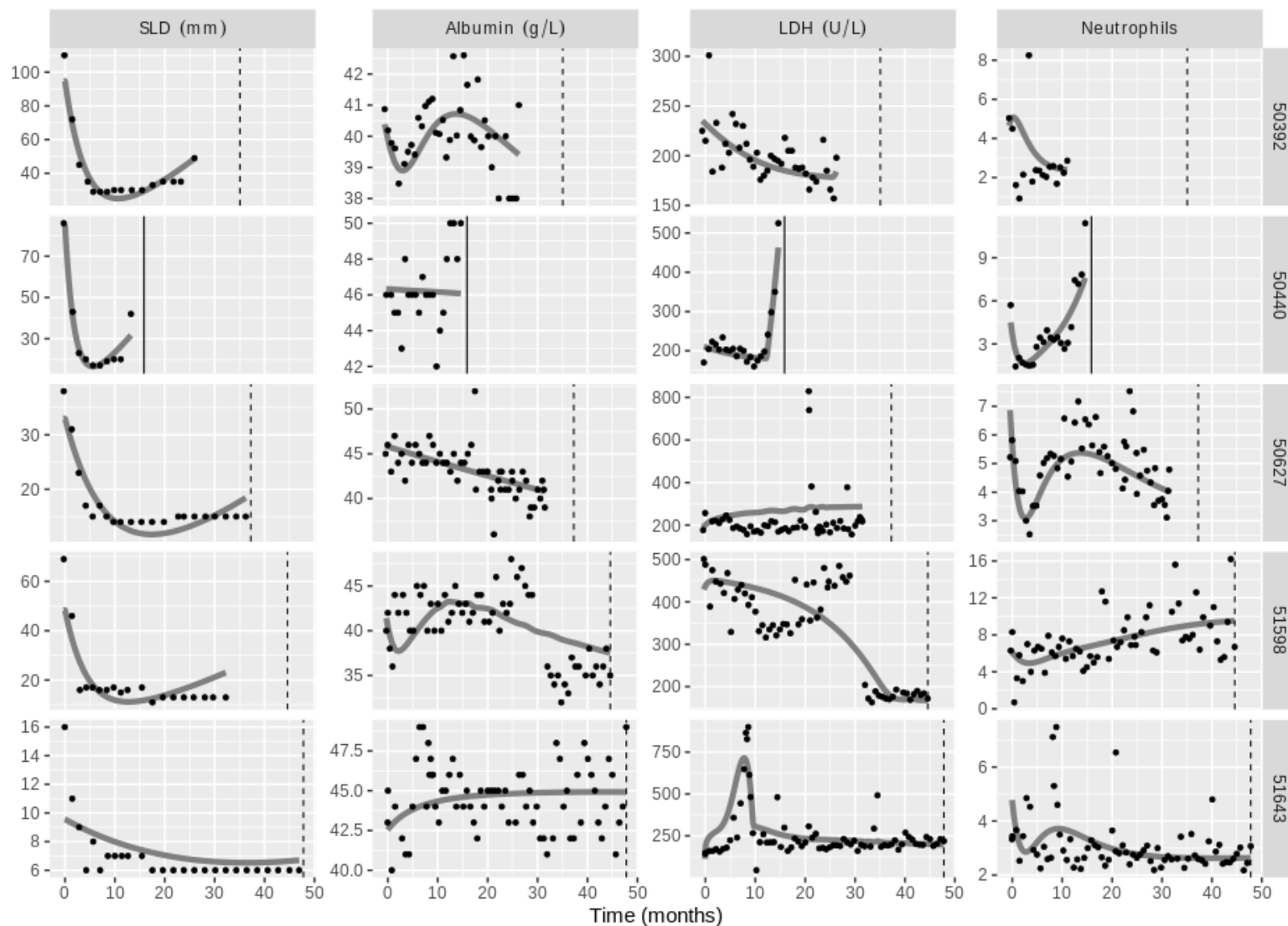
$$\begin{cases} \frac{dS_1}{dt} = (\alpha - d_1) S \\ \frac{dS_2}{dt} = (\alpha - d_1) S \\ S(t) = S_1(t) + S_2(t) \end{cases}$$

$$\begin{cases} \frac{dP}{dt} = k_{PROL} \left( 1 + n_{INFL} \frac{S}{1+S} \right) P \left( \frac{N_{eq}}{N} \right)^{\gamma_N} - k_{TR} P \\ \frac{dT_i}{dt} = k_{TR} T_{i-1} - k_{TR} T_i \\ \frac{dN}{dt} = k_{TR} T_3 - k_{CIRC} N \end{cases}$$

**Data:** clinical trials for atezolizumab (ATZ) :

- 3 phase 2 **monotherapy** **862 patients**  
44,911 data points
- 1 phase 3 **combinotherapy** ATZ, Bevacizumab and Carboplatin + Paclitaxel **1115 patients**  
67,507 data points

# Individual fits (Combotherapy)



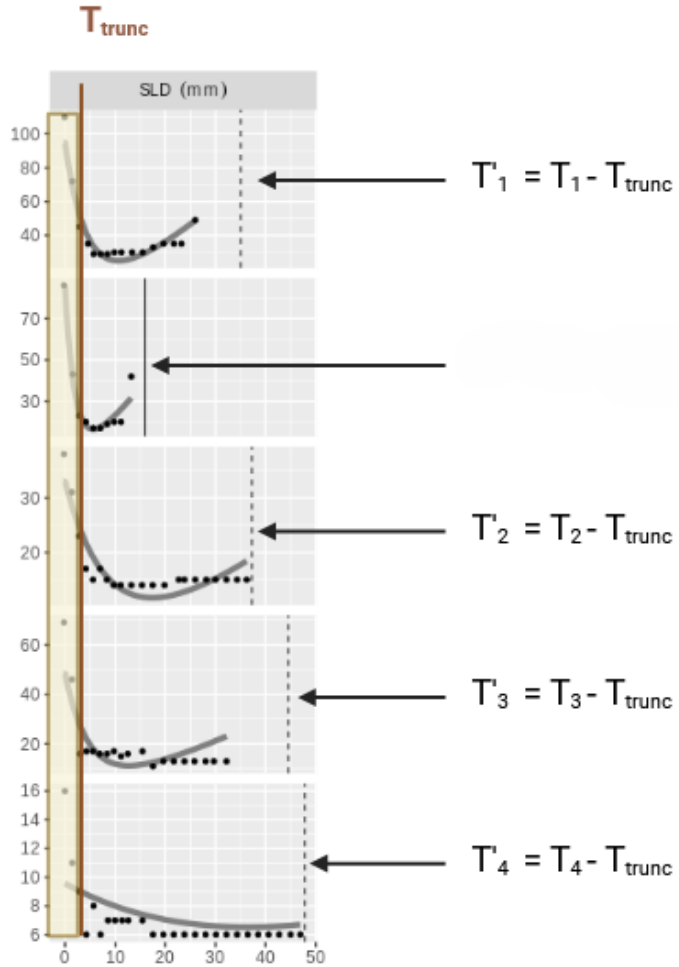
- **Fits heterogenous BK profiles**

- Albumin early drops
- LDH early peaks
- Late sharp increase of LDH (death shortly after)

- **Identifiability**

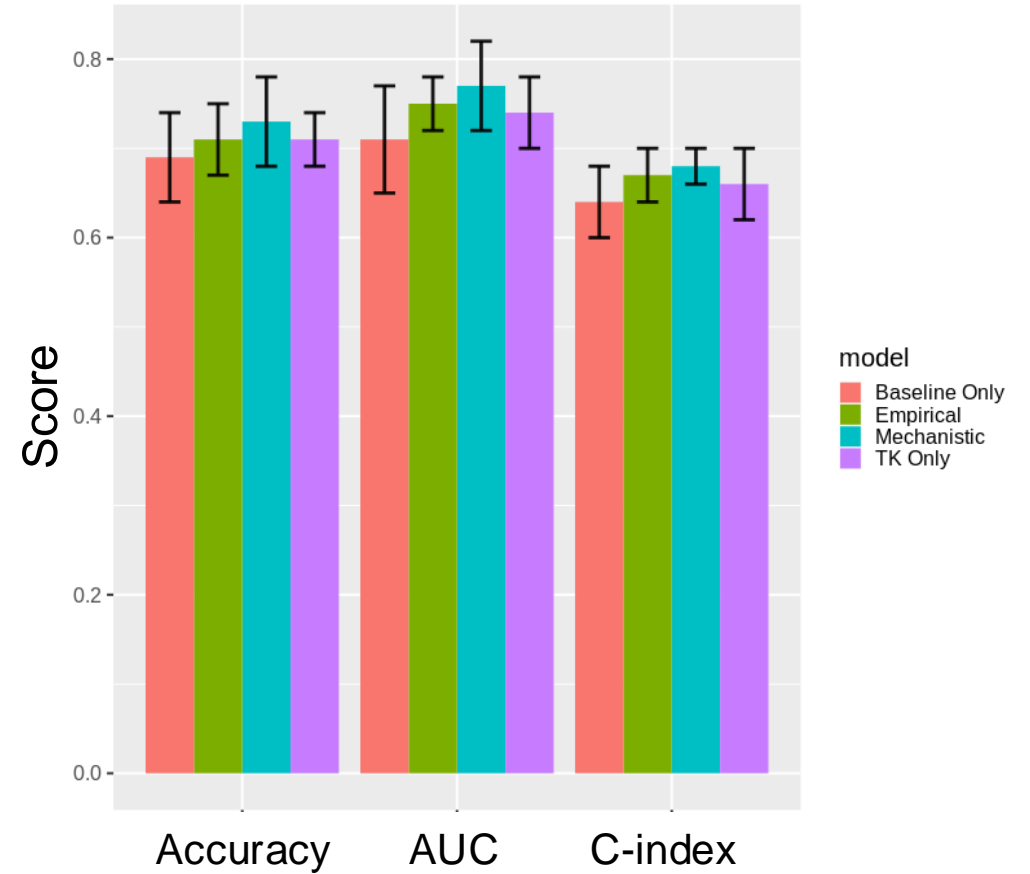
- All RSEs < **35%**
- despite the large number of parameters (14)
- Very large variability of some parameters
- **Small  $\eta$  - shrinkage** < 11%

# Post-cycle 4 prediction of individual survival



Longitudinal data considered

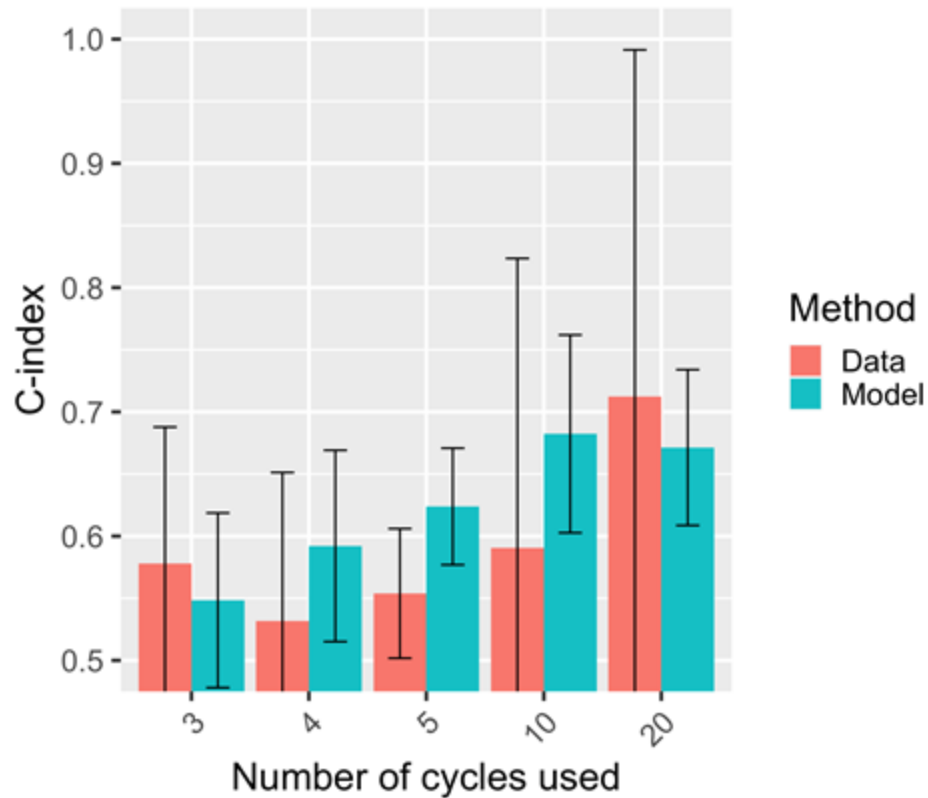
10 fold Cross-validation, AUC = accuracy at 12 months



**COMBO**

# Model-based outperforms data-based relative change for prediction

Compared CV c-indices for C3D1 relative change



What is the **added value** of NLME modeling compared to just observed data-based kinetic metrics?

- ⇒ Model-based metrics clearly have both **better predictive power** and **narrower uncertainty**
- ⇒ Dynamic modeling allows to **capture the kinetics and correct for intra-individual stochasticity** (noise)

Note: 10-fold cross-validation on FIR, BIRCH and POPLAR (train data set). Only relative change from baseline (at cycle 3 pre-dose) used as model metric. Other model-based parameters ignored here for fair comparison

# Summary

- **Kinetic models** based on **simple, routinely available blood markers** have prognostic value
  - Could be used as **non-invasive** surrogate markers of response
  - Modeling >> data only
- The **combination of NLME and ML** allowed to take the best of the two approaches in order to predict survival
  - NLME to account for **longitudinal** data
  - ML and data science methods to build **multivariable** models from a large number of features
  - **Minimal signature**: 11 baseline clinical features + TK (1 variable – 2 parameters) + BK (4 variables – 8 parameters)
  - **Mechanistic modeling** was shown to **improve predictions**
- Application to clinical oncology could improve **personalized health care**
- Application to **early prediction of trial outcome** could:
  - Save a lot of effort (..... **and money!**)
  - Be used to better stratify patients
  - Identify signal in early-phase trials

# **Beyond classical biomarkers: PIONeeR – QUANTIC and SChISM**

# THE PIONEER AND QUANTIC PROJECTS

Precision Immuno-Oncology for advanced Non-small cell lung cancer patients with PD-1 ICI Resistance



Assistance Publique  
Hôpitaux de Marseille

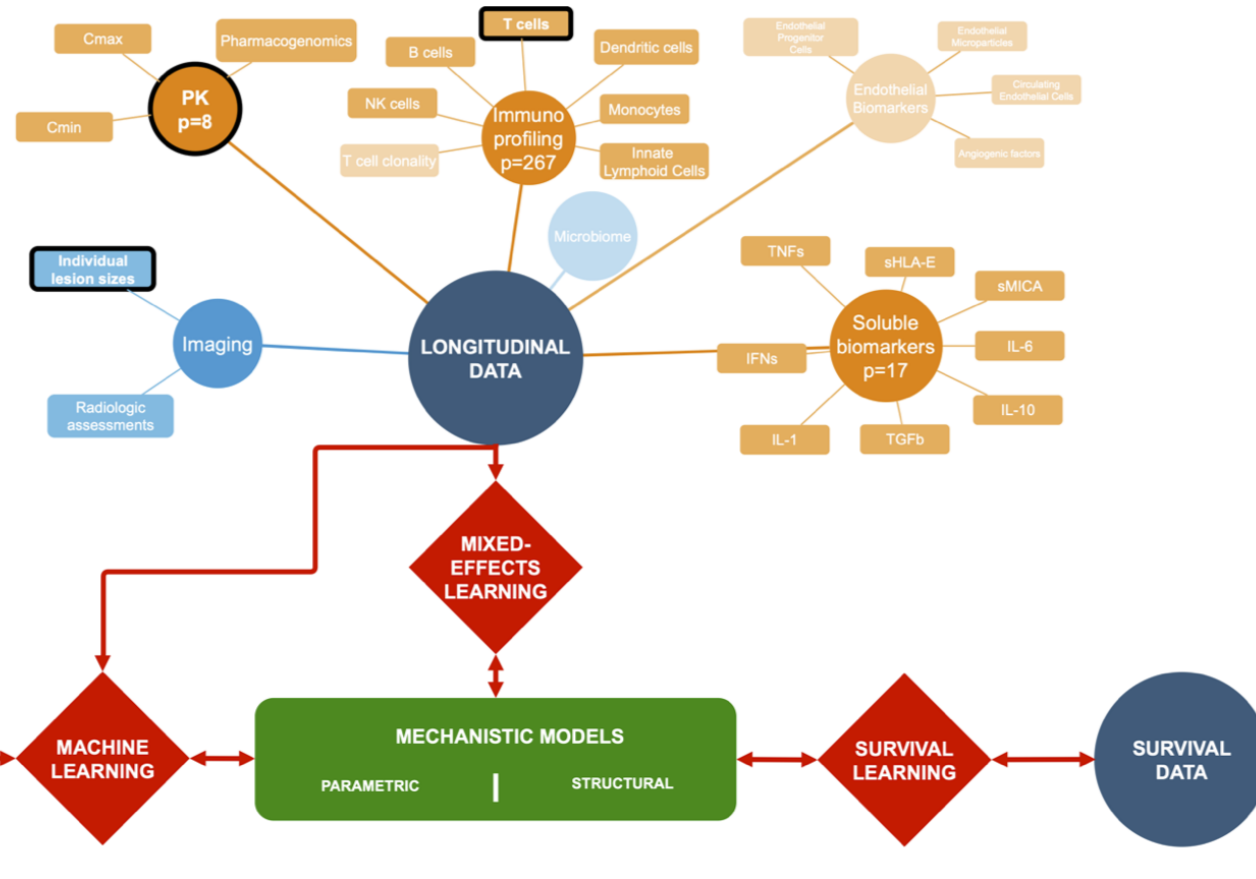
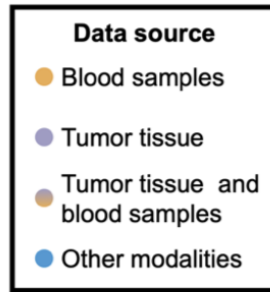


Inserm  
Institut national  
de la santé et de la recherche médicale



aviesan  
alliance nationale  
pour les sciences de la vie et de la santé  
ITMO Cancer

450 pts



24,287 variables

1,581 markers

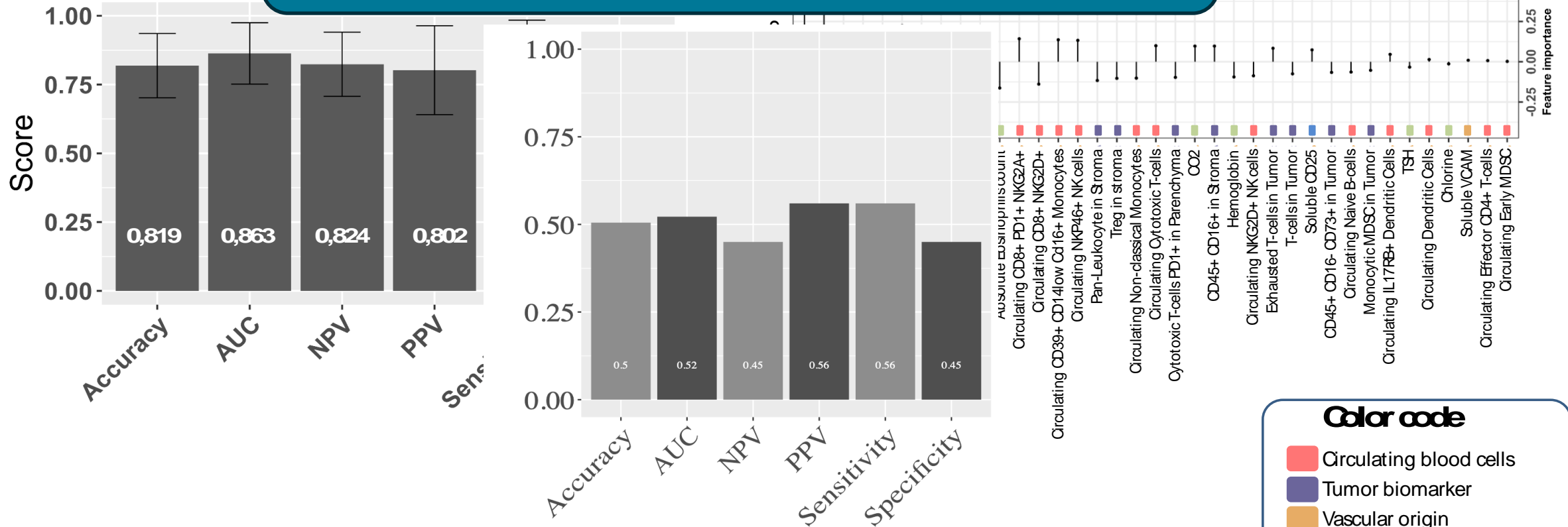
# 2+ LINE RESULTS (CROSS-VALIDATION ON THE TRAIN SET)

## DATA LOCK 2 - N = 127 - OUTCOME - EARLY PROGRESSION

**BUT.... results vanished on the test set !**

Predictive me

markers selected by LASSO



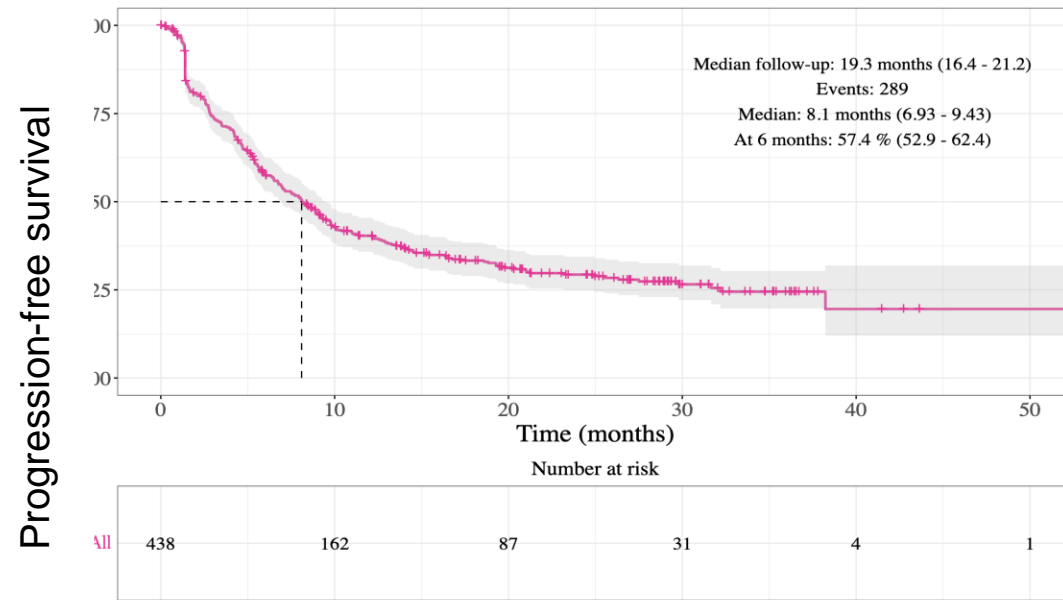
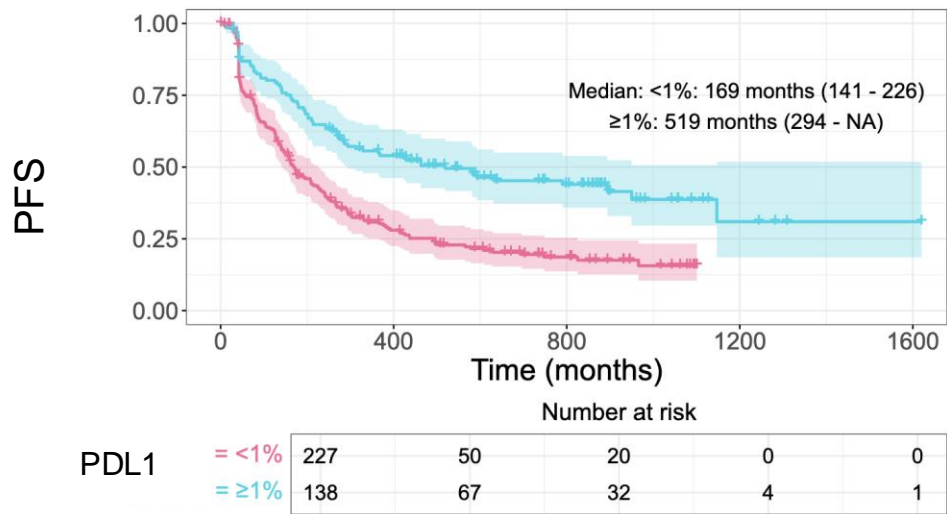
... partly due to the small number of patients added at data 3 (28)

Greillier et al., AACR 2022

# OUTCOME DEFINITION

## WHAT ENDPOINT?

Outcome	Time of PD (mths)	Non-evaluable	Resistant / Non-resistant (%)
Early progression (EP)	< 3	18	111/310 (26/74)
<b>Primary resistance<sup>1</sup></b>	<b>≤ 6</b>	<b>4</b>	<b>172/263 (40/60)</b>
Secondary resistance <sup>1</sup>	> 6	272	58/109 (35/65)



**PDL1 does not explain everything**

In PDL1 < 1% patients, 50% of primary resistant

Non-evaluable = patients with not enough follow-up (< 6 weeks) or censored



# CHALLENGES IN PIONEER BM ANALYSIS

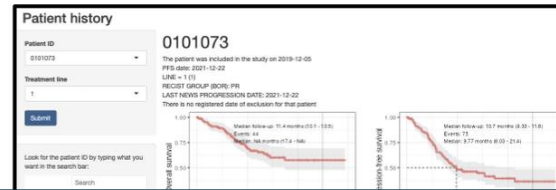
## MACHINE LEARNING

Difficulties	Remedies
Overfitting	Optimism-correction
Many, correlated data	Feature selection
A priori complex interactions	Nonlinear models
Missing data	Imputation
Small $n$ , Large $p$	Generative modelling

Tackle these problems one-by-one. Use theory if available, do simulations otherwise.

# AUTOMATIC REPORTING, PACKAGE DEVELOPMENT CONTINUOUS INTEGRATION / CONTINUOUS DEVELOPMENT

**PIONeer**  
Project ID: 44 [Leave project](#)  
4,787 Commits 18 Branches 31 Tags 43.1 GB Project Storage 3 Releases

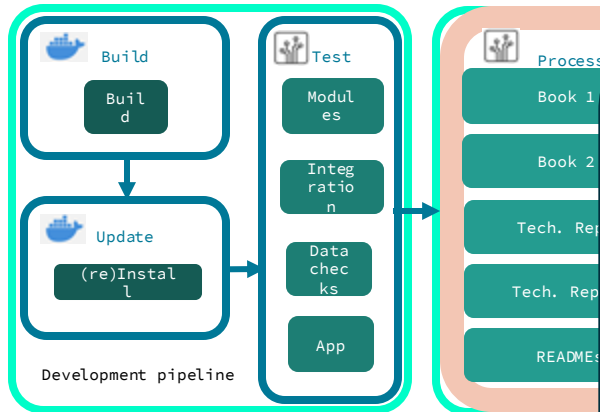


Dashboard Statistical reports Machine learning reports About

## Statistical reports

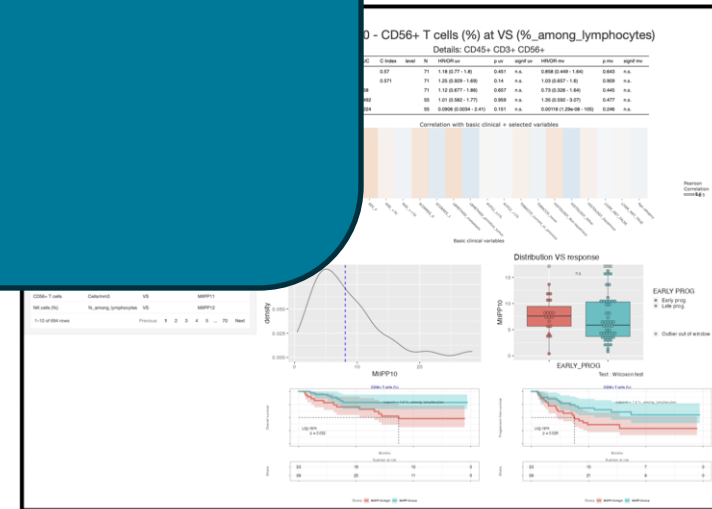
g patients

primary resistance / PFS



⇒ Reproducibility and traceability

git docker GitLab R markdown tidyverse ggplot2 bookdown



Human.months involved for preprocessing only = **28 h.m**

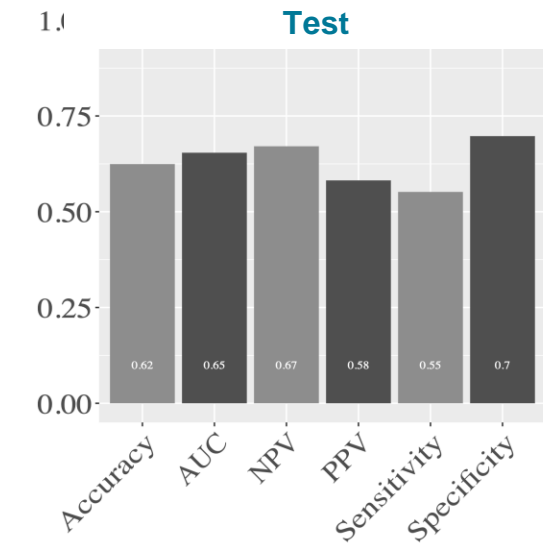
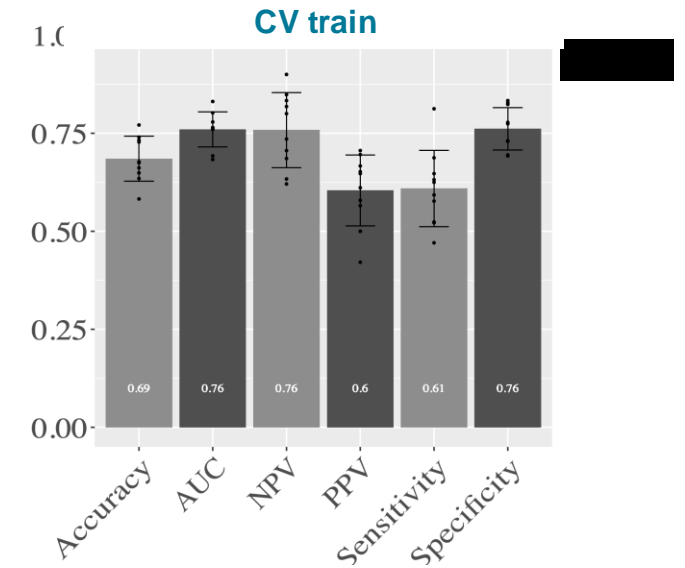
# METHODS ML

- Problem = overfitting in train / test
  - Because of  $p \sim n$  ( $N = 269 + 170 = 439$ ,  $p = 447$ )
  - Because of **NA imputation** ?
  - Because of **feature selection** ?

*Collins et al., Evaluation of clinical prediction models (part 1): from development to external validation, BMJ, 2024*

Splitting data at the moment of model development should generally be avoided as it discards data leading to a more unreliable model, whilst leaving too few data to reliably evaluate its performance

- Solution = **drop the train / test approach**
  - ⇒ Define a method (FS + ML)
  - ⇒ Apply to the **entire dataset** (apparent metrics)
  - ⇒ Evaluate and report **optimism** (= overfitting) using bootstraps



# SIGNATURE METHODOLOGY

- Problem = signature instability
- Reasons: missing values, correlations and  $p \sim N$

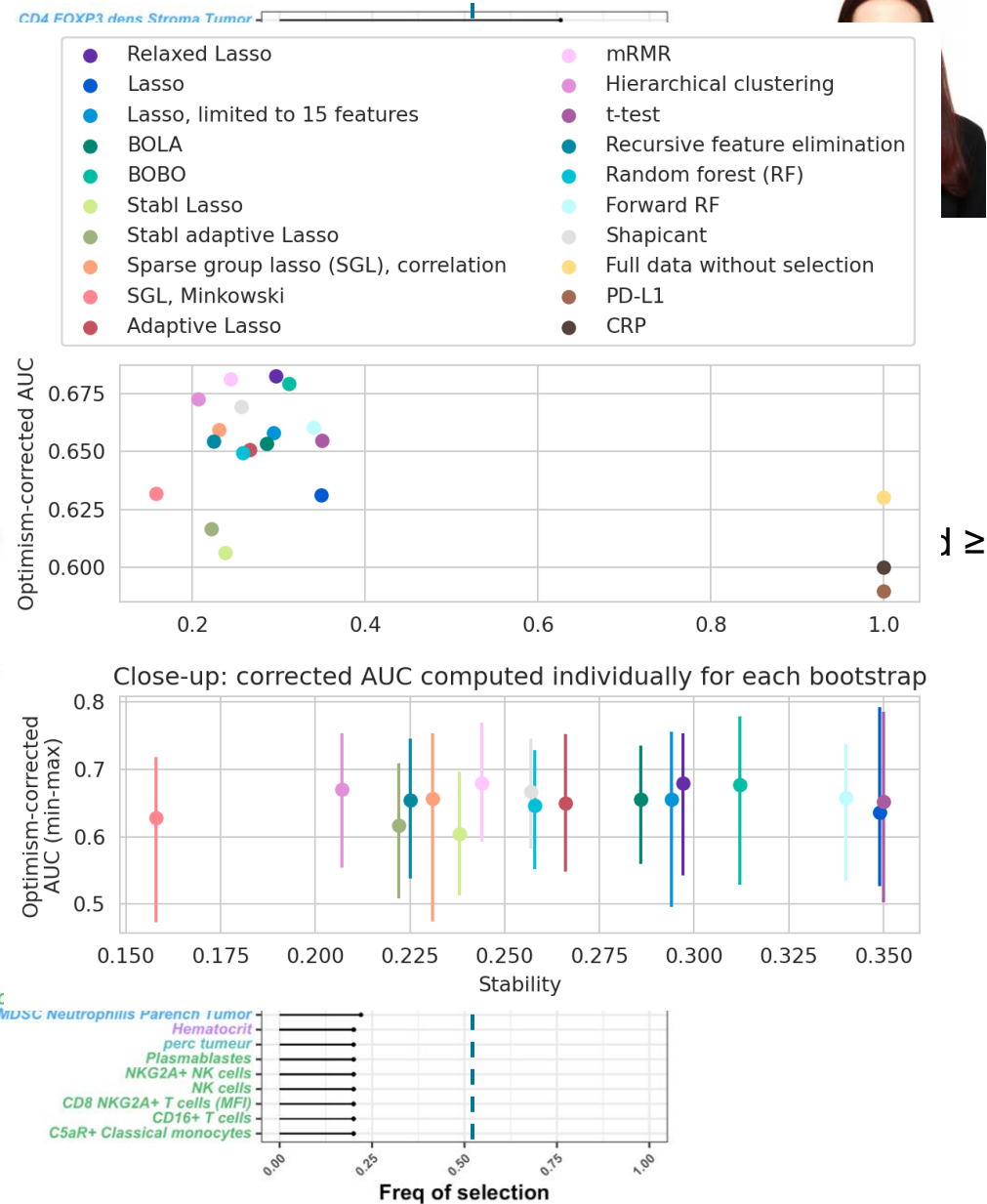
Benchmark of 18 FS methods: all failing 🤦

- Need more stringent filtering of variables for the signature
- At the same time, don't discard variables that are strong in but have high NA rate

Best solution = **simple** = use **univariate** stats and **FDR adjustment**

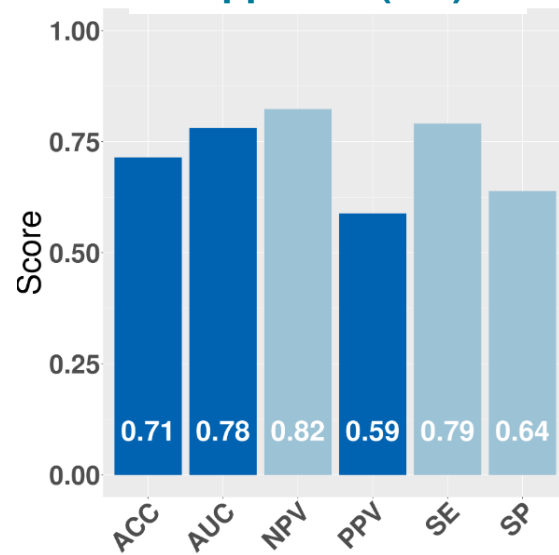
**$p.adjust\ mv < 0.5$  PRIM\_RES + PFS + AUC > 0.6 + C-index > 0.6**

## LASSO

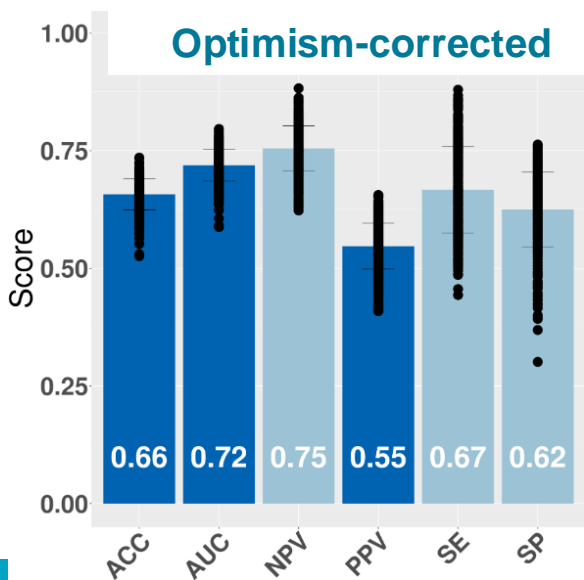


# PREDICTIVE MODEL RESULTS

Apparent (full)



Optimism-corrected

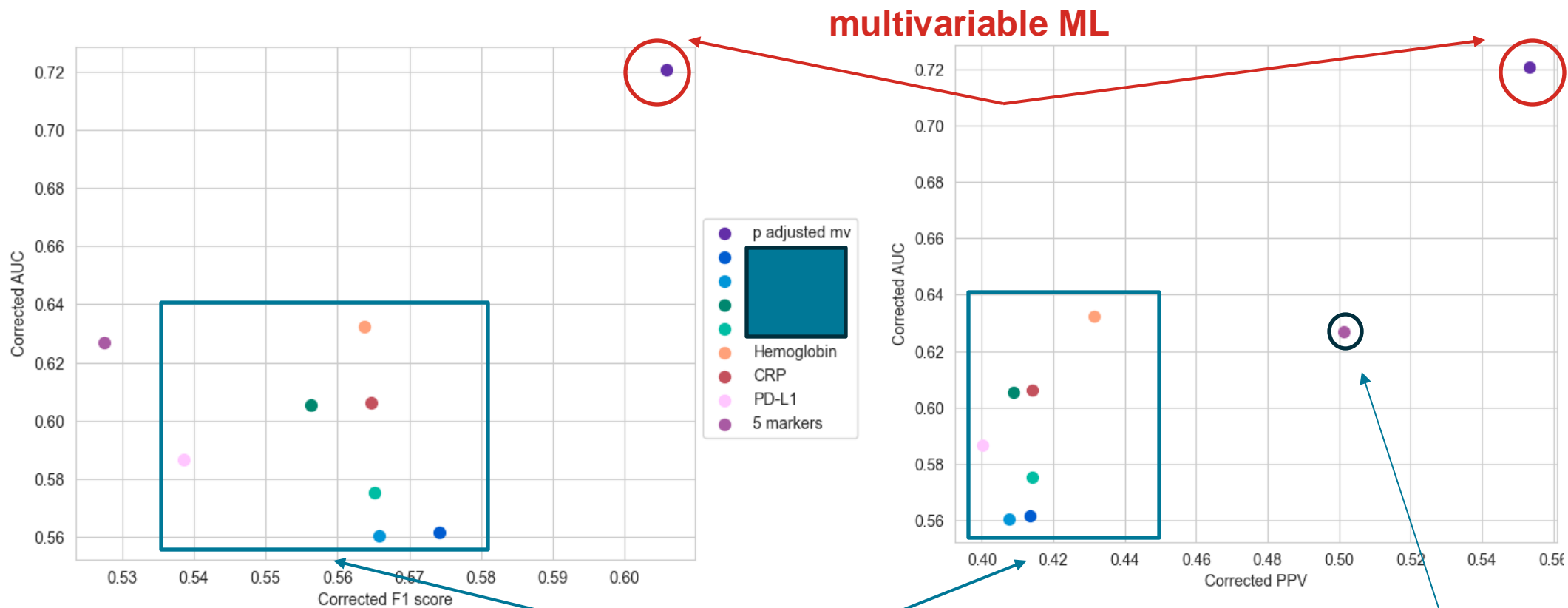


Classifier	AUC corr	AUC bias	SE corr	SE bias	PPV corr	PPV bias
RandomForestClassifier	0.721	0.06	0.669	0.121	0.553	0.035
HistGradientBoostingClassifier	0.72	0.17	0.557	0.211	0.593	0.193
ExtraTreesClassifier	0.718	0.028	0.737	0.007	0.544	0.049
LinearDiscriminantAnalysis	0.715	0.055	0.7	0.079	0.536	0.034
MLPClassifier	0.714	0.154	0.611	0.296	0.564	0.073
LogisticRegression	0.713	0.057	0.691	0.111	0.533	0.023
GradientBoostingClassifier	0.712	0.224	0.538	0.299	0.596	0.246
ExtraTreesClassifier	0.709	0.087	0.643	0.14	0.546	0.079
BaggingClassifier	0.707	0.293	0.289	0.711	0.68	0.32

- 14 ML models tested (linear, tree ensembles, neural networks)
- **Best = random forest**

# PREDICTIVE MODEL

## AUC AND POSITIVE PREDICTIVE VALUE



$$F1 = \frac{2}{\frac{1}{SE} + \frac{1}{PPV}}$$

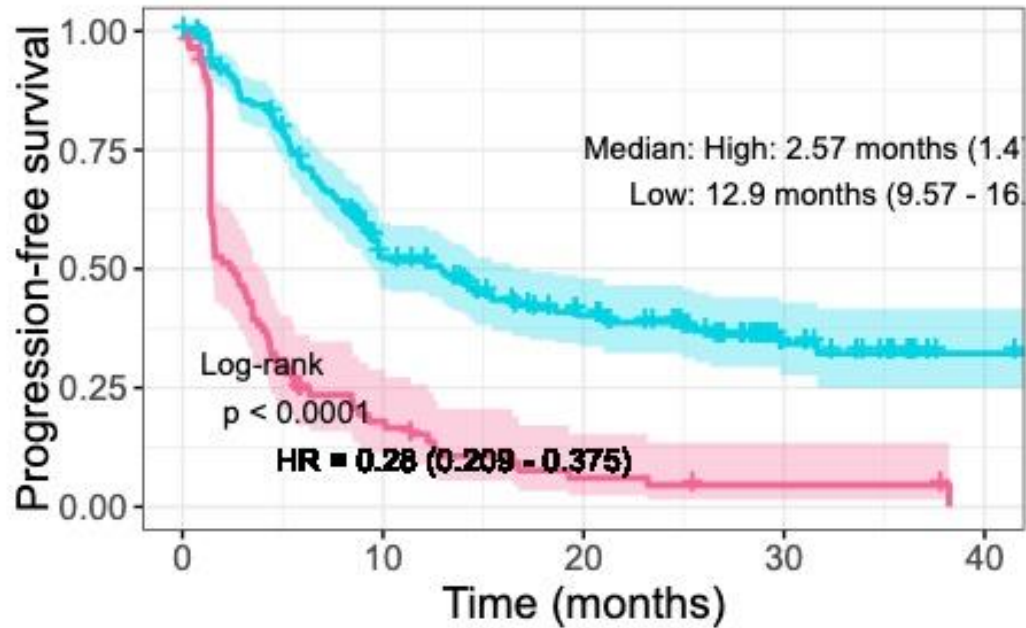
# PREDICTIVE MODEL

## PFS STRATIFICATION

Train

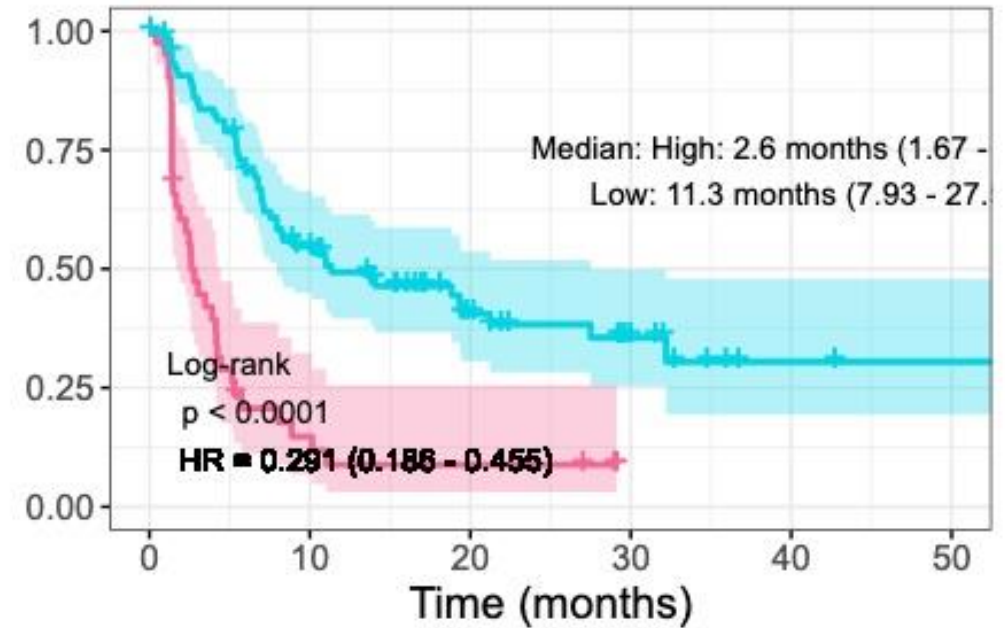
Test

ML\_score = High ML\_score = Low



Number at risk

86	13	4	2	0
212	97	57	19	2

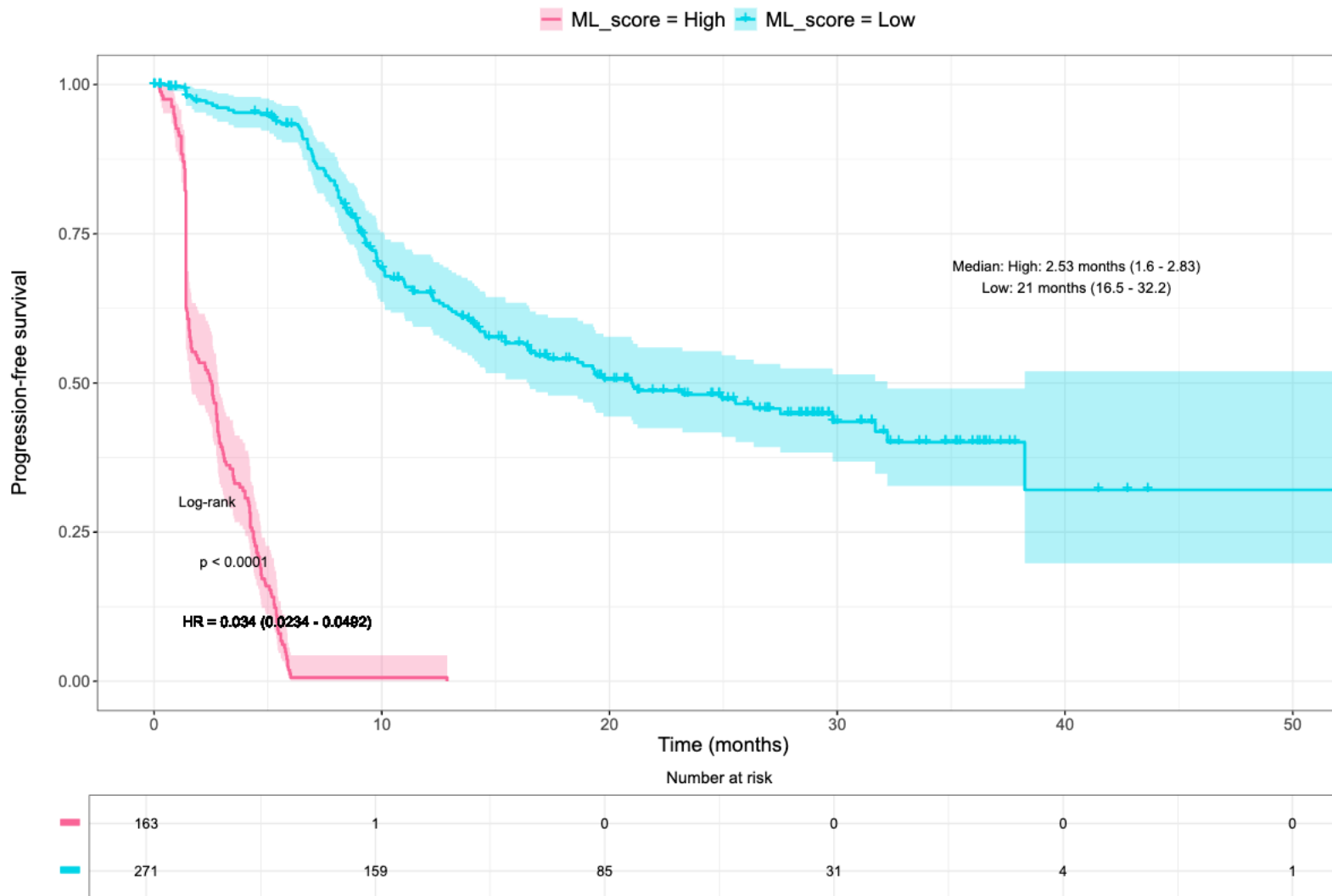


Number at risk

41	5	3	0	0	0
88	43	19	10	2	1

# PREDICTIVE MODEL

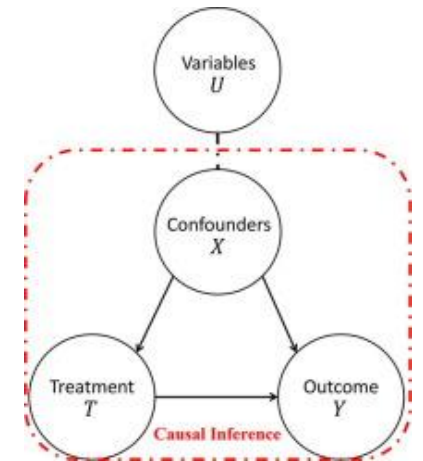
## PFS STRATIFICATION: FULL DATA



cutpoint defined at the proportion of PRIM\_RES on the full data set

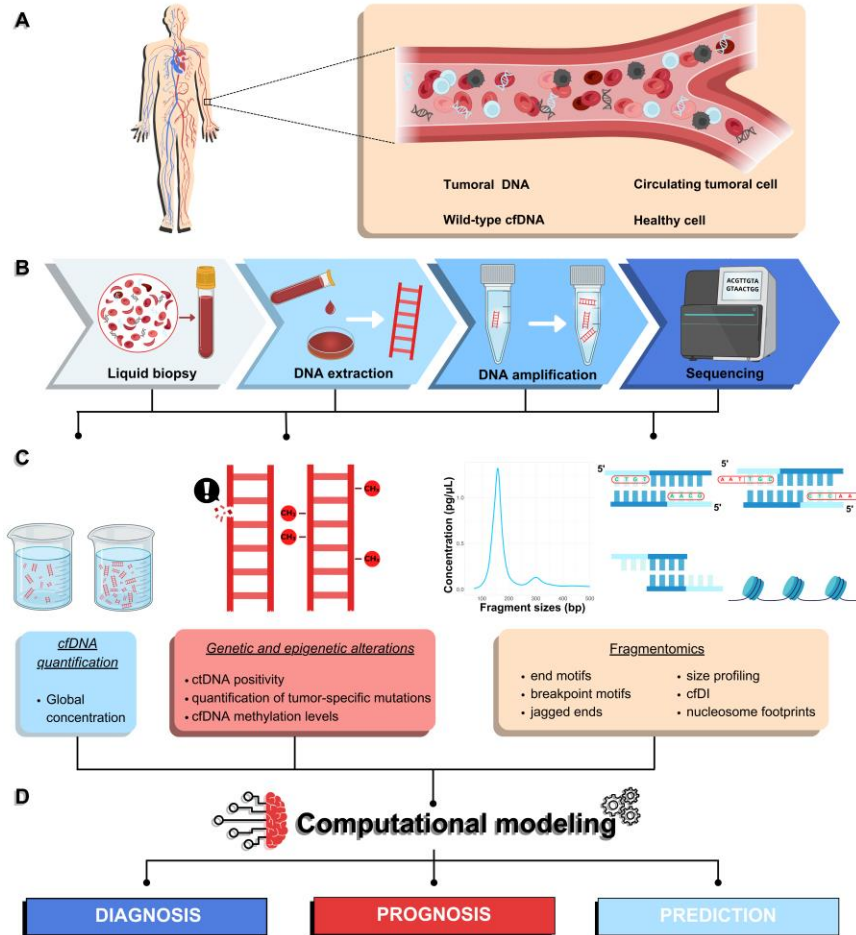
# LESSONS

- Preprocessing / **data cleaning** is essential but **time consuming**
- **Software development** agile methods are very useful to handle large data science projects
- Feature selection for biomarker studies is a **hard problem**
  - **ML literature**: a lot of methods but datasets are either synthetic or with  $p \ll N$
  - **Biomedical literature**:  $p \gg N$  (e.g., omics) but **stability** rarely assessed and **reproducibility** is an issue
- Even in large studies, we have  **$p \sim N$ : big fat data**
  - Don't waste data for model development: **drop the train / test** approach
  - Use **optimism-correction**
  - **Validate** the model in a subsequent study (ideally performed by another group)
- Results with **brute force** methods were disappointing. Methodological advances are needed
  - **Feature selection**
  - **NA imputation**
  - **Causal inference**
  - **Mechanistic modeling**



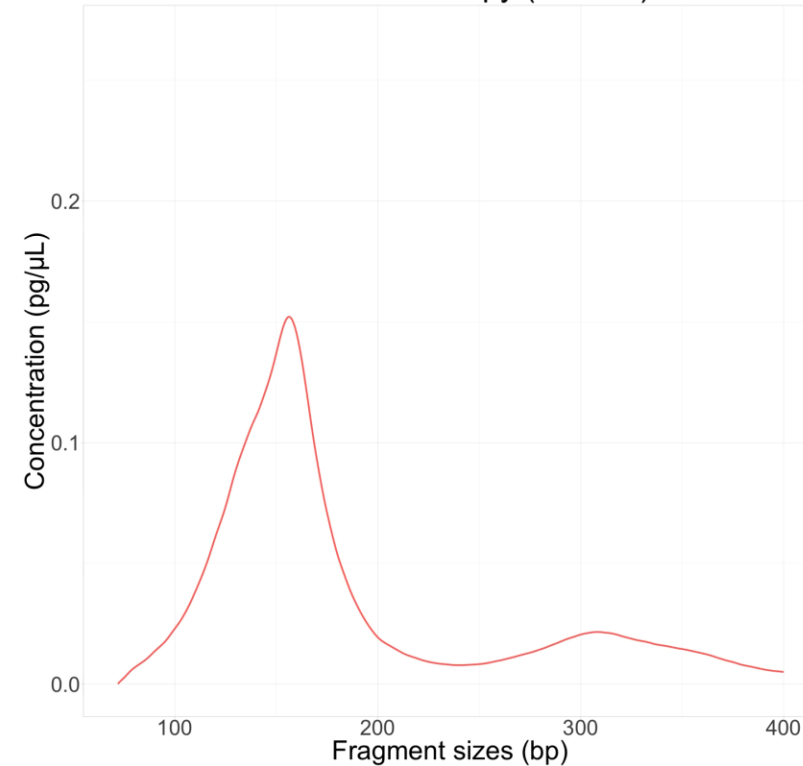
# Size CfDNA Immunotherapies Signature Monitoring

APHM: S. Salas, Inria / Inserm: S. Benzekry, L. Nguyen, Adelis

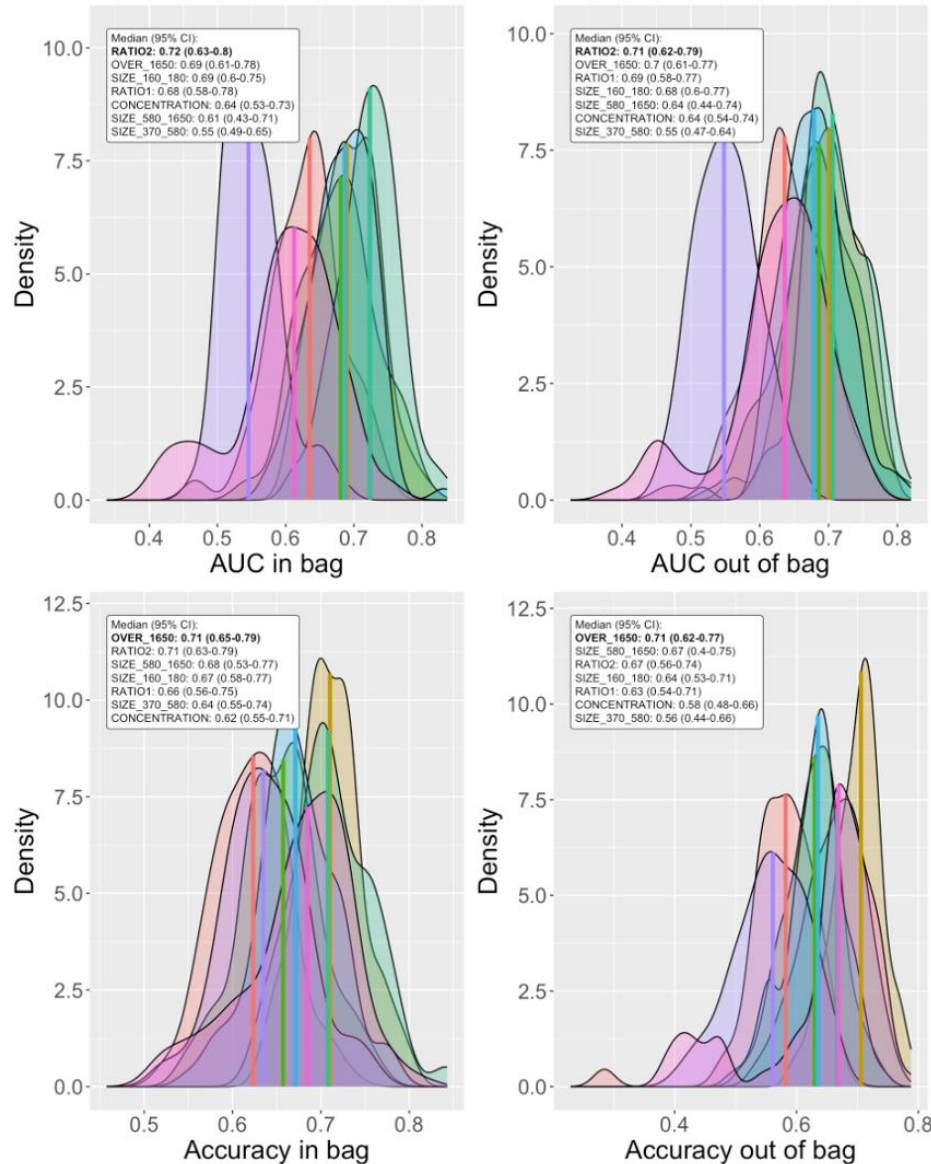


Patient 700-13 (Total N = 300)

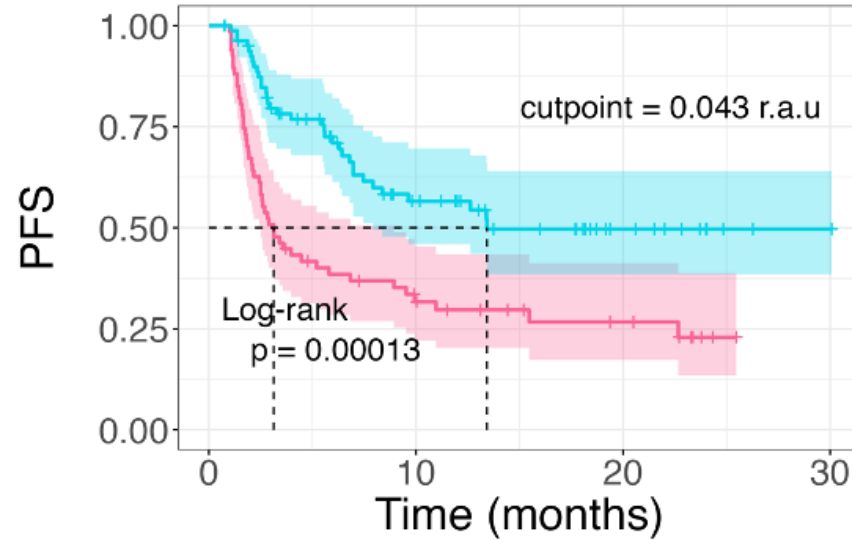
Time from first immunotherapy (months): 0



# Predictive value of baseline cfDNA markers



## Relative concentration of cfDNA >1650 bp (%)

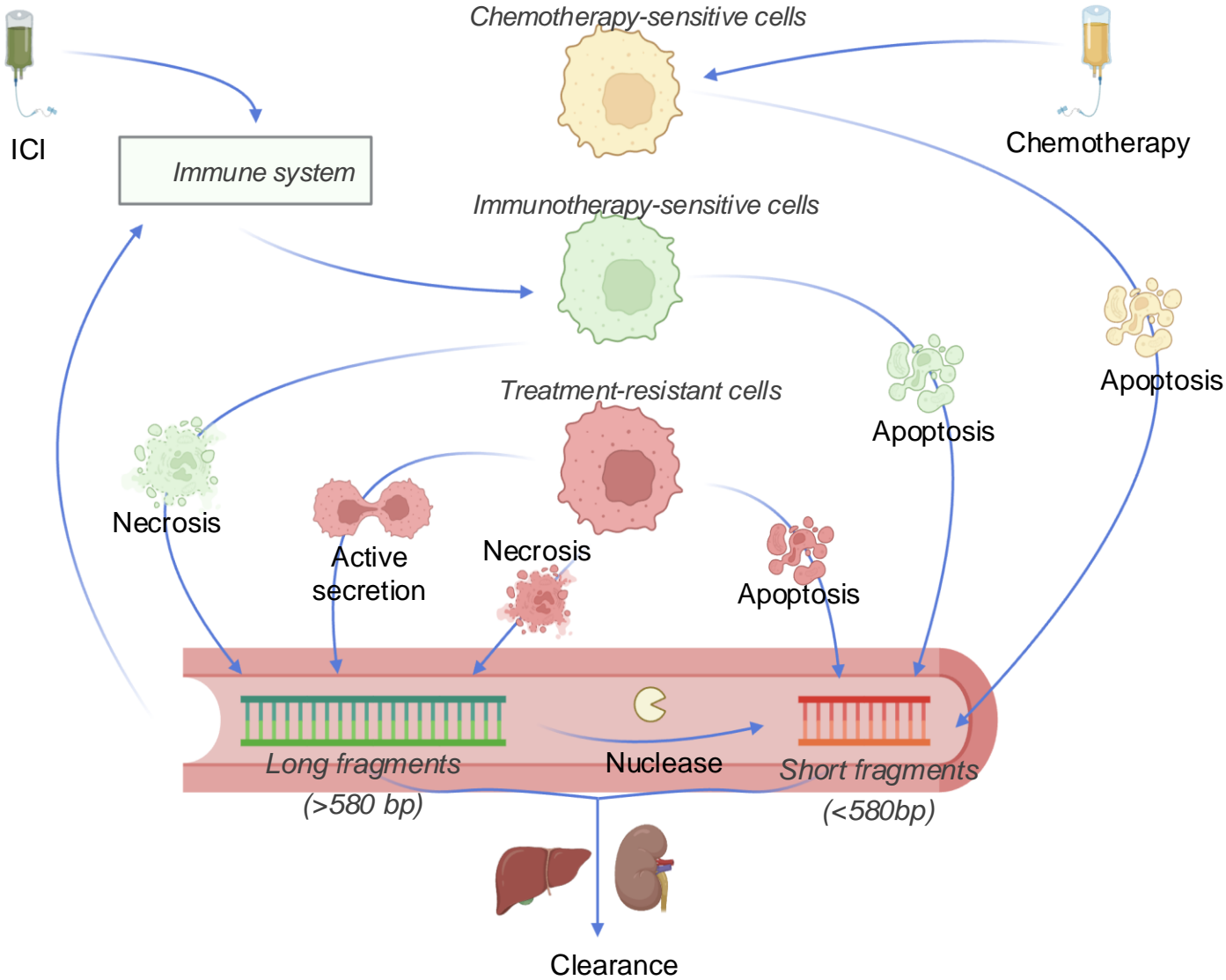


Number at risk

	0	10	20	30
low	67	17	8	0
high	82	31	10	1

+ OVER\_1650 = low   
 + OVER\_1650 = high

# Different processes of cfDNA release



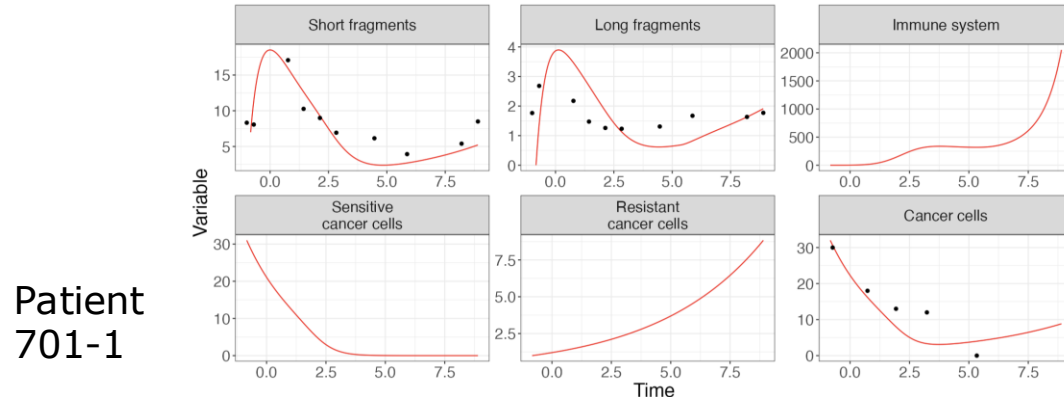
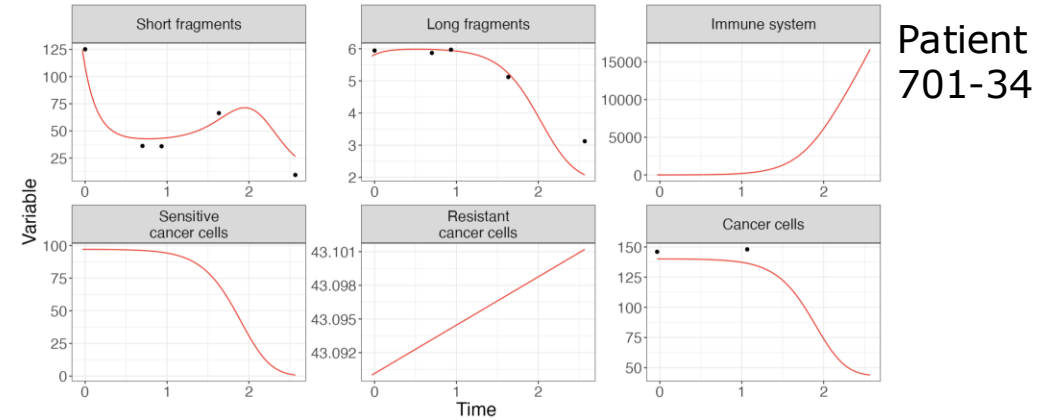
$$\frac{dS_S}{dt} = -\epsilon_C \cdot 1_C \cdot S_S - \epsilon_I \cdot I \cdot S_S$$

$$\frac{dS_R}{dt} = K_G \cdot S_R$$

$$\frac{dD_S}{dt} = \lambda_X \cdot (\epsilon_C \cdot 1_C \cdot S_S + \epsilon_I \cdot I \cdot S_S) + \lambda_S \cdot S - k_D \cdot D_S + kn \cdot D_L$$

$$\frac{dD_L}{dt} = \lambda_L \cdot S' \cdot 1_{S' > 0} + \lambda_{AS} \cdot S - k_D \cdot D_L - kn \cdot D_S$$

$$\frac{dI}{dt} = \beta \cdot D_L \cdot I - k_I \cdot I$$



# Acknowledgments

Thank you for your attention!

[Roche / Genentech](#)

R. Bruno  
P. Chanu  
F. Mercier  
C. Jamois



Open  
postdoc position

[sebastien.benzekry@inria.fr](mailto:sebastien.benzekry@inria.fr)

<https://team.inria.fr/compo/>



Patients and families

Computational

A. Bakhmach  
C. Bigarre  
M. Boussena  
P. Dufosse  
M. Karlsen  
L. Nguyen – Phuong  
R. Taieb  
A. Vaglio

Clinical oncology

D. Boulate  
J. Ciccolini  
L. Greillier  
X. Muracciole  
S. Salas

F. Barlesi

