



Integrative kinetics and machine learning modeling for prediction of outcome following immunotherapy in lung cancer

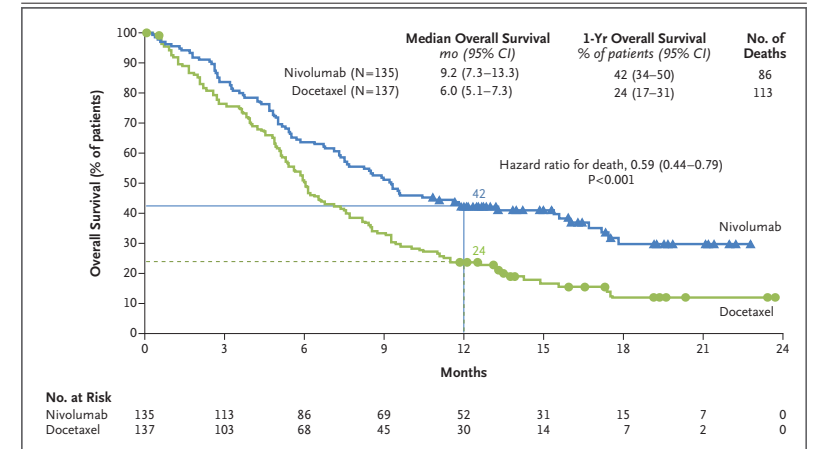
S. Benzekry

SMB annual meeting, Seoul, July 2024



Clinical challenge

- Lung cancer is the **leading cause** of cancer-related death (25.3 % of all cancer deaths, 13.6 % of new cancer cases)
 - 57 % are metastatic at diagnosis (stage IV), 5-year overall survival rate is poor (18.6 %)
 - Introduction of new immunotherapies (immune checkpoint inhibitors) in the 2010's has **revolutionized treatment of advanced lung cancer**
 - **20 – 30 % of patients have long term response** (> 10 years)
- Prediction of overall survival from baseline or early data could help
 - guide treatment decision during clinical **drug development** (e.g., combo trials, predict OS from tumor phase I/II data)
 - inform **personalized health care**
- Current state of the art
 - Only predictive biomarker used in the clinic: PDL1 expression from immunohistochemistry¹ (AUC = 0.601)
 - Predictive model from **baseline** clinical and biological data: ROPRO score²
 - 27 variables
 - Developed on 120k+ patients over multiple cancer types
 - **C-index = 0.69** on OAK as independent test set
 - Predictive value of **transcriptomic and mutation** data is unclear
 - **Tumor kinetics** model parameter growth rate (KG) has important predictive power of hazard ratio (HR)³
 - Predictive value of kinetics of **blood markers** (lab tests) is unclear



Brahmer et al., NEJM 2015

¹ Rizvi, H. et al., J Clin Oncol Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients with Non-Small-Cell Lung Cancer Profiled with Targeted Next-Generation Sequencing. J. Clin. Oncol. 2018, 36, 633–641..

² 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).

³ 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 24, 3292–3298 (2018).

kinetics – Machine learning (kML) for prediction of OS

Raw baseline data

Clinical characteristics
p = 73

Genomic and Transcriptomic
p = 395 and 58,311



Minimal signature

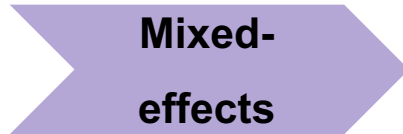
Clinical characteristics

Genomic and Transcriptomic

Raw longitudinal data

Tumor Kinetics (TK)
k = 8,635

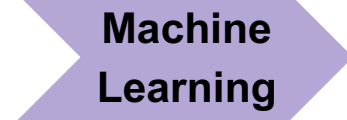
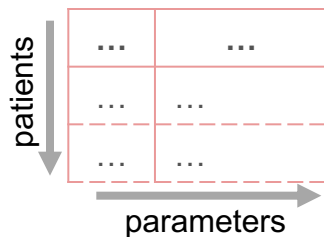
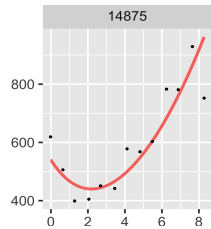
Blood markers Kinetics (BK)
k = 108,551
Albumin, CRP, Neutrophils, LDH



Model-based individual parameters

TK

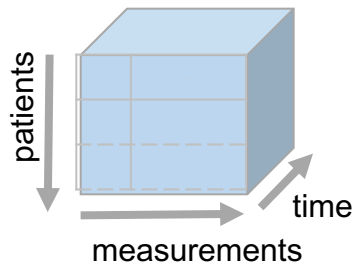
BK



Survival outputs

- Individual prognosis
- C-index
 - Survival

- Clinical trial
- Survival curve
 - Hazard ratio



Outline

1. Data
2. Modeling the kinetics of longitudinal data using nonlinear mixed-effects modeling (NLME)
 1. Tumor size kinetics (TK)
 2. Blood markers kinetics (BK)
3. Machine learning (kML)
 1. Model development
 2. Individual predictions
 3. Trial-level predictions
4. Mechanistic modeling
5. Beyond classical biomarkers
 1. cfDNA fragmentomics
 2. PIONeeR biomarkers (1000+)

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

Train

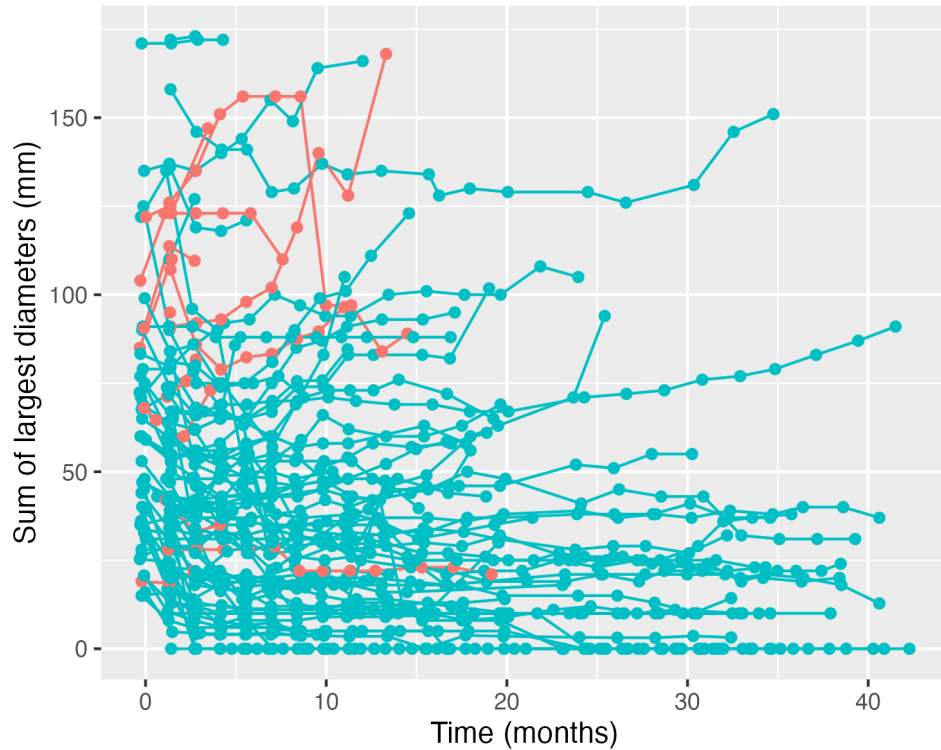
Test

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) 5
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)

Tumor kinetics

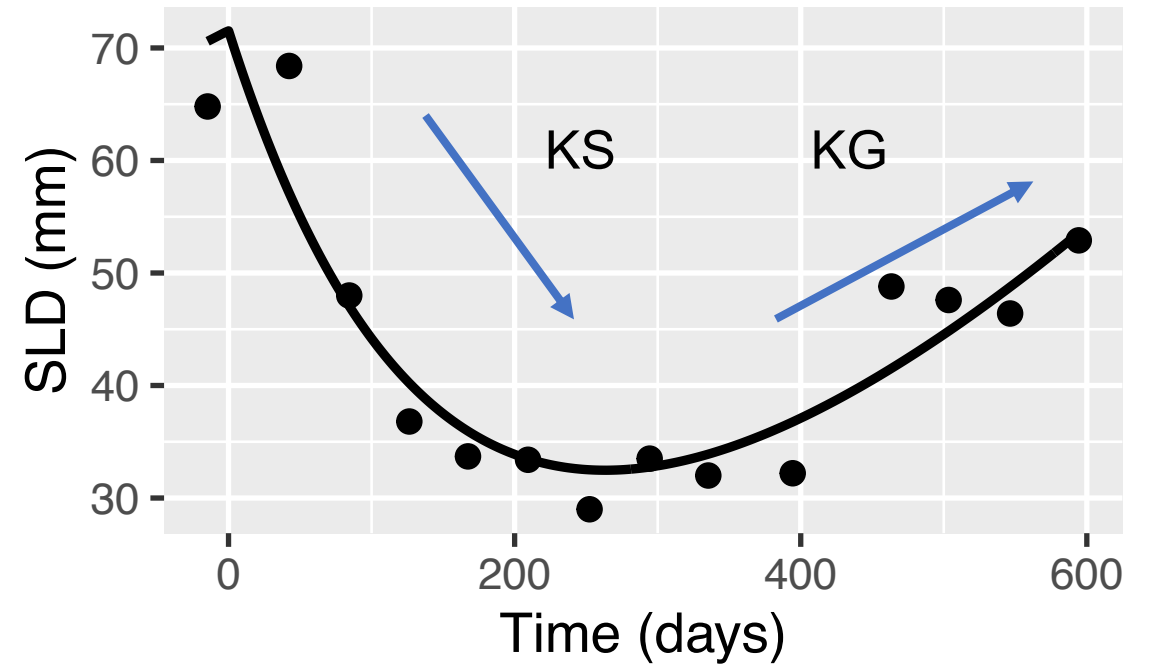
100 random patients



Best response
—●— CR + PR + SD
—●— PD

Individual kinetics

Double (sum of two) exponential model



$$S^i(t) = S0^i \left(e^{-KS^i t} + e^{KG^i t} - 1 \right)$$

Mixed-effects modeling for population longitudinal data

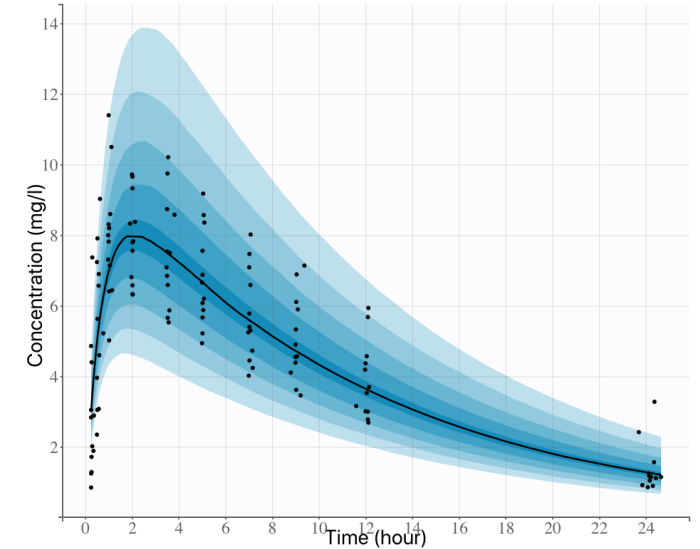
- Observation model

$$y_j^i = M(t_j^i; \theta^i) + \varepsilon_j^i, \varepsilon_j^i \sim \mathcal{N}(0, \sigma_j^i)$$

- Inter-individual variability model

$$\ln(\theta^i) = \ln(\theta_{\text{pop}}) + \eta^i, \eta^i \sim \mathcal{N}(0, \omega^2)$$

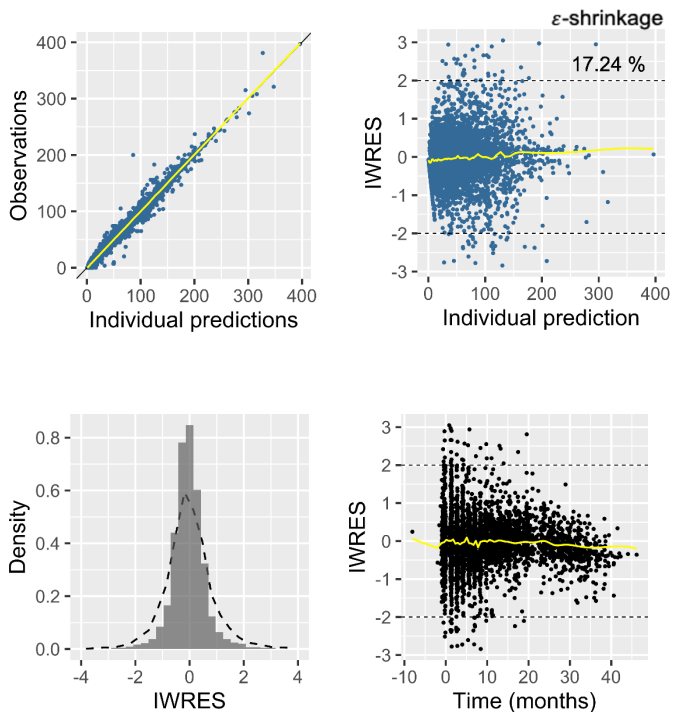
M = structural model
 y_j^i = observation at time t_j^i in patient i
 ε_j^i = error model (constant (TK) or proportional (BK))



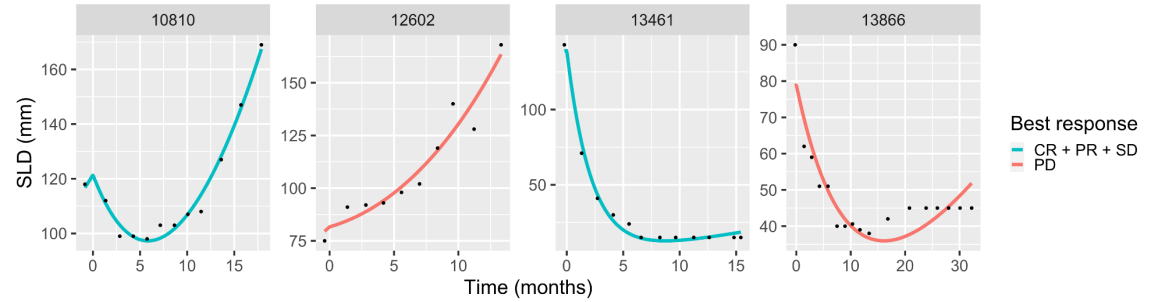
- Estimation of the population parameters using the SAEM algorithm for likelihood maximization
- Estimation of the individual empirical Bayes estimates (EBEs) from the maximum a posteriori estimator
- Fits performed using Monolix 2020R1 and scripted using the R Monolix API

The double exponential model was able to describe individual TK

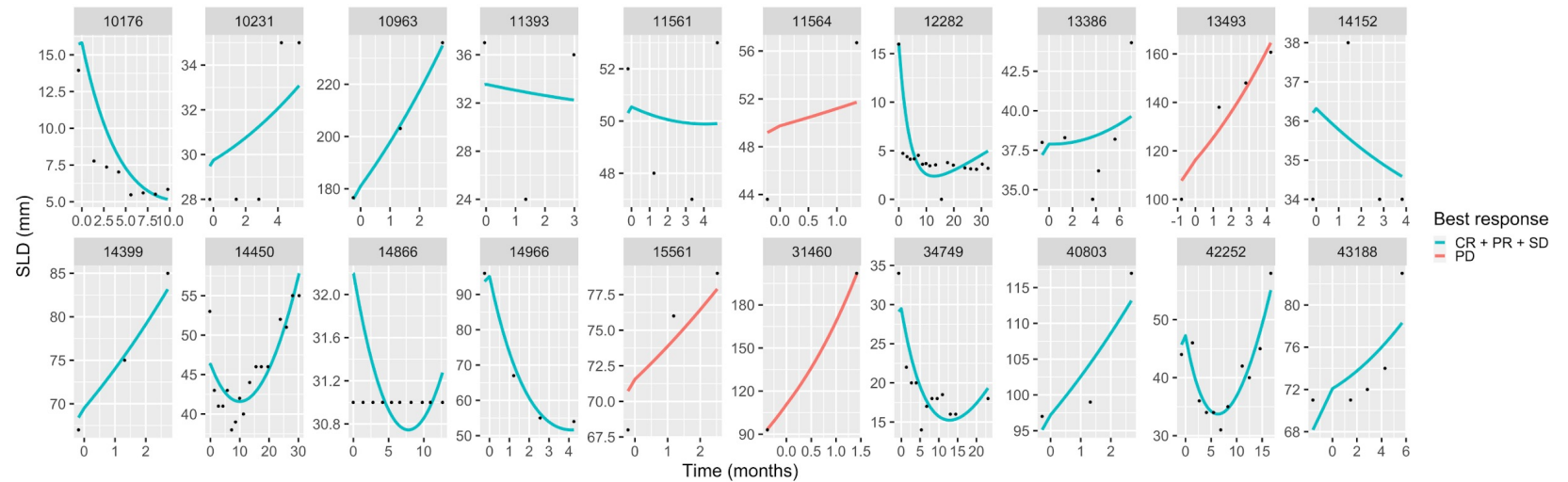
Individual goodness-of-fit (GOF) plots



2 best fits in each response group



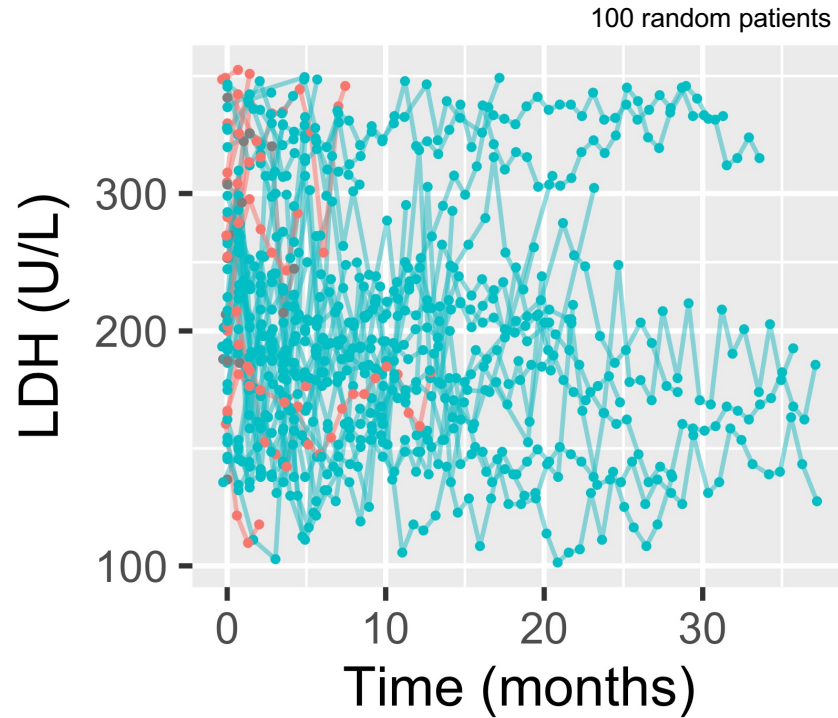
20 randomly selected individual fits



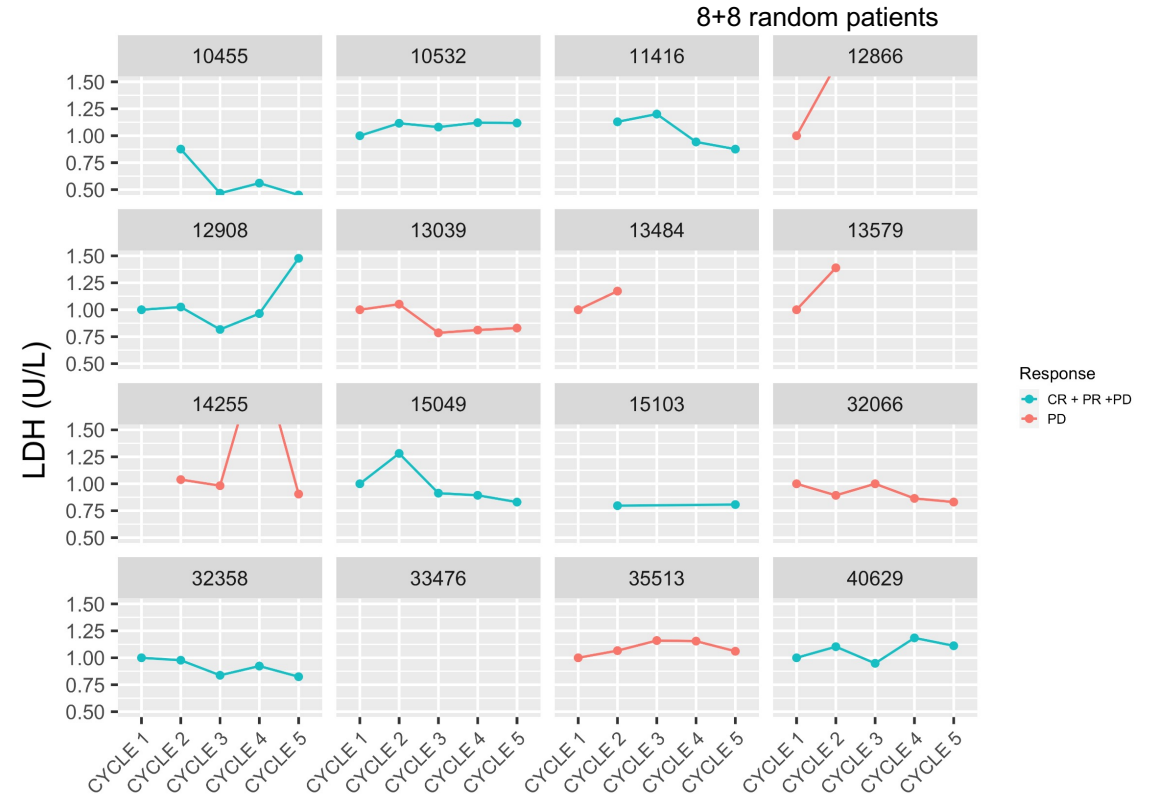
Goodness-of-fit quantitatively assessed with residuals-based metric

Blood markers kinetics (BK)

Spaghetti plot



Early cycles (normalized)



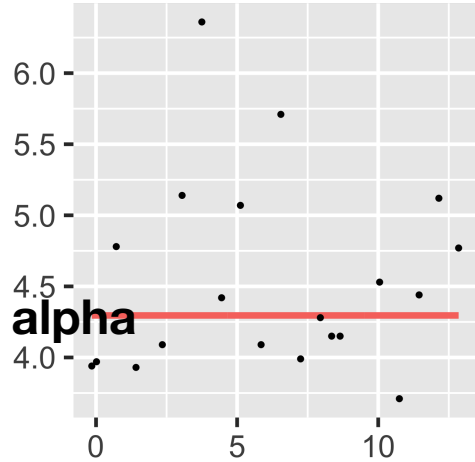
Is there any kinetic pattern ?

Does it carry predictive value ?

Can we describe dynamics with a mechanistic model ?

Empirical structural models to test for BK

$H_0 = \text{Constant}$

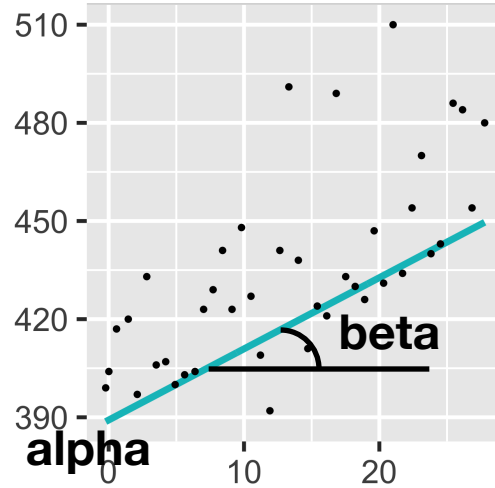


$$y^i = \alpha^i$$

Equivalent of no significant kinetics

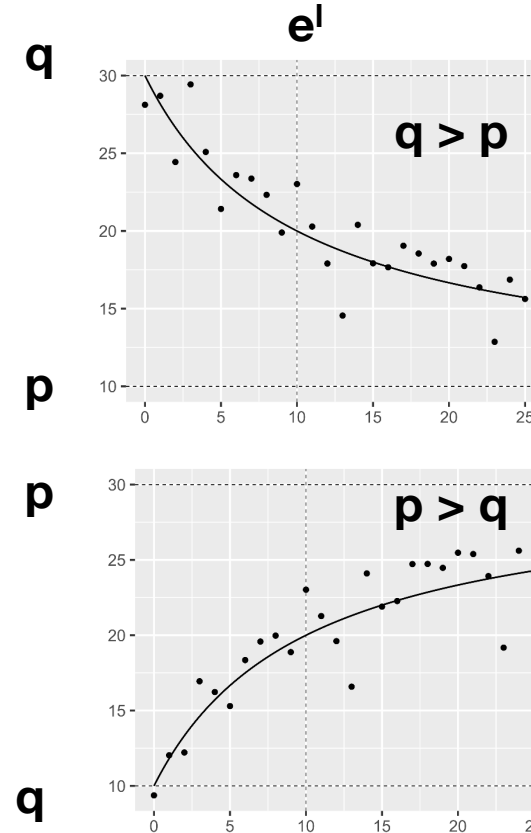
Used as a benchmark

Linear



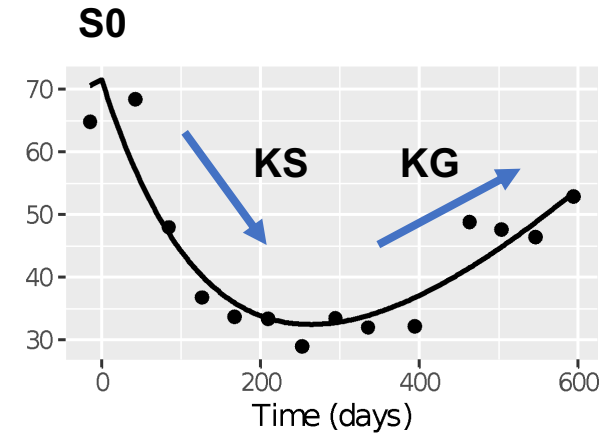
$$y^i = \alpha^i + \beta^i t$$

Hyperbolic



$$y^i = p^i + e^{li} \frac{q^i - p^i}{t + e^i}$$

dexp



$$y^i = \begin{cases} S0^i e^{KG^i t}, & t \leq 0 \\ S0^i (e^{-KS^i t} + e^{KG^i t} - 1), & t > 0 \end{cases}$$

Allows change of monotonicity

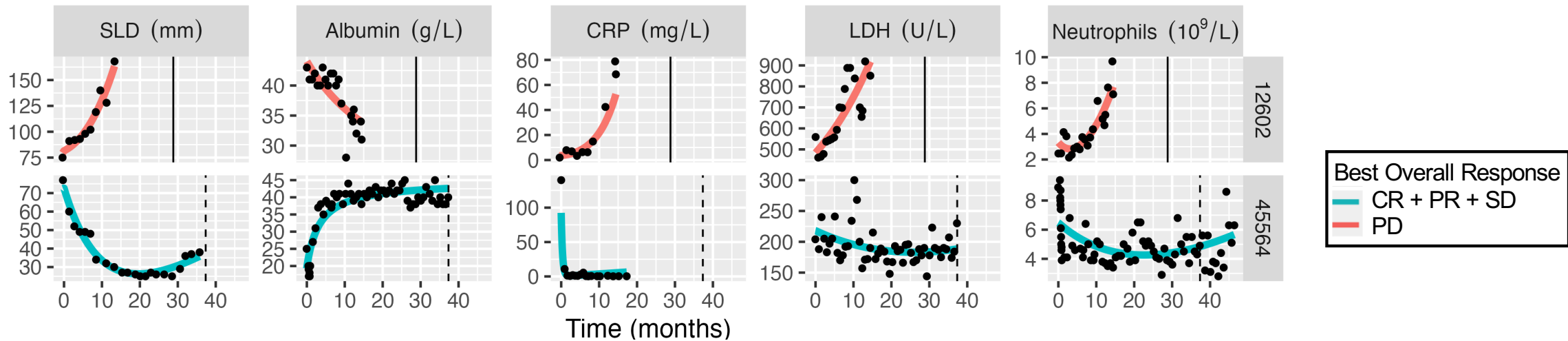
Stein et al., 2008 (PSA)
Claret et al. 2018 SLD atezo

Gavrilov et al. 2020 (NLR atezo)
Jackson and MacCluer et al. 2020

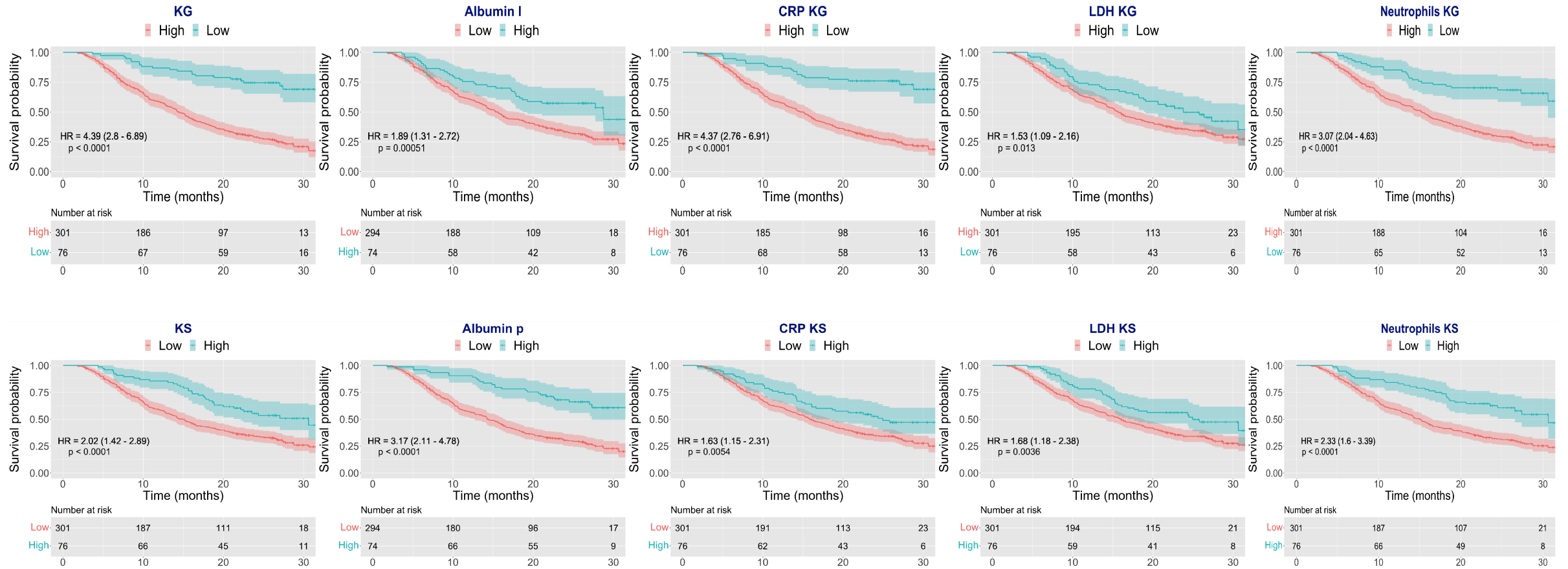
Is there any kinetic pattern? Yes !

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 |
| Double-exponential | 48.4 | 0.058 | 28.8 | 0.21 | 39.9 | 0.56 | 102 | 0.14 |
| Hyperbolic | 48.0 | 0.056 | 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 |

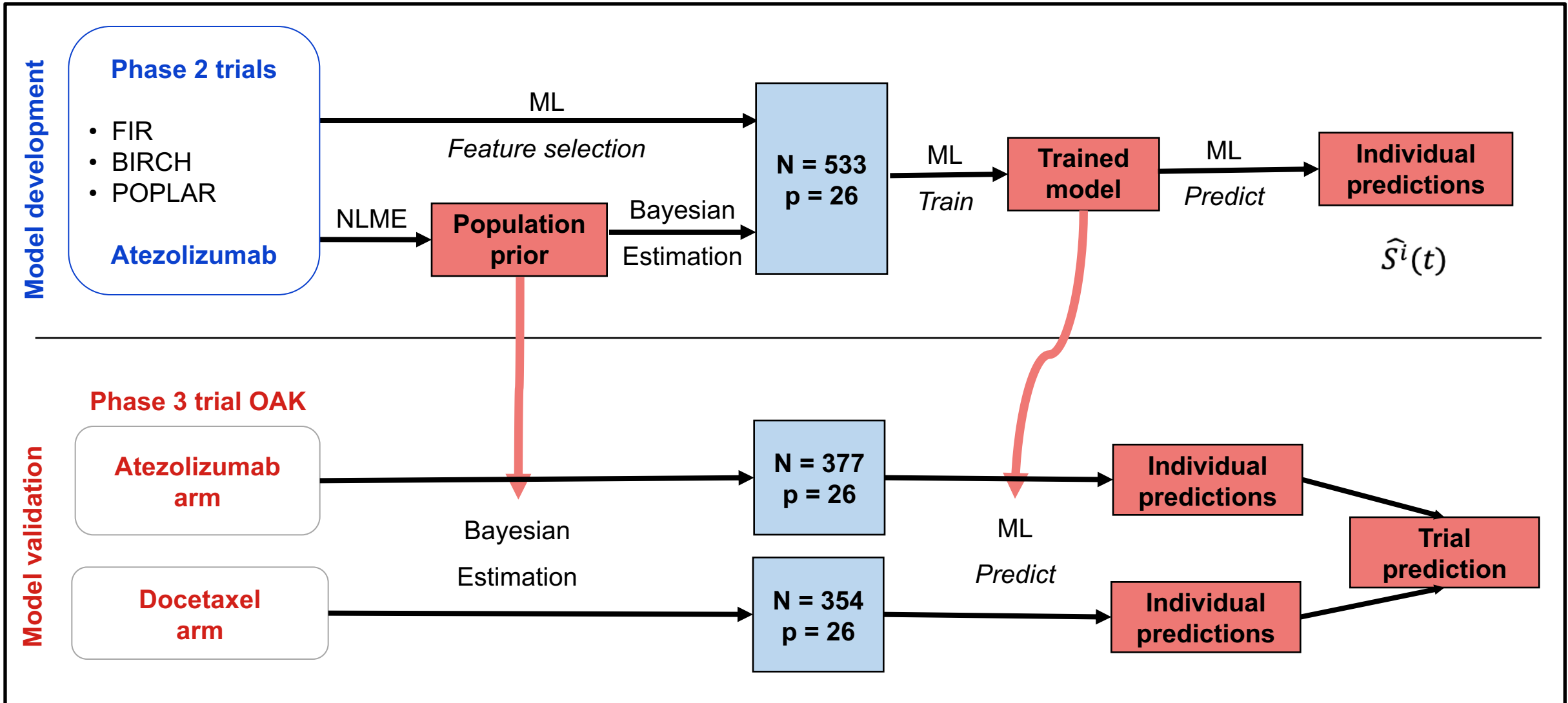


Does it carry predictive value? Yes !



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

Model development and validation



Methodology - Machine learning model development

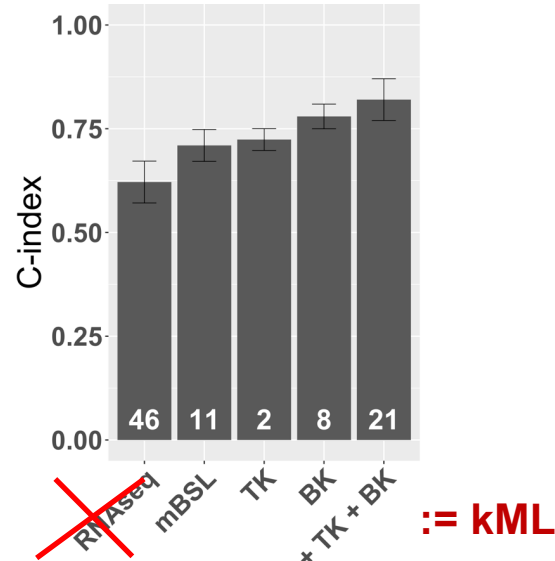
- Preprocess
 - Drop zero-variance or >25% NA columns, dummification, NA imputation, scaling
- Dimensionality reduction for RNAseq data (bootstrap LASSO)
- Features selection:
 - 5 methods: LASSO, random survival forest (RSF) importance, Cox-based and stepwise forward/backward
 - 3 strategies : i) all variables, ii) per feature set and iii) pooled selected sets
- 4 survival algorithms tested:
 - Cox, Cox and accelerated failure time with gradient boosting and random survival forest (RSF)
- Evaluation of machine learning models
 - Model development: 10-fold cross-validation, C-index, calibration curves and 12-months survival classification metrics
 - Study-level predictions: survival curves, hazard ratios

Working principle: need for a **minimal signature** model with limited number of easily measurable variables

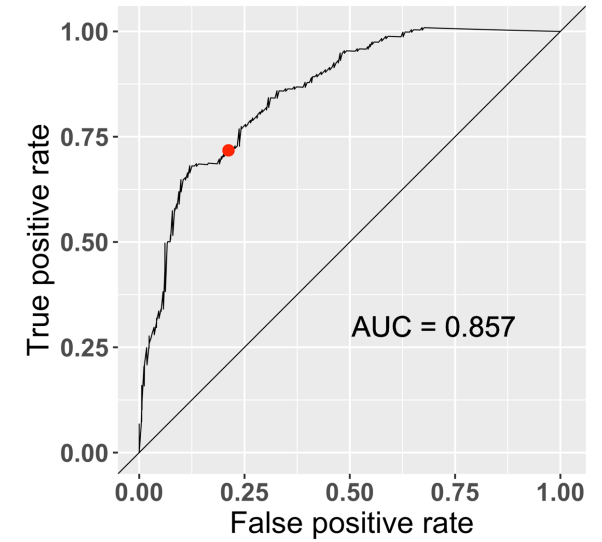
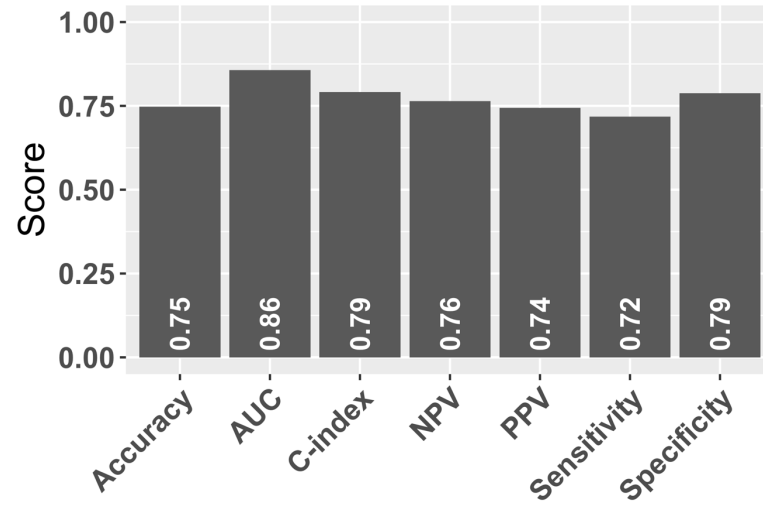
kML individual predictions

mBSL = minimal baseline

Cross-val on train



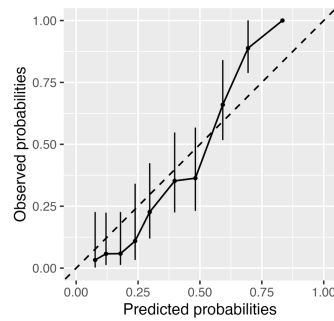
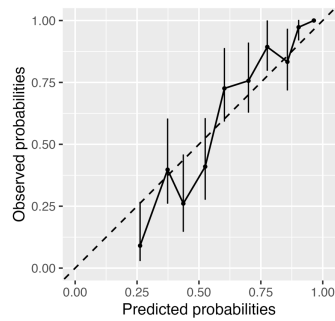
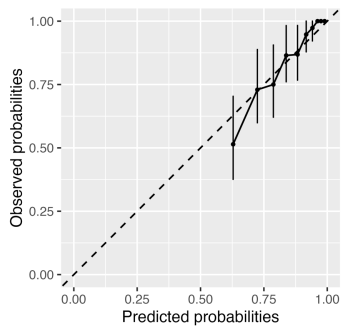
Test set



6 months

12 months

24 months



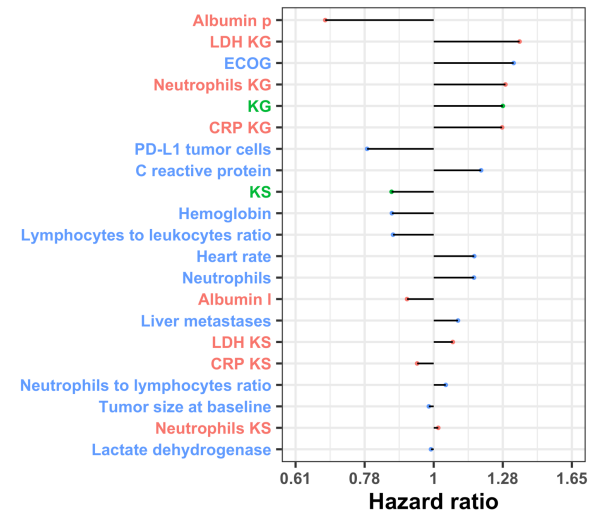
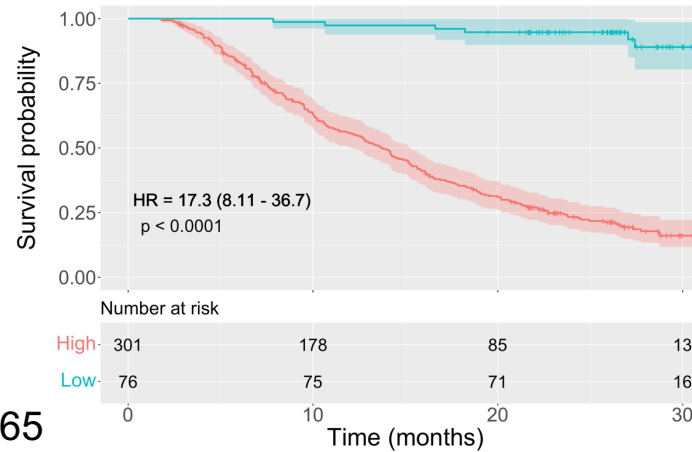
c-index = 0.79
AUC-12 months = 0.86

PDL1 : AUC = 0.65

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

kML

High Low





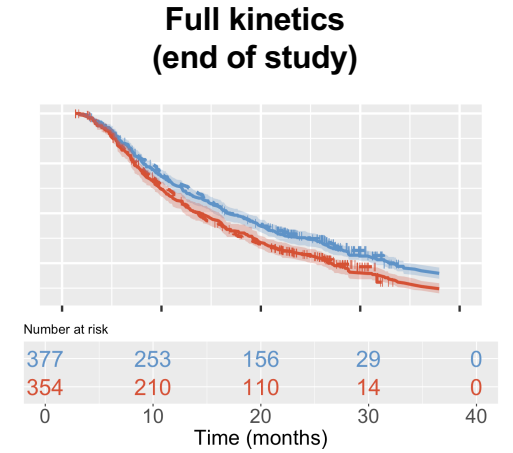
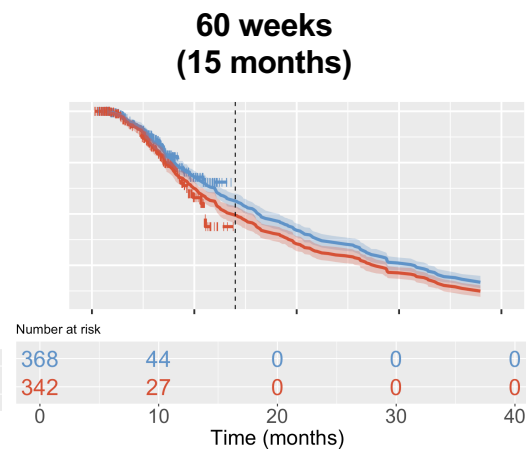
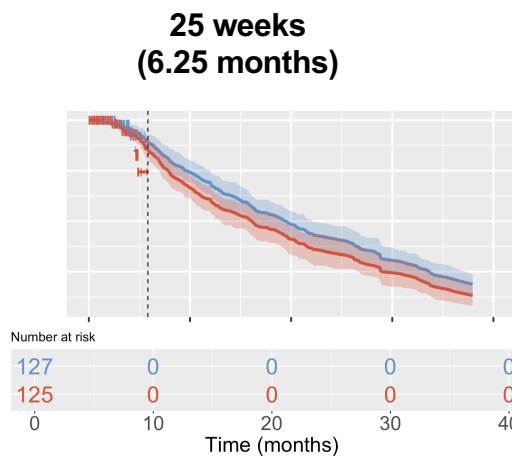
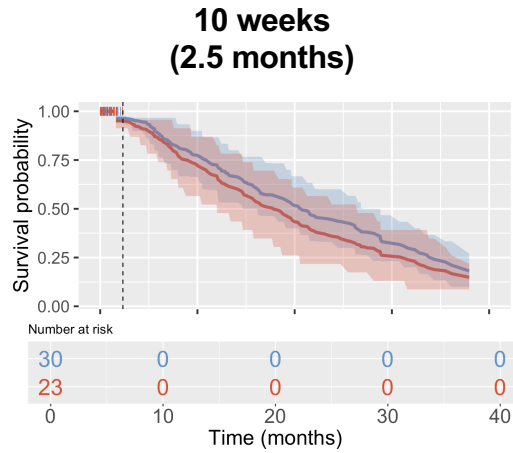
kML predicts phase 3 trial results from early data

Arm

+ Atezolizumab
+ Docetaxel

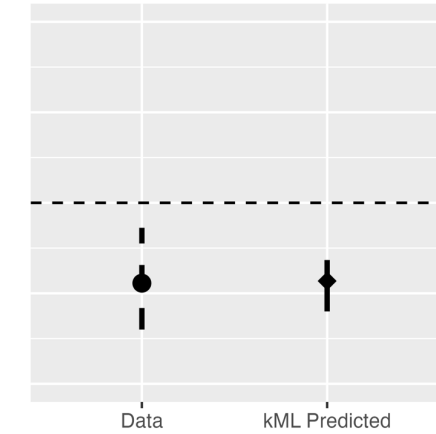
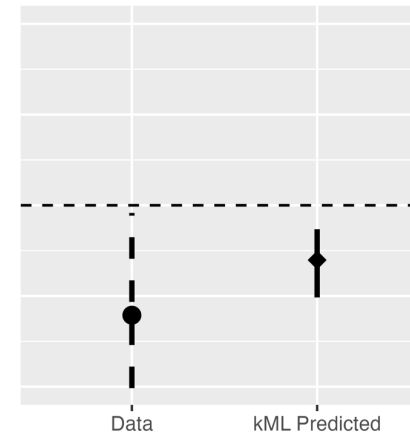
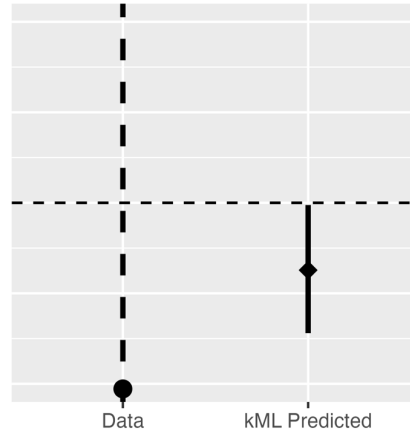
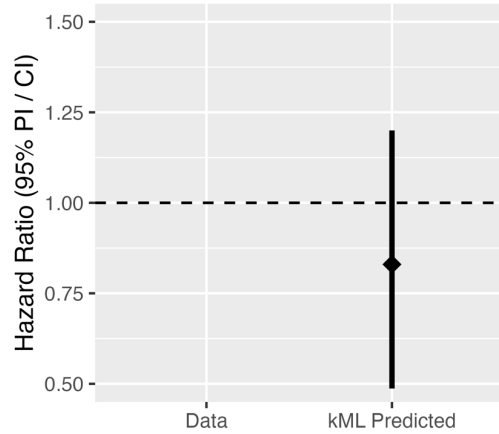
— Model prediction mean
- - Data

95% PI Atezolizumab
95% PI Docetaxel



Hazard ratio

● Data
● kML Predicted

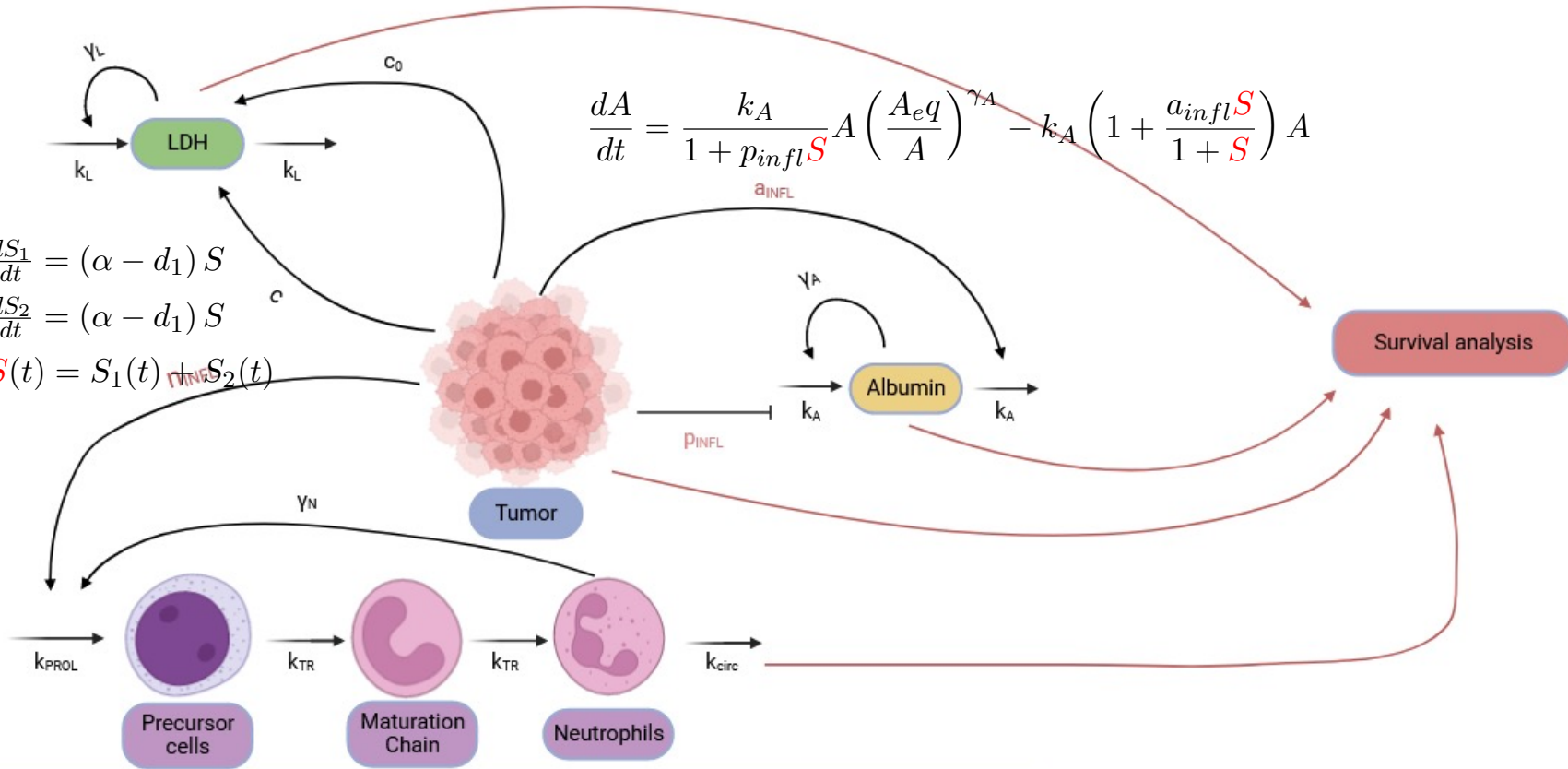


Mechanistic model

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

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

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

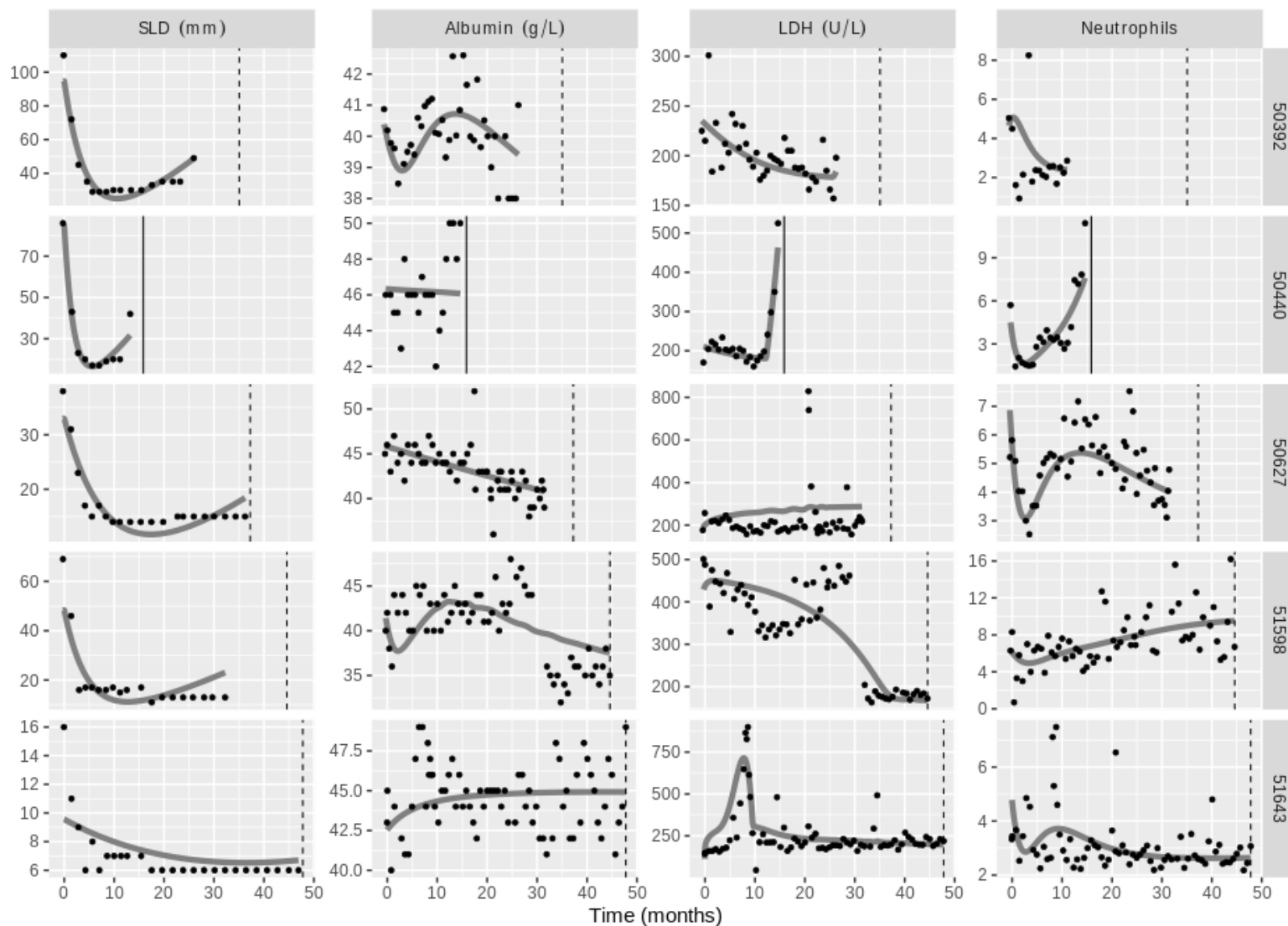


$$\begin{cases} \frac{dP}{dt} = k_{PROL} \left(1 + n_{INFL} \frac{S}{1+S} \right) P \left(\frac{N_{eq}}{N} \right)^{\gamma_N} - k_{TR} P \\ \frac{dT_i}{dt} = k_{TR} T_{i-1} - k_{TR} T_i \\ \frac{dN}{dt} = k_{TR} T_3 - k_{CIRC} N \end{cases}$$

Data: clinical trials for atezolizumab (ATZ) :

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

Individual fits (Combotherapy)



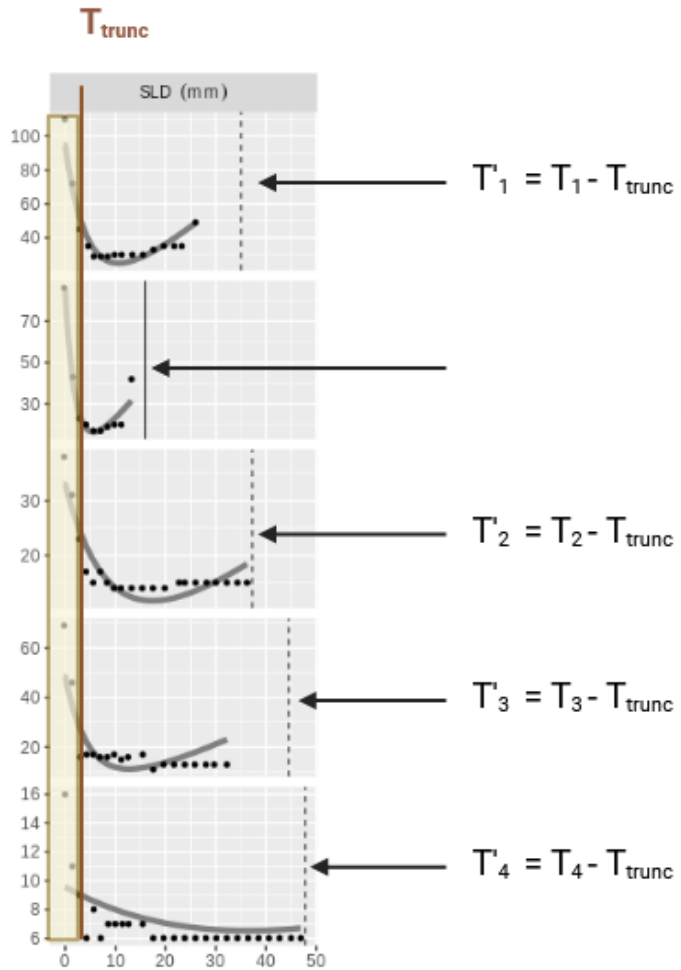
- Fits heterogenous** BK profiles

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

- Identifiability**

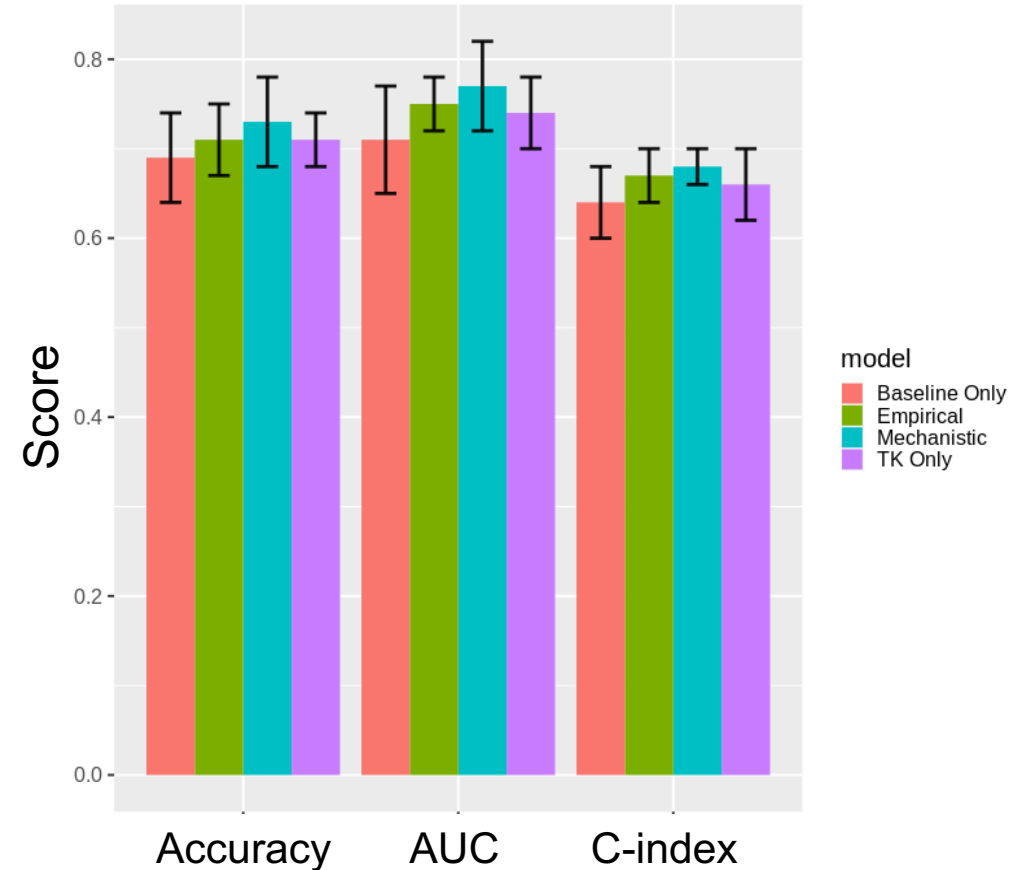
- All RSEs < **35%**
- despite the large number of parameters (14)
- Very large variability of some parameters
- **Small η - shrinkage** < 11%

Post-cycle 4 prediction of individual survival



Longitudinal data considered

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



COMBO

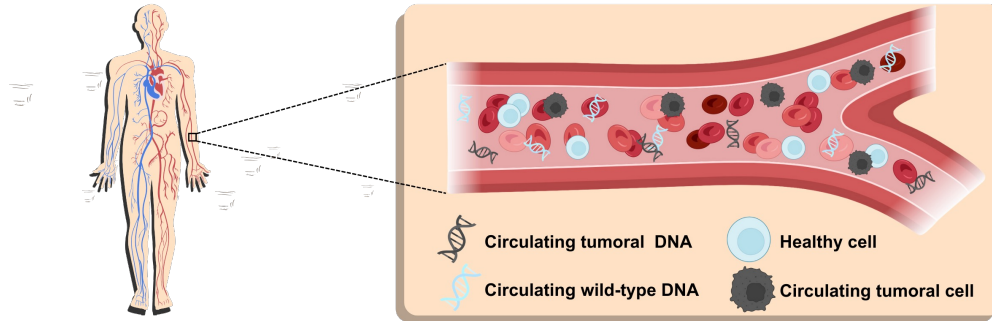
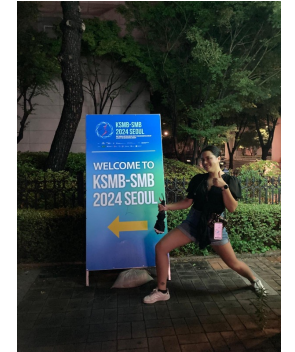
Summary

- **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

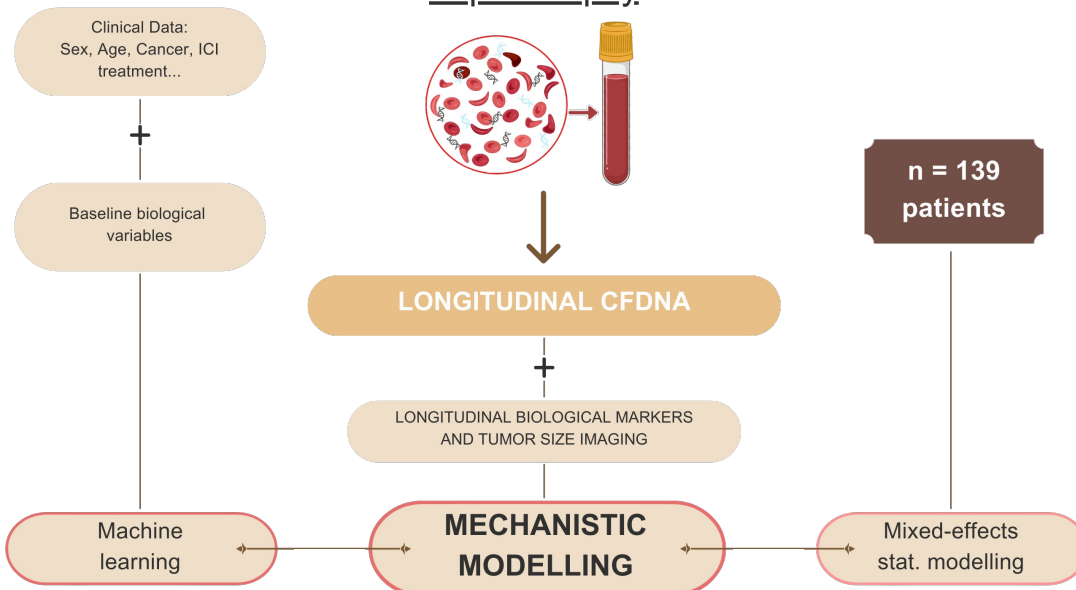
Beyond classical biomarkers: PIONeeR – QUANTIC and SChISM

Size CfDNA Immunotherapies Signature Monitoring

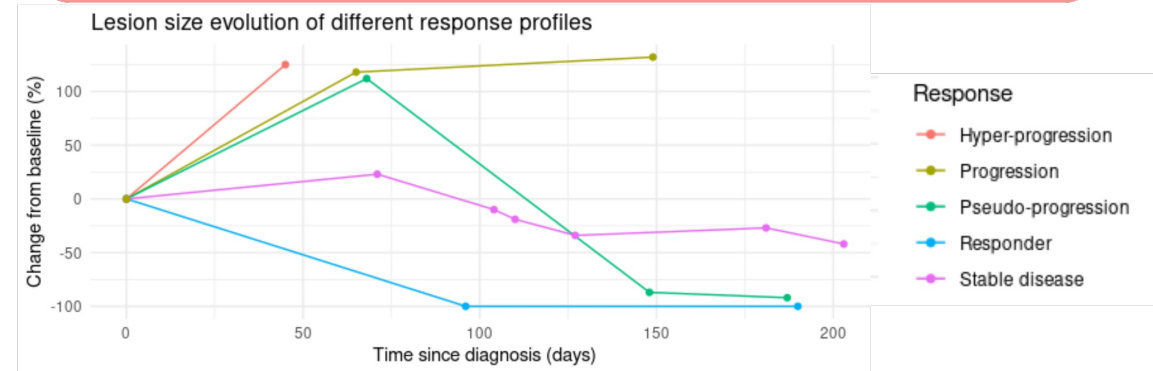
APHM: S. Salas, Inria / Inserm: S. Benzekry, L. Nguyen, Adelis



Liquid biopsy.



Different response profiles and survival prediction

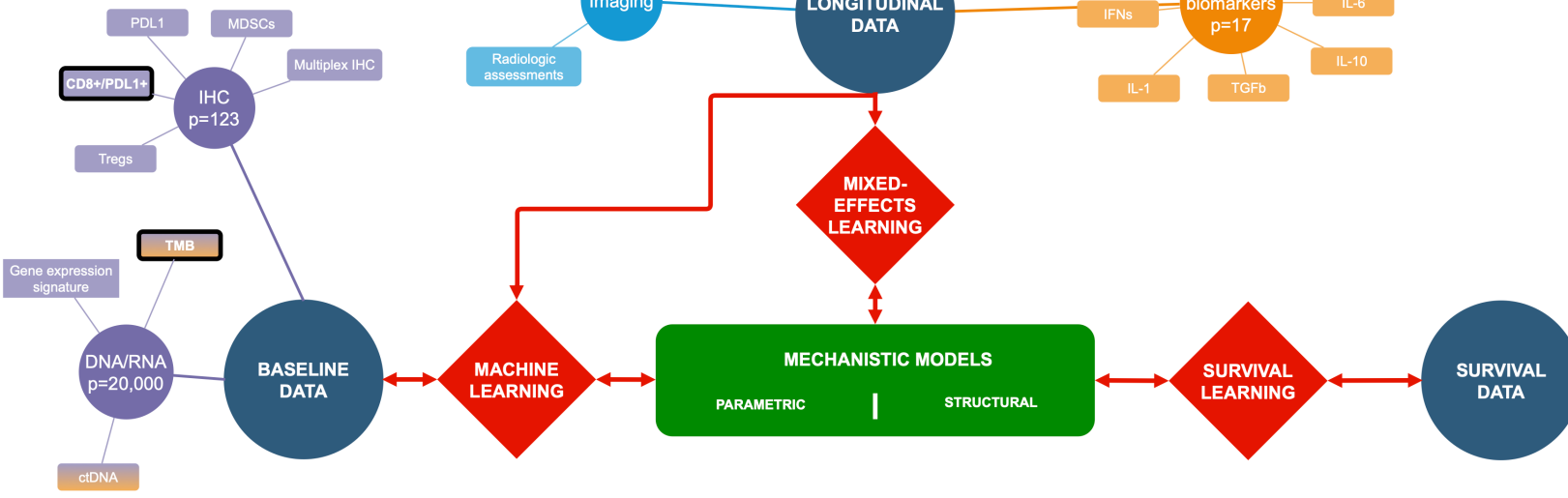
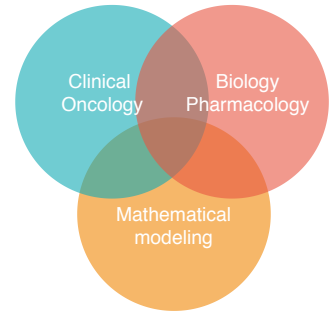
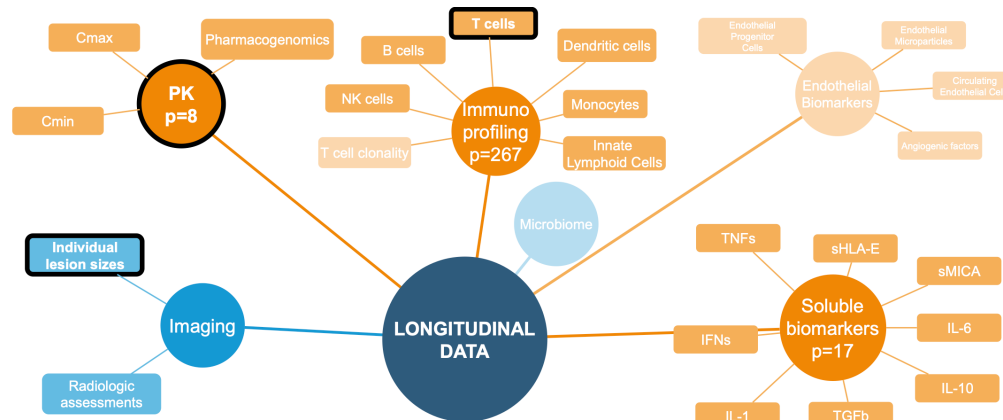
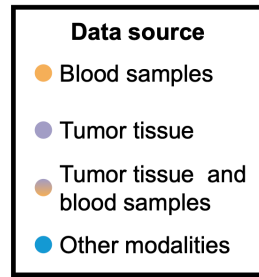
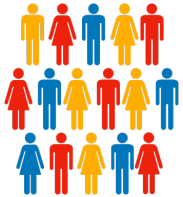


The PIONeeR and QUANTIC project

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



450 pts



24,287 variables

1,581 markers

Acknowledgments

Thank you for your attention!

Roche / Genentech

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



**Open PhD /
postdoc positions**

sebastien.benzekry@inria.fr



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



Patients and families

Computational

C. Bigarre
M. Karlsen
L. Nguyen – Phuong
R. Taieb

Clinical oncology

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

F. Barlesi
F. Bertucci
L. Padovani
P. Tomasini

....