Integrating large- to high- dimension markers in mecanistic models

Mélanie Prague, *Auriane Gabaut, Lisa Crépin, Boris Hejblum, Rodolphe Thiébaut and Cécile proust-Lima*

Pharmacometrics in France – 19 Sept. 2025

BORDEAUX POPULATION HEALTH | Centre de Recherche - U1219

SISTM / Statistiques pour la médecine translationnelle

VACCINE RESEARCH INSTITUTE

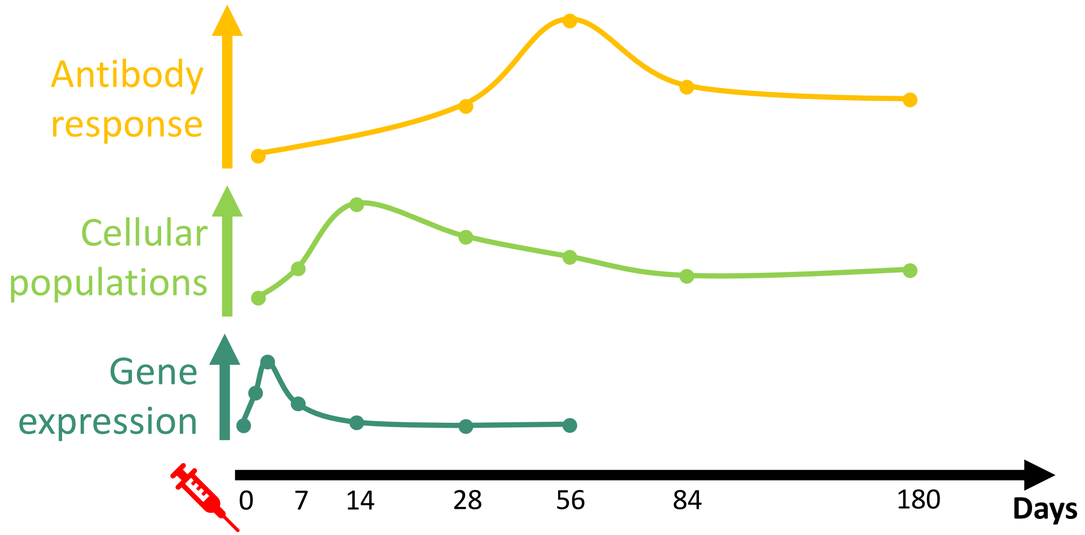
01

INTRODUCTION AND MOTIVATION

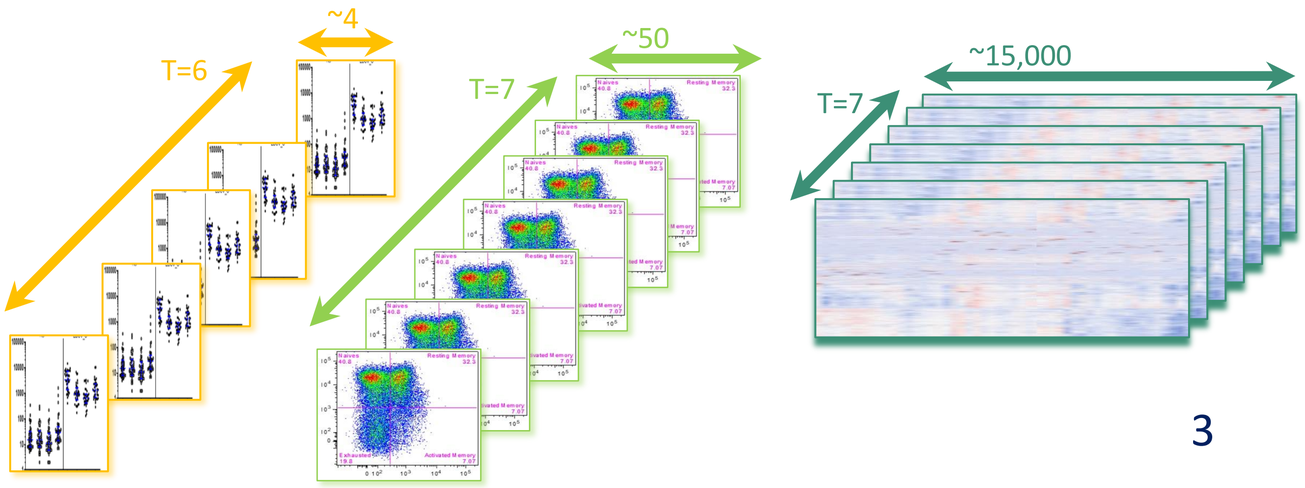
Motivations - Vaccinometrics

Starting point : **Vaccine development for treatment and control of infectious diseases**

- How it works ?
- How long ?
- How to predict the response in each individual ?
- What is the optimal vaccination strategy ?



Sparse longitudinal noisy measures



Mechanistic models – Population approach

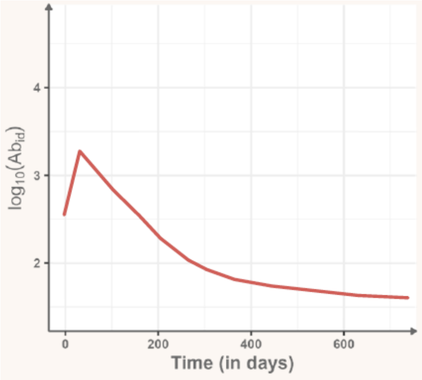
- **Structural model** (ODE-based)

$$\begin{cases} \dfrac{dX_k}{dt} = f_k(X_1, \dots, X_K, \theta, t), & X_k \in X \\ X_k(t=0) = X_{k,0}, & k \in \{1, \dots, K\} \end{cases}$$

Mechanistic models – Population approach

- **Structural model** (ODE-based)

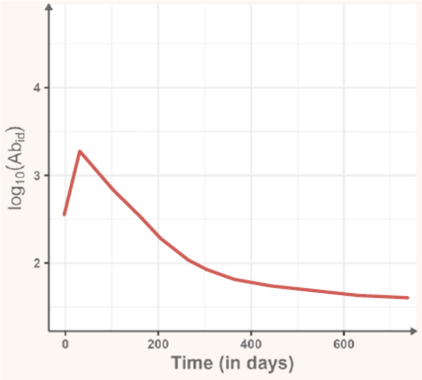
$$\begin{cases} \dfrac{dX_k}{dt} = f_k(X_1, \dots, X_K, \theta, t), & X_k \in X \\ X_k(t=0) = X_{k,0}, & k \in \{1, \dots, K\} \end{cases}$$

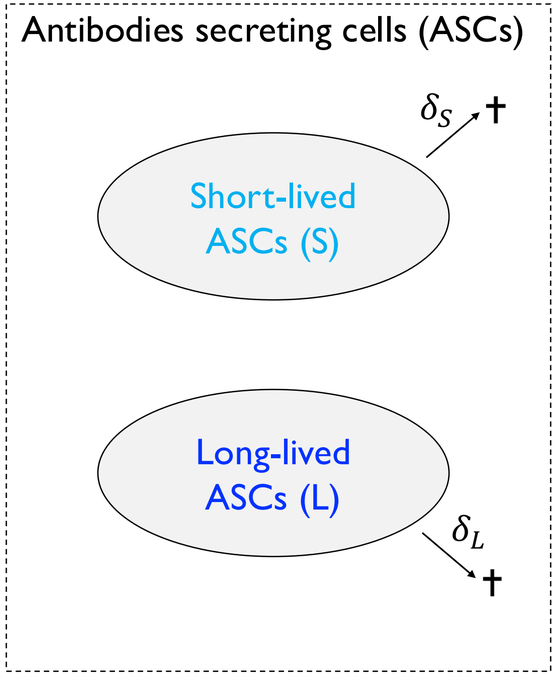


Mechanistic models – Population approach

- **Structural model** (ODE-based)

$$\begin{cases} \dfrac{dX_k}{dt} = f_k(X_1, \dots, X_K, \theta, t), & X_k \in X \\ X_k(t=0) = X_{k,0}, & k \in \{1, \dots, K\} \end{cases}$$





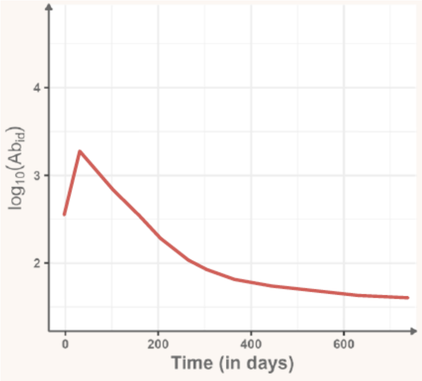
$$\frac{dS}{dt} = -\delta_S\, S \qquad S(t=0) = S_0$$

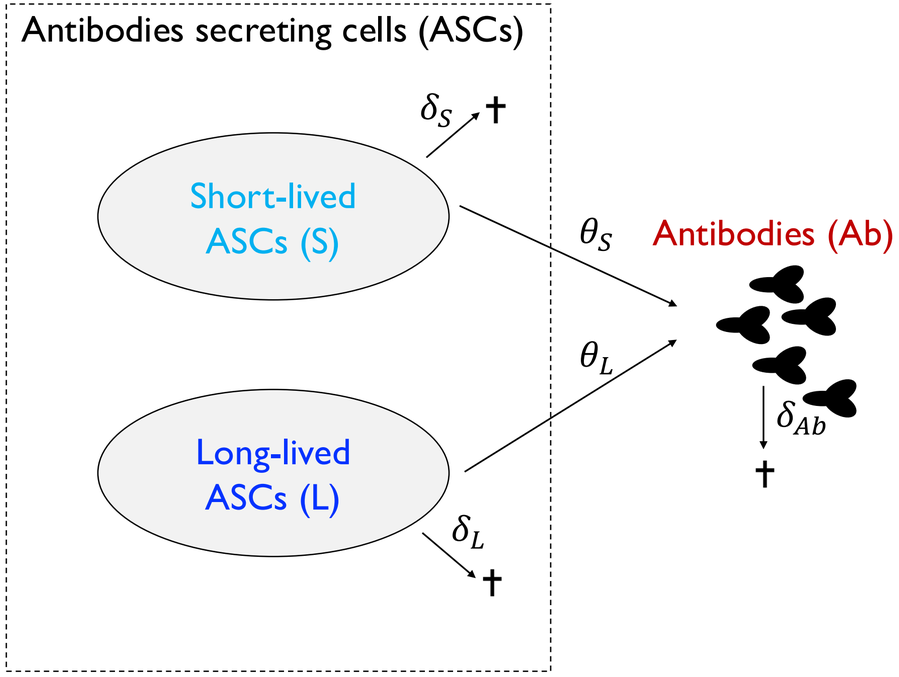
$$\frac{dL}{dt} = -\delta_L\, L \qquad L(t=0) = L_0$$

Mechanistic models – Population approach

- **Structural model** (ODE-based)

$$\begin{cases} \dfrac{dX_k}{dt} = f_k(X_1, \dots, X_K, \theta, t), & X_k \in X \\ X_k(t=0) = X_{k,0}, & k \in \{1, \dots, K\} \end{cases}$$





$$\frac{dS}{dt} = -\delta_S S \qquad S(t=0) = S_0$$

$$\frac{dL}{dt} = -\delta_L L \qquad L(t=0) = L_0$$

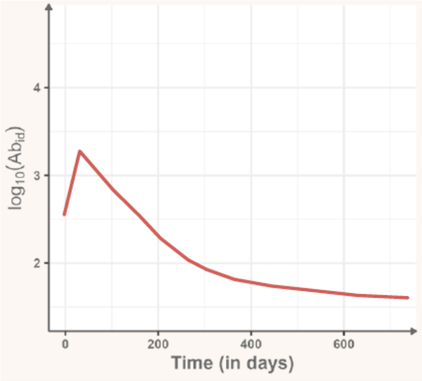
$$\frac{dAb}{dt} = \theta_S S + \theta_L L - \delta_{Ab} Ab \qquad Ab(t=0) = 0$$

$$\theta = (\delta_S, \delta_L, \theta_S, \theta_L, \delta_{Ab})$$

Mechanistic models – Population approach

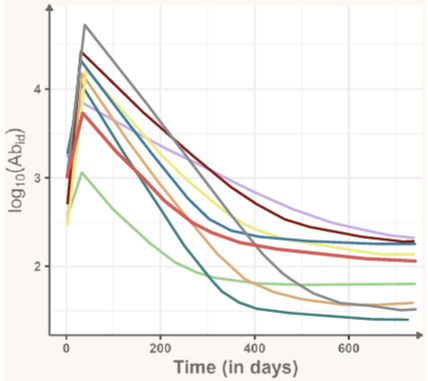
- **Structural model** (ODE-based)

$$\begin{cases} \dfrac{dX_k}{dt} = f_k(X_1, \dots, X_K, \theta, t), & X_k \in X \\ X_k(t=0) = X_{k,0}, & k \in \{1, \dots, K\} \end{cases}$$



- **Statistical model** (linear mixed effect model)

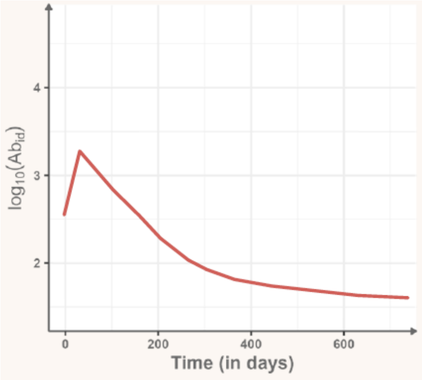
$$g\left(\theta_i(t)\right) = g(\theta_0) + \phi Z_i(t) + u_i$$
$$u_i \sim N(0, \Omega), \qquad i \in \{1, \dots, N\}$$



Mechanistic models – Population approach

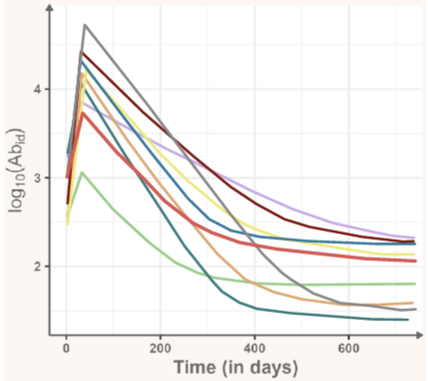
- **Structural model** (ODE-based)

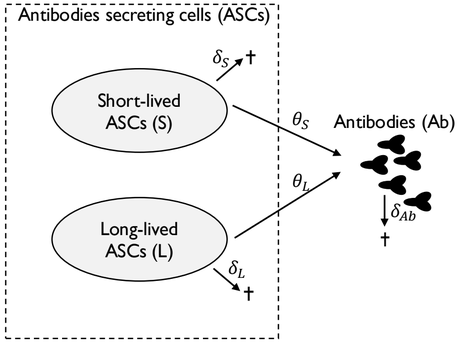
$$\begin{cases} \dfrac{dX_k}{dt} = f_k(X_1, \dots, X_K, \theta, t), & X_k \in X \\ X_k(t=0) = X_{k,0}, & k \in \{1, \dots, K\} \end{cases}$$



- **Statistical model** (linear mixed effect model)

$$g\big(\theta_i(t)\big) = g(\theta_0) + \phi Z_i(t) + u_i$$
$$u_i \sim N(0, \Omega), \qquad i \in \{1, \dots, N\}$$





Ensure positiveness of rates

Effect of covariates

$$log(\theta_{S\,i}) = \log(\theta_{S\,0}) + \beta_{AGE}\, AGE_i + u_i$$

$u_i \sim N(0, \omega)$

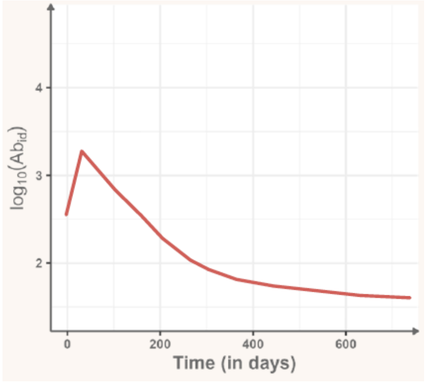
Residual heterogeneity
Random effect

Population average value

Mechanistic models – Population approach

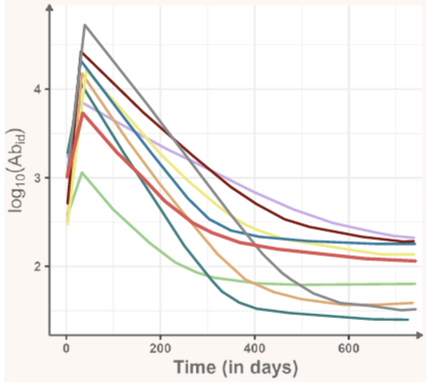
- **Structural model** (ODE-based)

$$\begin{cases} \dfrac{dX_k}{dt} = f_k(X_1, \dots, X_K, \theta, t), & X_k \in X \\ X_k(t=0) = X_{k,0}, & k \in \{1, \dots, K\} \end{cases}$$



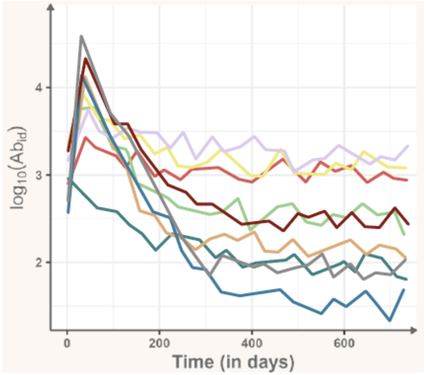
- **Statistical model** (linear mixed effect model)

$$g\left(\theta_i(t)\right) = g(\theta_0) + \phi Z_i(t) + u_i$$
$$u_i \sim N(0, \Omega), \qquad i \in \{1, \dots, N\}$$



- **Observation model** (Error model)

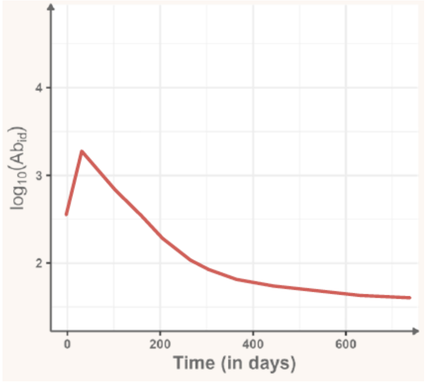
$$Y_{ij} = Y(t_{ij}) = h\left(X(t_{ij}, \theta)\right) + g\left(X(t_{ij}, \theta), \Sigma\right)\epsilon_{ij}$$
$$\epsilon_{ij} \sim N(0,1)$$



Mechanistic models – Population approach

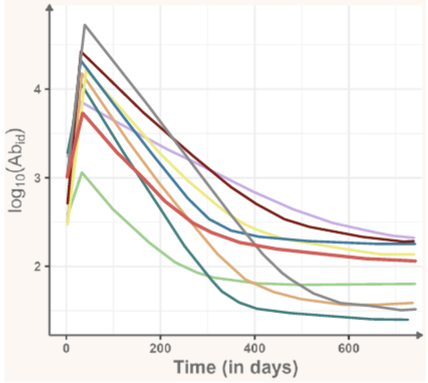
- **Structural model** (ODE-based)

$$\begin{cases} \dfrac{dX_k}{dt} = f_k(X_1, \dots, X_K, \theta, t), & X_k \in X \\ X_k(t=0) = X_{k,0}, & k \in \{1, \dots, K\} \end{cases}$$



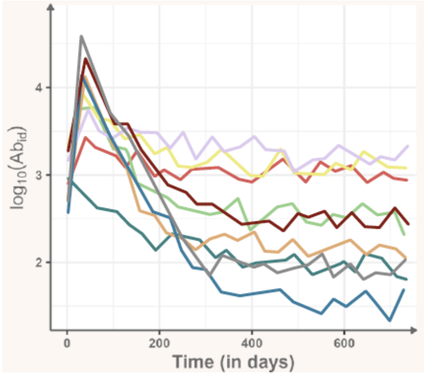
- **Statistical model** (linear mixed effect model)

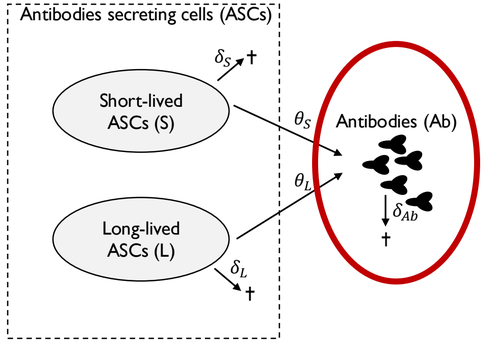
$$g\big(\theta_i(t)\big) = g(\theta_0) + \phi Z_i(t) + u_i$$
$$u_i \sim N(0, \Omega), \qquad i \in \{1, \dots, N\}$$



- **Observation model** (Error model)

$$Y_{ij} = Y(t_{ij}) = h\Big(X(t_{ij}, \theta)\Big) + g\big(X(t_{ij}, \theta), \Sigma\big)\epsilon_{ij}$$
$$\epsilon_{ij} \sim N(0,1)$$





$$Y_{ij} = \log_{10}\big(Ab(\theta_i, t)\big) + \epsilon_{ij}$$
$$\epsilon_{ij} \sim N(0, \sigma)$$

Mechanistic models – Population approach

- **Structural model** (ODE-based)

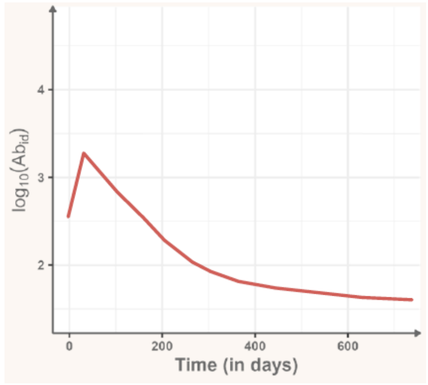
$$\begin{cases} \dfrac{dX_k}{dt} = f_k(X_1, \dots, X_K, \theta, t), & X_k \in X \\ X_k(t=0) = X_{k,0}, & k \in \{1, \dots, K\} \end{cases}$$

- **Statistical model** (linear mixed effect model)

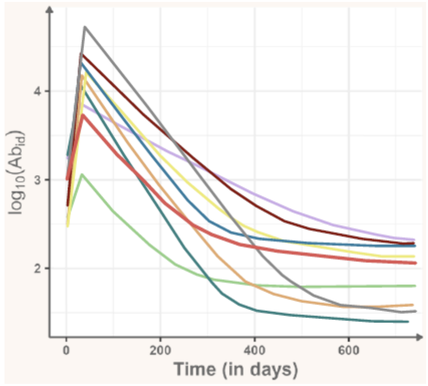
$$g\big(\theta_i(t)\big) = g(\theta_0) + \phi Z_i(t) + u_i$$
$$u_i \sim N(0, \Omega), \qquad i \in \{1, \dots, N\}$$

- **Observation model** (Error model)

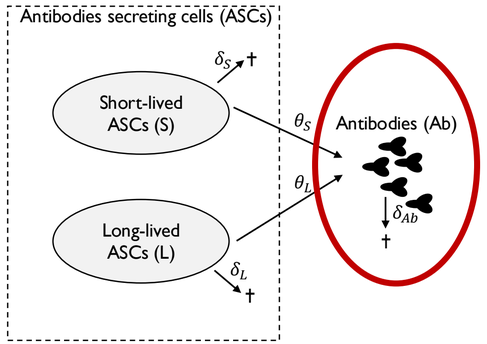
$$Y_{ij} = Y(t_{ij}) = h\Big(X(t_{ij}, \theta)\Big) + g\big(X(t_{ij}, \theta), \Sigma\big)\epsilon_{ij}$$
$$\epsilon_{ij} \sim N(0,1)$$



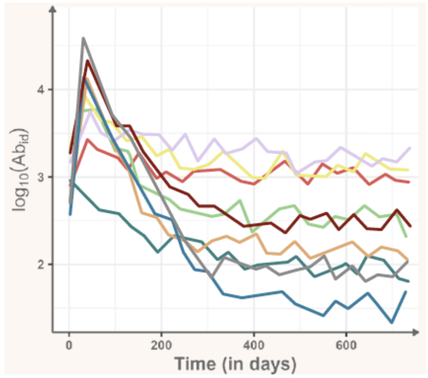
...



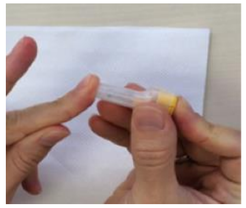
Inverse



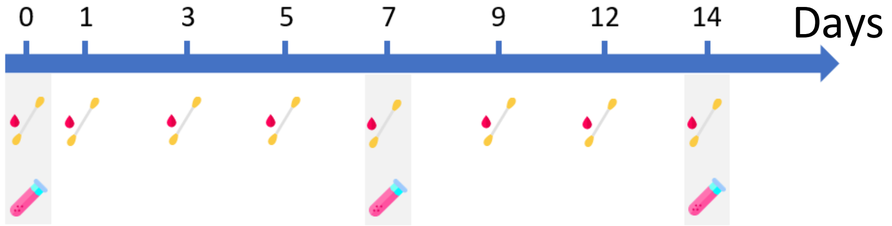
$$Y_{ij} = \log_{10}\big(Ab(\theta_i, t)\big) + \epsilon_{ij}$$
$$\epsilon_{ij} \sim N(0, \sigma)$$

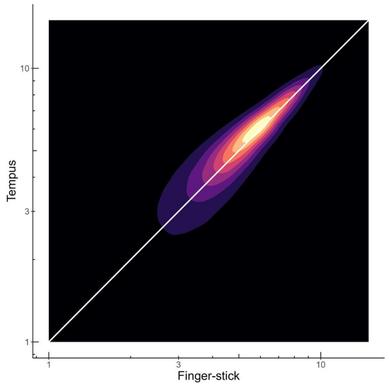


Rationale & Objective

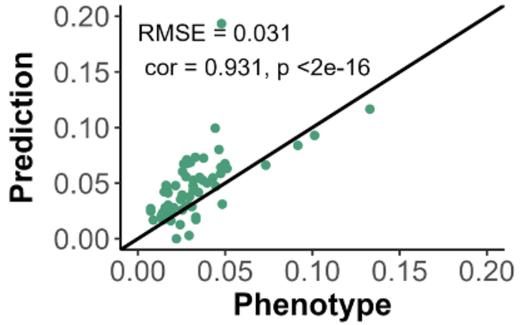


High frequency transcriptomics (RNA-seq) from self collected finger-prick blood
Study: COVERAGE-Immuno





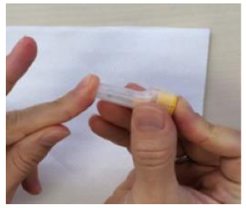
Good concordance between the **whole-blood** Tempus and Finger **prick-test**



Reasonable concordance between **cell abundance** by ICS and from transcriptomic data after deconvolution

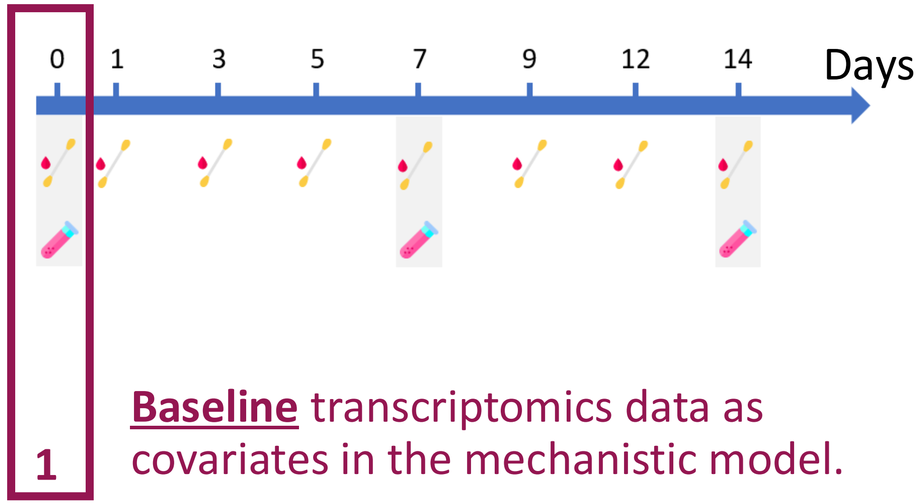
Lhomme et al. Submitted

Rationale & Objective



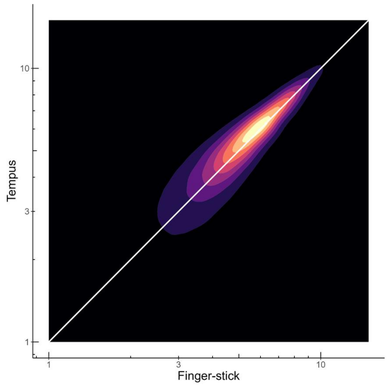
High frequency transcriptomics (RNA-seq) from self collected finger-prick blood

Study: COVERAGE-Immuno

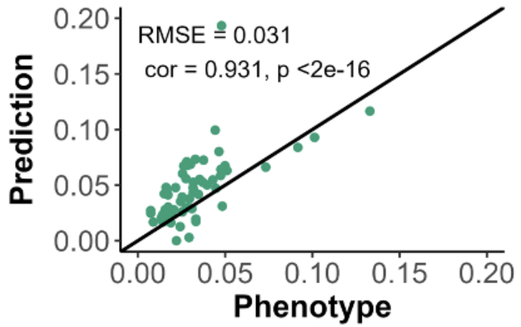


Baseline transcriptomics data as covariates in the mechanistic model.

R package Lasso-SAMBA



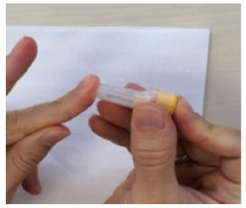
Good concordance between the **whole-blood** Tempus and Finger **prick-test**



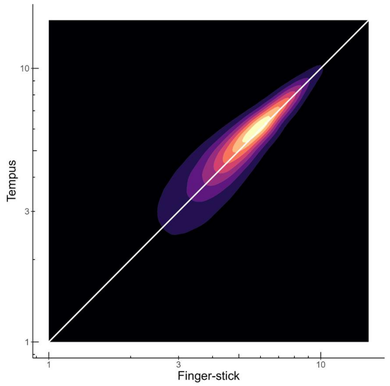
Reasonable concordance between **cell abundance** by ICS and from transcriptomic data after deconvolution

Lhomme et al. Submitted

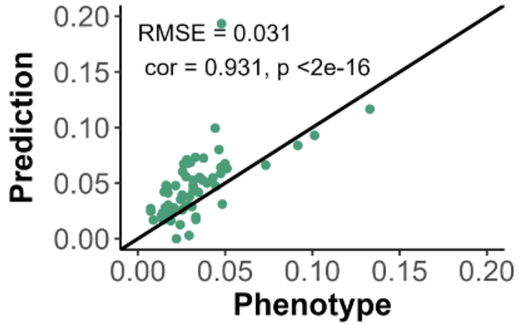
Rationale & Objective



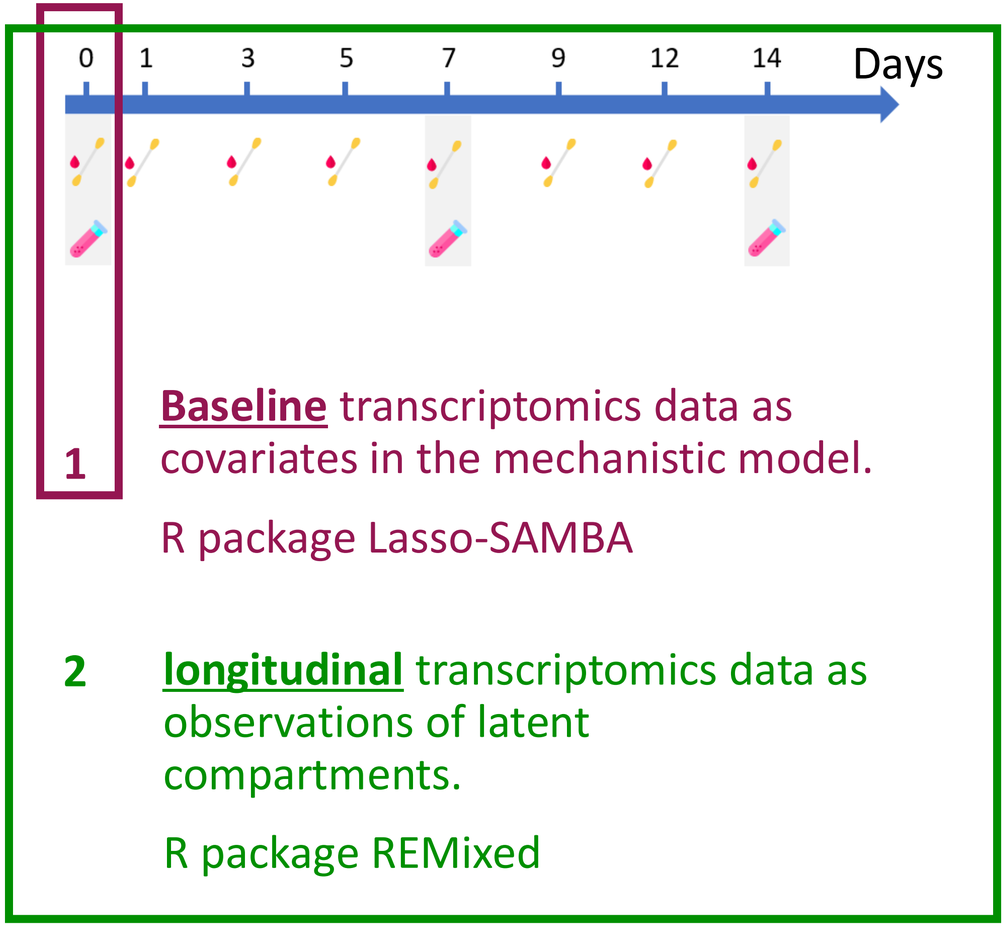
High frequency transcriptomics (RNA-seq) from self collected finger-prick blood
Study: COVERAGE-Immuno



Good concordance between the **whole-blood** Tempus and Finger **prick-test**



Reasonable concordance between **cell abundance** by ICS and from transcriptomic data after deconvolution



Lhomme et al. Submitted

02

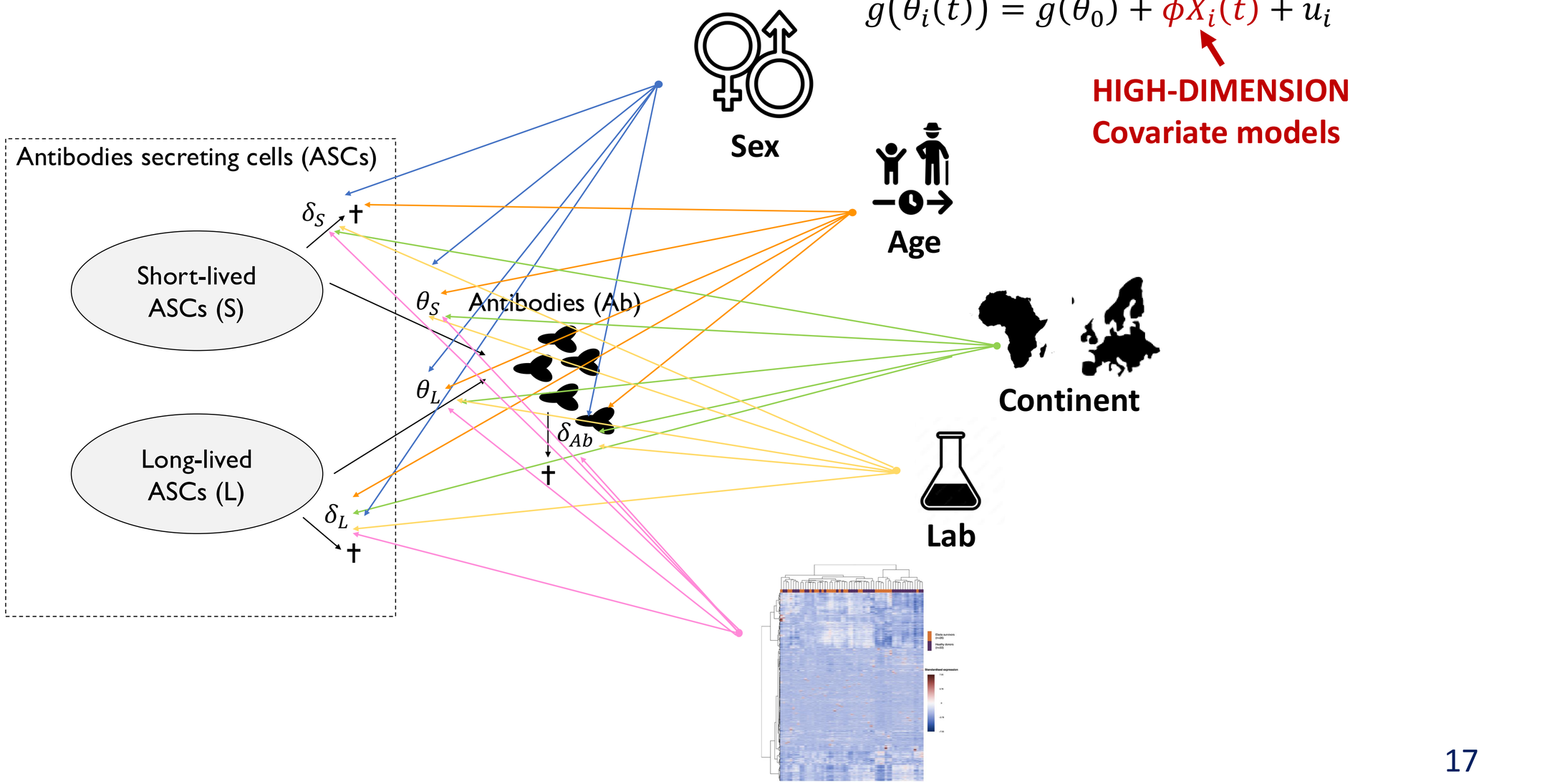
Using transcriptomic data as explanatory covariates (lasso-SAMBA Package)

Model Building strategy

Use **baseline** Gene expression as explanatory covariates

$$g(\theta_i(t)) = g(\theta_0) + \phi X_i(t) + u_i$$

HIGH-DIMENSION
Covariate models

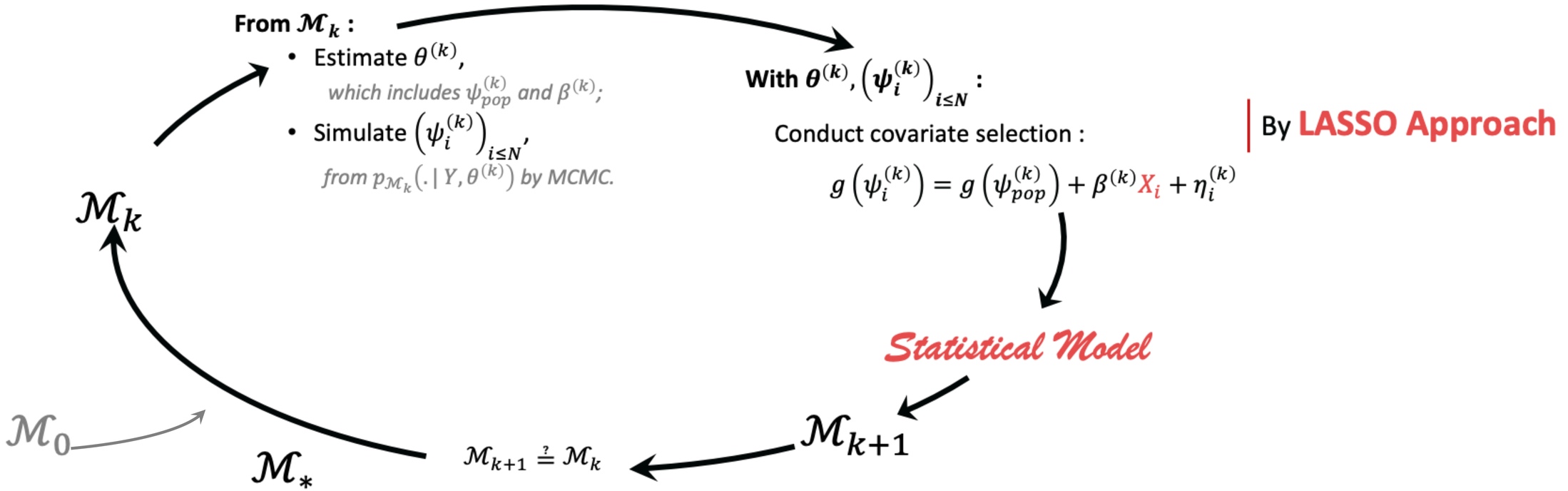


Method

Algorithms :

- SCM, *Stepwise Covariate Modeling* (Svensson and Jonsson, 2022) ;
- COSSAC, *COnditional Sampling use for Stepwise Approach based on Correlation tests* (Ayral and al. 2021) ;
- **SAMBA (Prague and Lavielle, 2022).**

SAMBA, *Stochastic Approximation for Model Building Algorithm*, is an **iterative** algorithm, learning from a previous "worse" model in order to move towards a relevant model (Prague and Lavielle, 2022).



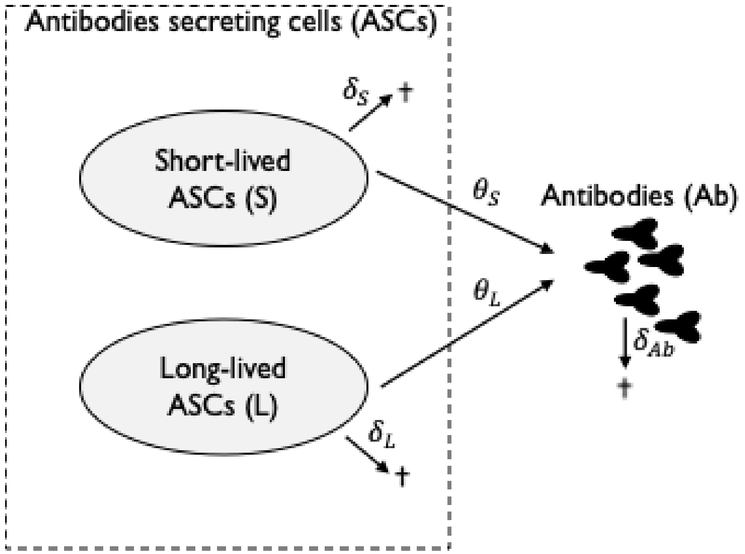
Tibshirani et al. 1996; Meinshausen et al. 2010; Bodiner et al. 2023

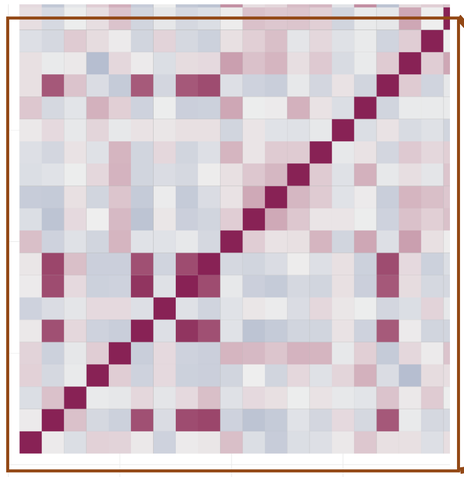
Simulations

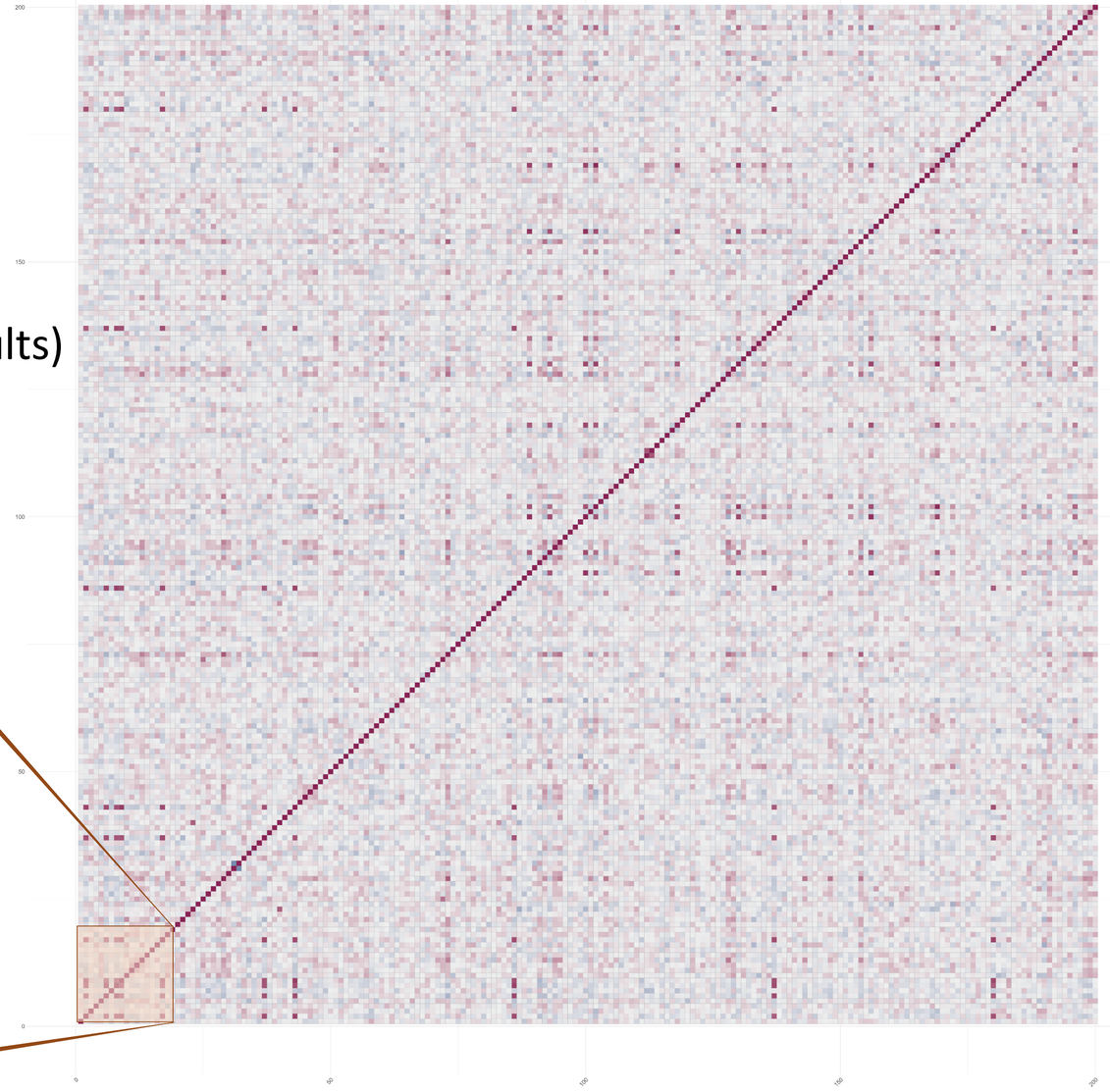
Covariates effetcs :

- $AGE \sim \mathcal{N}(35,4^2)$ on φ_S,
- $G_1 \sim \mathcal{N}(0,1)$ on φ_L
- $G_2 \sim \mathcal{N}(0,1)$ on δ_{Ab}

P = 200 gaussian **correlated** covariates (P=1000 similar results)
R = 100 replicates,
N = 100 individuals (N=20 similar results).







Pasin et al. 2020; Alexandre et al. 2024

Simulations

Parameter-Covariate link	Selected in the final model	NOT **selected in the final model**
In the generation model	True Positive (TP)	False Negative (FN)
NOT **in the generation model**	False Positive (FP)	True Negative (TN)

False Discovery Rate : $FDR = \frac{FP}{TP + FP}$

False Negative Rate : $FNR = \frac{FN}{TN + FN}$

F1-score : $F1_{score} = \frac{2TP}{2TP + FN + FP}$

Simulations

Error Rate Comparison Table

Rate	Median	Confidence Interval (quantiles 95%)
False Discovery Rate : - stepAIC with stat. test - Lasso : E[FDR]<10%	 72.7% 0.0%	 [50.0%;83.8%] [0.0%;25.0%]
False Negative Rate : - stepAIC with stat. test - Lasso : E[FDR]<10%	 0.0% 0.0%	 [0.0%;0.0%] [0.0%;0.0%]
F1 score : - stepAIC with stat. test - Lasso : E[FDR]<10%	 42.9% 100.0%	 [27.9%;66.7%] [85.7%;100.0%]

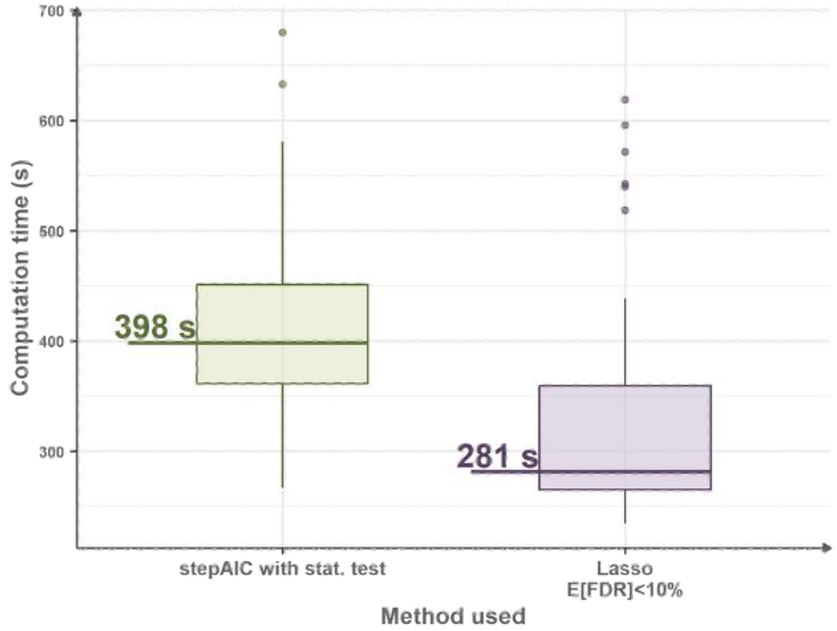
- Final Final model without any False Negatives :
  - stepAIC with stat. test : 99%
  - Lasso : E[FDR]<10% : 100%
- Final model is the true one :
  - stepAIC with stat. test : 0%
  - Lasso : E[FDR]<10% : 81%

Covariates presence in final model with parameters link

Among 100 simulated datasets of humoral immune response to prime-boost of Ad26.ZEBOV/MVA-BN-FILO vaccine against Ebola for 100 individuals, with 200 gaussian correlated covariates.

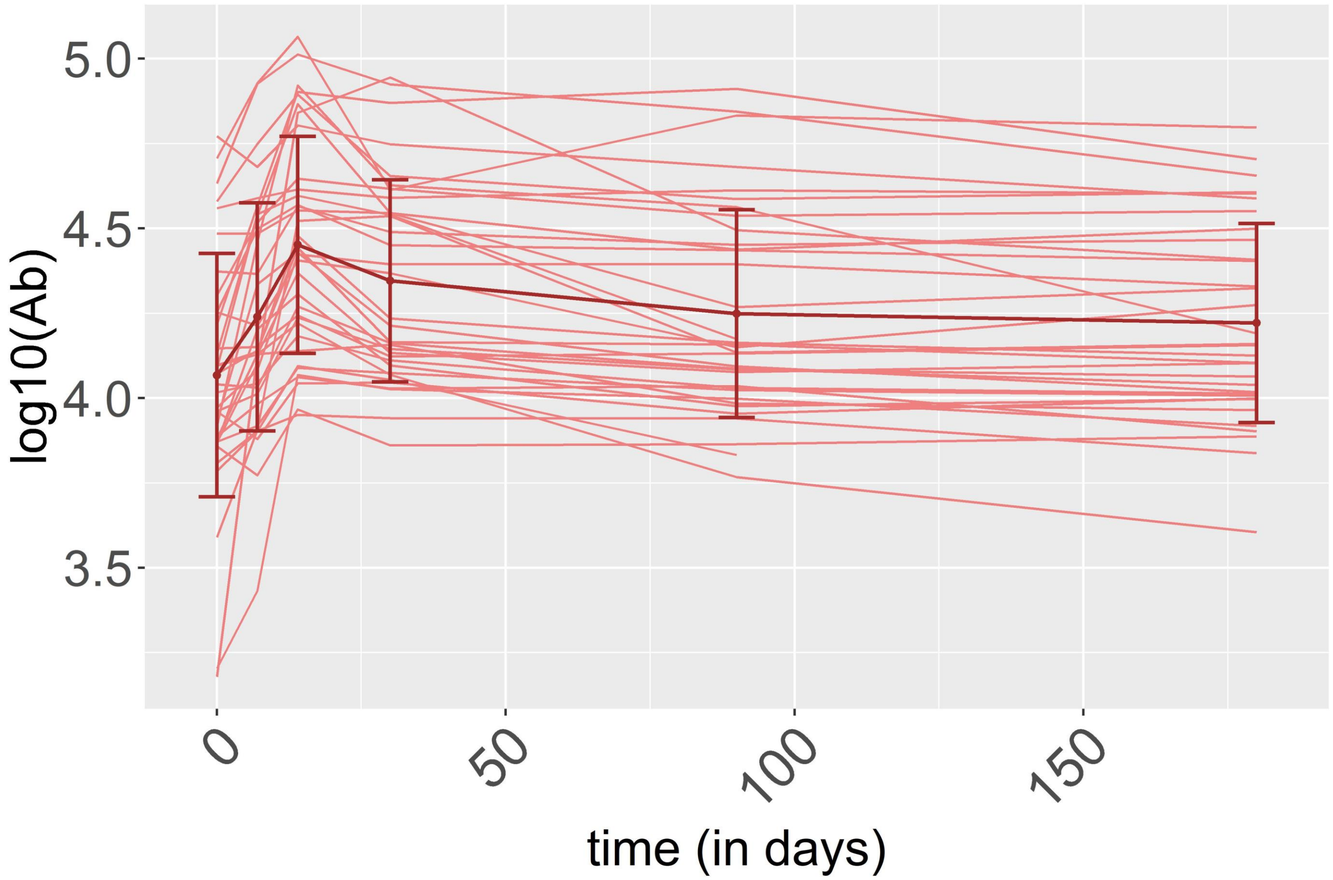
Computation Time Comparison

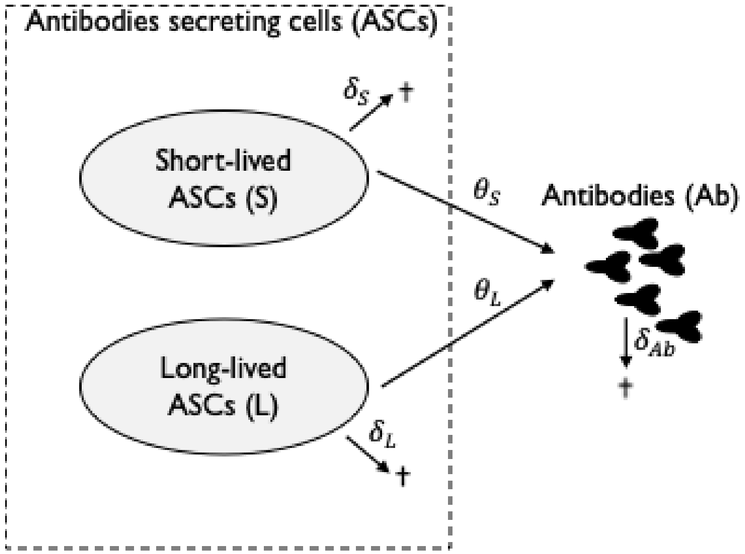
Among 100 simulated datasets of humoral immune response to prime-boost of Ad26.ZEBOV/MVA-BN-FILO vaccine against Ebola for 100 individuals, with 200 gaussian correlated covariates.



Application to vaccine study : VZV

- Clinical study of immune response to vaccination against the **Varicella-Zoster Virus (VZV)**
- **Gene expression and antibody response data** following immunization with ZOSTAVAX, a live attenuated vaccine.
- **35 adult** volunteers, **6 datapoints** at day 0, 7, 14, 30, 90, and 180

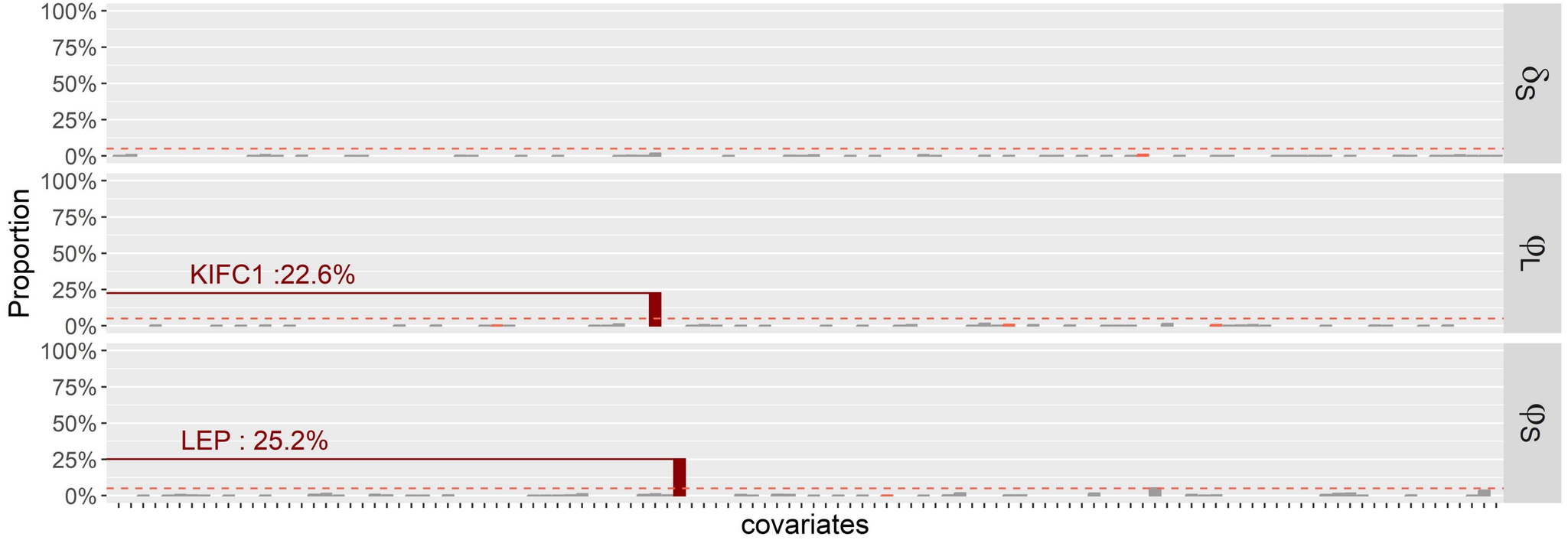




Application to vaccine study : VZV

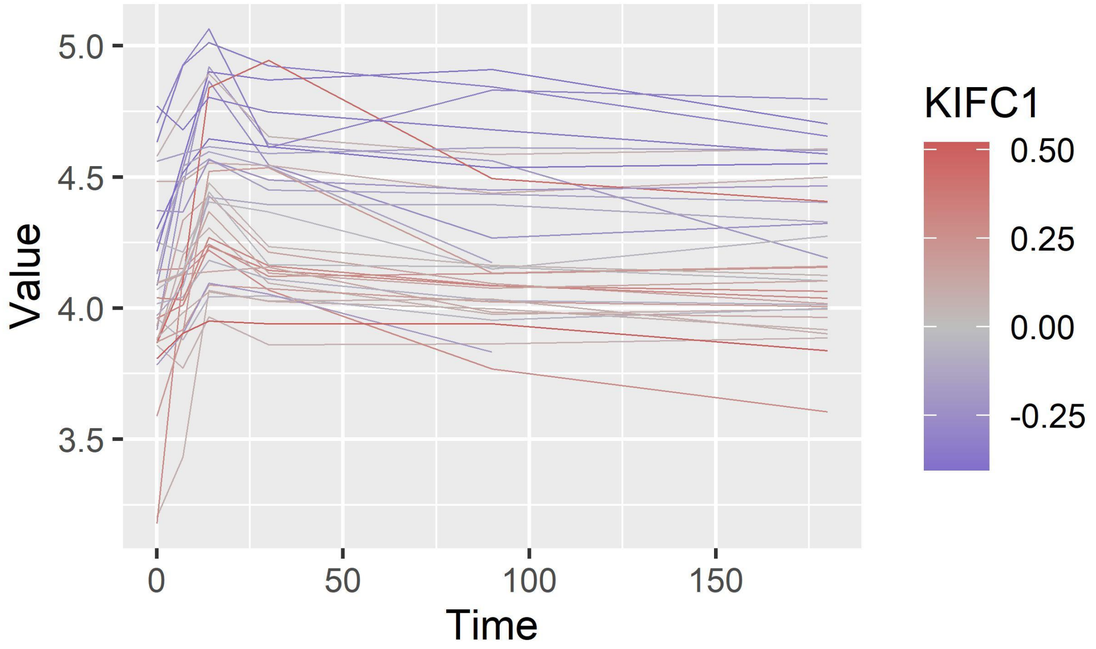
- **10,086 profiled genes**
- a subset of **784 protein-coding genes** was selected based on functional annotation with roles in
  - **Interferon signaling,**
  - **Type I Interferon response,**
  - **Neutrophil activation,**
  - **Inflammation,**
  - **Cytokine/chemokine activity,**
  - **and Cell cycle regulation**

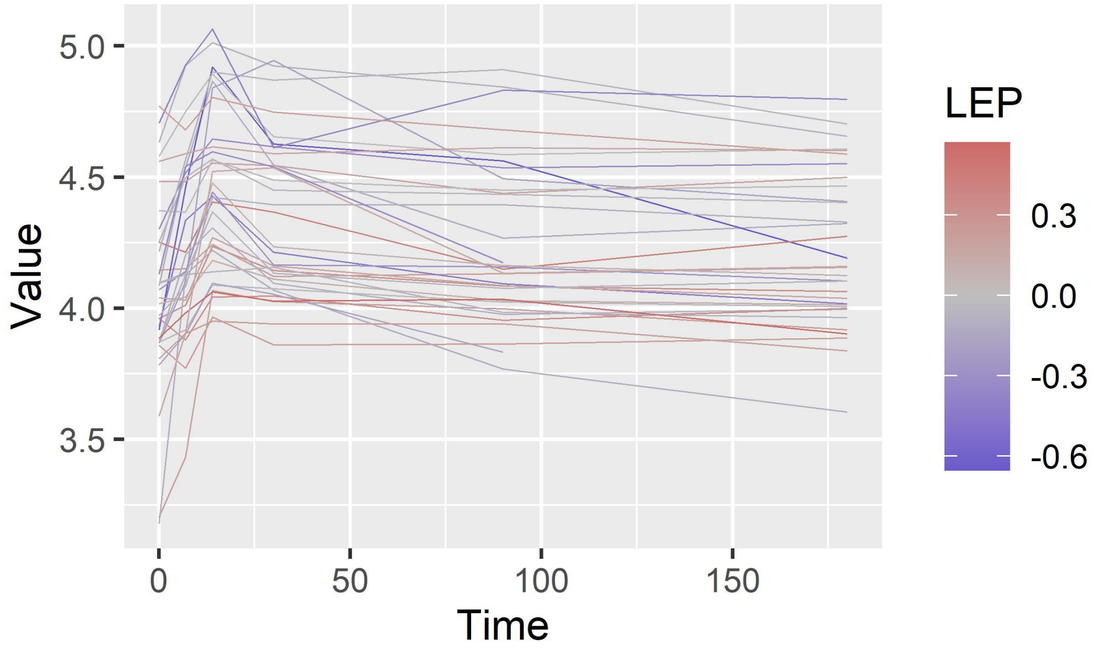
ASSOCIATED GENES : KIFC1 & LEP



Interpretation

- The LEP gene is a player in several common biological pathways involved in the immune response such as JAK-STAT or NFkB.
- KIFC1 is involved in cell proliferation and therefore is not specific of immune response. KIFC1 also has the ability to promote stable mitotic spindle formation during early B cell development where centriole duplication is frequent but must be tightly regulated





Francisco et al. Obesity, Fat Mass and Immune System: Role for Leptin. Frontiers in Physiology. 2018
Lam et al. Role of leptin in immunity. Cell Mol Immunol. 2007
Lucanus et al. Kinesin superfamily: roles in breast cancer, patient prognosis and therapeutics. Oncogene. 2018
Wu et al. An integrative pan-cancer analysis of kinesin family member C1 (KIFC1) in human tumors. Biomedicines. 2022
Hagan et al. Transcriptional atlas of the human immune response to 13 vaccines reveals predictor of vaccine-induced Ab responses. Nature Immunology. 2022

03

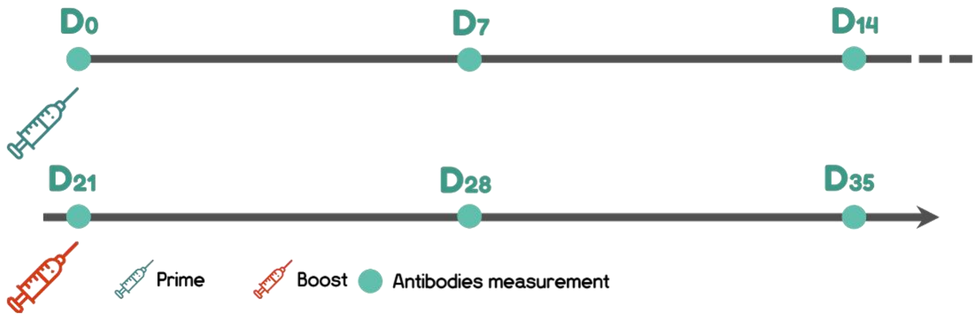
Toward high dimension mechanistic models using latent class models

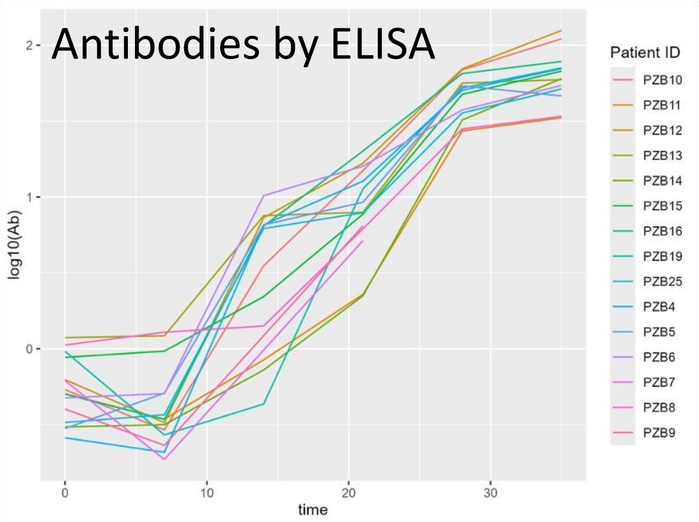
(REmixed Package)

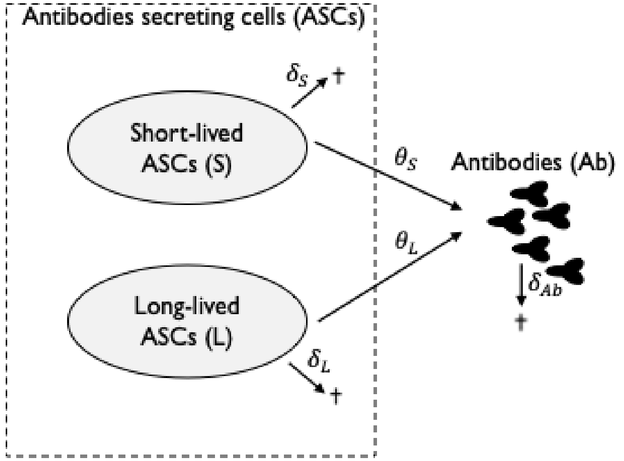
Motivating study

Evaluation of COVID-19 vaccine

15 adults receiving COVID-19 Pfizer vaccine





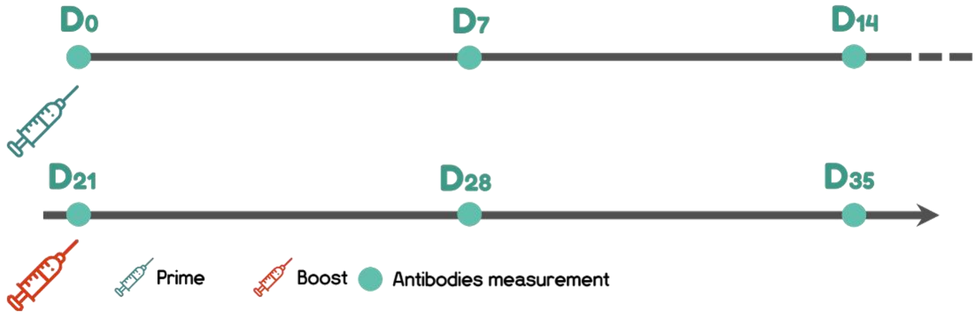


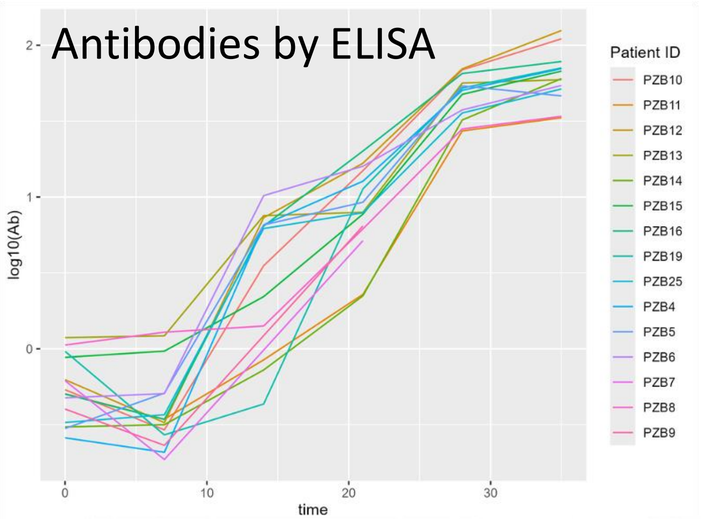
Rinchai et al. 2022

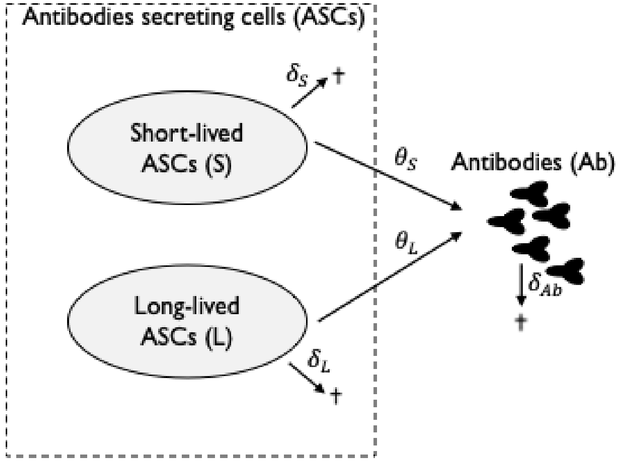
Motivating study

Evaluation of COVID-19 vaccine

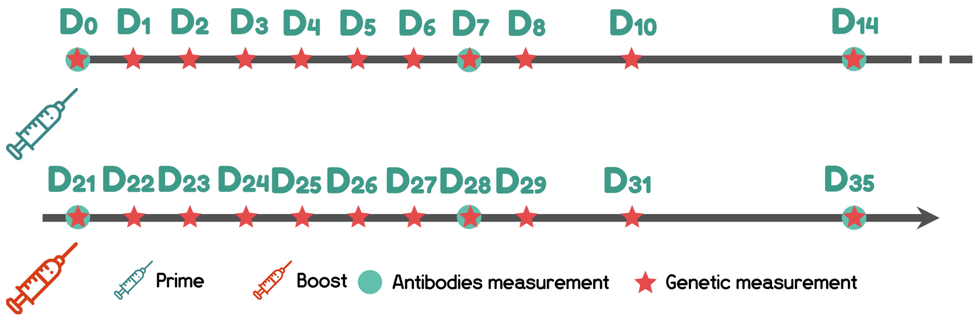
15 adults receiving COVID-19 Pfizer vaccine

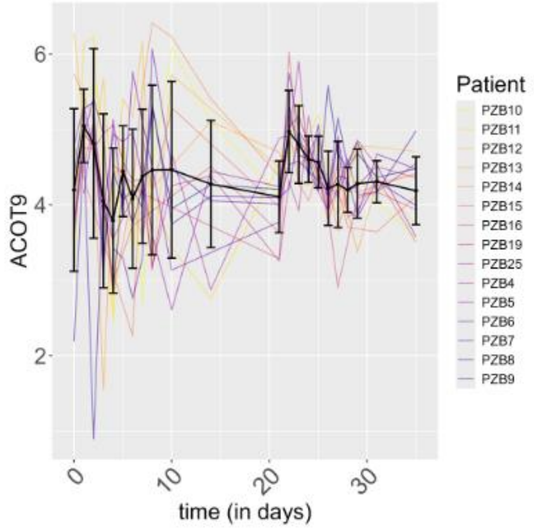


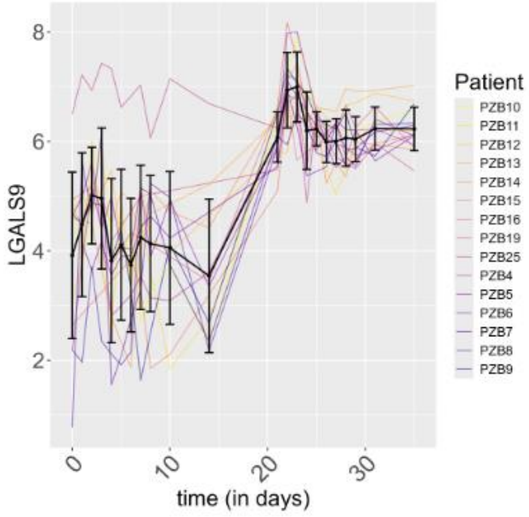




6000+ genes; 34 gene sets



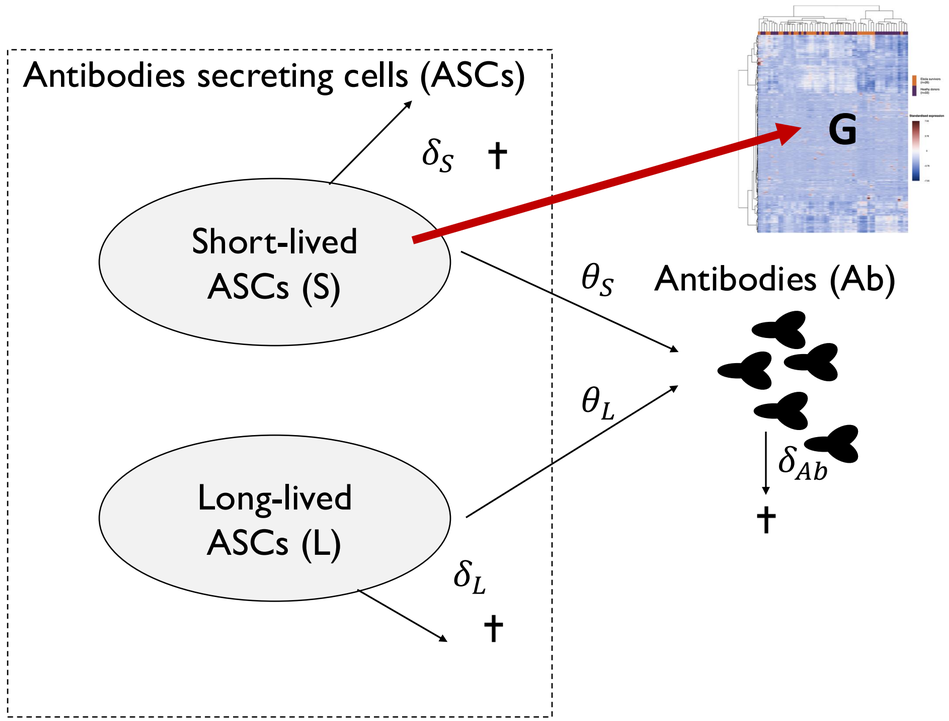




Rinchai et al. 2022

Method Idea

Use **longitudinal** Gene expression as observation of latent compartments



Observation Model

For each individual $i \leq N$, $k \leq K$ times $\left(t_{ij}\right)_{j \leq n_i}, \left(t_{ijk}\right)_{j \leq n_{ki}}$:

$$\boldsymbol{Y_{ij} = h\left(Ab_i\left(t_{ij}\right)\right) + \epsilon_{ij}}$$

$$\boldsymbol{G_{kij} = \alpha_{0k} + \alpha_{1k} S_i\left(t_{ijk}\right) + \varepsilon_{ijk}}$$

where $\epsilon = \left(\epsilon_{ij}\right)_{i \leq N, j \leq n_i} \sim \mathcal{N}(0, \Sigma^2)$

$$\varepsilon_k = \left(\varepsilon_{ijk}\right)_{i \leq N, j \leq n_{ik}} \sim \mathcal{N}\left(0, \sigma_k^2\right).$$

Simultaneous estimation of model parameters are $\theta = \left(\psi_{pop}, \beta, \Omega, \left(\sigma_k^2\right)_{k \leq K}, \left(\Sigma_p^2\right)_{p \leq P}, \left(\alpha_{0k}\right)_{k \leq K}\right)$ and the regularized parameters $\alpha = \left(\alpha_{1k}\right)_{k \leq K}$ by maximizing log-likelihood under lasso penalization :

$$\boldsymbol{LL_{pen}(\theta, \alpha) = LL(\theta, \alpha) - \lambda|\alpha|}$$

REMixed Algoritm – cyclic descent algorithm

$$\boldsymbol{LL_{pen}(\theta, \alpha) = LL(\theta, \alpha) - \lambda|\alpha|}$$

For a given penalty parameters λ, the iteration l in the estimation procedure correspond to :

At iteration l :

Current parameters are $\theta^{(l)}, \alpha_1^{(l)}$.

1. Update $\alpha_1^{(l+1)}$ for fixed $\theta = \theta^{(l)}$ using update formula derived from penalized log-ikelihood maximization.
2. Update $\theta^{(l+1)}$ for fixed $\alpha_1 = \alpha_1^{(l+1)}$ using SAEM algorithm through **Monolix** software.

We continue itérations until :

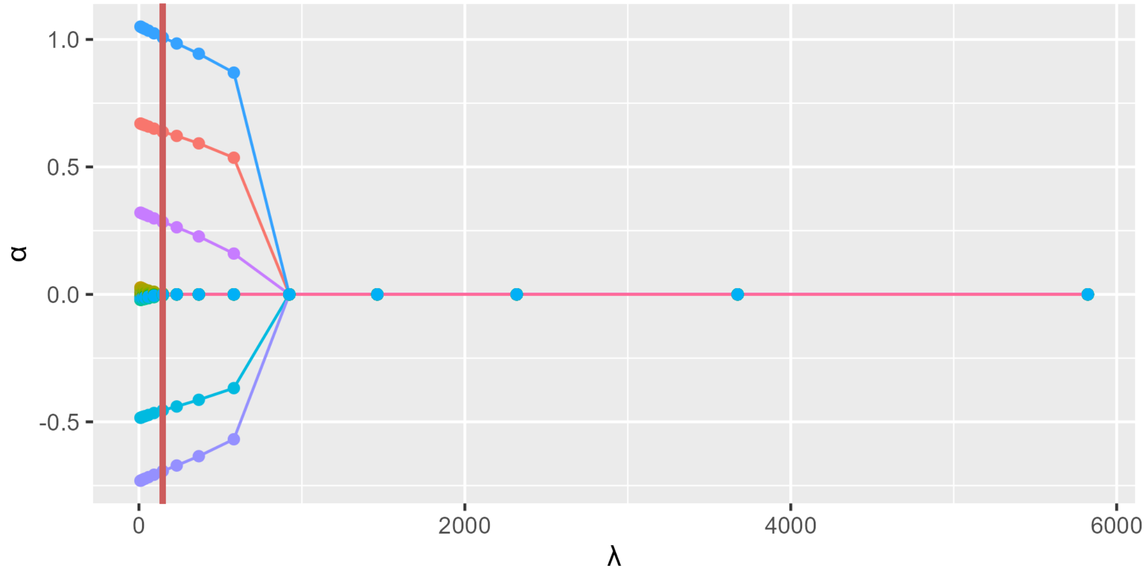
- $\left\| \left(\theta^{(l+1)}, \alpha_1^{(l+1)}\right) - \left(\theta^{(l)}, \alpha_1^{(l)}\right) \right\|_2 \leq \varepsilon_1$ (parameters stability)
- $\left| LL_{pen}\left(\theta^{(l+1)}, \alpha_1^{(l+1)}\right) - LL_{pen}\left(\theta^{(l)}, \alpha_1^{(l)}\right) \right| \leq \varepsilon_2$ (penalised log-likelihood stability)

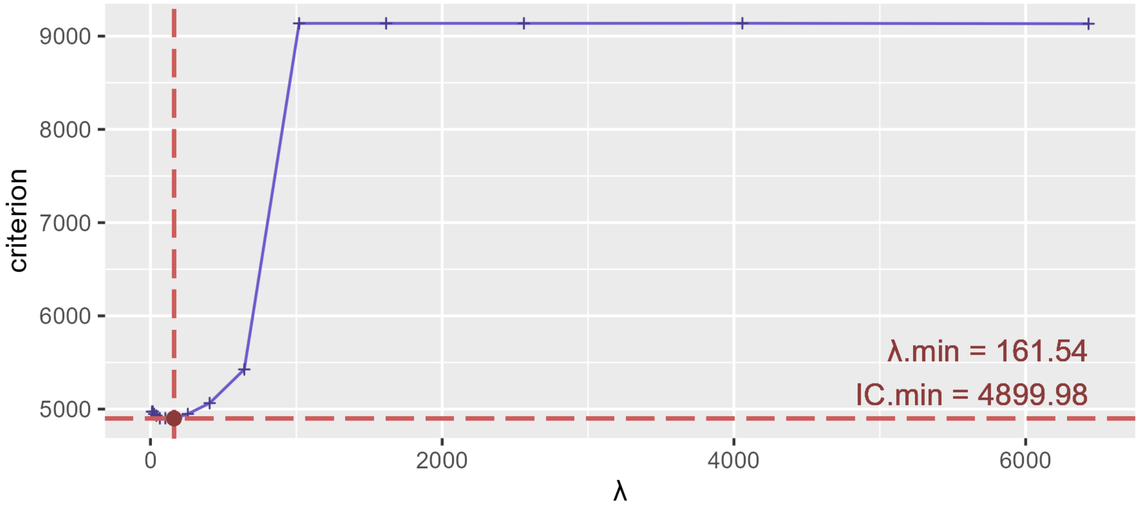
REMixed Algoritm – Choice of lambda

Presented procedure was given a penalty parameter λ.

For $\Lambda = \left\{\lambda_l = \lambda_{max} \times \alpha_\lambda^{\frac{l}{N_\lambda}}; 1 \leq l \leq N\right\}$, with $\lambda_{max} = \max\left(\partial_\alpha LL\left(\theta^{(0)}, \alpha\right)|_{\alpha = 0_K}\right)$,

$$BICc(\lambda) = -2LL(\theta^*, \alpha^*) + \log(N)\dim(\theta_R) + \log(n_{tot})\dim(\theta_F)$$





A final SAEM is then computed followed by statistical test to remove non-significant biomarkers.

Simulations setting

Structural Model 25 individuals, 50 biomarkers – 10 informative biomarkers , 200 replicates

$$\dot{S}_i(t) = -\delta_{S_i} S_i(t)$$
$$\dot{Ab}_i(t) = \varphi_{S_i} S_i(t) - \delta_{Ab_i} Ab_i(t)$$
$$S(t=0) = 5$$
$$Ab(t=0) = 1000$$

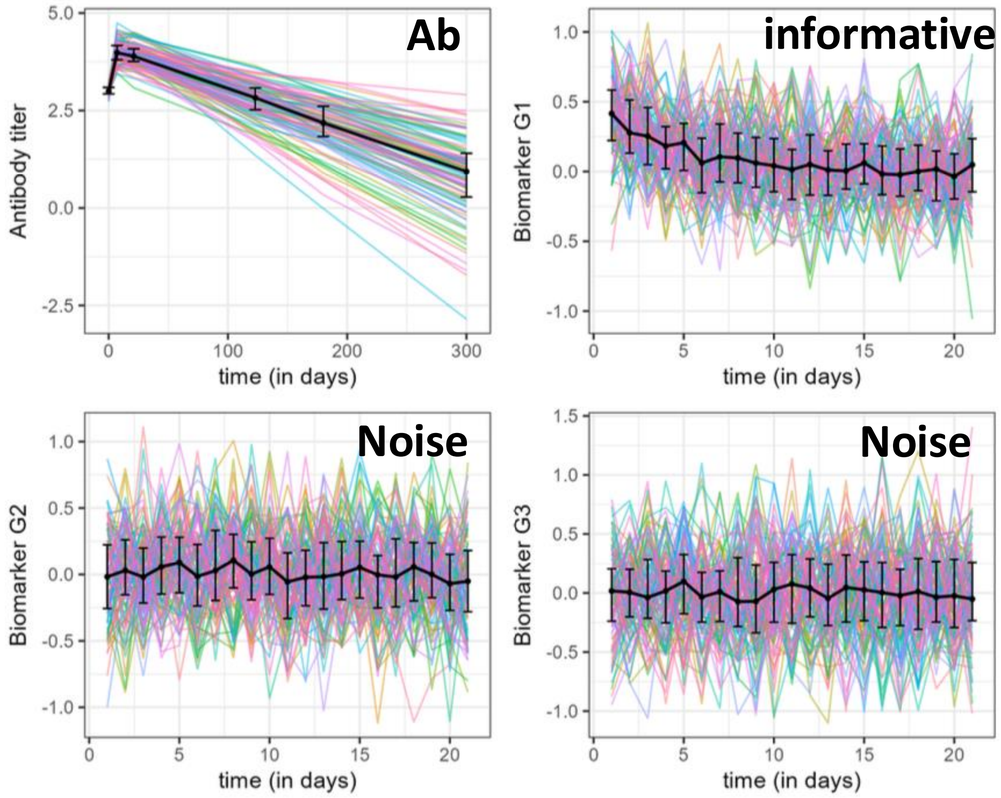
$$\log(\delta_{S_i}) = \log\left(\delta_{S_{pop}}\right) + \eta_i^S$$
$$\log(\delta_{Ab_i}) = \log\left(\delta_{Ab_{pop}}\right) + \eta_i^{Ab}$$
$$\log(\varphi_{S_i}) = \log\left(\varphi_{S_{pop}}\right) + \eta_i^{\varphi}$$

Statistical Model

$$\eta_i^{\chi} \sim \mathcal{N}(0, \omega_{\chi}^2), \qquad \chi \in \{S, Ab, \varphi\}$$

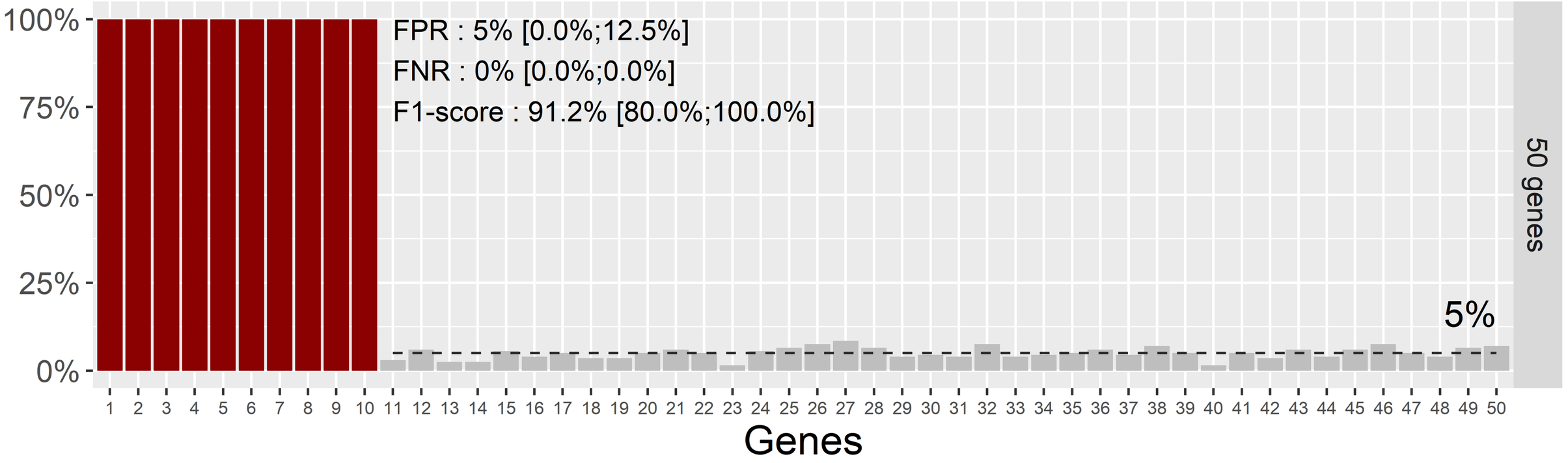
Observation Model $t_j \in \{0,7,21,123,180,300\}\,;\, t'_j \in \{0, \dots, 21\}$

$$Y_{ij} = \log_{10}\left(Ab_i(t_j)\right) + \varepsilon_{ij} \qquad (\varepsilon_{ij}) \sim \mathcal{N}(0, \sigma_{Ab}^2) \ \ iid$$
$$G_{kij} = \alpha_{0k} + \alpha_{1k} S_i(t'_j) + \varepsilon_{kij} \qquad (\varepsilon_{kij}) \sim \mathcal{N}(0, \sigma_{G_k}^2) \ \ iid$$



Pasin et al. 2020; Alexandre et al. 2024; Rinchai et al. 2024

Simulations results



04

Conclusions and perspectives

Perspectives

- Packages lasso-SAMBA and Remixed have been released on CRAN

CRAN lasso-SAMBA

CRAN REMixed

- Find relevant application of these methods (including in pharmacogenomics) and evaluate if assumptions can/have to be relaxed (ie. Linear relationship).

Lasso-SAMBA $g(\theta_i(t)) = g(\theta_0) + h(X_i(t)) + u_i$

REMixed $G_{kij} = h(S_i(t'_j), \varepsilon_{kij})$

Acknowledgment

This project was funded by RRI impulsion and EDCTP PREVAC

Inria/Inserm

Cécile Proust-Lima
Lisa Crépin
Boris Hejblum
Rodolphe Thiébaut
Auriane Gabaut

Vaccine Research Institute

Yves Levy
Christine Lacabaratz
Sylvain Cardinaud
Mathieu Surenaud
Séverin Coleon
Véronique Godot
Mireille Centlivre
Olivier Schwartz

Simulation+ / Lixoft Monolix Suite

Acknowledgements

Thank you !

WE ARE HIRING POSTDOCS IN BIOSTATISTICS / MODELING / PHARMACOMETRICS (MELANIE.PRAGUE@INRIA.FR)

2 Method : Lasso Selection

Individual Parameter regression :

We have a previoulsy build model $\mathcal{M}_k$ at iteration k, with parameters $\theta^{(k)}$, $\left(\psi_i^{(k)}\right)_{i\leq N}$; we write the regression model **for each parameter :**

$$g\left(\psi_i^{(k)}\right) = g_l\left(\psi_{pop}^{(k)}\right) + \beta X_i$$

Tibshirani, 1996

2 Method : Lasso Selection

Individual Parameter regression :

We have a previoulsy build model $\mathcal{M}_k$ at iteration k, with parameters $\theta^{(k)}$, $\left(\psi_i^{(k)}\right)_{i\leq N}$; we write the regression model **for each parameter :**

$$g\left(\psi_i^{(k)}\right) = g_l\left(\psi_{pop}^{(k)}\right) + \beta X_i$$

Lasso Regression :

We then compute the lasso estimator

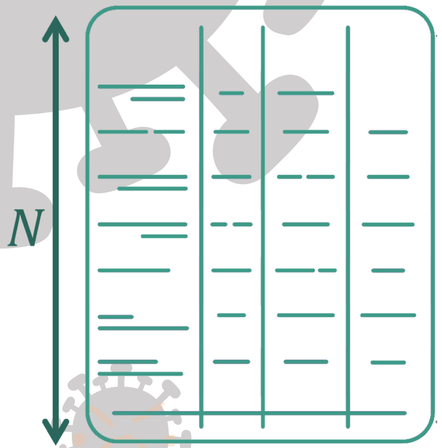
$$\hat{\beta} = \underset{\beta\in\mathbb{R}^n}{\arg\min}\left\{\sum_{i\leq N}\left(g\left(\psi_i^{(k)}\right) - g\left(\psi_{pop}^{(k)}\right) - \beta X_i\right)^2 + \lambda|\beta|_1\right\}$$

Where $\boldsymbol{\lambda > 0}$ is a data-driven penalization parameter.

Tibshirani, 1996

2 Method : Stability Selection

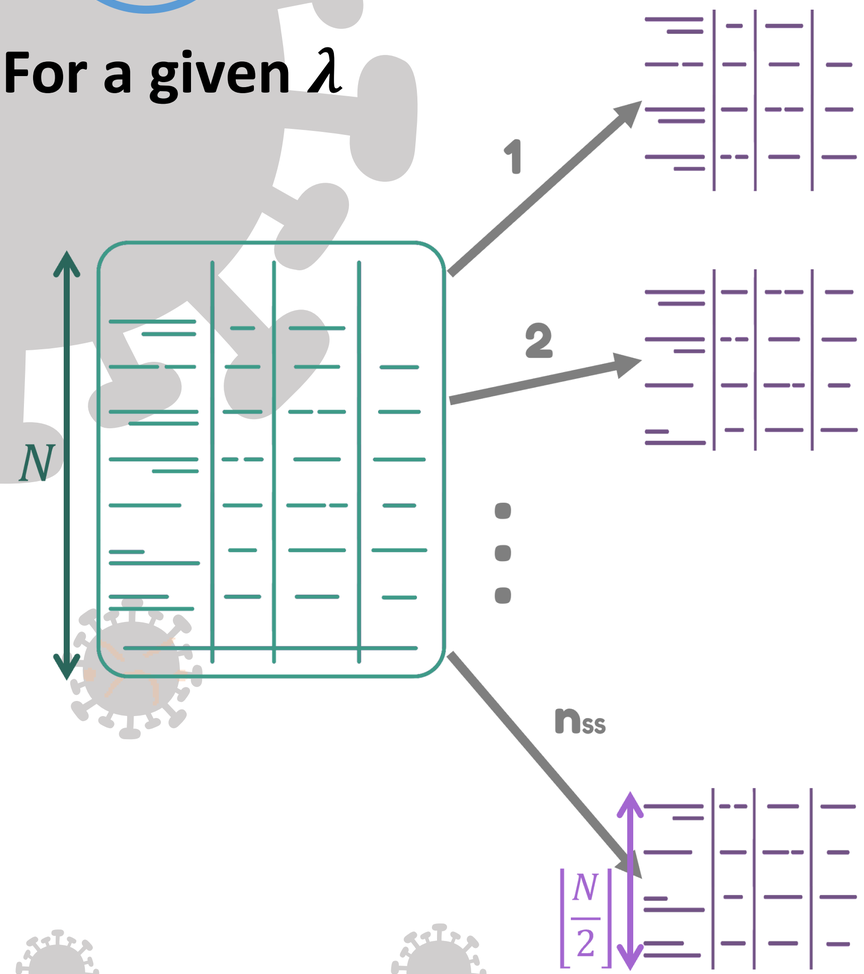
For a given λ



Meinshausen and Bühlmann, 2010

2 Method : Stability Selection

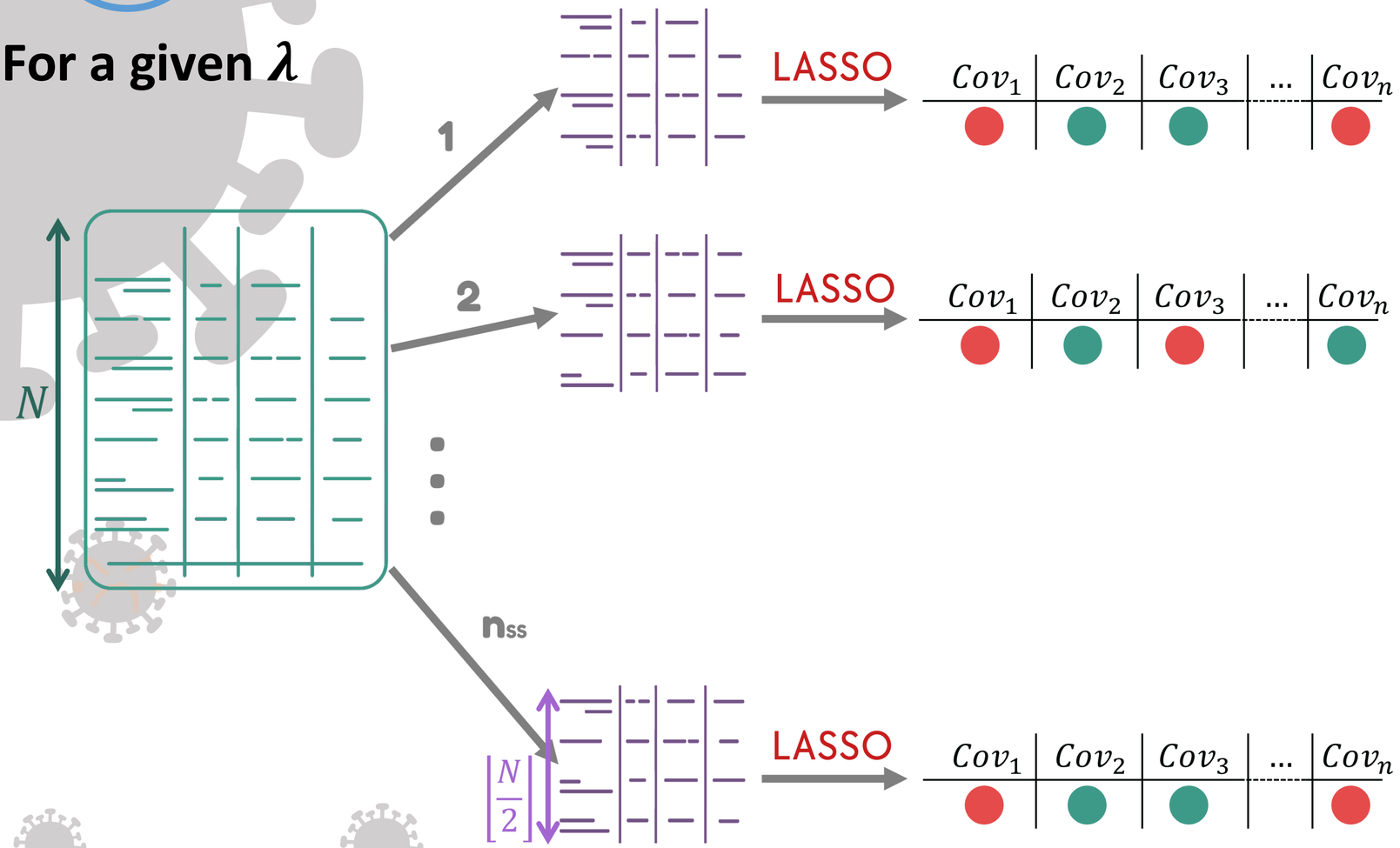
For a given λ



Meinshausen and Bühlmann, 2010

2 Method : Stability Selection

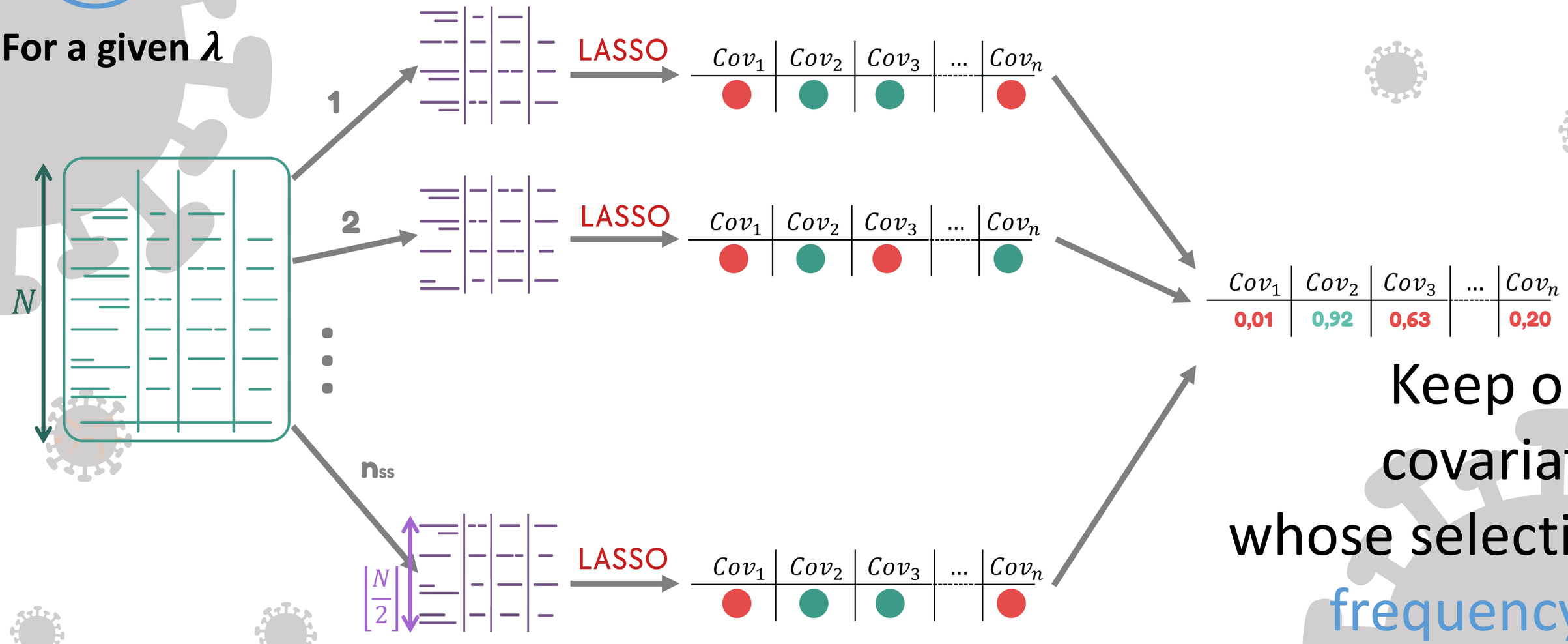
For a given λ



Meinshausen and Bühlmann, 2010

② Method : Stability Selection

For a given λ



Keep only covariates whose selection frequency is higher than a given **thresholds**

Meinshausen and Bühlmann, 2010

2 Method : Parameters calibration and IC

Lasso selection enhanced by stability selection depends on two parameters

$\boldsymbol{\lambda}$ the penalization parameter
$\boldsymbol{t_{ss}}$ the selection thresholds

Meinshausen and Bühlmann, 2010 – Bodinier and al., 2023

2 Method : Parameters calibration and IC

Lasso selection enhanced by stability selection depends on two parameters

λ the penalization parameter
t_{ss} the selection thresholds

! Decreasing IC in covariate selection step by stepAIC in SAMBA seems to allows algorithm to stop.

> Rather than one single model, we construct a set of relevant model to explore, searching for one decreasing the IC.

Meinshausen and Bühlmann, 2010 – Bodinier and al., 2023

2 Method : Parameters calibration and IC

Lasso selection enhanced by stability selection depends on two parameters

$\boldsymbol{\lambda}$ the penalization parameter
$\boldsymbol{t_{SS}}$ the selection thresholds

! Decreasing IC in covariate selection step by stepAIC in SAMBA seems to allows algorithm to stop.

> Rather than one single model, we construct a set of relevant model to explore, searching for one decreasing the IC.

Meinshausen and Bühlmann, 2010 :

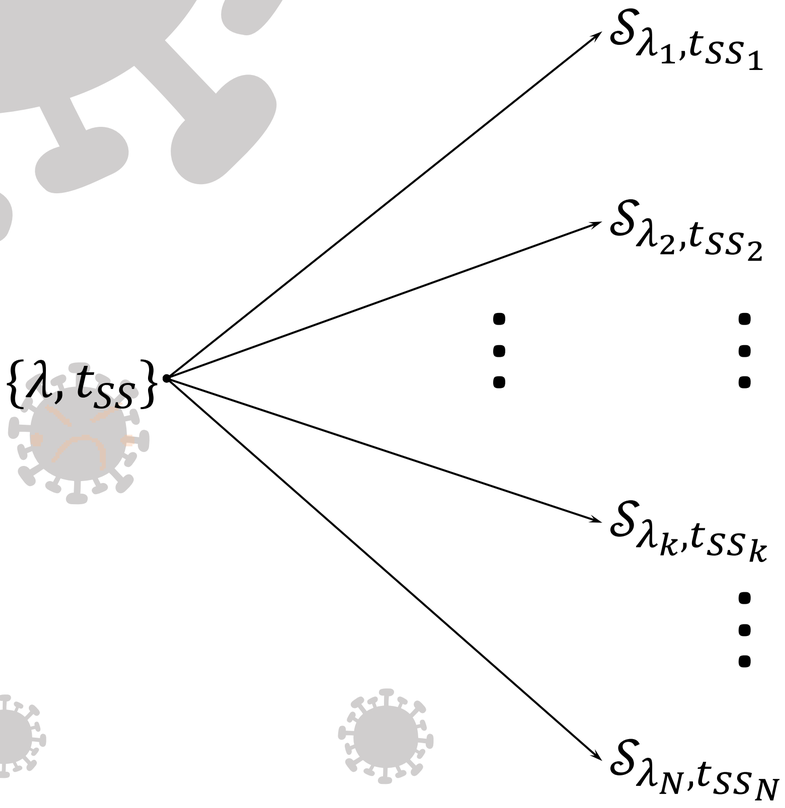
$$\mathbb{E}\left[\, \#FP\left(S^{\lambda, t_{SS}}\right)\right] \leq U_{\lambda, t_{SS}} = \frac{1}{2t_{SS} - 2} \times \frac{q_\lambda^2}{n}$$

Where q_λ is the average number of features that are selected at least once by the Lasso algorithm, and n the number of covariates.

Meinshausen and Bühlmann, 2010 – Bodinier and al., 2023

2 Method : Parameters calibration and IC

Selected set $\mathcal{S}_{\lambda,t_{SS}}$ of covariates by Lasso

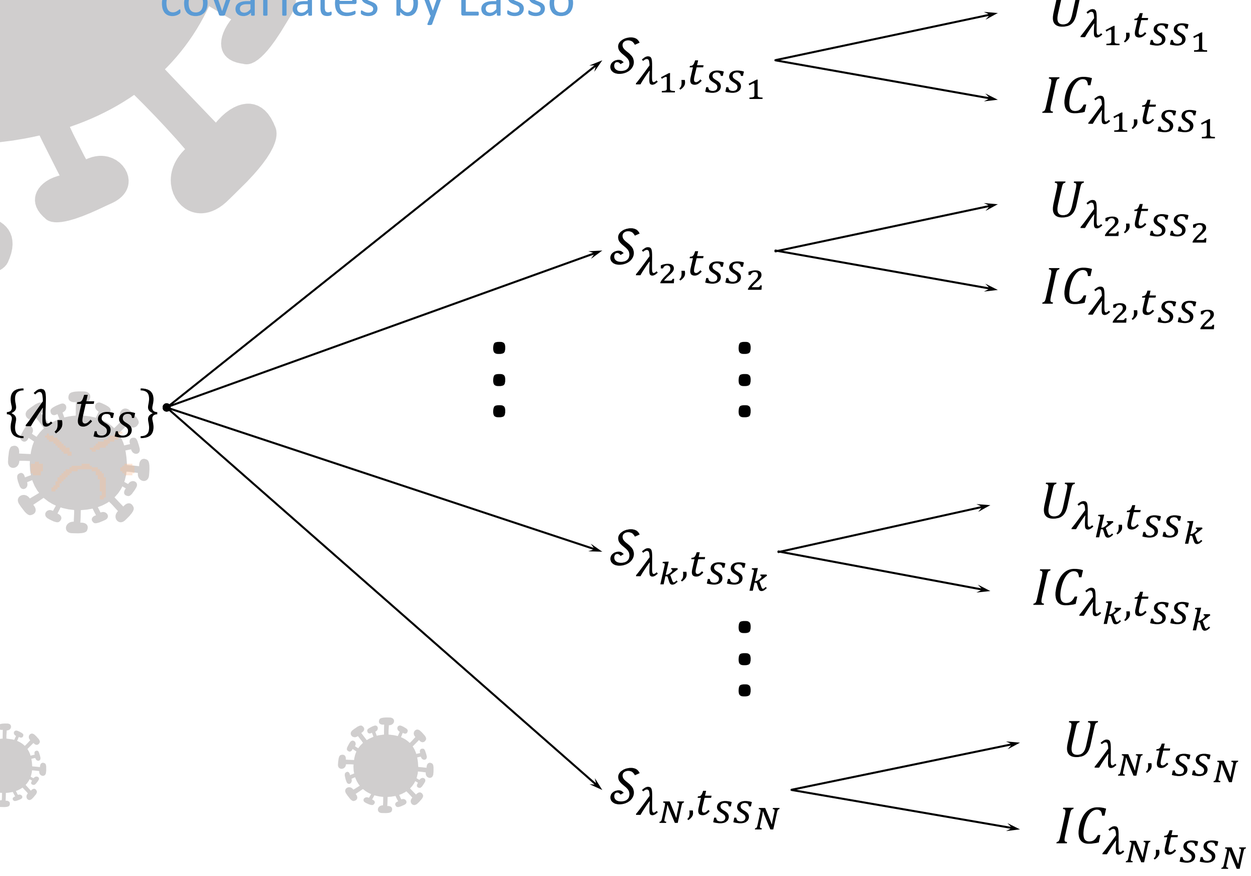


Meinshausen and Bühlmann, 2010 – Bodinier and al., 2023

2 Method : Parameters calibration and IC

Selected set $S_{\lambda,t_{SS}}$ of covariates by Lasso

Compute PFER upper-bound and IC

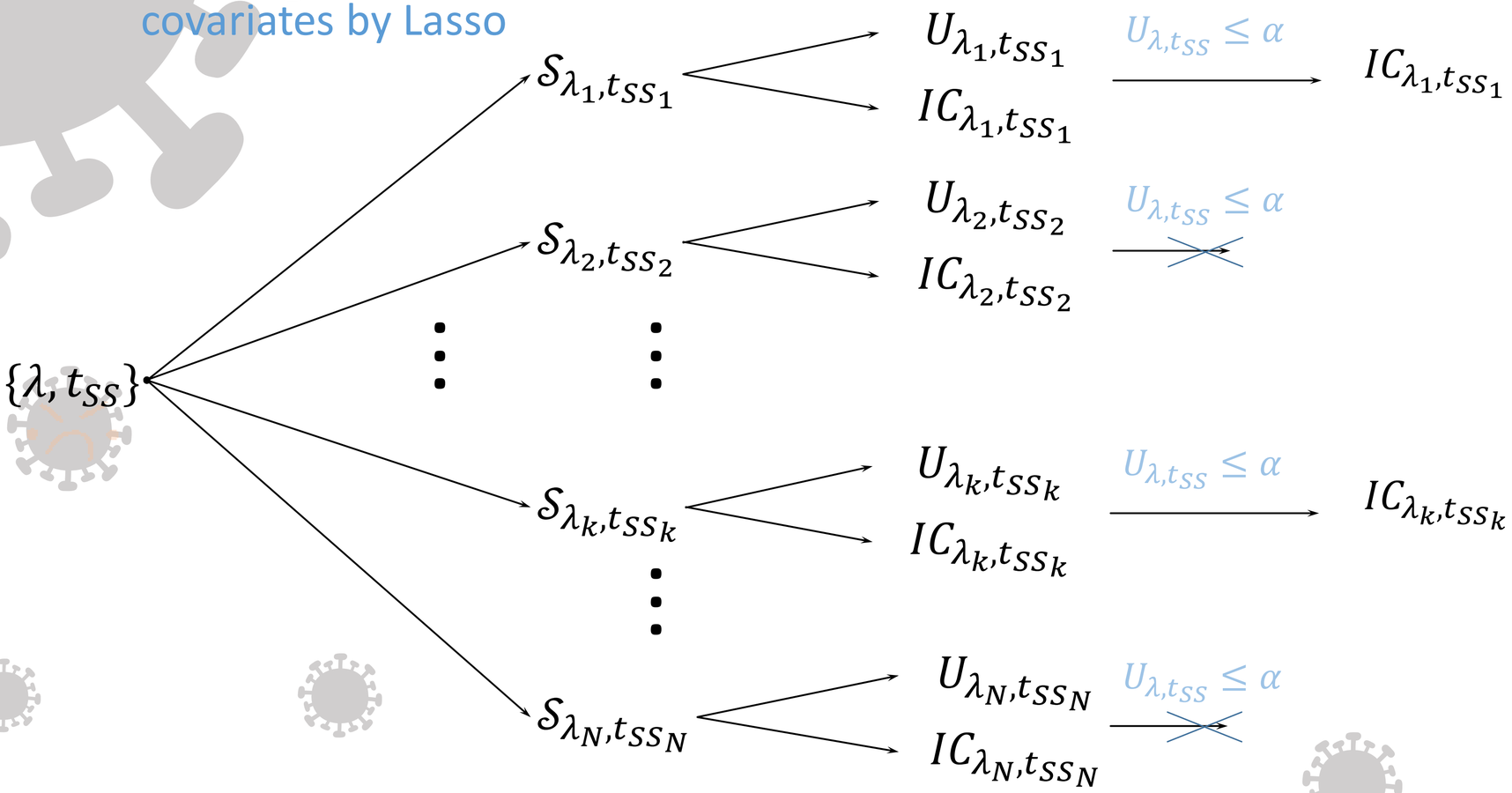


Meinshausen and Bühlmann, 2010 – Bodinier and al., 2023

2 Method : Parameters calibration and IC

Selected set $S_{\lambda,t_{SS}}$ of covariates by Lasso

Compute PFER upper-bound and IC

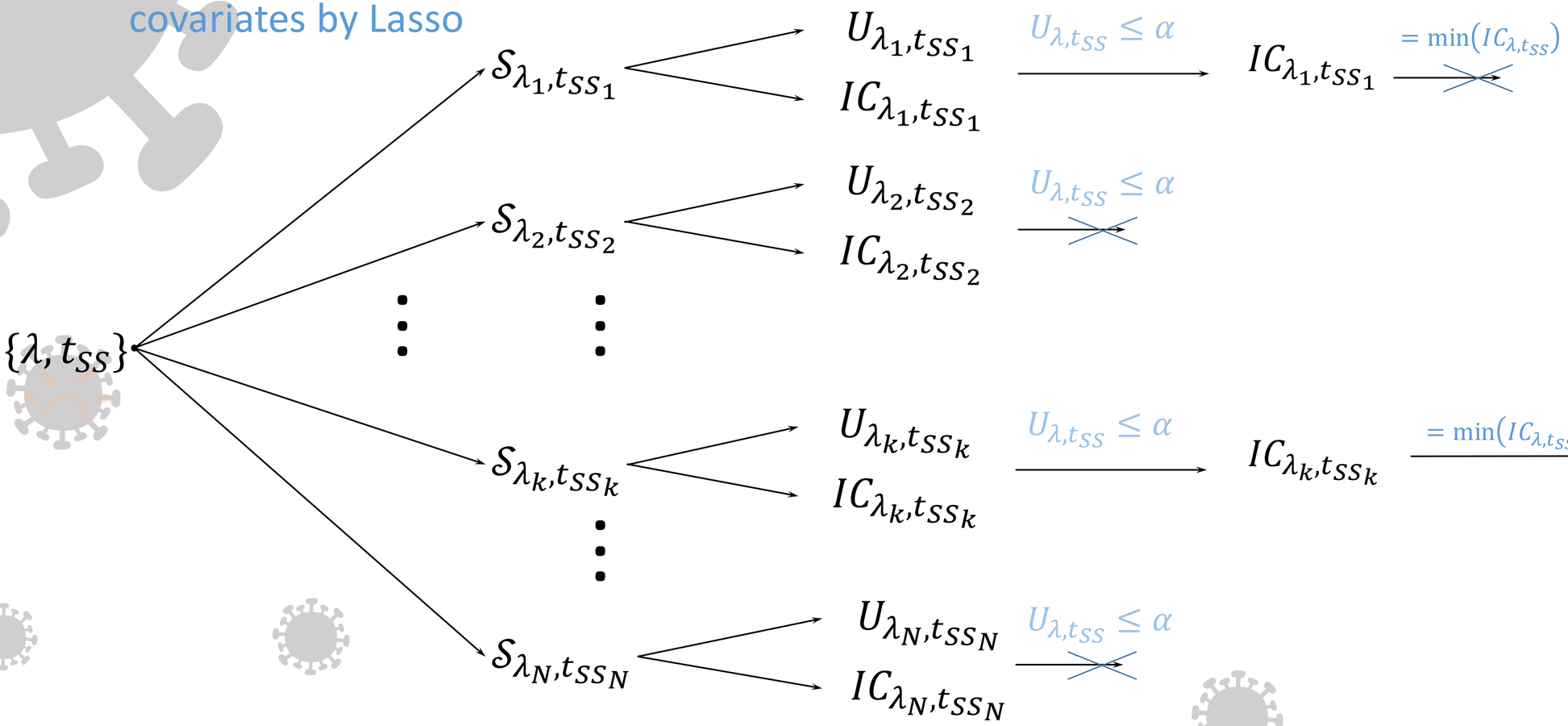


Meinshausen and Bühlmann, 2010 – Bodinier and al., 2023

② Method : Parameters calibration and IC

Selected set $S_{\lambda,t_{SS}}$ of covariates by Lasso

Compute PFER upper-bound and IC



Meinshausen and Bühlmann, 2010 – Bodinier and al., 2023

2 Method : REMix algorithm

Update formula of regularized parameters.

At this step, we suppose fixed $\theta^{(l)}$, and current estimate $\alpha^{(l)}$

We want to solve

$$\partial_{\alpha_1}\left(LL\left(\theta^{(l)},\alpha\right)-\lambda|\alpha|\right)=0$$

We use a second order Taylor development

$$LL\left(\theta^{(l)},\alpha\right)=LL\left(\theta^{(l)},\alpha^{(l)}\right)+\left(\alpha-\alpha^{(l)}\right)^T\partial_\alpha LL\left(\theta^{(l)},\alpha\right)\Big|_{\alpha=\alpha^{(l)}}-\frac{1}{2}\left(\alpha-\alpha^{(l)}\right)^T\partial_\alpha^2 LL\left(\theta^{(l)},\alpha\right)\Big|_{\alpha=\alpha^{(l)}}\left(\alpha-\alpha^{(l)}\right)$$

The update formula for regularized parameters α_{1k} :

$$\alpha_{1k}^{(l+1)}=\begin{cases}\dfrac{A+\lambda}{\partial^2_{\alpha_{1k}}LL(\theta^{(l)},\alpha)|_{\alpha=\alpha^{(l)}}}, & if\ A<-\lambda\\ \dfrac{A-\lambda}{\partial^2_{\alpha_{1k}}LL(\theta^{(l)},\alpha_1)|_{\alpha=\alpha^{(l)}}}, & if\ A<+\lambda\\ 0, & otherwise.\end{cases}$$

With A a function of gradient, hessian and current estimates.

Tibshirani, 1996 – Park et al., 2007

3 Simulation Studies

Initialization Strategy

- We randomly group genes by pairs and run a SAEM algorithm for each pair.
- The initialization yielding the best log-likelihood is selected to launch the final algorithm.