The Intertox project

Antonin Della Noce,
Antonio Di Meglio (PI), Paul-Henry Cournède (
Institut Gustave Roussy and CentraleSupélec
antonin.della-noce@gustaveroussy.fr

www.dataia.eu

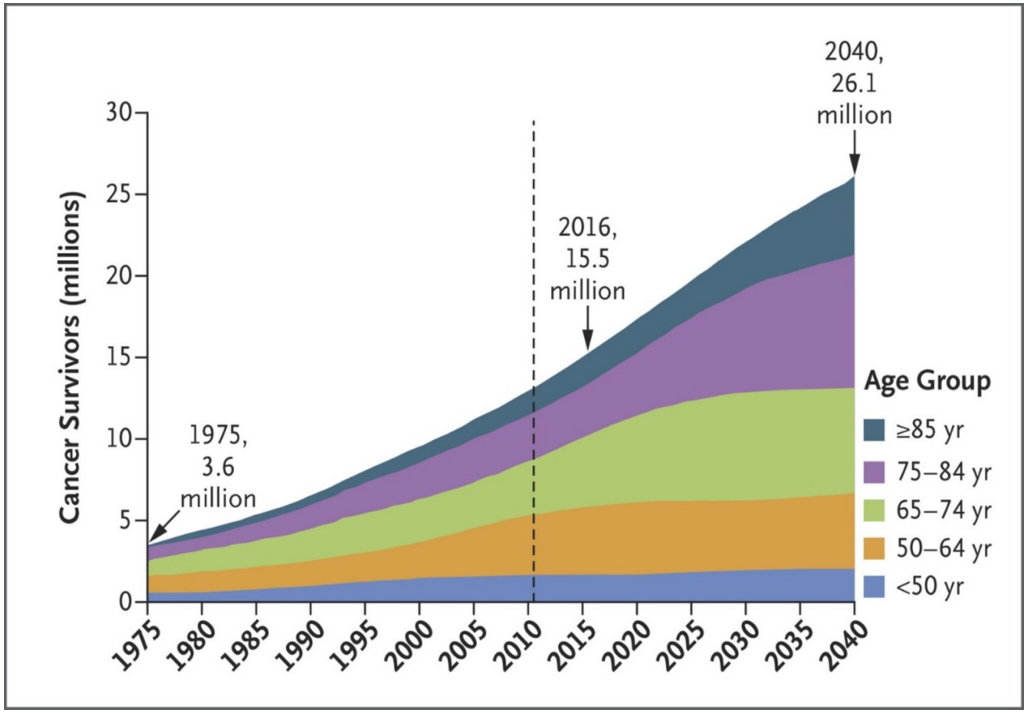
Context: Breast cancer survivorship

université PARIS-SACLAY
INSTITUT DATAIA Science des données, Intelligence & Société

Cancer survivorship

Definition and demography

- Cancer survivor: any person with history of cancer, from time of diagnosis and for the remainder of life.[1]
- Need to focus on minimizing the physical, psychological and social burden of surviving breast cancer



1. Rowland and Baker, 2005, Cancer

Cancer survivorship

The CANTO cohort for breast cancer toxicities

CANTO (CANcer Toxicities Cohort; NCT01993498)

- prospective longitudinal cohort started in 2012
- **26 French comprehensive cancer centers**
- dedicated **national network** sponsored by UNICANCER
- reached inclusion of **12012** patients in 2018

Inclusion criteria:

- 18+ years old at BC diagnosis
- Histologically confirmed invasive BC
- Stage I-II-III
- Untreated at time of inclusion

	Baseline	Follow-up after treatment					Long-term follow-up yearly for 5 years
Collected Information	**Diagnosis**	**Year-1**	**Year-2**	**Year-3**	**Year-4**	**Year-5**	
Inclusion criteria							**Prolonged and long-term toxicity Survival Outcomes**
Signed informed consent							
Clinical examination^							
Blood tests							
Paraclinical examination							
Questionnaires (PROs)*							
Biological samples							

Completion of treatment (surgery, chemo, or radio)

^Includes detailed assessment of supportive care consultations

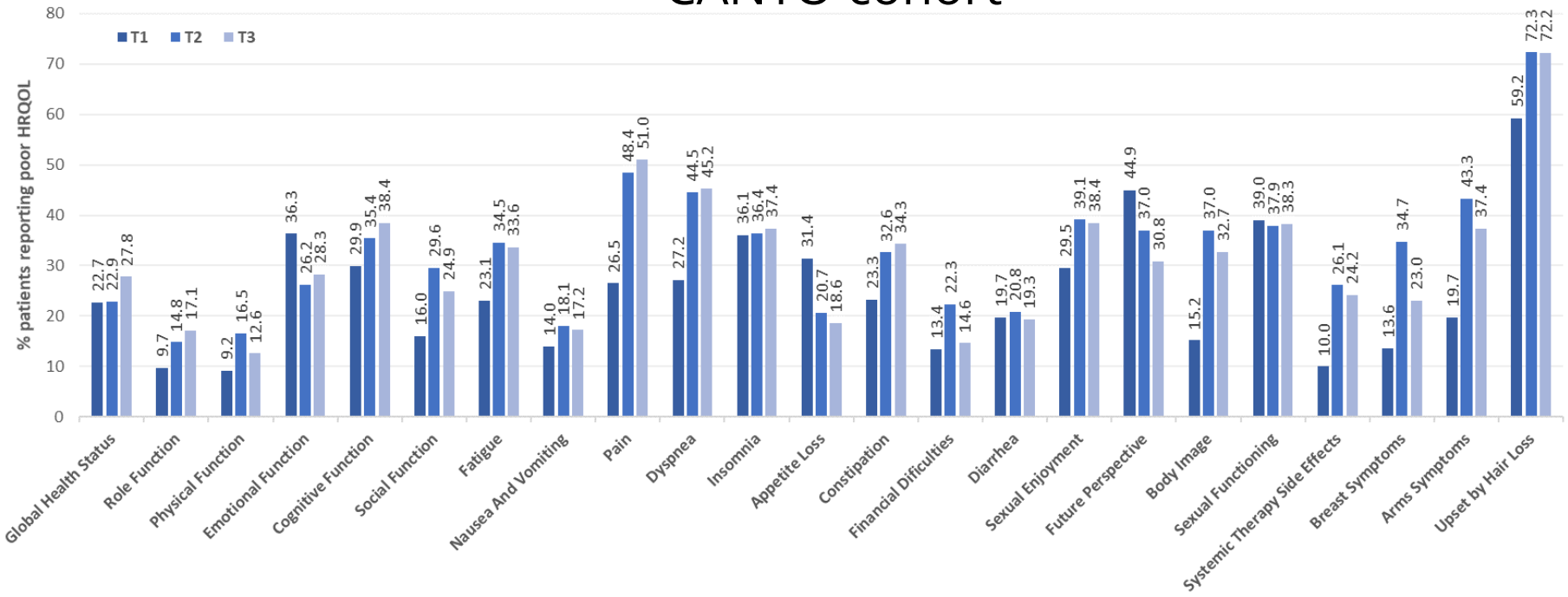
***EORTC-QLQ C30**, BR23, **FA12, GPAQ-16**, HADS, SF-12, IOCv2, social and financial reports

Ferreira et al., 2019, Annals of onco

Cancer survivorship

Prevalence of toxicities related to quality of life until 3 years after diagnosis

N= 4,262 BC patients from the French CANTO cohort



Ferreira et al., 2019, Annals of oncology

Interdisciplinary effort to better understand and communicate breast cancer toxicities
Intertox

université PARIS-SACLAY | INSTITUT DATAIA Science des données, Intelligence & Société

Objectives of the Intertox project

Aim 1

- Predict the evolution of the quality of life and its main dimensions
breast cancer (BC)
- Identification of biomarkers associated with deterioration of qualit
main dimensions after BC

Available data

- Outcome: Quantification of the quality of life and its dimensions u
EORTC-QLQ-C30 questionnaire.
- Variables:
    - Clinical data for 10,000 patients until 4 years after diagnosi
    - 21 inflammatory proteins measured in 1,500 patients befo
    - ~500 proteins quantified in two independent hyper-reactio
    mass spectrometry analyses conducted on the plasma of 462
    538 patients respectively, before treatment.
    - ~600,000 germline SNPs in 7,000 patients
    - Future metabolomics analysis on 1,000 patients

Objectives of the Intertox project

Aim 2: in collaboration with Maison des Sciences de l'Homme

- Communicate the risk of QOL degradation

Requirements

- Interpretability, or even explainability, of the predictive models
- Acceptability assessed by focus groups with patients and care

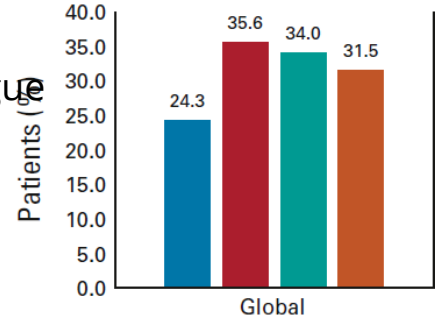
Case study: Prediction of cancer-related fatigue

université PARIS-SACLAY
INSTITUT DATAIA
Science des données, Intelligence & Société

Prediction of cancer-related fatigue

Prevalence of fatigue and clinical model

Prevalence of global fatigue at year 0,1,2,3



Logistic regression with clinical variables selected by augmented backward elimination

TABLE 2. Predictive Model of the Risk of Severe Fatigue at 2 Years After Diagnosis

Variable	OR	95% CI	β Coefficient	95% CI	*P*
Severe pretreatment fatigue,[a] yes versus no	3.191	2.704 to 3.767	1.160	0.995 to 1.326	< .0001
Age, continuous (for 1-year decrement)	1.015	1.009 to 1.022	–0.015	–0.021 to –0.0088	< .0001
BMI, continuous (for unit increment)	1.025	1.012 to 1.038	0.025	0.012 to 0.038	.0001
Tobacco use behavior, former versus never	1.243	1.055 to 1.463	0.217	0.053 to 0.381	.009
Tobacco use behavior, current versus never	1.552	1.291 to 1.866	0.440	0.256 to 0.624	< .0001
Anxiety,[b] doubtful case versus noncase	1.063	0.895 to 1.262	0.061	–0.110 to 0.233	.485
Anxiety,[b] case versus noncase	1.265	1.073 to 1.492	0.235	0.070 to 0.400	.005
Insomnia,[a] continuous (for unit increment)	1.005	1.003 to 1.007	0.0048	0.0026 to 0.0070	< .0001
Pain,[a] continuous (for unit increment)	1.014	1.010 to 1.017	0.014	0.010 to 0.017	< .0001
Intercept			–1.445	–1.912 to –0.978	< .0001
AUC (95% CI)			0.73 (0.72 to 0.75)		

Di Meglio et al., 2022, Journ of clinical oncol

Prediction of cancer-related fatigue

Using serum proteins

Observation model

$$y^{fatigue} | x^{clin}, x^{prot} \sim Bin(\sigma(\beta^{clin}.x^{clin} + \beta^{prot}.x^{prot}))$$

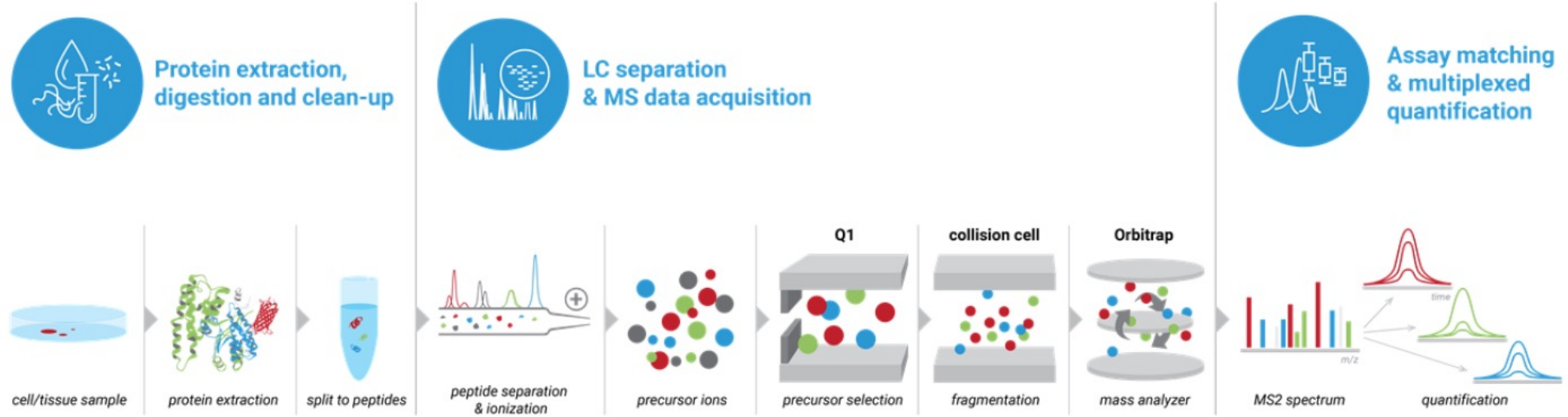
Variable selection: augmented backward elimination

AUC = 0.78 (0.75-0.8

Clinico-bio-behavioral model of pre-treatment predictors of severe global fatigue at T2, incorporating circulating inflammatory biomarkers (N=1153)	
	Adjusted OR* (95% CI)
Age, per additional 1 year	**0.98 (0.96-0.99)**
BMI, per additional unit	1.02 (0.99-1.06)
Current smoker, vs never	**2.27 (1.47-3.51)**
Former smoker, vs never	0.97 (0.64-1.46)
Anxiety case, vs normal	1.13 (0.75-1.70)
Doubtful anxiety, vs normal	**1.11 (0.73-1.68)**
Pre-treatment Insomnia, per additional 10 points**	**1.09 (1.04-1.15)**
Pre-treatment Pain, per additional 10 points**	**1.10 (1.01-1.18)**
Severe pre-treatment CRF **, vs no	**4.70 (3.13-7.05)**
IL6***	**1.72 (1.25-2.36)**
IL1RA***	1.24 (0.85-1.81)
IL2***	1.43 (0.99-2.08)
IFNg***	**0.54 (0.30-0.95)**
IL10***	**0.40 (0.18-0.87)**
IL4***	1.47 (0.67-3.20)
IL8***	1.15 (0.83-1.60)
OR= Odds Ratio; CI= Confidence Interval; *by all factors in Table plus anxiety (HADS); **QLQ-C30; ***per log-unit increase	

Parenthesis on HRM mass spectrometry

For discovery proteomics



Characteristics of the post-processed data:

- Relative abundance (with errors) of a given peptide in the samp
- 2nd source of error: mis-identification of a peptide
- Highly dimensional (7,324 quantified peptides for each patient)

Bruderer et al., 2015, Molecular and cellular pro

Prediction of cancer-related fatigue

Using proteomics and Group-adaptive Lasso

Peptides selection: Group-adaptive Lasso

$$\beta^* = \operatorname{argmin}\left\{\frac{1}{N_{pts}}\sum_{i=1}^{N_{pts}} \ell\left(y_i, \beta^{clin}.x_i^{clin} + \beta^{pep}.x_i^{pep}\right) + \lambda \sum_{p=1}^{N_{pep}} w_p \left|\beta_p^{pep}\right|\right\}$$

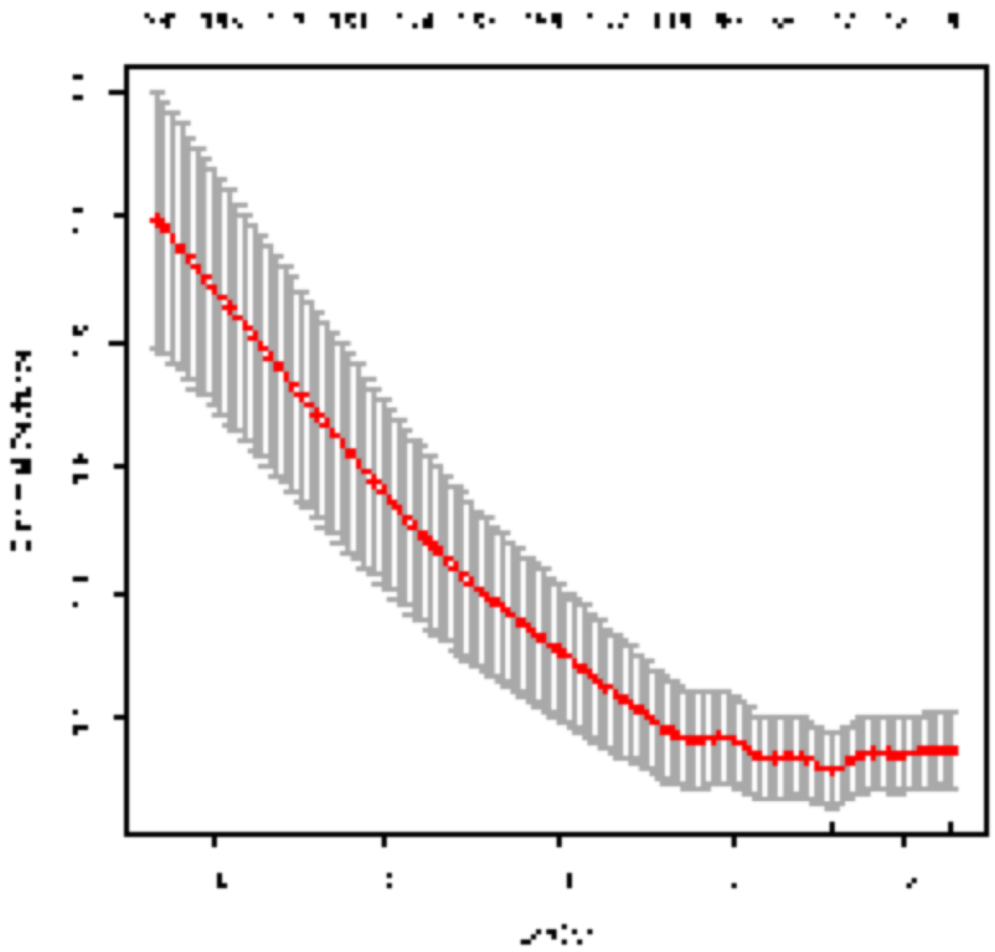
with w_p the weight of the peptide p associated with the protein π_p

$$w_p = \frac{|\mathrm{Pep}(\pi_p)|}{\sum_{p' \in Pep(\pi_p)} \frac{\widehat{\beta_p}}{se(\beta_p)}}$$

Belhechmi et al., 2020, Bioinformatics

Prediction of cancer-related fatigue

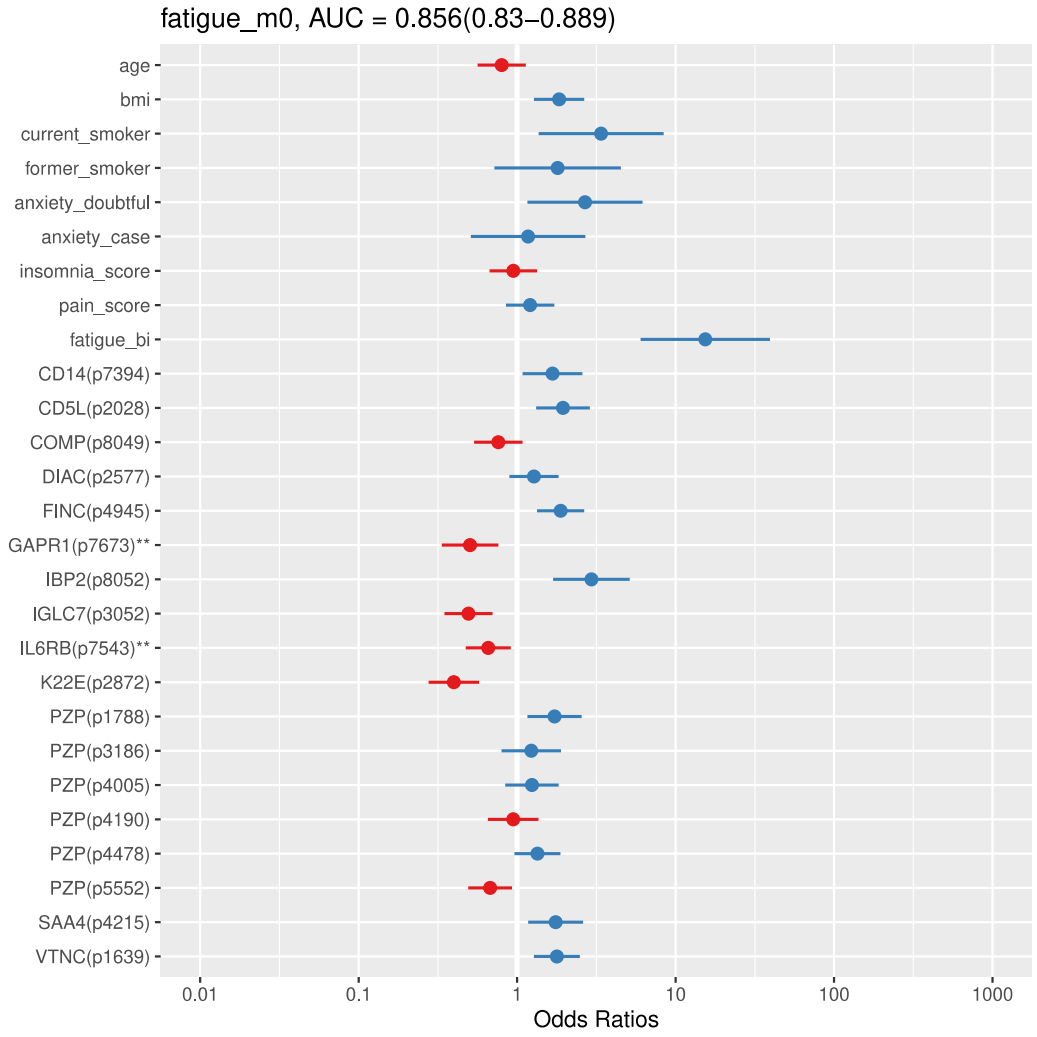
No peptides selected for fatigue at T2

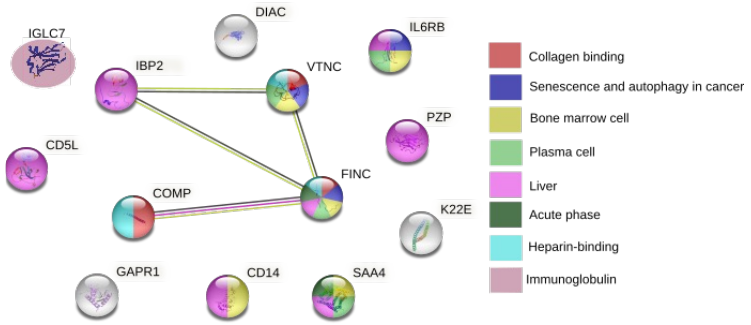


Leave-one-out cross-validat
of λ

Prediction of cancer-related fatigue

Fatigue at T1 predicted with selected peptides





Conclusion

Next steps concerning the ML part of the Intertox project

- Check the results on an independent analysis (problem of the i
undetected peptides)
- Multi-omics integration
- Combination of models for different outcomes (fatigue, QOL) ir
a decision aid tool.

université PARIS-SACLAY

INSTITUT DATAIA
Science des données, Intelligence & Société

INSTITUT DATAIA PARIS-SACLAY
Centre Inria Saclay - Île-de-France 1 rue Honoré d'Estienne d'Orves
Campus de l'Ecole Polytechnique - Bât Alan Turing 91120 Palaiseau
www.dataia.eu