

Le projet-phare HepatO'Twin

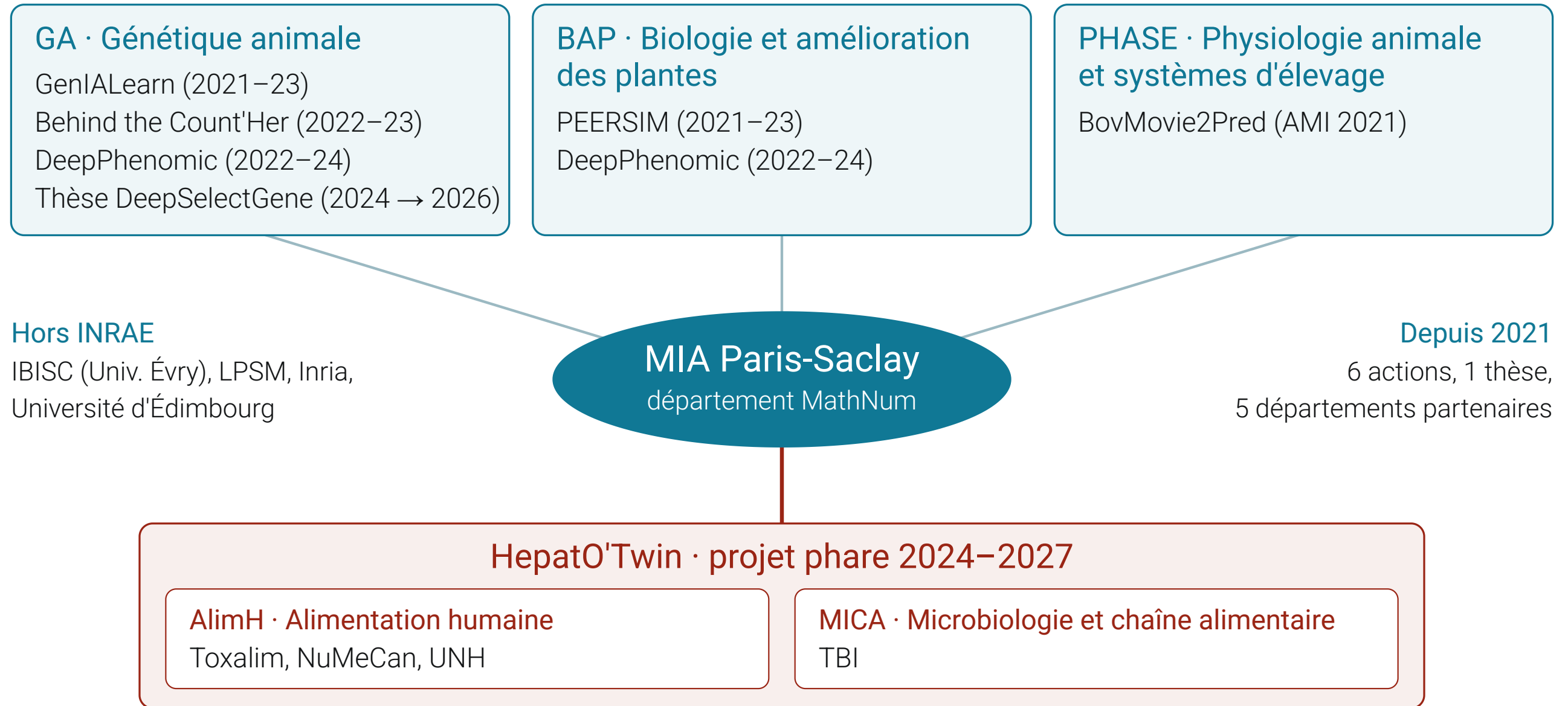
Un jumeau numérique du métabolisme hépatique, vu par un chercheur en apprentissage statistique

Projet coordonné par Nathalie Poupin (Toxalim)

Julien Chiquet (MIA Paris-Saclay, membre du COPIL du MP DIGIT-BIO)

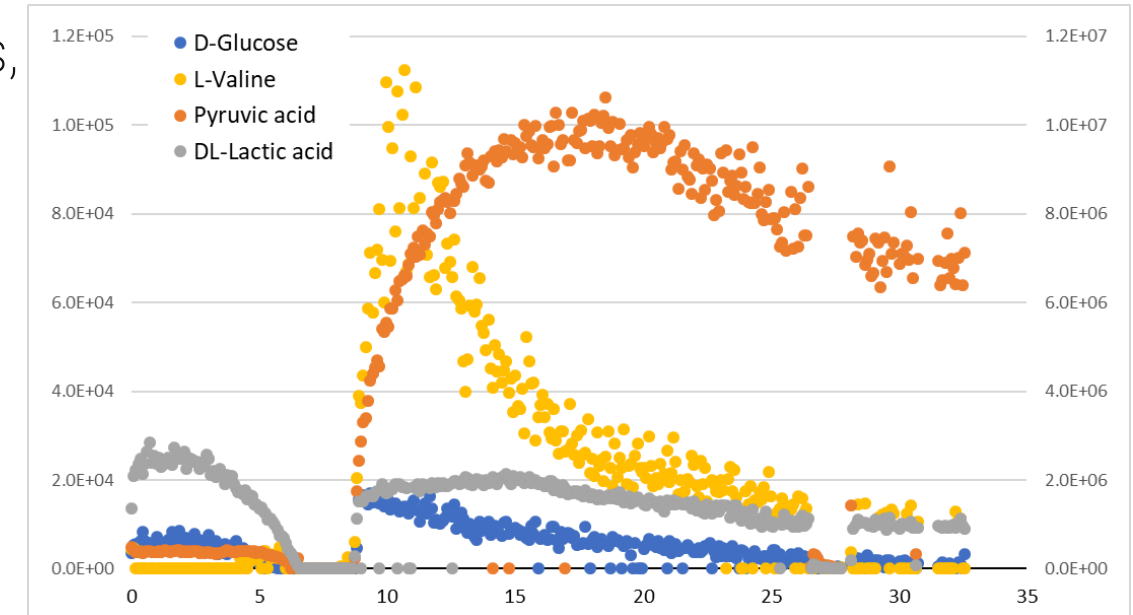
AG MathNum · 9 octobre 2026

MIA Paris-Saclay dans le MP DIGIT-BIO



HepatoTwin: la question biologique

- Contaminants alimentaires (bisphénols, pesticides, perfluorés) = **perturbateurs métaboliques**
- Hypothèse : exposition → foie moins apte à répondre à un **stress nutritionnel** → stéatose (accumulation de graisse), insulino-résistance
- Verrou : mesures **statiques** (24 h, 48 h) ; *quand* regarder ?



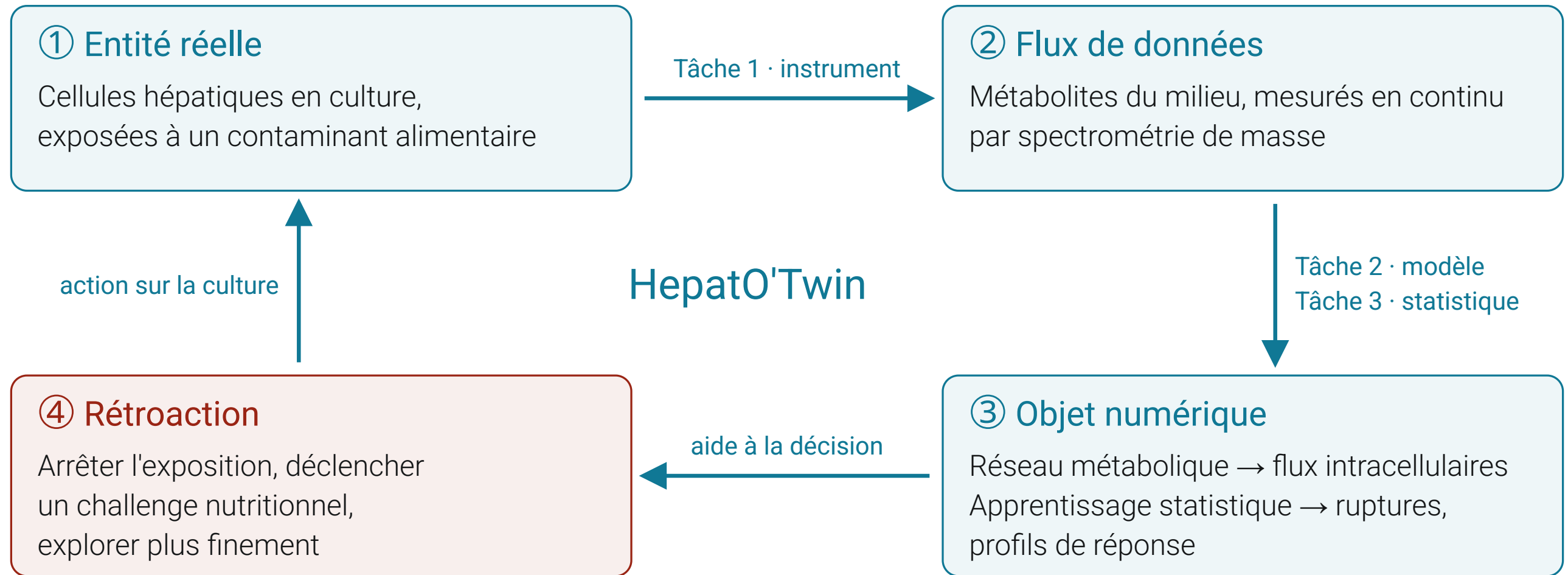
Suivi en continu de 4 métabolites. Figure : N. Poupin, Toxalim.

Deux questions des biologistes

1. Mesurer et modéliser la **dynamique** du métabolisme
2. **Détecter** ou prédire une altération des cinétiques

Question 2 = question de statistique, notre porte d'entrée dans le projet.

Le jumeau numérique : une boucle



« Modèle numérique conçu pour représenter aussi fidèlement que possible un objet ou un système réel et pour **le suivre au cours du temps**, afin d'aider aux **prises de décision** et d'**agir** sur l'objet ou le système observé. »

Rapport INRAE, Caranta et al., 2024

Le paroxysme 🚀 de l'interdisciplinarité 😄

	Unité et tutelles	Département	Apport
	Toxalim , Toulouse (coord.) INRAE · ENVT · INP-Purpan · Univ. Toulouse	AlimH	toxicologie, réseaux métaboliques
	TBI , Toulouse INSA Toulouse · CNRS · INRAE	MICA	métabolomique, instrumentation
	NuMeCan , Rennes Inserm · INRAE · Univ. Rennes	AlimH	physiologie du foie
	UNH , Clermont-Ferrand INRAE · Univ. Clermont Auvergne	AlimH	nutrition, maladies métaboliques
	Burgess group Université d'Édimbourg	—	métabolomique en temps réel
	MIA Paris-Saclay INRAE · AgroParisTech · Univ. Paris-Saclay	MathNum	apprentissage statistique



Le consortium. Photo : N. Poupin.

310 k€ du MP DIGIT-BIO

4 ans, sans salaires : expériences, petit équipement, missions, stages

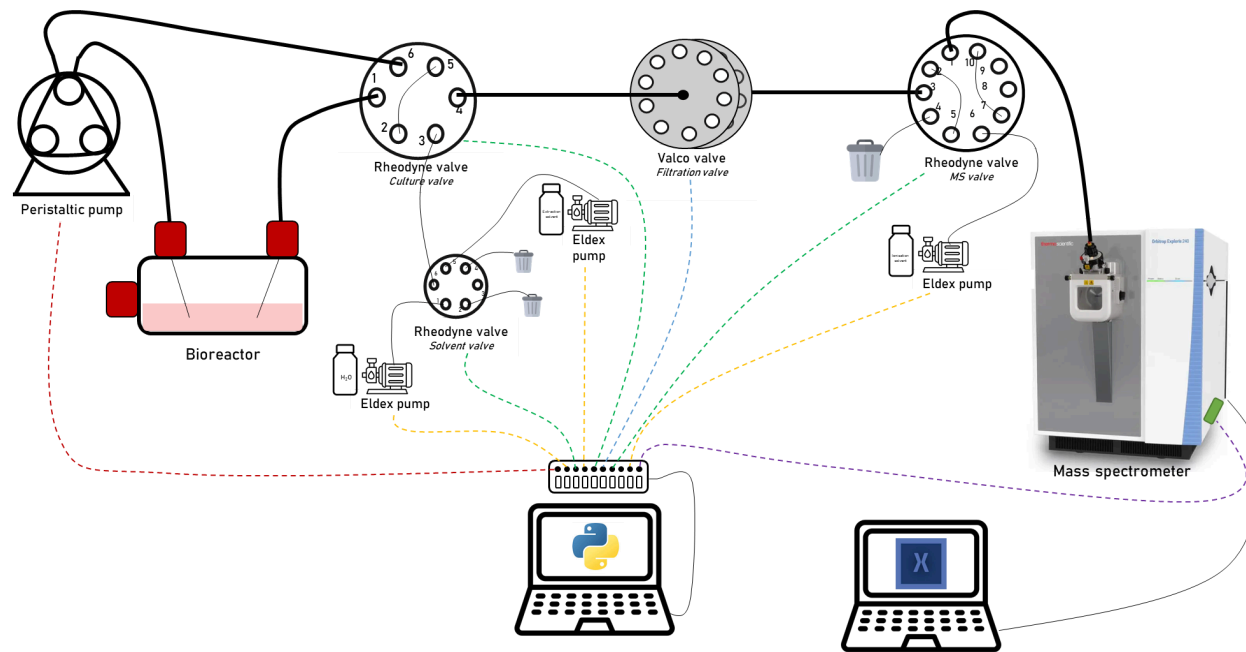
Moyens humains, hors MP

1 thèse (INRAE), 2 ans de CDD, 18 mois de post-doc, stages de M2 (fonds propres)

≈ 9 k€ pour MIA-PS

missions + 1 stage de M2 ; 3 encadrants à temps très partiel

Tâche 1 · L'instrument et source de données



Couplage culture cellulaire / spectromètre de masse, piloté en Python. Figures : Toxalim, TBI.

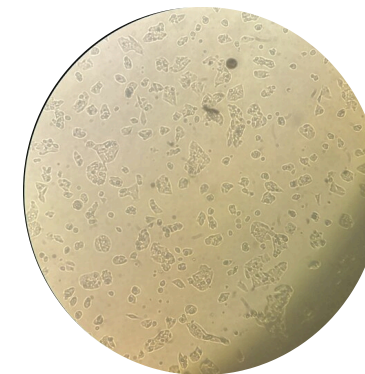
Pour nous : ≈ 100 séries courtes et bruitées, valeurs manquantes et aberrantes, au fil de l'eau.

Résultats

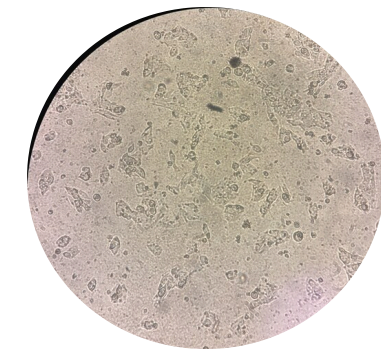
- Culture en flux continu, prélèvement automatique, injection directe dans le spectromètre
- Premiers essais : 1 prélèvement / h sur 48 h, **70 à 160 métabolites** détectés

En cours de résolution

- Stérilité de la chambre en ligne
- Extraction des signaux **en temps réel**, encore manuelle



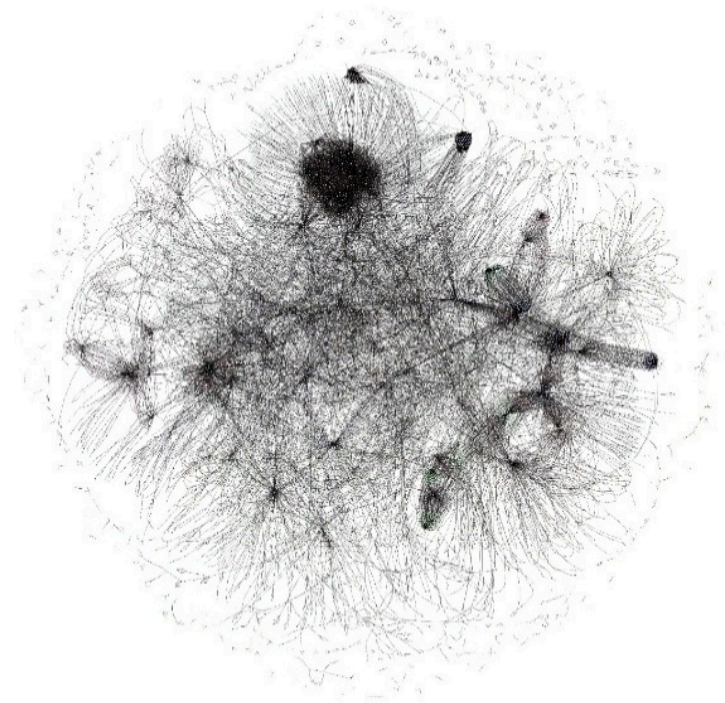
Hors ligne : culture saine



En ligne : contamination bactérienne

Tâche 2 · Le modèle : un réseau métabolique sous contraintes

Expertise Toxalim : prédire les flux **intracellulaires**, non observés, à partir des échanges avec le milieu.



Réseau Human1 : 13 417 réactions, 10 138
métabolites, 3 625 gènes (Robinson et al., 2020).

Bilan de matière : $\frac{dx}{dt} = S v \approx 0$, avec $S \in \mathbb{R}^{m \times r}$ matrice stœchiométrique et $v \in \mathbb{R}^r$ flux. Flux admissibles au temps t : un **polytope**

$$\mathcal{P}(t) = \{ v : S v = 0, \ell \leq v \leq u, v_{\text{éch}} = \hat{v}_{\text{éch}}(t) \},$$

où ℓ, u sont des bornes et $\hat{v}_{\text{éch}}(t)$ le flux d'échange mesuré dans le milieu.

- Le Polytope est exploré par **programmation linéaire** :
 - soit pour résolution $\max_{v \in \mathcal{P}(t)} c^T v$ pour un objectif c (e.g., production d'un métabolite cible)
 - soit pour évaluation de la plage de chaque flux $[\min_{\mathcal{P}(t)} v_j, \max_{\mathcal{P}(t)} v_j]$, étroite si le flux est identifiable
- **Dynamique** par morceaux : état quasi stationnaire par fenêtre de temps

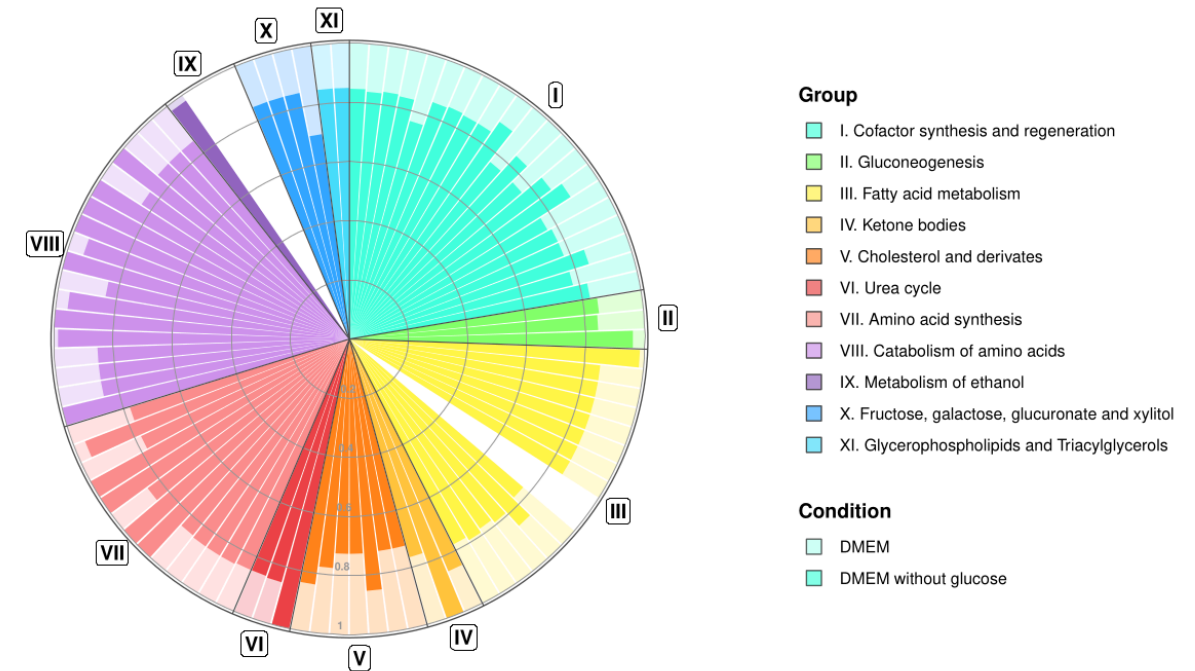
Tâche 2 · Rendre le modèle lisible et identifiable

Lisibilité : les fonctions métaboliques

- 13 000 flux : ininterprétable tel quel
- **Fonction métabolique** = capacité à produire un métabolite cible à partir de substrats donnés
- **94 fonctions**, 11 familles, tirées de la littérature ; 1 programme linéaire par fonction (package Python)

Identifiabilité : réduire le réseau

- 13 000 inconnues, $\approx 10/100$ métabolites mesurés
- Réduire à ≈ 100 réactions, en préservant **les 94 fonctions** (connaissance) et l'ajustement **aux mesures** (données)



Simulation: production maximale par fonction, milieu complet (clair) ou sans glucose (foncé). Figure : K. Rabemanantsoa.

Thèse de Koloina Rabemanantsoa (Toxalim)

À terme, décider en suivant 94 fonctions au cours du temps plutôt que 13 000 flux.

Tâche 3 · Le problème statistique

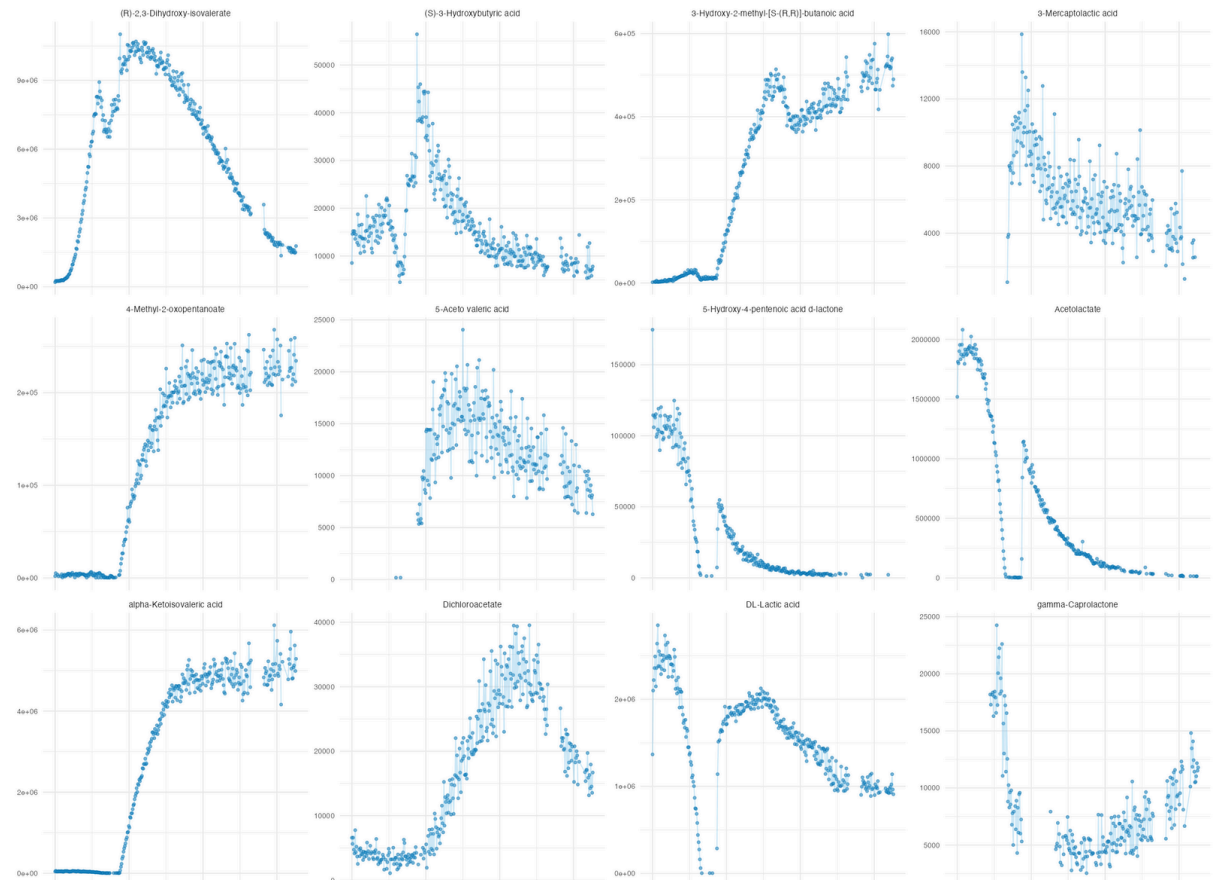
Objectif : détecter une perturbation métabolique à partir des cinétiques, si possible en temps réel

Deux scénarios

1. **Rupture nette**, partagée par beaucoup de métabolites
2. **Changement de profil** dans un groupe de métabolites

Données, en attendant le foie

- Fermentation d'*E. coli* suivie en ligne (K. Burgess, Édimbourg)
- 64 métabolites, 1 mesure / 5 min, ≈ 30 h
- Apport de glucose vers 9 h : rupture **connue**

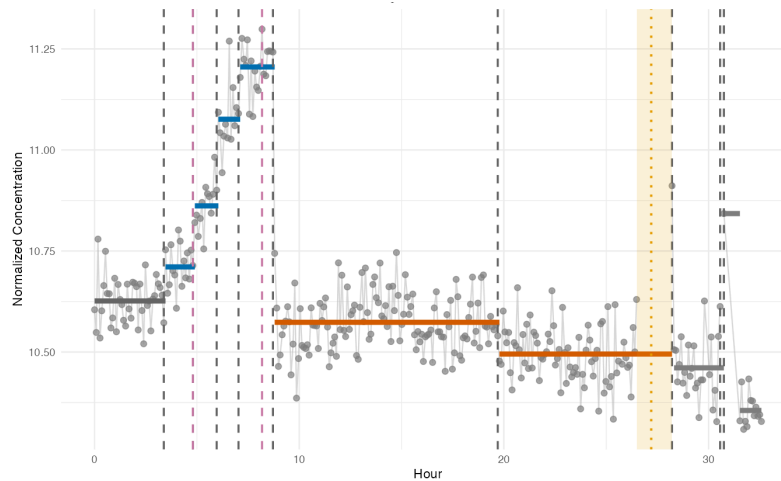


Équipe: G. Rigai (IPS2/MIA-PS, BAP), A. Leroy (GABI/MIA-PS, GA), J. Chiquet (MIA-PS, MathNum) · Z. Colas et E. Corcos (master MathSV), K. Stetsun (M2)

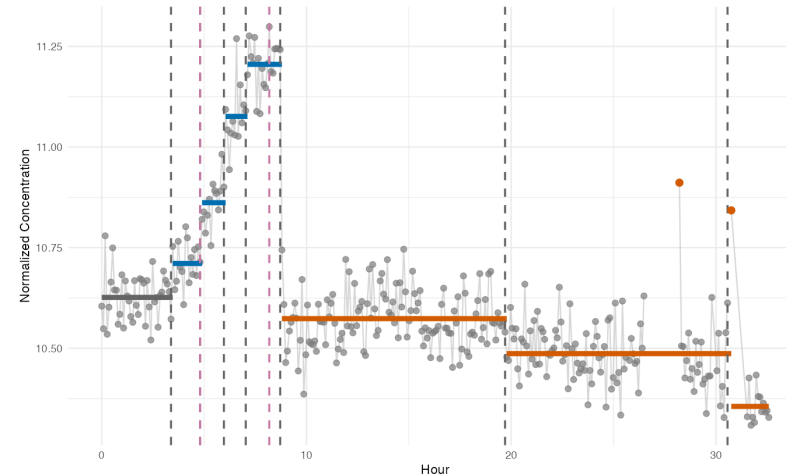
Détection de ruptures : choisir la bonne perte

avec Guillem Rigai (BAP, 50 % MIA-PS)

$$(\hat{D}, \hat{\tau}) = \arg \min_{D, \tau} \sum_{k=0}^D \left[\min_{\theta} \sum_{t=\tau_k+1}^{\tau_{k+1}} \gamma(y_t - \theta) + \beta \right], \quad \beta = 2\hat{\sigma}^2 \log n$$



PELT, $\gamma(u) = u^2$: 10 ruptures, certaines dues à des points isolés



Robseg, $\gamma_K(u) = \min(u^2, K^2)$: 7 ruptures, 2 points aberrants signalés

64 séries : 549 ruptures contre 592, ajustement quasi identique (R^2 moyen 0,86 contre 0,87). Perte bornée retenue.

Killick et al., JASA 2012 · Fearnhead et Rigai, JASA 2019 · programmation dynamique exacte, $K = 3\hat{\sigma}$, $\hat{\sigma}$ par l'estimateur de Hall.

Regrouper les cinétiques : processus gaussiens multi-tâches

avec Arthur Leroy (GA, 50 % MIA-PS)

Métabolite i = tendance de son groupe g + écart propre :

$$y_i(t) = \mu_g(t) + f_i(t) + \varepsilon_i(t) \quad \text{si } Z_i = g,$$

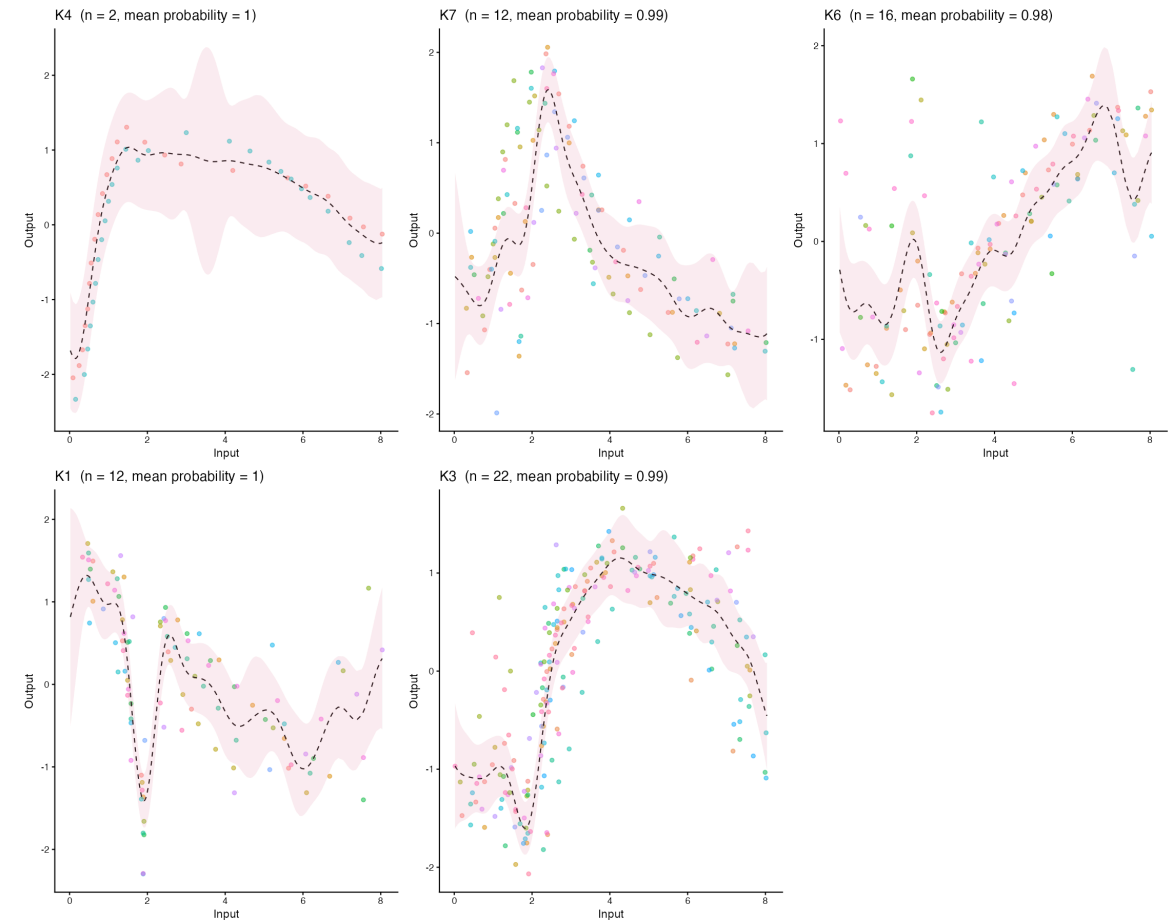
$$\mu_g \sim \mathcal{GP}(m_g, C_{\gamma_g}), \quad f_i \sim \mathcal{GP}(0, \Sigma_\theta).$$

- EM variationnel, incertitude quantifiée, échantillonnage irrégulier géré nativement

Apport du projet à l'outil

Hyperparamètres **partagés** entre séries,

$\hat{\theta} = \arg \max_{\theta} \sum_i \log L(y_i \mid \theta)$: clustering enfin stable sur profils hétérogènes



MagmaClust : Leroy, Latouche, Guedj et Gey, JMLR 2023 · 3 fonctions ajoutées à MagmaClustR pendant le stage.

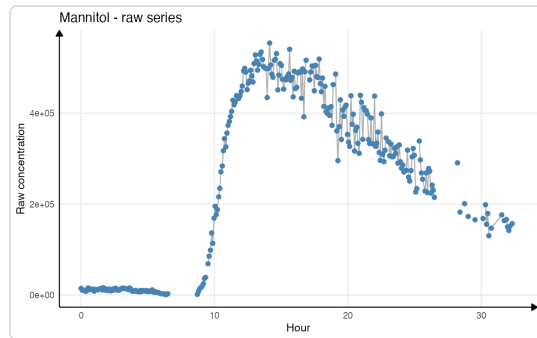
La chaîne complète, sur un métabolite

Zahra Colas et Ethan Corcos

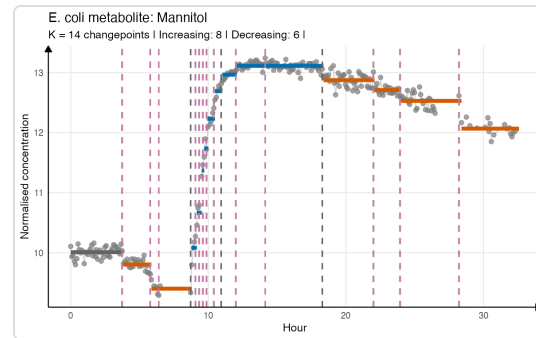
Projet de recherche, master MathSV (Paris-Saclay), 2025–26

Kateryna Stetsun

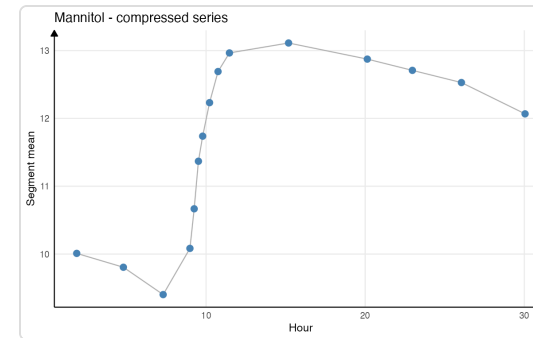
Stage de M2, master SSD (Montpellier), février–août 2026



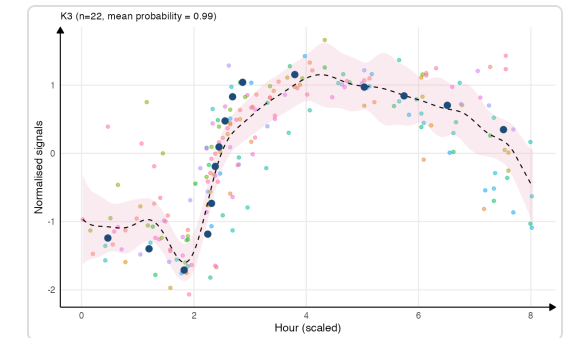
1. Série brute
variance croissante avec le niveau



2. $\log(x + a)$, puis ruptures
Robseg, perte bornée



3. Résumé
un point par segment



4. Classification
le mannitol dans son groupe

Acquis

- Chaîne reproductible documentée sur la forge
- 64 métabolites, 5 groupes de cinétiques
- 3 étudiant·es formé·es, 1 contribution logicielle

Limites, à mi-parcours

- Un seul jeu de données, pas du foie
- Pas encore de validation biologique
- Segmentation a posteriori, pas encore en ligne

Questions ouvertes côté méthodologie

Statistique

- **En ligne** : décider au fil de l'eau, retard de détection contrôlé
- **Multivarié** : ruptures partagées par un sous-ensemble de métabolites, dépendances entre séries
- **Comparer des cinétiques** entre conditions (exposé ou non) après challenge nutritionnel

InSi-liver (AMI CNRS-MITI / DIGIT-BIO, juin 2026, avec Toxalim et TBI), non retenu. **Simuler** des cinétiques à partir du réseau pour calibrer détection de ruptures et plan d'échantillonnage : dialogue entre modèle mécaniste et statistique.

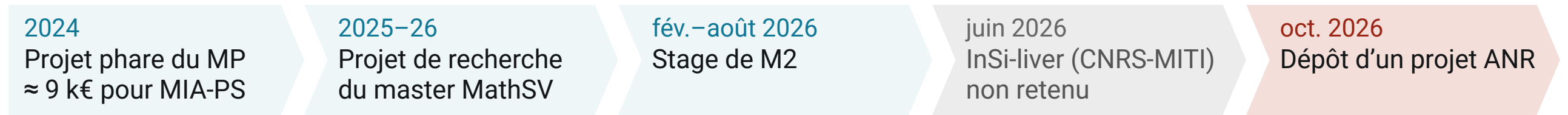
Modélisation hybride

- Statistique branchée sur le modèle mécaniste : détecter dans l'espace des **flux**, pas des concentrations
- Propager l'incertitude des mesures à travers $S v = 0$
- Réduire le réseau sans perdre les fonctions métaboliques

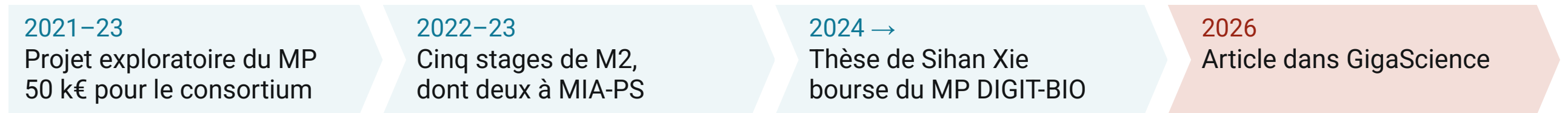
Projet ANR, dépôt en cours avec Nathalie Poupin. Métabolomique en ligne + challenges nutritionnels (sucres, acides gras) : **comparer et classer** les réponses cinétiques, avec ou sans exposition.

Effet tremplin de Digit-Biot

HepatO'Twin · Toxalim, TBI, AlimH, MICA



GenIALearn · GABI (GA), IBISC



- GenIALearn : modèle de référence des généticiens difficile à battre. Résultat négatif, qui a orienté la thèse vers les **modèles génératifs de génotypes** (VAE, diffusion, GAN)
- Même montage relancé avec l'UNH (Clermont), rencontrée dans HepatO'Twin : projet de master, puis stage de M2

Pourquoi participer à un projet de MP

1. **Ticket d'entrée modeste.** Une dizaine de k€ et du temps de chercheur : un consortium, des données originales, des questions qu'on n'aurait pas inventées.
2. **Un retour sur la production des données.** Savoir comment elles sont produites, en discuter avec ceux qui les produisent, pour affiner et ajuster les modèles : extrêmement précieux.
3. **Les questions des autres font avancer nos méthodes.** Segmentation robuste éprouvée sur données réelles, MagmaClustR amélioré, modèles génératifs pour génotypes.
4. **Les projets de MP amorcent** et préparent les suivants : thèse, projet ANR... 🏆 Europe 💰
5. **On devient identifiable.** Toxalim pour l'ANR, UNH pour un nouveau projet : le réseau s'étend de proche en proche.

Coût : temps d'acculturation, données qui tardent, résultats qui ne tombent pas dans l'année.

Merci à Nathalie Poupin et au consortium HepatO'Twin, à Guillem Rigai, Arthur Leroy, Kateryna Stetsun, Zahra Colas, Ethan Corcos, et au MP DIGIT-BIO.