

● BIOTWIN

SOIRÉE IA · QUÉBEC · MAI 2026

Du sang aux *modèles IA.*

BIOTWIN.AI · TWINME.AI



CTO · BIOTWIN

Moi, c'est *Pierrick Hauguel*.

- Arrivé à Québec il y a moins d'un an comme développeur, devenu CTO en septembre.
- Avant, en France: ma propre agence, des clients en santé et beaucoup de projets liés à l'IA.
- Aujourd'hui, entre la biologie, la donnée et le code.

Un *jumeau virtuel*, l'idée vient de l'aviation.

Chaque avion a un jumeau virtuel au sol, nourri par ses capteurs. On y rejoue les vols, on teste des scénarios, on anticipe l'usure, sans toucher à l'appareil réel.

Pareil pour une personne. Pas un avatar, pas un chatbot: une copie calculable de votre biologie, enrichie à chaque prélèvement.

L'intérêt: y *simuler* un changement, un régime, une trajectoire de maladie, au lieu d'attendre six mois pour voir le résultat sur soi.

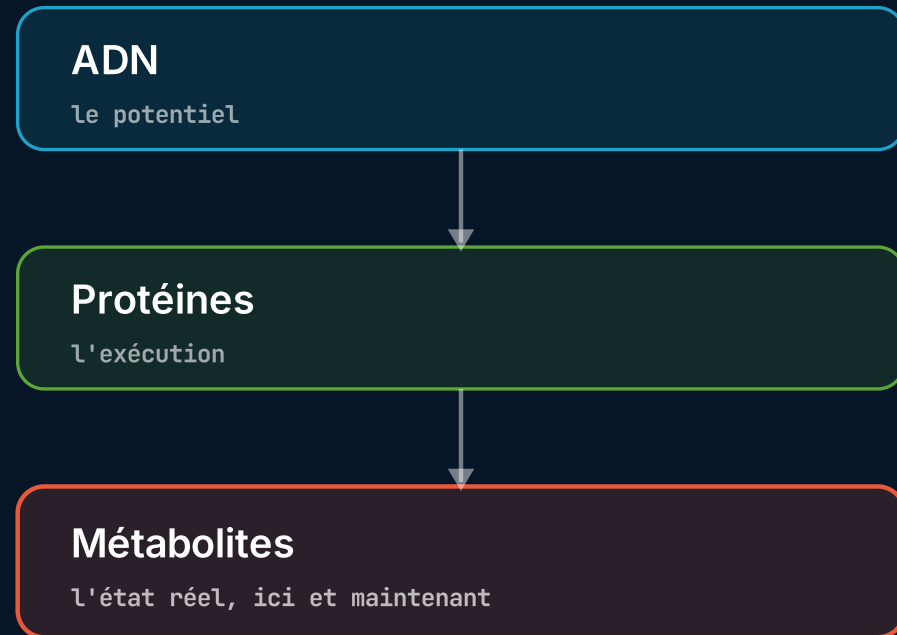


Tout part d'une *goutte de sang*, séchée sur une carte.

Pour simplifier: l'ADN, c'est le plan; les protéines, ce sont les machines qui lisent ce plan et font fonctionner les cellules; les métabolites, ce sont les petites molécules produites, consommées ou absorbées.

Dans cette goutte, on mesure surtout ces **métabolites**: sucres, lipides, hormones, traces de ce qu'on mange. C'est une photo chimique du corps à l'instant du prélèvement.

L'ADN dit ce qui *peut* arriver. Les métabolites disent ce qui se passe *maintenant*.



du potentiel à l'état réel

Du sang à la *donnée*.

Au coeur de la machine: un spectromètre de masse haute résolution, couplé à une chromatographie liquide. Il sépare les molécules par masse et par temps de rétention.

La sortie n'est pas un résultat médical. C'est un paysage de pics: la plupart sans nom, mais qu'on sait suivre et comparer.

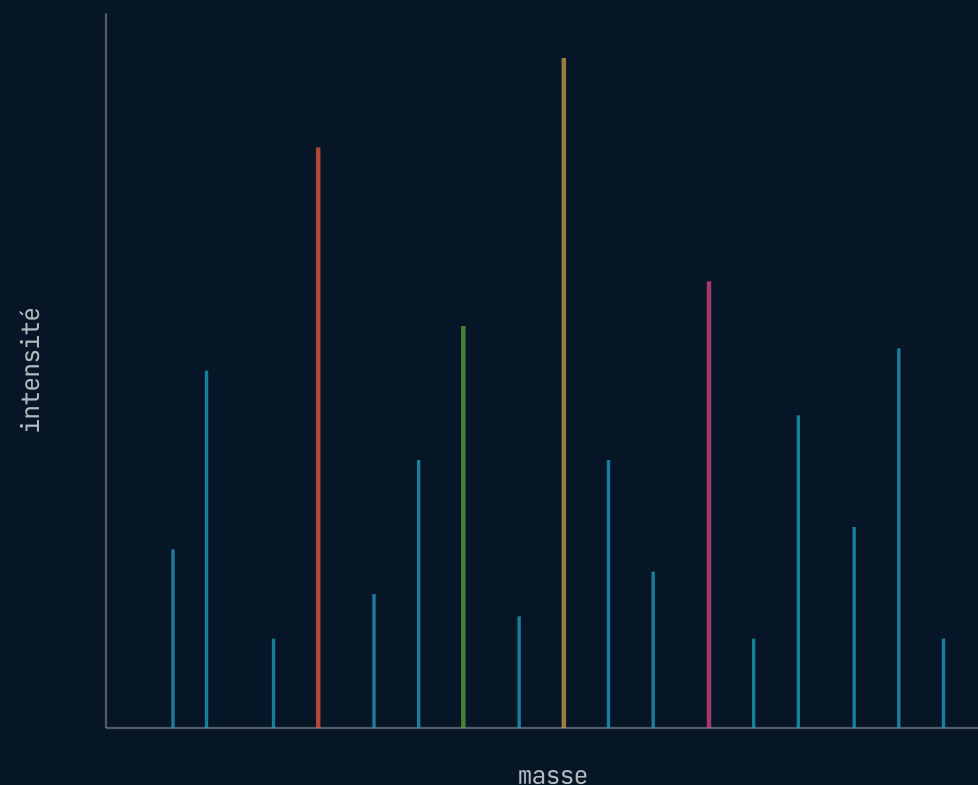
ORDRES DE GRANDEUR

~1 Go de données brutes par goutte

des dizaines de milliers de pics par échantillon

~10 000 molécules par échantillon

des millions vues au total, ~30 000 suivies



un échantillon · des dizaines de milliers de pics

Une prise de sang ne trouve que *ce qu'on lui demande*.

Chez le médecin, on cible 40 à 50 molécules: cholestérol, glucose, créatinine, testostérone. Tout le reste du sang part à la poubelle.

Six mois plus tard, un article décrit un nouveau marqueur. Pour savoir si vous l'aviez, il faut refaire une prise de sang. L'ancienne est perdue. On a trouvé ça absurde.

Alors on fait du *non-ciblé*: on ne cible rien, donc on voit tout. On mesure autant que la chimie le permet, et on garde le signal brut pour toujours. Une question posée demain peut tourner sur un échantillon d'hier.



• DÉMO LIVE

Du signal aux *scores*.

Une carte, des programmes, une trajectoire dans le temps. On passe du signal brut à quelque chose qu'un humain lit en quelques secondes.

Le même signal alimente *tout un catalogue de scores.*

Un programme n'est pas une page marketing. C'est une famille de modèles biology-first branchés sur la même base de signaux. Voici ce qui est déjà dans le portail.

OPTIMISATION HUMAINE ÉLITE

Sommeil · Stress · Nutrition · Activité physique · Substances · Connexion sociale

GESTION DU POIDS

Poids biologique · Masse musculaire · Composition en eau · Gras corporel (bientôt)

VIEILLISSEMENT DURABLE

Rythme de vieillissement biologique

RISQUE DE CANCER

Risque de cancer du sein

PERFORMANCE SPORTIVE APEX

Récupération · Préparation · Adaptation, en développement

5 programmes, une douzaine de scores en production, tous sur la même carte. Le produit devient extensible parce que l'entrée ne change pas.

Le signal est riche. Il est aussi *piégé*.

Haute dimension, petit n

~10 000 variables, ~2 000 personnes. Le surapprentissage guette: il est facile de trouver un signal qui n'existe pas.

Fuite de batch

Chaque session de labo a sa signature. Un mauvais découpage train/test apprend le lot, pas la biologie.

Pics non annotés

La plupart des molécules n'ont pas de nom. On les suit quand même, comme des variables stables, comparables dans le temps.

Généraliser dans le temps

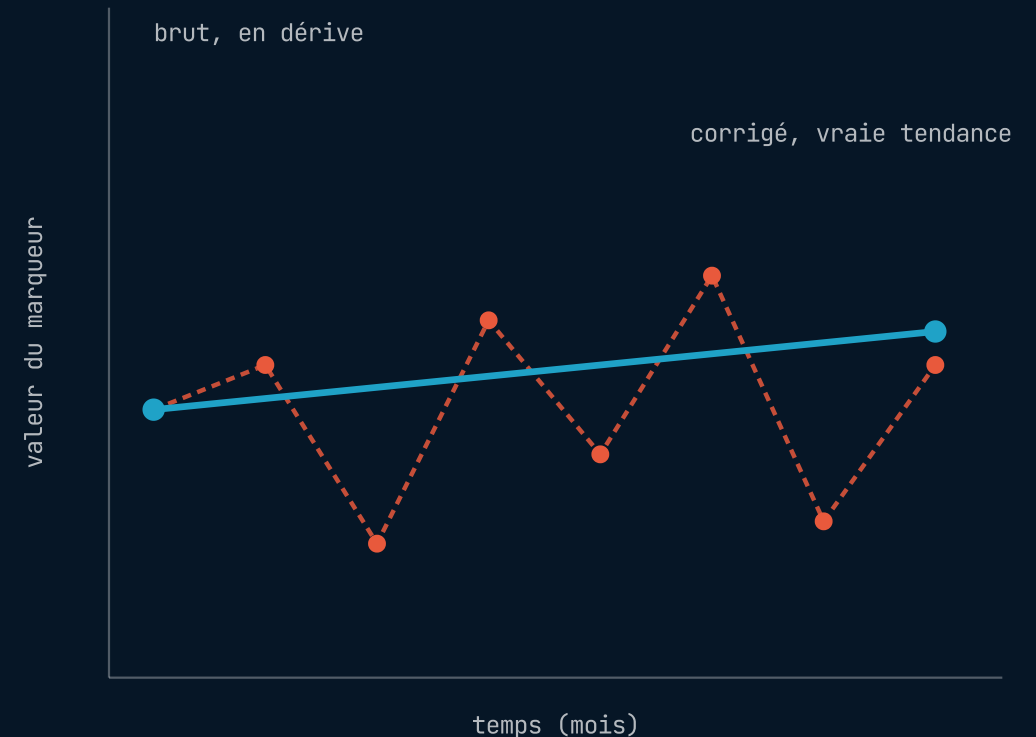
Un modèle doit tenir sur des lots futurs jamais vus, pas seulement bien marcher sur le passé.

La *même personne*, à deux mois d'écart. Deux lectures différentes.

La biologie n'est pas une webcam. Chaque session de labo dérive: température, calibration, âge de l'instrument.

Sans corriger cette dérive, la même personne ressemble à une autre d'une session à l'autre. On rend les lectures comparables, échantillon par échantillon, année après année.

Le piège: si le découpage train/test laisse fuiter le lot, le modèle apprend le lot. Sur une de nos études, un découpage naïf gonflait un score de *près de 93 points*.



même personne · huit échantillons · dérive retirée

On part de la *biologie*, puis on cherche dans tous les sens.

On ne lâche pas un algorithme sur des étiquettes de maladie en espérant une corrélation.

On choisit **à l'avance** les marqueurs décrits dans la littérature, on les normalise dans la population du même lot, et on construit des signatures interprétables.

Sur le cancer du sein, le panel de **39 métabolites était figé avant de voir les données**. Zéro sélection de variables sur les résultats.

Plusieurs modèles en parallèle

On teste plusieurs familles en même temps: linéaires, arbres, ensembles. Jamais un seul modèle, jamais un seul réglage.

Validation honnête

On garde des lots entiers à part pour tester. On mesure la vraie performance, pas le bruit de la machine.

// PREMIÈRE PREUVE

Cancer du sein: notre travail le plus avancé.

Un panel de **39 métabolites figé d'avance**, six familles de modèles comparées, une validation qui respecte les lots pour ne pas se mentir sur la performance.

Cleveland Clinic Abu Dhabi mène un pilote sur ce signal, et notre outil compagnon est autorisé aux Émirats par le Department of Health Abu Dhabi.

Ce n'est pas un dépistage. Le programme ne détecte pas le cancer et ne remplace pas le suivi clinique.

2 734

PARTICIPANTS · 114 CAS, 2 620
TÉMOINS

0,93

AUC · SÉPARE BIEN CAS ET TÉMOINS

~80%

DES CAS REPÉRÉS, À 5% DE FAUX
POSITIFS

AUC: 0,5 = hasard, 1 = parfait

Votre métabolome est une *empreinte*: on reconnaît la personne.

Au départ, une question de contrôle qualité: peut-on relier deux cartes au même individu juste par leur chimie? La réponse a dépassé nos attentes.

Sur **1 257 participants** et 18 288 échantillons, un classifieur retrouve la bonne personne avec **94% de justesse**, et **96%** sur des lots futurs jamais vus, contre un hasard à 0,09%.

C'est cette étude qui nous a appris la fuite de batch: un découpage naïf gonflait la justesse de près de 93 points.

94%

JUSTESSE, VALIDATION SANS FUITE DE LOT

96%

SUR DES LOTS FUTURS JAMAIS VUS

signatures: acides aminés, transport des acides gras, sphingolipides

Fatigue chronique: notre chantier scientifique du moment.



La fatigue chronique est mal mesurée: surtout des questionnaires, peu de biologie. On l'aborde avec la même méthode biology-first.

Sur **1784 adultes**, une signature métabolique sépare le phénotype de fatigue post-effort des témoins, et surtout elle **réplique**, sur sang séché à domicile, des voies déjà décrites sur plasma: navette de la carnitine, bêta-oxydation, cycle de Krebs, voie kynurénine.

Le score monte avec la sévérité déclarée, ce qui est rassurant: ce n'est pas un hasard de corrélation.

Travail en cours, pas encore soumis. Vocabulaire prudent: validation de construit, pas un test diagnostique.

Ce qu'on trouve, on le *publie*.

ONCOLOGIE

PRÉPRINT MEDRXIV · 2026

Metabolomic Profiling of Dried Blood Spots for Breast Cancer Detection: A Multi-Classifer Validation Study in 2,734 Participants

N. Anctil, P. Hauguel, L.-P. Noel · BioTwin

Panel de 39 métabolites figé avant acquisition, six familles de modèles, validation respectant les lots. AUC 0,93 à 0,95, environ 80% de sensibilité à 95% de spécificité sur 114 cas et 2 620 témoins.

IDENTIFICATION

PRÉPRINT · 2026

Metabolomic Fingerprinting from Dried Blood Spots Enables Individual Identification Across 1,257 Participants at 94% User-Level Accuracy

P. Hauguel, N. Anctil, S. Grobmyer, L.-P. Noel

18 288 échantillons, 134 lots, 15 mois. 94% de justesse sans fuite de lot, 96% sur lots futurs. Démontre aussi comment un découpage naïf gonfle la performance de près de 93 points.

FATIGUE CHRONIQUE

EN PRÉPARATION · 2026

Dried blood spot metabolomic signatures of post-exertional fatigue replicate pathway-level ME/CFS findings in 1,784 adults

P. Hauguel, N. Anctil, L.-P. Noel · BioTwin

Réplication, sur sang séché à domicile, de voies décrites sur plasma (carnitine, bêta-oxydation, Krebs, kynurénine). Score qui s'aligne sur la sévérité déclarée. Travail en cours.

Citations complètes et liens sur [BioTwin.ai/science](https://www.biotwin.ai/science)

Le sang est la première entrée forte. Le jumeau virtuel devient multimodal.

Aujourd'hui. Profilage sanguin large, répétable, comparable, relié à une douzaine de scores.

Maintenant. Biométries des objets connectés reliées au signal sanguin.

Ensuite. Le génome comme prior lent, pas comme verdict déterministe.

Bientôt. Des surfaces utilisables par les cliniciens et les chercheurs, sur le même moteur.

Plus tard. Des cohortes de jumeaux virtuels pour la recherche et les essais, avec des données longitudinales réelles.

QUESTIONS

Du sang aux *modèles IA.*

BIOTWIN.AI TWINME.AI

