

Titre: Développement d'algorithmes d'analyse multi-omique pour la détection précoce et le ciblage thérapeutique du cancer de l'ovaire
Title: Early detection and therapeutic targeting of ovarian cancer using multi-omic analysis algorithms

Auteur: Felix Habor
Author:

Date: 2025

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Habor, F. (2025). Développement d'algorithmes d'analyse multi-omique pour la détection précoce et le ciblage thérapeutique du cancer de l'ovaire [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/65706/>

Document en libre accès dans PolyPublie Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/65706/>
PolyPublie URL:

Directeurs de recherche: Mario Jolicoeur, & Jorgelindo Da Veiga Moreira
Advisors:

Programme: Génie biomédical
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Développement d'algorithmes d'analyse multi-omique pour la détection
précoce et le ciblage thérapeutique du cancer de l'ovaire**

FELIX HABOR

Institut de génie biomédical

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie biomédical

Mai 2025

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Développement d'algorithmes d'analyse multi-omique pour la détection précoce et le ciblage thérapeutique du cancer de l'ovaire

présenté par **Felix HABOR**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Michel PERRIER, président

Mario JOLICOEUR, membre et directeur de recherche

Jorgelindo DA VEIGA MOREIRA, membre et codirecteur de recherche

Moncef CHIOUA, membre

REMERCIEMENTS

J'aimerais remercier Pr. Mario Jolicoeur et Dr. Da Veiga Moreira pour leur accompagnement tout au long de ma maîtrise. Leur disponibilité, leurs conseils et leur expertise m'ont grandement aidé pour mener à terme mon projet de recherche. J'aimerais également souligner le travail de Raphaël Ngassam lors de son stage d'été. Son engagement, sa curiosité et son attitude positive ont permis d'apporter de nouvelles idées et ont favorisé une dynamique de travail motivante. Enfin, j'aimerais remercier mes parents, Silvia et Claudio, qui m'ont appuyé durant chaque étape de mon parcours académique et à qui je dois une grande part de ma réussite.

RÉSUMÉ

Le cancer de l'ovaire se caractérise par des symptômes souvent imperceptibles à ses premiers stades, retardant ainsi son diagnostic et contribuant à un taux de mortalité élevé. En effet, une détection précoce est fortement corrélée à de meilleurs taux de survie, mais les méthodes actuelles, telles que l'imagerie médicale, restent coûteuses et difficilement accessibles à l'ensemble de la population. Il est donc crucial de développer des approches systématiques, abordables et facilement déployables, reposant sur des analyses simples d'échantillons biologiques courants. Dans ce contexte, ce projet propose une méthode de détection précoce du cancer de l'ovaire à partir de données d'expression génique issues de biopsies de patientes. Bien que ces données proviennent actuellement de tissus cancéreux, des profils transcriptomiques d'une patiente pourraient être obtenus à partir d'échantillons sanguins, ouvrant ainsi la voie à un dépistage plus accessible et moins invasif. En combinant ces données transcriptomiques à des algorithmes d'apprentissage supervisé, cette approche vise à différencier les patientes atteintes d'un cancer de l'ovaire de celles qui sont saines. Les algorithmes ayant été testés incluent *Logistic Regression* (LR), *Linear Support Vector Classification* (Linear SVC), *K-Nearest Neighbors* (KNN) et *Random Forest* (RF). À partir de l'analyse de 13 110 niveaux d'expression génique d'une base de données pour 578 patientes, les algorithmes LR et KNN ont atteint une exactitude balancée de 99,9% pour la détection du cancer de l'ovaire. Par ailleurs, en se concentrant uniquement sur les patientes saines et celles atteintes d'un cancer de l'ovaire au stade I, et en réduisant l'analyse à un sous-ensemble de 112 gènes du métabolisme central du carbone, l'algorithme RF s'est démarqué avec une exactitude balancée de 97,5%, faisant de lui le plus performant pour la détection précoce.

En plus de distinguer avec une très grande exactitude une patiente saine d'une patiente atteinte du cancer de l'ovaire, les algorithmes développés dans ce projet permettent également de prédire la croissance tumorale et les variations métaboliques des cellules cancéreuses. Un modèle métabolique impliquant 35 réactions et 45 métabolites a été utilisé pour représenter les principales voies du métabolisme central du carbone. Dans un deuxième temps, l'intégration des données transcriptomiques à un modèle métabolique a été explorée afin de mieux décrire l'évolution du cancer de l'ovaire. Un modèle dynamique, initialement basé sur des données métaboliques provenant d'un modèle murin, avec une xénogreffé de tumeur pulmonaire humaine, a été enrichi par l'intégration de données transcriptomiques humaines aux cinétiques enzymatiques, et adapté afin de mieux refléter la croissance tumorale chez l'humain. Ces améliorations ont permis de

simuler l'évolution tumorale et les fluctuations métaboliques du cancer de l'ovaire sur une période d'au moins 270 jours. L'analyse des simulations métaboliques révèle notamment que les niveaux d'expression génique influencent la croissance tumorale et les concentrations métaboliques sans altérer leur dynamique générale, et que le volume tumoral initial influence à la fois la prolifération des cellules tumorales et les concentrations des métabolites circulants.

Grâce à l'établissement de règles gène-protéine-réaction (GPR), qui lie chaque gène à son enzyme correspondante et à la réaction enzymatique qu'elle catalyse, le modèle métabolique a également été identifié comme un outil potentiel afin d'identifier des gènes à cibler pour la conception de traitements contre le cancer de l'ovaire. En effet, le niveau d'expression d'un gène peut être modulé par un coefficient afin d'inclure l'effet de sa suppression ou de son activation sur la croissance tumorale. Ainsi, deux algorithmes de recherche de solution ont été conçus afin de repérer les combinaisons de gènes offrant la plus grande réduction de la croissance tumorale. Parmi les gènes identifiés comme cibles thérapeutiques, certains ont démontré devoir être inhibés (GPT, G6PD, PGLS, PGD, RPIA et FASN), tandis que d'autres (GOT1, GOT2, BCAT1 et BCAT2) doivent être activés pour induire un ralentissement de la croissance tumorale. Par exemple, la suppression simultanée des gènes GPT et FASN réduit jusqu'à 28,0% la croissance tumorale au jour 270 de la simulation métabolique, tandis que l'activation conjointe des gènes GOT2 et BCAT2 entraîne une diminution pouvant atteindre 38,4% sur la même période.

Ce projet de recherche propose une approche innovante et prometteuse pour la détection précoce, le profilage métabolique et le ciblage thérapeutique du cancer de l'ovaire. En intégrant l'apprentissage supervisé et un modèle métabolique amélioré, les algorithmes développés dans le cadre de ce projet permettent non seulement de distinguer avec précision les patientes atteintes du cancer de celles en bonne santé, mais aussi de prédire l'évolution tumorale et les fluctuations métaboliques associées. De plus, l'application des principes de l'intelligence artificielle pour l'identification de cibles thérapeutiques ouvre de nouvelles perspectives pour des stratégies de traitement plus ciblées et efficaces contre le cancer de l'ovaire. Ce projet a donc pour objectif principal d'offrir de nouvelles solutions visant à améliorer la prise en charge du cancer de l'ovaire et ainsi à augmenter les chances de survie des patientes.

ABSTRACT

Ovarian cancer is characterized by symptoms that are often imperceptible in its early stages, delaying its diagnosis and contributing to a high mortality rate. Indeed, early detection is strongly correlated with better survival rates, yet current methods, such as medical imaging, remain costly and are not easily accessible to the entire population. It is therefore crucial to develop systematic, affordable, and easily deployable approaches based on simple analyses of routine biological samples to improve screening. In this context, this project proposes a method for the early detection of ovarian cancer using gene expression data from patient biopsies. Although these data currently come from cancerous tissues, transcriptomic profiles from a patient could be obtained from blood samples, thus paving the way for a more accessible and less invasive screening approach. By combining these transcriptomic data with supervised learning algorithms, this approach aims to differentiate patients with ovarian cancer from healthy ones. The tested algorithms include Logistic Regression (LR), Linear Support Vector Classification (Linear SVC), K-Nearest Neighbors (KNN), and Random Forest (RF). Based on the analysis of 13 110 gene expression levels from 578 patients, LR and KNN achieved a balanced accuracy of 99.9% for ovarian cancer detection. Furthermore, when focusing exclusively on healthy individuals and stage I ovarian cancer patients and narrowing the analysis to a subset of 112 genes, RF emerged as the most effective algorithm for early detection, achieving a balanced accuracy of 97.5%.

Beyond accurately distinguishing healthy individuals from ovarian cancer patients, the algorithms developed in this project also enable the prediction of tumor growth and metabolic variations in cancerous cells. A metabolic model involving 35 reactions and 45 metabolites was used to represent the main pathways of the central carbon metabolism. In a second phase, the integration of transcriptomic data into a metabolic model was explored to better describe the evolution of ovarian cancer. A dynamic model, initially based on metabolic data from a mouse model with xenografted human lung tumors, was enhanced by integrating human transcriptomic data into the enzymatic kinetics and adapted to better reflect tumor growth in humans. These improvements enabled the simulation of tumor evolution and metabolic fluctuations in ovarian cancer over a period of at least 270 days. The analysis of the metabolic simulations notably reveals that gene expression levels influence tumor growth and metabolite concentrations without altering their general dynamics, and that the initial tumor volume influences both tumor cell proliferation and circulating metabolite concentrations.

Through the establishment of gene-protein-reaction (GPR) rules, linking each gene to its corresponding enzyme and the enzymatic reaction it catalyzes, the metabolic model was also identified as a potential tool for identifying genes to target in the design of treatments for ovarian cancer. Indeed, the expression level of a gene can be modulated by a coefficient to include the effect of its suppression or activation on tumor growth. Thus, two solution search algorithms were designed to identify combinations of genes that offer the greatest reduction in tumor growth. Among the genes identified as therapeutic targets, some showed the need to be inhibited (GPT, G6PD, PGLS, PGD, RPIA, and FASN), while others (GOT1, GOT2, BCAT1, and BCAT2) should be activated to induce a slowdown in tumor growth. For example, the simultaneous suppression of the GPT and FASN genes reduces tumor growth by up to 28.0% on day 270 of the metabolic simulation, while the joint activation of the GOT2 and BCAT2 genes results in a reduction of up to 38.4% over the same period.

This research project introduces an innovative and promising approach to early detection, metabolic profiling and therapeutic targeting of ovarian cancer. By integrating supervised learning algorithms with an enhanced metabolic model, the developed methodologies not only allow for highly accurate differentiation between healthy individuals and cancer patients but also enable predictions of tumor progression and associated metabolic fluctuations. Furthermore, the application of artificial intelligence principles for the identification of therapeutic targets using principles of artificial intelligence opens new perspectives for more targeted and effective ovarian cancer treatments. Therefore, the main objective of this project is to provide new solutions to improve ovarian cancer management and thus increase patient survival rates.

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ.....	IV
ABSTRACT	VI
LISTE DES TABLEAUX	XI
LISTE DES FIGURES	XIII
LISTE DES SIGLES ET ABRÉVIATIONS	XV
LISTE DES ANNEXES	XVIII
CHAPITRE 1 INTRODUCTION.....	1
CHAPITRE 2 REVUE DE LA LITTÉRATURE.....	5
2.1 Méthodes pour la détection précoce du cancer de l'ovaire	5
2.2 Cibles et traitement du cancer de l'ovaire	7
2.3 Modélisation du cancer de l'ovaire	9
CHAPITRE 3 MÉTHODOLOGIE.....	12
3.1 Données transcriptomiques utilisées	12
3.2 Modèle métabolique.....	12
3.2.1 Réactions biochimiques.....	13
3.2.2 Règles gène-protéine-réaction.....	13
3.2.3 Cinétiques enzymatiques.....	16
3.2.4 Vitesses maximales de réaction.....	16
3.2.5 Conditions métaboliques initiales	22
3.2.6 Volume tumoral initial	23
3.2.7 Taux de croissance tumoral chez l'humain	25
3.2.8 Flux sanguin tumoral chez l'humain	26

3.2.9	Optimisation du taux de croissance tumoral et du flux sanguin tumoral	27
3.2.10	Résumé des modifications apportées au modèle métabolique	30
3.3	Apprentissage supervisé	31
3.3.1	Métriques d'évaluation des algorithmes d'apprentissage supervisé	31
3.3.2	Méthode d'évaluation des algorithmes d'apprentissage supervisé	33
3.3.3	Optimisation de l'algorithme de sur-échantillonnage	35
3.3.4	Optimisation des algorithmes d'apprentissage supervisé.....	36
3.4	Ciblage thérapeutique.....	37
3.4.1	Évaluation d'une solution.....	39
3.4.2	Recherche en profondeur de solutions (<i>Depth Search</i>).....	41
3.4.3	Recherche intelligente de solutions (<i>Intelligent Search</i>).....	42
CHAPITRE 4	RÉSULTATS	44
4.1	Apprentissage supervisé	44
4.1.1	Hyperparamètre de l'algorithme de sur-échantillonnage	44
4.1.2	Hyperparamètres des algorithmes d'apprentissage supervisé	45
4.1.3	Impact des variables d'apprentissage sur la détection du cancer	47
4.1.4	Détection précoce du cancer	50
4.1.5	Impact des variables d'apprentissage sur la classification du stade.....	51
4.1.6	Impact des variables d'apprentissage sur l'identification de la récurrence	54
4.2	Modélisation métabolique	56
4.2.1	Croissance tumorale et fluctuations métaboliques selon le stade du cancer	57
4.2.2	Effet de la modification du volume de sang <i>Vblood</i>	60
4.2.3	Comparaison des cinétiques enzymatiques explorées.....	63
4.2.4	Coefficients <i>Fold Change</i> (FC).....	66

4.2.5	Limitation sur la durée de la simulation métabolique	67
4.3	Ciblage thérapeutique.....	68
4.3.1	Suppression et activation d'un seul gène	68
4.3.2	Suppression et activation d'une paire de gènes.....	72
4.3.3	Recherche intelligente	74
CHAPITRE 5 DISCUSSION		76
5.1	Apprentissage supervisé	76
5.2	Modélisation métabolique	79
5.3	Ciblage thérapeutique.....	81
5.4	Contributions du projet et état de la recherche.....	84
CHAPITRE 6 CONCLUSION ET RECOMMENDATIONS.....		85
RÉFÉRENCES		87
ANNEXES		92

LISTE DES TABLEAUX

Tableau 3.1 Variations estimées des vitesses maximales de réaction selon la présence d'ATP (f _{ATP(t)}), de NAD (f _{NAD(t)}) ou de NADP (f _{NADP(t)}) en tant que cofacteurs enzymatiques ...	19
Tableau 3.2 Volumes tumoraux initiaux chez l'humain.....	25
Tableau 3.3 Optimisation de la vitesse maximale du taux de croissance tumoral selon le flux sanguin tumoral pour obtenir un temps de doublement tumoral de 2,9 mois	28
Tableau 3.4 Principales modifications apportées au modèle métabolique.....	30
Tableau 3.5 Hyperparamètres testés des algorithmes d'apprentissage supervisé.....	36
Tableau 4.1 Optimisation de l'hyperparamètre de SMOTE	45
Tableau 4.2 Optimisation des hyperparamètres des algorithmes d'apprentissage supervisés.....	46
Tableau 4.3 Comparaison des exactitudes balancées des algorithmes d'apprentissage supervisé avec les hyperparamètres par défaut et les hyperparamètres optimisés	47
Tableau 4.4 Évaluation des algorithmes d'apprentissage supervisé pour la détection du cancer selon les variables d'apprentissage utilisées	48
Tableau 4.5 Évaluation des algorithmes d'apprentissage supervisé pour la détection précoce du cancer en ne considérant que les patientes saines et au stade I	50
Tableau 4.6 Évaluation des algorithmes d'apprentissage supervisé pour la classification du stade selon les variables d'apprentissage utilisées	52
Tableau 4.7 Évaluation des algorithmes d'apprentissage supervisé pour l'identification de la récurrence du cancer de l'ovaire selon les variables d'apprentissage utilisées	54
Tableau 4.8 Analyse des coefficients FC	66
Tableau 4.9 Meilleures cibles thérapeutiques par la suppression totale ($f_m = 0$) d'un seul gène... ..	70
Tableau 4.10 Meilleures cibles thérapeutiques par l'activation ($f_m = 2$) d'un seul gène.....	71
Tableau 4.11 Meilleures cibles thérapeutiques par la suppression ($f_m = 0$) et/ou l'activation ($f_m = 2$) d'une paire de gènes	72
Tableau 4.12 Réactions reliées à la production de G6P.....	73

Tableau 4.13 Meilleures cibles thérapeutiques identifiées par l'algorithme <i>Intelligent Search</i> en considérant l'activation et la suppression d'un triplet de gènes.....	75
Tableau 5.1 Cibles thérapeutiques identifiées grâce aux algorithmes de recherche	81
Tableau A.1 Métabolites et composés impliqués dans le modèle métabolique	92
Tableau A.2 Réactions et règles GPR du modèle métabolique.....	93
Tableau A.3 Noms complets des gènes associés au modèle métabolique	95
Tableau A.4 Cinétiques enzymatiques des vitesses de réactions	98
Tableau A.5 Vitesses maximales de réaction.....	100
Tableau A.6 Constantes cinétiques du modèle métabolique	102
Tableau A.7 Équations différentielles du modèle métabolique	103
Tableau B.1 Résumé des résultats et limitations des volets de la détection du cancer de l'ovaire, de la modélisation métabolique et du ciblage thérapeutique	106
Tableau C.1 Gènes associés, coefficients de Spearman, codons sélectionnés et fréquences relatives moyennes pour le design de l'ARN interférent de certains gènes cibles.....	111

LISTE DES FIGURES

Figure 1.1 Schématisation du projet de recherche	3
Figure 3.1 Réseau métabolique illustrant les métabolites (en vert), les cofacteurs enzymatiques et petites molécules (en gris) ainsi que les gènes associés aux réactions du modèle métabolique (en bleu).....	15
Figure 3.2 Variations de la vitesse maximale de réaction causée par la fluctuation en ATP, en NAD et en NADP durant le cycle cellulaire.....	20
Figure 3.3 Effet du flux sanguin tumoral (F) sur le volume tumoral total (VT) et sur le glucose extracellulaire (EGLC).....	29
Figure 3.4 Méthode d'évaluation des algorithmes d'apprentissage supervisé	34
Figure 3.5 Schématisation de l'algorithme de recherche intelligente (<i>Intelligent Search</i>).....	43
Figure 4.1 Matrices de confusion moyennant cinq K-Folds et utilisant les 112 gènes du modèle métabolique pour la détection du cancer.....	49
Figure 4.2 Matrices de confusion moyennant cinq K-Folds et utilisant les 13 110 gènes du TCGA pour la détection du cancer.....	49
Figure 4.3 Matrices de confusion moyennant cinq K-Folds pour la détection précoce du cancer	51
Figure 4.4 Matrices de confusion moyennant cinq K-Folds et utilisant les 112 gènes du modèle métabolique pour la classification du stade.....	53
Figure 4.5 Matrices de confusion moyennant cinq K-Folds et utilisant les 13 110 gènes du TCGA pour la classification du stade	53
Figure 4.6 Matrices de confusion moyennant cinq K-Folds et utilisant les 112 gènes du modèle métabolique pour l'identification de la récurrence du cancer	55
Figure 4.7 Matrices de confusion moyennant cinq K-Folds et utilisant les 13 110 gènes du TCGA pour l'identification de la récurrence du cancer	56
Figure 4.8 Volumes tumoraux et ratios énergétiques selon le stade du cancer de l'ovaire	58
Figure 4.9 Première série de métabolites selon le stade du cancer de l'ovaire.....	58

Figure 4.10 Deuxième série de métabolites selon le stade du cancer de l'ovaire.....	59
Figure 4.11 Troisième série de métabolites selon le stade du cancer de l'ovaire	59
Figure 4.12 Quatrième série de métabolites selon le stade du cancer de l'ovaire	60
Figure 4.13 Effet de la modification de <i>Vblood</i> sur le ratio <i>Vblood/VT</i>	61
Figure 4.14 Effet de la modification de <i>Vblood</i> sur les volumes tumoraux.....	62
Figure 4.15 Effet de la modification de <i>Vblood</i> sur le métabolite intracellulaire PYR et le métabolite extracellulaire HIS.....	62
Figure 4.16 Volumes tumoraux et ratios énergétiques selon différentes cinétiques enzymatiques	64
Figure 4.17 Première série de métabolites selon différentes cinétiques enzymatiques	64
Figure 4.18 Deuxième série de métabolites selon différentes cinétiques enzymatiques	65
Figure 4.19 Troisième série de métabolites selon différentes cinétiques enzymatiques.....	65
Figure 4.20 Quatrième série de métabolites selon différentes cinétiques enzymatiques	66
Figure 4.21 Volume tumoral total et concentrations en cofacteurs NAD et ATP sur 1095 jours.	67
Figure 4.22 Concentrations des métabolites intracellulaires G6P et GAP sur 1095 jours.....	68
Figure 4.23 Effet de la suppression du gène GPT pour une patiente au stade IV.....	69
Figure 4.24 Effet de l'activation du gène GOT2 pour une patiente au stade IV.....	72
Figure 4.25 Effet de la suppression simultanée des gènes HK1 et TKT pour une patiente au stade II	74

LISTE DES SIGLES ET ABRÉVIATIONS

ACLY ATP citrate lyase

ALA Alanine

ARN Acide ribonucléique

ARNic ARN interférent court

ARNm ARN messager

ARNseq Séquençage de l'ARN

ARNsh ARN en épingle à cheveux courts

ARNt ARN de transfert

ASP Aspartate

ASPG Asparaginase

BCAT1 Branched chain amino acid transaminase 1

BCAT2 Branched chain amino acid transaminase 2

BRCA1 BRCA1 DNA repair associated

BRCA2 BRCA2 DNA repair associated

CA125 Antigène tumoral 125

cfRNA ARN libre circulant

CHO Chinese hamster ovary

CTCs Cellules tumorales circulantes

DCNN Réseau neuronal convolutif profond

DHEA Dehydroepiandrosterone

EGLN Glutamine extracellulaire

FASN Fatty acid synthase

FBA Analyse de bilan de flux

FC Fold Change

FDA Food & Drug Administration

F6P Fructose 6-phosphate

GAP Glycéraldéhyde 3-phosphate

GEM Modèle métabolique à l'échelle du génome,

GOT1 Glutamic-oxaloacetic transaminase 1

GOT2 Glutamic-oxaloacetic transaminase 2

GPR Gène-protéine-réaction

GPT Glutamic-pyruvic transaminase

G6P Glucose 6-phosphate

G6PD Glucose-6-phosphate dehydrogenase

HGNC HUGO Gene Nomenclature Committee

HGSOC Cancer séreux de haut grade

HK1 Hexokinase 1

HK2 Hexokinase 2

IRM Imagerie par résonance magnétique

KNN K-Nearest Neighbors

LGSOC Cancer séreux de bas grade

LR Logistic Regression

MCS Minimal cut sets

MD-SCs Cellules suppressives d'origine myéloïde

miARN Micro-ARN

MUC16 Mucin 16 cell surface associated

NB Naive Bayes

NCI National Cancer Institute

NHGRI National Human Genome Research Institute

PGD Phosphogluconate dehydrogenase

PGLS Phosphogluconolactonase

PGM1 Phosphoglycerate mutase 1

SMOTE Synthetic Minority Over-sampling Technique

SLC2A1 Solute carrier family 2 member 1

SVC Support Vector Classification

SVM Support Vector Machine

RF Random Forest

RISC RNA-Induced Silencing Complex

RNN Réseau de neurones récurrents

ROCA Risk of Ovarian Cancer Algorithm

ROMA Risk of Ovarian Malignancy Algorithm

RPIA Ribose-5-phosphate isomerase A

TCA Acide tricarboxylique

TCGA The Cancer Genome Atlas

TDM Tomodensitométrie

TDT Temps de doublement tumoral

TP53 Tumor protein p53

TKT Transketolase

XGBoost Extreme Gradient Boosting

2-DG 2-Deoxy-D-glucose

6-AN 6-Aminonicotinamide

LISTE DES ANNEXES

ANNEXE A Modèle métabolique.....	92
ANNEXE B Résumé des résultats et limitations du projet.....	106
ANNEXE C Design d'un ARN interférent	108

CHAPITRE 1 INTRODUCTION

La Société canadienne du cancer estime qu'en 2024, le cancer de l'ovaire entraînera près de 2000 décès au Canada et que près de 3000 nouveaux cas seront diagnostiqués [1]. En raison de l'absence de symptômes spécifiques apparaissant aux stades précoces, le cancer de l'ovaire est diagnostiqué plus de 75% du temps à un stade avancé [2]. Il en résulte des taux de survie à cinq ans pour des cancers diagnostiqués aux stades III et IV respectivement de 42% et 26% [3], mais de 93% lorsque détecté à un stade initial [3]. Compte tenu que les options de protocoles approuvés pour le traitement du cancer de l'ovaire se limitent principalement à la chirurgie et la chimiothérapie, des moyens très invasifs, il est devenu prioritaire en recherche biomédicale d'améliorer les moyens de détection précoce du cancer de l'ovaire afin d'améliorer le taux de survie mais également la qualité de vie des patientes.

Étant donné que le cancer est parfois relié à des mutations génétiques survenant notamment durant la réPLICATION de l'ADN, il est en théorie possible de repérer des marqueurs du cancer dans le transcriptome, c'est-à-dire à partir de l'analyse des ARN d'un individu. L'étude du transcriptome permet en effet d'identifier des altérations dans l'expression des gènes impliqués dans le développement tumoral [4]. L'utilisation de l'intelligence artificielle, via des algorithmes d'apprentissage supervisé, permettrait d'analyser ces altérations géniques afin d'identifier des marqueurs transcriptomiques associés aux premiers stades du cancer de l'ovaire, améliorant ainsi sa détection précoce. Cette approche permettrait ainsi un gain de précision et un complément à l'imagerie médicale. En effet, bien que cette dernière soit efficace pour détecter des tumeurs à un stade avancé, elle ne permet pas toujours une détection fiable lorsque la masse tumorale est réduite et localisée [5]. À l'inverse, une analyse sanguine couplée à des biopsies peut s'appuyer sur des données transcriptomiques et métaboliques pour identifier des marqueurs tumoraux. En effet, les ARN messager (ARNm), libérés par les cellules, circulent dans le sang et peuvent être analysés, à la fois issus du tissu épithelial et dans la circulation sanguine, par séquençage d'ARN (ARNseq) afin de détecter des altérations spécifiques pouvant être associées aux premiers stades du cancer de l'ovaire [6] [7]. Cette approche pourrait ainsi assurer un diagnostic précoce et augmenter les chances de survie des patientes à la suite d'un traitement moindrement invasif et hautement efficace. De plus, si les méthodes d'analyses gagnent en sensibilité, il est envisageable, pour les prochaines années, de limiter ces analyses à de simples dépistages sanguins.

Au-delà des altérations géniques, les cellules cancéreuses présentent également des modifications métaboliques qui favorisent leur mode prolifératif et leur survie. L'étude du métabolome, soit l'ensemble des molécules impliquées dans le métabolisme cellulaire, permettrait de mieux comprendre ces dérèglements et d'identifier des voies métaboliques significatives dans l'occurrence et le développement tumoral [8]. En intégrant des données transcriptomiques et métabolomiques, il devient possible de modéliser le métabolisme des cellules tumorales, de façon spécifique. Cette modélisation permettrait également d'identifier des cibles thérapeutiques potentielles en révélant les gènes et les enzymes traduites qui ont un grand impact sur la survie et la prolifération des cellules cancéreuses.

Objectifs du projet

L'amélioration du taux de survie du cancer de l'ovaire repose sur des avancées tant dans le diagnostic que dans le traitement de cette maladie. Ce projet cherche d'abord à explorer une méthode de détection précoce fondée sur l'analyse du profil transcriptomique des patientes. Ce profil permet non seulement de détecter la maladie, mais aussi d'enrichir un modèle métabolique du développement tumoral en ajustant les vitesses de réaction en fonction des niveaux d'expression génique. Ce modèle sera ensuite utilisé pour identifier des cibles thérapeutiques potentielles, à savoir des gènes à moduler afin de réduire la croissance tumorale. Les objectifs spécifiques de ce projet sont les suivants :

1. Développer une méthode algorithmique de détection précoce du cancer de l'ovaire basée sur l'analyse de données transcriptomiques.
2. Intégrer les données transcriptomiques dans les cinétiques enzymatiques d'un modèle métabolique.
3. Transposer un modèle métabolique murin vers un modèle métabolique humain du cancer de l'ovaire.
4. Concevoir un algorithme basé sur l'intelligence artificielle afin d'identifier des gènes cibles thérapeutiques démontrant une réduction de la croissance tumorale.

Organisation du mémoire

Le présent mémoire commence par un premier chapitre portant sur la revue de la littérature et présentant les avancées réalisées en matière de détection, de traitement et de modélisation du cancer de l'ovaire. Un deuxième chapitre présente la méthodologie utilisée pour les approches

développées dans les trois volets principaux du projet: la détection précoce du cancer à partir de données transcriptomiques, la modélisation du métabolisme tumoral et l'identification de cibles thérapeutiques. Un chapitre regroupe par la suite les résultats obtenus pour chacun de ces volets, qui sont présentés et analysés afin d'évaluer leur portée et leur pertinence. Un chapitre élaborant une discussion des résultats et des approches utilisées approfondit les analyses effectuées et examine les contraintes et limitations du projet. Enfin, un chapitre finalise ce mémoire par une conclusion qui résume les contributions du travail de recherche et propose des perspectives pour des recherches futures. À l'Annexe C, nous présentons une approche exploratoire visant à exploiter les outils développés dans nos travaux pour permettre l'identification de cibles thérapeutiques, soit des gènes dont la modulation du niveau d'expression par des ARN interférents permet de réduire la croissance tumorale. Cette approche se base sur la conception d'un ARN interférent visant à surcharger et épuiser les ressources métaboliques requises pour la traduction du gène cible. Toutefois, cette approche est strictement exploratoire et ne figure pas comme volet principal du projet, mais plutôt comme perspective pour relier le ciblage thérapeutique à la conception d'un traitement personnalisé pour les patientes.

Afin de mieux saisir la conduite de ce projet de recherche, la Figure 1.1 illustre les liens entre les différents volets et résultats du projet.

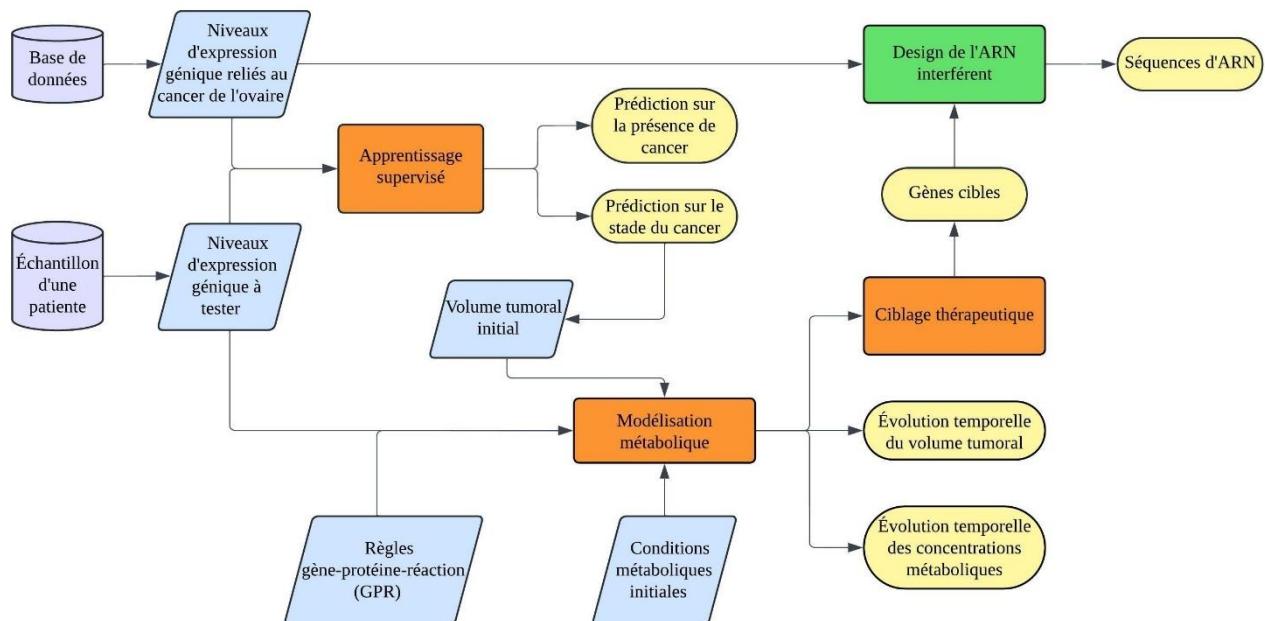


Figure 1.1 Schématisation du projet de recherche

Le schéma illustre les trois volets principaux (en orange) et le volet exploratoire (en vert) du projet, ainsi que leurs interactions avec les entrées (en bleu), les données transcriptomiques (en violet) et les sorties (en jaune). L'apprentissage supervisé permet d'analyser les niveaux d'expression génique, issus d'une base de données (ensemble d'entraînement) ou d'un échantillon de cellules prélevées de patientes (échantillon de test), afin de prédire la présence et le stade du cancer de l'ovaire. Il est important de préciser qu'aucun échantillon issu de biobanques locales n'a été utilisé dans ce projet. Les niveaux d'expression génique en mains proviennent exclusivement d'une base de données. La mention d'un échantillon prélevé d'une patiente dans le schéma sert uniquement à illustrer l'application potentielle des algorithmes dans un contexte clinique. Ensuite, les niveaux d'expression génique, combinés à un volume tumoral initial défini selon le stade du cancer, servent de base à la modélisation métabolique. Celle-ci s'appuie sur des règles GPR et des conditions métaboliques initiales pour simuler l'évolution du volume tumoral et des concentrations métaboliques. La modélisation métabolique permet ensuite d'identifier des cibles thérapeutiques, soit des gènes jugés comme essentiels à la progression tumorale. Ces gènes peuvent être exploités pour le design d'un ARN interférent, dont l'objectif est de réduire ou d'inhiber leur expression. En sortie, les algorithmes développés dans ce projet fournissent des prédictions diagnostiques, des simulations de l'évolution tumorale et métabolique, du niveau d'expression effective des gènes cibles thérapeutiques, et, potentiellement des séquences d'ARN interférent optimisées.

CHAPITRE 2 REVUE DE LA LITTÉRATURE

La revue de la littérature vise à synthétiser les avancées réalisées en matière de détection, de traitement et de modélisation du cancer de l'ovaire. Les références ont été sélectionnées en fonction de leur pertinence par rapport aux objectifs de ce projet, en privilégiant des études récentes pour refléter les tendances actuelles dans la prise en charge de ce cancer.

2.1 Méthodes pour la détection précoce du cancer de l'ovaire

À titre de cancer agressif, la détection précoce du cancer de l'ovaire est étroitement liée aux chances de survie des patientes. Ainsi, de nombreux groupes de recherche ont développé et testé diverses méthodes de diagnostic précoce. Parmi celles-ci, plusieurs méthodes sont basées sur l'imagerie médicale, notamment l'échographie transvaginale étant une des modalités préférées pour imager les structures anatomiques confinées au pelvis, en raison du faible coût et du niveau de confort de la patiente [9]. D'autres techniques telles que la tomodensitométrie (TDM) ainsi que l'imagerie par résonance magnétique (IRM) sont particulièrement utiles pour la planification chirurgicale [10]. Il est ainsi devenu récemment possible d'utiliser l'apprentissage machine basé sur ce type d'images médicales pour améliorer la détection du cancer de l'ovaire. Par exemple, Gao *et al.* (2022) ont développé un réseau neuronal convolutif profond (DCNN) pour assister les radiologues dans le diagnostic du cancer de l'ovaire à partir d'images échographiques. Leur modèle a démontré une exactitude diagnostique significativement améliorée lorsque l'évaluation est assistée par DCNN, atteignant un taux de 87,6% comparé au taux de 78,3% obtenu par les radiologues seuls [11].

D'autres méthodes reposent sur l'identification de biomarqueurs moléculaires dont l'expression est altérée; soit surexprimée, soit sous-exprimée chez une patiente atteinte du cancer de l'ovaire. L'antigène tumoral 125 (CA125), une glycoprotéine codée par le gène *mucin 16 cell surface associated* (MUC16), est le biomarqueur sérique le plus étudié pour le diagnostic du cancer de l'ovaire, car il est surexprimé chez plus de 80% des patientes atteintes de ce cancer [12]. Ainsi, plusieurs algorithmes ont été conçus pour analyser les concentrations sanguines de CA125 dans le cadre de la détection et du suivi du cancer de l'ovaire. Parmi eux, le *Risk of Ovarian Cancer Algorithm* (ROCA), un essai statistique basé sur plusieurs mesures de concentrations du CA125

dans le sang, et le *Risk of Ovarian Malignancy Algorithm* (ROMA) [13], un autre essai statistique qui intègre les concentrations sanguines de CA125, de la protéine 4 de l'épididyme humain sérique (HE4) et du statut ménopausique, ont été développés. Selon le niveau de risque ainsi calculé, les patientes peuvent être orientées vers des examens d'imagerie, tels que l'échographie transvaginale, ou vers une consultation classique [14]. Akazawa et Hashimoto (2020) ont notamment testé cinq algorithmes de classification, soient le *Support Vector Machine* (SVM), le *Random Forest* (RF), le *Naive Bayes* (NB), le *Logistic Regression* (LR) et l'*Extreme Gradient Boosting* (XGBoost), selon leur capacité à détecter le cancer de l'ovaire à partir de 16 variables incluant la concentration en CA125 et quelques autres biomarqueurs moléculaires. La meilleure exactitude a été obtenue avec l'algorithme XGB, avec une exactitude atteignant jusqu'à 80% [15]. De leur côté, Lu *et al.* (2020) ont testé trois algorithmes de détection du cancer de l'ovaire, à savoir le LR, un arbre de décision à deux niveaux et le ROMA, à partir de 49 variables comprenant entre autres les données d'une analyse sanguine de routine et plusieurs marqueurs tumoraux comme le CA125. Leur étude a révélé que l'algorithme RF offrait la meilleure exactitude, atteignant jusqu'à 97.4% [13].

Finalement, l'utilisation de larges bases de données comme celle du programme *The Cancer Genome Atlas* (TCGA), a permis d'entrainer et de tester des algorithmes d'apprentissage machine à partir de plusieurs types de données. Liñares-Blanco *et al.* (2021) ont notamment répertorié les algorithmes et les types de données utilisés pour la détection du cancer dans plus de 100 études ayant utilisé une base de données du TCGA. Ils ont constaté que les algorithmes les plus utilisés, par ordre d'importance, étaient les variantes du SVM, suivies des algorithmes basés sur des arbres de décision, puis des modèles linéaires généralisés. De plus, les données d'expression génique se sont révélées être le type de données le plus fréquemment utilisé, probablement en raison de leur disponibilité, comparativement aux données issues des images, des micro-ARN (miARN) et de la protéomique. Étant donné que de telles bases de données comportent, dans le cas de données omiques, beaucoup plus de variables que d'échantillons, les types d'algorithmes favorisés sont ceux n'étant pas très sensibles à une faible proportion d'échantillons comme le SVM, le RF et le LR [16]. En revanche, Liñares-Blanco *et al.* stipulent que les travaux utilisant des données d'imagerie ont plus tendance à recourir aux réseaux de neurones, car la proportion d'échantillons est généralement plus élevée [16]. Par ailleurs, la recherche menée par Arfiani et Rustam (2019), qui s'appuie sur des données issues de l'UCI Machine Learning Repository, comprenant 15 154 expressions de gènes pour 266 échantillons, a mis en évidence l'efficacité de l'algorithme RF et de

la méthode bagging (*bootstrap aggregating*) pour détecter le cancer de l'ovaire. La méthode bagging consiste à entraîner plusieurs modèles sur des sous-ensembles aléatoires des données d'entraînement et à combiner leurs prédictions pour améliorer la stabilité et l'exactitude globale [17]. Les exactitudes obtenues, en utilisant 90% des données pour l'entraînement, étaient respectivement de 98,2% pour l'algorithme RF et de 100% pour la méthode bagging [17].

En constante évolution, intégrant les nouveaux développements en intelligence artificielle, les méthodes de détection précoce du cancer de l'ovaire sont nombreuses, variées, exploitant des variables issues d'images médicales, de biomarqueurs spécifiques au cancer de l'ovaire et d'expressions géniques. Ces approches permettent de tirer parti de différentes sources de données pour améliorer la précision des diagnostics et identifier le cancer à des stades précoces et ce, afin de maximiser les chances de survie des patientes.

2.2 Cibles et traitement du cancer de l'ovaire

Le cancer de l'ovaire ne constitue pas une maladie unique. En réalité, plusieurs sous-types ont été répertoriés, et sont d'abord classés selon le type cellulaire affecté, en plus des caractéristiques morphologiques du développement tumoral. Le cancer ovarien apparaît principalement dans trois types cellulaires: les cellules épithéliales, les cellules stromales et les cellules germinales. Le carcinome épithelial de l'ovaire est le plus fréquent et représente de 85 à 90% des tumeurs malignes ovariennes [18], tandis que les tumeurs stromales et germinales malignes ne représentent respectivement que moins de 5% et de 2% des cas [19]. Parmi les sous-types du carcinome épithelial de l'ovaire, on distingue le cancer séreux de haut grade (HGSOC), représentant environ 68% des cas de cancer de l'ovaire [20], le cancer séreux de bas grade (LGSOC), le carcinome endométrioïde, le carcinome à cellules claires et le carcinome mucineux [19]. Étant le sous-type le plus fréquent et l'un des plus agressifs du cancer de l'ovaire, le HGSOC constitue le principal sujet d'étude de ce projet. Ce cancer se caractérise par une forte instabilité génomique, notamment par la présence de mutations du gène suppresseur *tumour tumor protein p53* (TP53) dans 80% des cas [20]. Par ailleurs, les mutations des gènes *BRCA1 DNA repair associated* (BRCA1) et *BRCA2 DNA repair associated* (BRCA2), impliqués dans la réparation de l'ADN, sont retrouvées dans 90% des cas héréditaires de HGSOC, soulignant leur rôle clé dans la prédisposition à cette maladie [20].

Le traitement standard du cancer de l'ovaire combine une chirurgie cytoréductrice visant à retirer un maximum de la tumeur, suivie d'une chimiothérapie à base de platine et de taxanes, comme le carboplatine et le paclitaxel, afin d'éliminer les cellules tumorales résiduelles [21]. Environ 80% des patientes atteintes de cette maladie suivent ce traitement standard. Cependant, près de 70 % d'entre elles rechutent, le cancer devenant souvent résistant à la chimiothérapie à base de platine [22]. Par conséquent, afin d'améliorer la survie des patientes, compte tenu du faible taux de succès du traitement standard, l'exploration de nouvelles approches est requise, et les thérapies métaboliques représentent une des voies les plus prometteuses par leur haute spécificité d'action pouvant réduire significativement les effets secondaires problématiques.

La reprogrammation du métabolisme énergétique, reconnue comme une caractéristique émergente du cancer par Hanahan et Weinberg (2011) [23], engendre des vulnérabilités métaboliques spécifiques au cancer qui peuvent être ciblées pour le développement de nouvelles thérapies [24]. L'effet Warburg, décrit par l'Allemand Otto Heinrich Warburg au début des années 1900, désigne la tendance des cellules cancéreuses à privilégier la glycolyse aérobie à la phosphorylation oxydative pour produire de l'énergie, même en présence d'oxygène [25] [26] [27]. Contrairement à la phosphorylation oxydative, la glycolyse aérobie produit moins d'ATP [26], mais entraîne une augmentation de la prolifération cellulaire, une absorption accrue de glucose et la production de lactate [27]. En ce sens, le transporteur de glucose GLUT1, codé par le gène *solute carrier family 2 member 1* (SLC2A1), ainsi que l'enzyme de la voie glycolytique *hexokinase 2* (HK2), tous deux surexprimés dans le HGSOC, sont considérés comme de potentielles cibles thérapeutiques contre le cancer de l'ovaire [26] [27]. Les inhibiteurs de GLUT1 et de HK2, tels que le BAY-876 et le *2-deoxy-D-glucose* (2-DG), ont notamment été proposés comme traitements potentiels [27]. En parallèle de la glycolyse, le métabolisme des acides gras et la voie des pentoses phosphates sont également altérés dans le cancer de l'ovaire, identifiant ainsi d'autres potentielles cibles thérapeutiques. Dans le HGSOC, la synthèse des acides gras est amplifiée [26] et l'accumulation de lipides est associée à un mauvais pronostic [27]. Une cible thérapeutique prometteuse est la *fatty acid synthase* (FASN), une enzyme clé de la lipogenèse qui convertit l'acétyl-CoA en acides gras comme le palmitate [27]. Alors que son expression est faible chez les adultes en bonne santé [26], la FASN est fortement surexprimée dans le HGSOC [27]. Des inhibiteurs comme le Compound 34 et la cérolénine montrent un bon potentiel pour cibler cette enzyme [26] [27]. De plus, l'inhibition de l'*ATP citrate lyase* (ACLY) par l'ARN en épingle à

cheveux courts (ARNsh), une enzyme reliant la glycolyse au métabolisme des acides gras et surexprimée dans le HGSOC, supprime la prolifération des cellules cancéreuses de l'ovaire [27]. La voie des pentoses phosphates, dérivée de la glycolyse, est surrégulée dans plusieurs cancers, dont le HGSOC, et soutient la prolifération tumorale en fournissant à la fois des pentoses pour la synthèse des nucléotides et du cofacteur réactionnel NADPH [27]. La surexpression de la *Glucose-6-phosphate dehydrogenase* (G6PD), enzyme limitante de la voie des pentoses phosphates, est notamment impliquée dans l'initiation et la progression du HGSOC et est liée à la résistance à certains agents chimiothérapeutiques tel que le paclitaxel [27]. Les inhibiteurs de la G6PD comme le *6-aminonicotinamide* (6-AN) et le *dehydroepiandrosterone* (DHEA) peuvent donc limiter la progression du HGSOC [27]. Enfin, la metformine, un médicament réduisant l'accessibilité en glucose et modulant la différenciation et l'activation de cellules immunitaires [27], a également été étudiée comme adjuvant au traitement du cancer de l'ovaire [25] [26] [27].

En résumé, le cancer de l'ovaire regroupe plusieurs sous-types, dont le HGSOC, le plus fréquent et le plus agressif. Malgré un traitement standard combinant chirurgie et chimiothérapie, les taux de rechute restent élevés en raison de la résistance aux traitements. L'exploration de nouvelles approches, telles que les thérapies métaboliques, offre des perspectives prometteuses en ciblant des vulnérabilités spécifiques du métabolisme tumoral, notamment la glycolyse, le métabolisme des acides gras et la voie des pentoses phosphates.

2.3 Modélisation du cancer de l'ovaire

Les modèles développés pour reproduire et analyser les processus biologiques du cancer se répartissent en trois grandes catégories : les modèles *in vitro*, les modèles *in vivo* et les modèles computationnels [28]. Les modèles *in vitro*, basés sur la culture cellulaire en deux ou trois dimensions, permettent d'étudier les mécanismes cellulaires et moléculaires du cancer dans un environnement contrôlé, facilitant ainsi l'étude du comportement des cellules cancéreuses et de tumeurs, ainsi que l'évaluation de traitements potentiels. Les modèles *in vivo*, qui reposent sur l'injection ou le déclenchement d'une masse de cellules cancéreuses dans un organisme vivant, comme la souris, offrent une reproduction plus fidèle de la progression tumorale grâce aux interactions complexes avec le microenvironnement de l'organisme. Enfin, les modèles computationnels s'appuient sur des approches mathématiques et informatiques pour simuler la

dynamique du cancer, intégrant plusieurs données expérimentales afin de caler le modèle sur la réalité biologique, pour ultimement prédire l'évolution tumorale. Dans le cadre de ce projet, les modèles computationnels sont privilégiés, car ils permettent de générer rapidement des données sur l'évolution du cancer et d'identifier de potentielles cibles thérapeutiques avec une grande efficacité.

Parmi les modèles computationnels utilisés pour modéliser le cancer, on retrouve notamment les approches d'optimisation tels que l'analyse de bilan de flux métaboliques (FBA) et les modèles dynamiques intégrant des cinétiques enzymatiques. D'une part, le FBA est une méthode mathématique qui analyse les flux métaboliques en utilisant un modèle métabolique jusqu'à l'échelle du génome (GEM) [29]. Cette méthode requiert l'élaboration d'une fonction objectif, étant généralement définie comme la production de biomasse, c'est-à-dire la croissance cellulaire, dans le cas de la modélisation du cancer. Les flux métaboliques sont ensuite optimisés selon cette fonction objectif, ce qui peut restreindre la modélisation à un seul processus cellulaire [30]. Le FBA présente l'avantage d'être peu coûteux en temps de calcul et en quantité d'informations requises, en supposant l'optimalité et un état stationnaire dans lequel les concentrations en métabolites restent constantes [24]. Cette approche permet donc d'intégrer des GEM assez complexes comme le modèle *Human1* comprenant 13 096 réactions biochimiques. En utilisant ce modèle, Meeson et Schwartz (2024) [30] ont notamment pu estimer les différences métaboliques entre deux sous-types du cancer ovarien, soient le HGSOC et LGSOC. Ils ont également intégré les données transcriptomiques du *Broad DepMap Portal* pour restreindre certaines réactions. Ainsi, à l'aide de l'approche FBA, on optimise les flux métaboliques en leur imposant des bornes maximales et minimales, et les niveaux d'expression génique issus des données transcriptomiques permettent d'imposer des contraintes encore plus strictes sur ces bornes [30].

D'autre part, les modèles dynamiques intégrant des cinétiques enzymatiques reposent sur l'élaboration d'équations différentielles transitoires pour décrire l'évolution temporelle des réactions. Ils utilisent généralement la loi de Michaelis-Menten et, dans certains cas, celle de Hill, pour estimer les vitesses réactionnelles en fonction des concentrations en substrats et cofacteurs. Contrairement au FBA, ces modèles dynamiques sont plus coûteux en ressources computationnelles, car ils nécessitent, en plus de la résolution d'équations différentielles de premier ordre, un nombre élevé de constantes biocinétiques afin de capturer l'évolution temporelle des

cinétiques métaboliques. Ainsi, la carte métabolique décrite par les modèles dynamiques comporte généralement moins de réactions que celle décrite par l'approche FBA. Par exemple, da Veiga Moreira *et al.* (2019) ont développé un modèle dynamique de 35 réactions enzymatiques basé sur la cinétique de Michaelis-Menten pour simuler la croissance d'une tumeur pulmonaire implantée par xénogreffe chez la souris. Leur modèle a également permis de décrire l'effet d'une thérapie métabolique utilisant un cocktail de quatre composés sur la progression tumorale [25]. De leur côté, Bejjak *et al.* (2021) ont développé le premier modèle dynamique recensé intégrant des données transcriptomiques dans la cinétique enzymatique de Michaelis-Menten [31]. Leur modèle, composé de 36 réactions couvrant plusieurs voies métaboliques dont la glycolyse et le cycle des acides tricarboxyliques (TCA), a permis d'améliorer l'estimation des concentrations métaboliques pour un procédé opéré en mode cuvée-alimenté grâce à l'intégration de la transcriptomique.

En résumé, les modèles computationnels du cancer, comme le FBA et les modèles dynamiques intégrant des cinétiques enzymatiques, constituent des approches complémentaires permettant d'examiner différents aspects du métabolisme tumoral. Le FBA optimise les flux métaboliques sous contraintes selon une fonction objectif directrice, tandis que les modèles dynamiques simulent l'évolution temporelle des taux des réactions enzymatiques ainsi que des concentrations en métabolites intracellulaires et extracellulaires grâce à un système d'équations différentielles. Ces modèles permettent d'améliorer notre compréhension des mécanismes métaboliques du cancer et peuvent mener à la découverte de cibles thérapeutiques beaucoup plus rapidement que ne le permettent les modèles *in vitro* et *in vivo*. Les modèles computationnels peuvent ainsi être utilisés en amont des modèles *in vitro* et *in vivo* pour en accélérer l'identification de pistes de thérapies.

CHAPITRE 3 MÉTHODOLOGIE

Ce chapitre présente la méthodologie développée pour atteindre les objectifs du projet de recherche. Il décrit d'abord les données transcriptomiques utilisées, qui sont ensuite intégrées aux cinétiques enzymatiques du modèle métabolique sélectionné. Il détaille également l'application des algorithmes d'apprentissage supervisé pour la détection du cancer et la classification de son stade, ainsi que la conception d'algorithmes de recherche pour l'identification de cibles thérapeutiques.

3.1 Données transcriptomiques utilisées

Les données transcriptomiques utilisées dans le cadre de ce projet sont issues de la base de données de l'étude *curatedOvarianData* [32] du programme *The Cancer Genome Atlas* (TCGA) mené par le *National Cancer Institute* (NCI) et le *National Human Genome Research Institute* (NHGRI) [33]. Ces données comprennent les niveaux d'expression d'ARNm de 13 110 gènes issus d'échantillons prélevés sur 578 patientes par ARNseq, ainsi que des informations relatives à la présence ou non du cancer de l'ovaire et à son stade de développement le cas échéant. Parmi ces 578 patientes, 570 ont été diagnostiquées avec un cancer de l'ovaire: 16 au stade I, 27 au stade II, 436 au stade III, et 84 au stade IV, tandis que pour sept patientes aucune information n'a été répertoriée concernant le stade d'évolution de la maladie. Ces niveaux d'expression génique mesurés ont nécessité des biopsies pour être obtenus. En utilisant des cellules tumorales circulantes (CTCs) [6] ou de l'ARN libre circulant (cfRNA) [7] extrait des exosomes plasmatiques, il serait possible d'obtenir des profils transcriptomiques similaires à partir d'un simple échantillon sanguin, offrant ainsi une approche non invasive pour l'évaluation du cancer de l'ovaire et de son stade de développement. L'utilisation de cette base de données s'est faite en raison de sa disponibilité, de l'efficacité élevée des données issues par ARNseq pour détecter le cancer de l'ovaire [17] et de la possibilité de jumeler les données transcriptomiques aux données métabolomiques, permettant une analyse intégrée des interactions entre l'expression génique et le métabolisme tumoral.

3.2 Modèle métabolique

La modélisation métabolique consiste à représenter et analyser les réseaux biochimiques impliqués dans le métabolisme cellulaire. Elle permet de mieux comprendre la dynamique des

réactions métaboliques ainsi que leur régulation, notamment dans un contexte d'une pathologie telle que le cancer de l'ovaire. Le modèle métabolique utilisé dans ce projet est issu du travail de recherche mené par da Veiga Moreira *et al.* [25]. Ce modèle décrit des cellules en croissance, incluant à la fois des cellules tumorales et des cellules souches, issues de xénogreffes de tumeur pulmonaire injectée chez la souris. De plus, il modifie certains modèles développés pour décrire des cellules de *Chinese hamster ovary* (CHO) [34] [35] [36] [37] et des cellules suppressives d'origine myéloïde (MD-SCs) de souris [38], en s'appuyant sur la notion de *minimal cut sets* (MCS) pour les réseaux biochimiques [39]. Le choix d'utilisation du modèle développé par da Veiga Moreira *et al.* repose notamment sur sa disponibilité, sur la pertinence de la modélisation de cellules CHO, les cellules CHO étant des cellules immortalisées d'ovaires de hamster chinois, ainsi que sur la pertinence de la modélisation des voies métaboliques de la glycolyse, des pentoses phosphates et de la lipogenèse, toutes identifiées comme potentielles cibles thérapeutiques.

3.2.1 Réactions biochimiques

Le modèle métabolique utilisé est composé de 35 réactions biochimiques et implique 45 métabolites. Comme décrit dans le travail de da Veiga Moreira *et al.*, l'ensemble des réactions modélise les voies principales du métabolisme central du carbone telles la glycolyse, le cycle des acides tricarboxyliques, la voie des pentoses phosphates, le métabolisme des acides aminés, la lipogenèse et la phosphorylation oxydative [25], soit le cœur réactionnel de l'ensemble des cellules eucaryotes, dont les cellules de mammifère. Parmi les métabolites étudiés, 25 sont classés comme intracellulaires et 20 comme extracellulaires. La liste complète des réactions est présentée au Tableau A.2 de l'Annexe A et la liste détaillée des métabolites impliqués dans ces réactions est détaillée au Tableau A.1 de l'Annexe A.

3.2.2 Règles gène-protéine-réaction

Afin d'intégrer adéquatement les données transcriptomiques d'une patiente dans les modèles de cinétiques enzymatiques, il faut établir des règles gène-protéine-réaction (GPR). Les règles GPR représentent les relations entre les données d'expression génique, les enzymes codées par ces gènes et les réactions catalysées par ces enzymes. En d'autres termes, pour chaque réaction

du modèle métabolique, il faut définir le ou les gènes codant l'enzyme impliquée dans la catalyse de cette réaction. Ainsi, les données transcriptomiques de ce ou ces gènes seront intégrées directement à la réaction enzymatique correspondante. L'objectif de l'inclusion des règles GPR dans le modèle métabolique est de pouvoir moduler les cinétiques enzymatiques selon les niveaux d'expression des gènes d'une patiente. Plus un gène est exprimé et plus la concentration de l'enzyme codée par ce gène est élevée. Il en résulte donc une augmentation de la vitesse maximale de la réaction catalysée par cette enzyme.

Le Tableau A.2 situé à l'Annexe A décrit également les règles GPR utilisées dans le modèle métabolique. Les gènes associés à chaque réaction sont nommés selon les acronymes utilisés dans la base de données du TCGA (voir Tableau A.3 de l'Annexe A pour les noms complets des gènes). Les règles GPR indiquent la donnée transcriptomique à intégrer dans chaque réaction du modèle métabolique en prenant, pour les réactions liées à plus d'un gène, une combinaison de la moyenne (opération intitulée « moy »), du minimum (opération intitulée « min ») ou du maximum (opération intitulée « max ») des niveaux d'expression des gènes associés à la réaction. Le choix des opérations est dicté par la nature du ou des gènes nécessaires à coder l'enzyme catalysant la réaction. Si la même enzyme ou une isoenzyme peut être synthétisée à partir de plusieurs gènes différents, l'opération « max » est utilisée [31], car la concentration de l'enzyme sera principalement influencée par le gène le plus exprimé. Cependant, si plusieurs gènes différents sont requis pour coder l'intégralité d'une enzyme (un gène peut coder une sous-unité de l'enzyme), l'opération « min » est utilisée [31], et la concentration de l'enzyme sera alors limitée par le gène le moins exprimé. Puisque certaines réactions du modèle métabolique sont simplifiées (une réaction du modèle métabolique généralise, dans certains cas, un ensemble de réactions biochimiques), elles nécessitent parfois plusieurs enzymes. En ce sens, pour une réaction simplifiant un ensemble de réactions, plusieurs gènes sont généralement considérés. Dans ce cas, l'opération « min » est également utilisée, car la réaction généralisée sera limitée par l'enzyme en moindre quantité. Finalement, la réaction *Vgrowth* modélisant la croissance du volume de la tumeur tient compte de gènes suppresseurs de tumeur et d'oncogènes et donc, deux opérations « moy » sont utilisées, car, par hypothèse, le niveau d'activité des gènes suppresseurs de tumeur est inversement proportionnel à la vitesse de réaction et celui des oncogènes est directement proportionnel à cette vitesse. Grâce à ces opérations, pour chaque réaction, une seule donnée transcriptomique est ainsi obtenue et celle-ci est directement intégrée au niveau de la vitesse

maximale de la réaction correspondante. La Figure 3.1 représente l'ensemble des réactions biochimiques du modèle métabolique et met en évidence l'association des gènes aux différentes réactions.

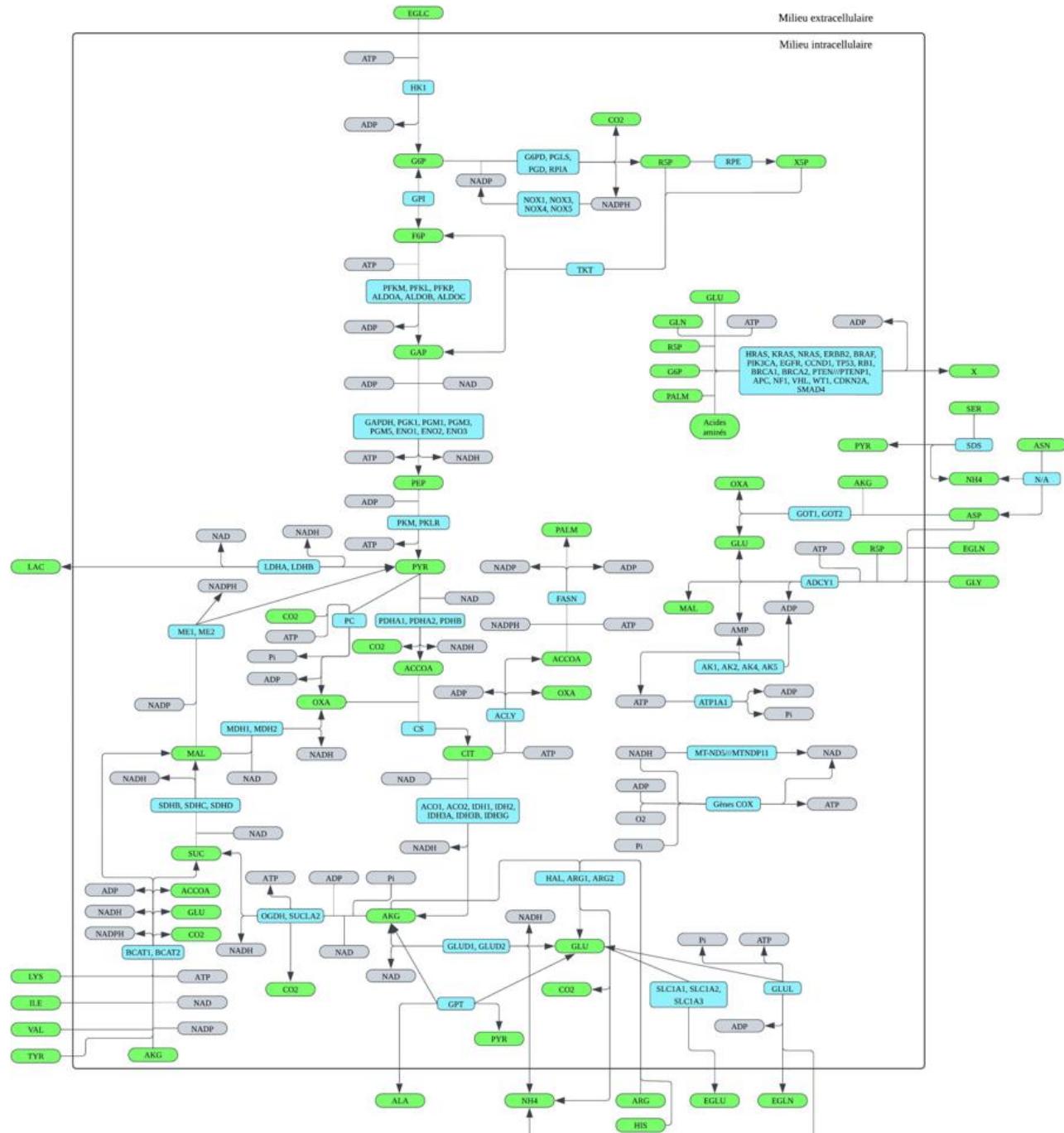


Figure 3.1 Réseau métabolique illustrant les métabolites (en vert), les cofacteurs enzymatiques et petites molécules (en gris) ainsi que les gènes associés aux réactions du modèle métabolique (en bleu)

3.2.3 Cinétiques enzymatiques

Chaque réaction du modèle métabolique possède une vitesse de réaction qui est estimée selon un produit de cinétiques enzymatiques de Michaelis-Menten (équation 3.1) pour chaque substrat enzymatique et cofacteur, qu'il s'agisse de cofacteurs énergétiques comme l'ADP ou l'ATP, ou de cofacteurs d'oxydoréduction tels que le couple NADH/NAD ou NADPH/NADP, impliqués dans la réaction.

$$v = \frac{v_{max}[S]}{K_m + [S]} \quad (3.1)$$

De plus, les réactions impliquant un mécanisme d'inhibition sont modélisées selon l'équation 3.2 et celles impliquant un phénomène d'activation suivent le modèle d'activation non-essentielle (équation 3.3) tel que présenté par Robitaille *et al.* (2015) [34].

$$v = v_{max} \frac{K_I}{K_I + [S]} \quad (3.2)$$

$$v = \frac{v_{max}[S](1 + \beta[A]/\alpha K_A)}{K_s(1 + [A]/K_A) + [S](1 + [A]/\alpha K_A)} \quad (3.3)$$

Il est à noter que les expressions mathématiques des cinétiques enzymatiques sont simplifiées. Elles reposent sur des équations de type Michaelis-Menten pour chaque substrat et cofacteur, sans pour autant dériver d'un modèle mécanistique global. Les vitesses de réaction sont toutes exprimées en millimolaire par heure (mM/h) à l'exception de la réaction *Vgrowth*, exprimée en h⁻¹, soit les unités d'un taux de croissance spécifique. Le Tableau A.4 de l'Annexe A présente les équations des vitesses de réaction pour l'ensemble des réactions du modèle métabolique.

3.2.4 Vitesses maximales de réaction

Afin d'intégrer les données transcriptomiques dans les cinétiques enzymatiques du modèle métabolique et d'adapter les vitesses de réaction aux variations de la demande en cofacteurs énergétiques et d'oxydoréduction, les vitesses maximales de réaction théoriques ($v_{max,th}$), soient celles obtenues à partir du travail de recherche effectué par da Veiga Moreira *et al.* [25], sont modulées par un coefficient *Fold Change* (FC) et une fonction temporelle F(t) (équation 3.4).

$$v_{max} = v_{max,th} * FC * F(t) \quad (3.4)$$

Tout d'abord, le coefficient FC module la vitesse maximale d'une réaction en fonction du niveau d'expression du ou des gènes lui étant associés. Le FC est défini comme le rapport entre la donnée transcriptomique de la patiente cancéreuse (DT_{cancer}) et la moyenne des données transcriptomiques des n patientes saines ($DT_{i,sain}$) (équation 3.5). Les données transcriptomiques sont calculées pour chaque réaction en suivant les règles GPR présentées au Tableau A.2 de l'Annexe A.

$$FC = \frac{DT_{cancer}}{\left(\frac{1}{n} * \sum_{i=1}^n DT_{i,sain}\right)} \quad (3.5)$$

L'équation 3.5 s'appuie sur les travaux de Bezjak *et al.* [31] qui ont intégré des données transcriptomiques dans la cinétique enzymatique de Michaelis-Menten en ajustant la vitesse maximale de réaction à l'aide d'un coefficient FC. Toutefois, leur étude portait sur des cultures cellulaires de cellules CHO, où le FC était défini comme le rapport entre la donnée transcriptomique mesurée à un jour j et la donnée transcriptomique initiale. Cette approche a permis d'adapter la vitesse maximale de réaction en fonction des variations d'expression génique observées au cours de la croissance des cellules cancéreuses. L'équation 3.5 applique un principe similaire en ajustant la vitesse maximale d'une réaction en fonction des variations d'expression génique observées lors du développement tumoral.

Puis, pour une cellule humaine typique, le cycle cellulaire dure environ 24 heures et se divise en quatre phases :

- La phase G1, où la cellule croît et prépare la réPLICATION de l'ADN (environ 11 heures)
- La phase S, durant laquelle l'ADN est repliqué (environ huit heures)
- La phase G2, dédiée aux préparatifs de la mitose (environ quatre heures)
- La phase M, correspondant à la division cellulaire (environ une heure) [40].

Il est à noter que les durées respectives de chacune des phases du cycle cellulaire peuvent varier en fonction du type cellulaire et des conditions extracellulaires. Par exemple, une cellule peut entrer dans un état de repos, appelé phase G0, dans le cas de conditions extracellulaires défavorables [41]. Au cours du cycle cellulaire, les ratios ATP/ADP, NAD/NADH et NADP/NADPH fluctuent pour ajuster l'équilibre entre les réactions anaboliques et cataboliques. Ces oscillations métaboliques permettent aux cellules eucaryotes d'adapter leur production

énergétique en fonction des exigences de chaque phase du cycle [42]. Ainsi, les constantes cinétiques peuvent présenter des variations typiques allant jusqu'à 20% lors d'un tel comportement oscillatoire [43].

Dans la phase G0, les cellules quiescentes possèdent un métabolisme oxydatif basal, c'est-à-dire une activité métabolique minimale pour maintenir leurs fonctions vitales. Ce métabolisme oxydatif basal résulte en des ratios ATP/ADP et NADP/NADPH plutôt faibles et un ratio NAD/NADH élevé [42]. La phase G1, quant à elle, est principalement régulée par la voie métabolique de la glycolyse et est caractérisée par la synthèse de protéines nécessaires à la réPLICATION de l'ADN. Au cours de cette phase, les ratios ATP/ADP et NAD/NADH baissent tandis que le ratio NADP/NADPH augmente [42]. Dans la phase S, dédiée à la réPLICATION de l'ADN, le ratio ATP/ADP augmente et les ratios NAD/NADH et NADP/NADPH diminuent [42]. La phase G2 est une phase de préparation à la mitose, marquée par la synthèse des lipides nécessaires aux membranes cellulaires. Dans cette phase, les ratios ATP/ADP, NAD/NADH et NADP/NADPH augmentent [42]. Enfin, pendant la mitose, soit la phase de division cellulaire, les chromosomes répliqués se séparent à travers les étapes de la prophase, de la métaphase, de l'anaphase et de la télophase. Puis, la cytocinèse divise le cytoplasme pour former deux cellules filles. Au cours de la mitose, le ratio ATP/ADP diminue en raison d'une hydrolyse intensive de l'ATP [42], libérant l'énergie nécessaire à la formation du fuseau mitotique, à la séparation des chromatides et à la cytocinèse.

Conséquemment, afin de tenir compte des variations des ratios ATP/ADP, NAD/NADH et NADP/NADPH au cours du cycle cellulaire, une fonction temporelle $F(t)$ module la vitesse maximale d'une réACTION lorsque celle-ci dépend de ces métabolites. En effet, pour chacun de ces cofacteurs enzymatiques, une concentration plus élevée favorise généralement la réACTION. Ainsi, une fonction temporelle spécifique à chaque métabolite vient moduler la vitesse maximale de la réACTION (équation 3.6).

$$F(t) = 1 + \left(\sum_{\text{métabolites}} f_{\text{moy,métabolite}}(t) \right) \quad (3.6)$$

Cependant, les fonctions $f_{\text{ADP}}(t)$, $f_{\text{NADH}}(t)$ et $f_{\text{NADPH}}(t)$ sont directement dérivées respectivement des fonctions $f_{\text{ATP}}(t)$, $f_{\text{NAD}}(t)$ et $f_{\text{NADP}}(t)$ en prenant leur opposé (équations 3.7 à 3.9). En effet, lorsqu'un métabolite tel que l'ATP, le NAD ou le NADP agit comme cofacteur

enzymatique, son « conjugué » correspondant, soit l'ADP, le NADH ou le NADPH, apparaît du côté des produits.

$$f_{ADP}(t) = -f_{ATP}(t) \quad (3.7)$$

$$f_{NADH}(t) = -f_{NAD}(t) \quad (3.8)$$

$$f_{NADPH}(t) = -f_{NADP}(t) \quad (3.9)$$

Selon les fluctuations des ratios ATP/ADP, NAD/NADH et NADP/NADPH rapportées par da Veiga Moreira *et al.* (2015) et en respectant une variation maximale des constantes cinétiques de 20% [43], les effets sur la vitesse maximale d'une réaction lorsque l'ATP, le NAD ou le NADP agissent comme cofacteurs enzymatiques peuvent être estimés (Tableau 3.1). Ces estimations reposent donc sur une analyse qualitative des fluctuations des ratios énergétiques, sans calculs directs.

Tableau 3.1 Variations estimées des vitesses maximales de réaction selon la présence d'ATP ($f_{ATP}(t)$), de NAD ($f_{NAD}(t)$) ou de NADP ($f_{NADP}(t)$) en tant que cofacteurs enzymatiques

	$f_{ATP}(t)$	$f_{NAD}(t)$	$f_{NADP}(t)$
Début de la phase G1 (t = 0 h)	-5%	10%	-5%
Début de la phase S (t = 11 h)	-10%	0%	5%
Début de la phase G2 (t = 19 h)	0%	-10%	-15%
Début de la phase M (t = 23 h)	10%	0%	-10%

À l'aide de l'interpolateur *PchipInterpolator* importé depuis la classe *interpolate* de la bibliothèque *SciPy*, disponible librement pour le langage de programmation Python [44], les variations des vitesses maximales de réactions au début de chacune des phases du cycle cellulaire sont interpolées pour obtenir les fonctions continues $f_{ATP}(t)$, $f_{NAD}(t)$ et $f_{NADP}(t)$ oscillant périodiquement selon les durées des phases du cycle cellulaire (Figure 3.2). À des fins de simplification, la phase de repos G0 n'est pas considérée et les cellules sortant de la phase M entrent directement dans la phase G1.

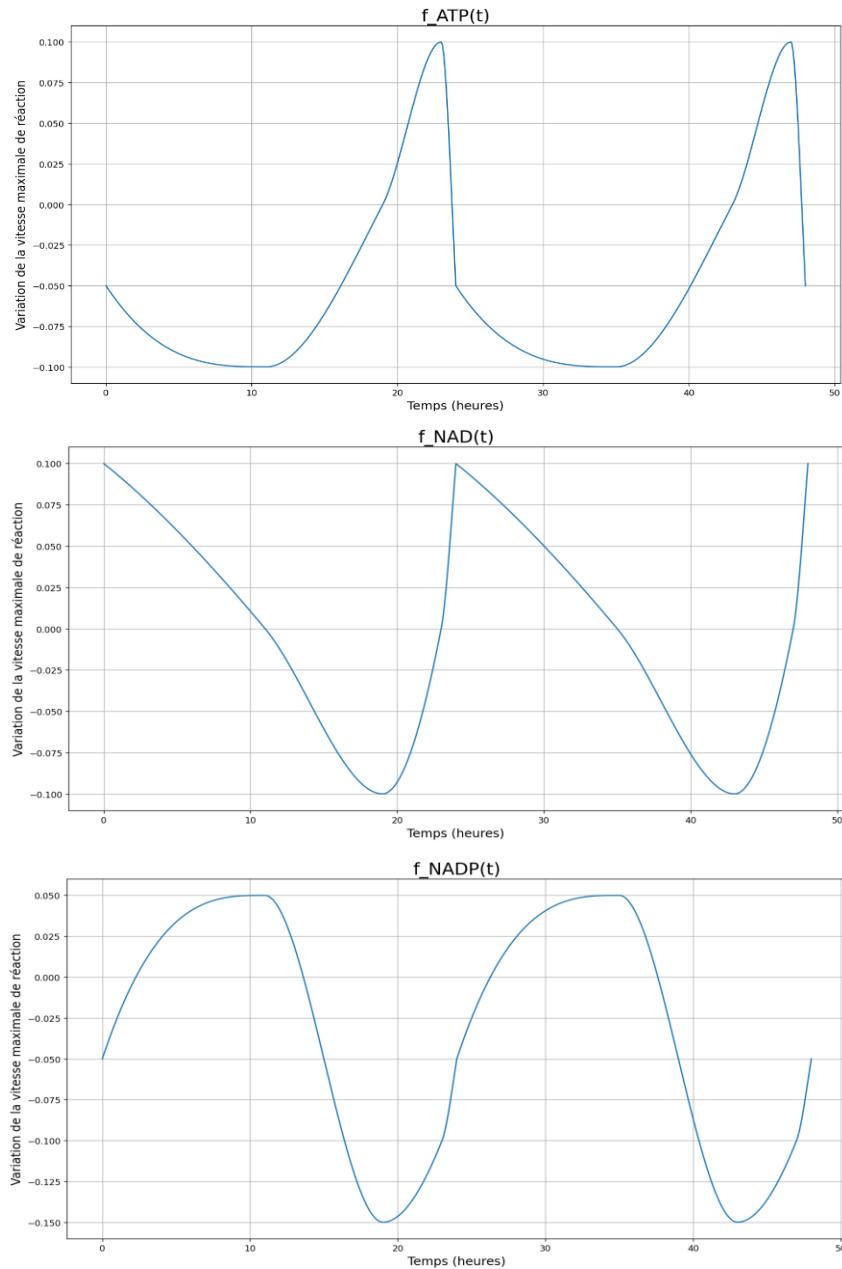


Figure 3.2 Variations de la vitesse maximale de réaction causée par la fluctuation en ATP, en NAD et en NADP durant le cycle cellulaire

Telles qu'elles sont décrites, les fonctions $f_{ATP}(t)$, $f_{ADP}(t)$, $f_{NAD}(t)$, $f_{NADH}(t)$, $f_{NADP}(t)$ et $f_{NADPH}(t)$ supposent qu'au temps zéro, la cellule entre en phase G1. Toutefois, en modélisant un ensemble de cellules cancéreuses, il n'est pas réaliste de supposer que toutes les cellules commencent leur phase G1 au temps zéro. Ainsi, en hypothétisant que la proportion de cellules dans chaque phase du cycle cellulaire est proportionnelle à la durée respective de chaque phase, il est

possible d'effectuer une moyenne pondérée de ces fonctions pour déterminer $f_{moy,ATP}(t)$, $f_{moy,ADP}(t)$, $f_{moy,NAD}(t)$, $f_{moy,NADH}(t)$, $f_{moy,NADP}(t)$ et $f_{moy,NADPH}(t)$ (équation 3.10).

$$f_{moy,métab.}(t) = f_{métab,G1}(t) + f_{métab,S}(t) + f_{métab,G2}(t) + f_{métab,M}(t) \quad (3.10)$$

Ces nouvelles fonctions moyennes sont obtenues en estimant que 50% des cellules commencent en phase G1 (durée de 11 heures), 30% commencent en phase S (durée de huit heures), 15% commencent en phase G2 (durée de quatre heures) et 5% commencent en phase M (durée d'une heure) (équations 3.11 à 3.14).

$$f_{métab,G1}(t) = 0,5 * f_{métab.}(t) \quad (3.11)$$

$$f_{métab,S}(t) = 0,3 * f_{métab.}(t + 11) \quad (3.12)$$

$$f_{métab,G2}(t) = 0,15 * f_{métab.}(t + 19) \quad (3.13)$$

$$f_{métab,M}(t) = 0,05 * f_{métab.}(t + 23) \quad (3.14)$$

Par exemple, pour quantifier à l'instant $t = 100$ h la fonction $F(t)$ associée à la réaction VPGK (telle que définie au Tableau A.5 de l'Annexe A), on utilise l'expression suivante :

$$F_{VPGK}(t) = 1 + f_{moy,ADP}(t) + f_{moy,NAD}(t)$$

Il faut d'abord évaluer, pour l'ADP à $t = 100$ h :

$$f_{ADP,G1}(t = 100) = 0,5 * f_{ADP}(100) = 0,5 * f_{ADP}(4) \approx 0,5 * 0,087 \approx 0,044$$

$$f_{ADP,S}(t = 100) = 0,3 * f_{ADP}(100 + 11) = 0,3 * f_{ADP}(15) \approx 0,3 * 0,067 \approx 0,020$$

$$f_{ADP,G2}(t = 100) = 0,15 * f_{ADP}(100 + 19) = 0,15 * f_{ADP}(23) = 0,15 * -0,10 = -0,015$$

$$f_{ADP,M}(t = 100) = 0,05 * f_{ADP}(100 + 23) = 0,05 * f_{ADP}(3) \approx 0,05 * 0,081 \approx 0,0041$$

Afin d'obtenir :

$$\begin{aligned} f_{moy,ADP}(t = 100) &= f_{ADP,G1}(100) + f_{ADP,S}(100) + f_{ADP,G2}(100) + f_{ADP,M}(100) \\ &\approx 0,044 + 0,020 - 0,015 + 0,0041 = 0,053 \end{aligned}$$

Puis, il faut répéter ces calculs pour le NAD à $t = 100$ h :

$$f_{NAD,G1}(t = 100) = 0,5 * f_{NAD}(100) = 0,5 * f_{NAD}(4) \approx 0,5 * 0,068 \approx 0,034$$

$$f_{NAD,S}(t = 100) = 0,3 * f_{NAD}(100 + 11) = 0,3 * f_{NAD}(15) \approx 0,3 * -0,061 \approx -0,018$$

$$f_{NAD,G2}(t = 100) = 0,15 * f_{NAD}(100 + 19) = 0,15 * f_{NAD}(23) = 0,15 * 0 = 0$$

$$f_{NAD,M}(t = 100) = 0,05 * f_{NAD}(100 + 23) = 0,05 * f_{NAD}(3) \approx 0,05 * 0,077 \approx 0,0039$$

On peut donc obtenir :

$$\begin{aligned} f_{moy,ADP}(t = 100) &= f_{ADP,G1}(100) + f_{ADP,S}(100) + f_{ADP,G2}(100) + f_{ADP,M}(100) \\ &\approx 0,034 - 0,018 + 0 + 0,0039 = 0,020 \end{aligned}$$

Finalement, on peut calculer F(t) :

$$F_{VPGK}(t = 100) = 1 + f_{moy,ADP}(100) + f_{moy,NAD}(100) \approx 1 + 0,053 + 0,020 = 1,073$$

Dans cet exemple, la vitesse maximale de la réaction VPGK à la centième heure est donc augmentée d'environ 7,3% en raison des fluctuations en ADP et en NAD au cours du cycle cellulaire.

En résumé, les vitesses maximales des réactions du modèle métabolique sont modulées à la fois par les niveaux d'expression des gènes associés, via le coefficient FC, ainsi que par les fluctuations des cofacteurs enzymatiques, à travers les fonctions temporelles F(t). Cette approche permet d'intégrer pleinement les données transcriptomiques au cœur du modèle métabolique, tout en faisant varier dynamiquement ces vitesses au cours du temps.

3.2.5 Conditions métaboliques initiales

Tous les métabolites intracellulaires et extracellulaires sont exprimés en mM. À titre informatif, pour convertir les concentrations des métabolites intracellulaires de mmol par million de cellules en mM, le facteur de conversion de $5,4 \times 10^{-9}$ litres par cellule avait été utilisé. Les conditions initiales correspondent à celles utilisées dans l'étude de da Veiga Moreira *et al.* [25] et sont présentées au Tableau A.1 de l'Annexe A.

3.2.6 Volume tumoral initial

Le volume tumoral est défini par trois variables dans le modèle métabolique : X étant le volume de la masse des cellules tumorales, V_{blood} étant le volume de sang à même le volume tumoral et alimentant les cellules tumorales, et VT étant le volume tumoral total caractérisé par la somme de ces deux composantes. Les conditions initiales utilisées dans le travail de da Veiga Moreira *et al.* sont 104 mm^3 pour X , 26 mm^3 pour V_{blood} et donc 130 mm^3 pour VT . Cependant, ces conditions initiales sont celles de la tumeur injectée chez la souris (ayant une masse moyenne de 21.15 g) [25], et ne sont pas représentatives du volume tumoral chez l'humain, particulièrement dans les cas de cancers à un stade avancé. Ainsi, le volume tumoral doit être ajusté en fonction de la physiologie humaine et du stade d'avancement du cancer. Selon l'*American Cancer Society*, les deux systèmes utilisés pour la classification du stade du cancer de l'ovaire sont le système de l'*International Federation of Gynecology and Obstetrics* (FIGO) et le système TNM de l'*American Joint Committee on Cancer* (AJCC) [45]. Ces deux systèmes sont complémentaires dans leur approche; le système TNM fournit d'abord une description précise de la taille et de l'étendue de la tumeur primaire (« T »), de l'atteinte des ganglions lymphatiques régionaux (« N ») et de la présence ou de l'absence de métastases à des sites éloignés (« M »). À partir de ces informations, le système FIGO classe la maladie en stades allant de I à IV.

Selon le rapport de la FIGO en 2021, le stade I correspond à une tumeur limitée aux ovaires ou aux trompes de Fallope, sans extension au-delà de ces structures (T1 - N0 - M0). Le stade II reflète une extension tumorale dans le pelvis, impliquant des structures telles que l'utérus ou d'autres tissus pelviens (T2 - N0 - M0). Le stade III indique une propagation de la tumeur au-delà du pelvis, incluant des métastases péritonéales ou une atteinte des ganglions lymphatiques rétropéritonéaux (T1 à T3 - N0 à N1 - M0). Enfin, le stade IV désigne la présence de métastases à distance, telles que dans les organes extra-abdominaux ou le parenchyme pulmonaire (tout T – tout N – M1) [46]. Cependant, il est très difficile d'estimer des volumes tumoraux précis pour chaque stade, car ces derniers sont principalement définis en fonction de l'étendue de la maladie plutôt que de la taille de la tumeur. De plus, en présence de métastases, la tumeur maligne ne se limite plus à un seul volume, mais correspond plutôt à une somme de volumes localisés dans différents tissus. L'étude de Petru *et al.* (2018), menée auprès de 553 patientes atteintes de cancer de l'ovaire classées selon le système FIGO entre les stades I et III, révèle que la médiane du diamètre maximal de la tumeur primaire invasive est de 6,7 cm pour le stade I et de 8,3 cm pour le stade III. De plus,

chez les patientes au stade III, la médiane du diamètre maximal des métastases intrapéritonéales est de 5,8 cm [47]. Aucune donnée spécifique n'a été trouvée pour les stades II et IV. En ce sens, en appliquant la formule du volume d'une sphère, le volume tumoral au stade I peut être estimé à 150 cm³ et celui au stade III à 400 cm³ (équations 3.15 et 3.16).

$$V_{stade\ I} = \frac{4\pi r^3}{3} = \frac{4 * \pi * (3,35\ cm)^3}{3} \approx 150\ cm^3 \quad (3.15)$$

$$V_{stade\ III} = V_{ primaire } + V_{métastase}$$

$$V_{stade\ III} = \frac{4 * \pi * (4,15\ cm)^3}{3} + \frac{4 * \pi * (2,9\ cm)^3}{3} \approx 400\ cm^3 \quad (3.16)$$

Ensuite, en supposant une croissance exponentielle du volume tumoral malin en fonction du stade, il est possible d'estimer approximativement les volumes tumoraux pour les stades II et IV. On peut d'abord déterminer le paramètre k de l'équation exponentielle en considérant Δx comme la différence entre les numéros des stades (équations 3.17 et 3.18).

$$V = V_{stade\ I} * e^{k\Delta x} \quad (3.17)$$

$$k = \frac{\ln(V_{stade\ III}/V_{stade\ I})}{\Delta x} = \frac{\ln(400\ 000\ mm^3/150\ 000\ mm^3)}{2} \approx 0,49 \quad (3.18)$$

Puis, avec le paramètre k égal à 0.49 on peut estimer des volumes tumoraux de 245 cm³ pour le stade II et de 650 cm³ pour le stade IV (équations 3.19 et 3.20).

$$V_{stade\ II} = V_{stade\ I} * e^{k\Delta x} = 150\ cm^3 * e^{0,49*1} \approx 245\ cm^3 \quad (3.19)$$

$$V_{stade\ IV} = V_{stade\ I} * e^{k\Delta x} = 150\ cm^3 * e^{0,49*3} \approx 650\ cm^3 \quad (3.20)$$

Ces nouveaux volumes tumoraux totaux sont donc d'un ordre de grandeur de 10³ fois plus élevés que celui utilisé chez la souris [25], ce qui est cohérent lorsque l'on considère que l'humain a également une masse environ 10³ fois supérieure à celle de la souris. De plus, l'étude d'Aviram *et al.* (2008) utilisant l'échographie transvaginale pour mesurer le volume de l'ovaire chez 466 patientes a montré une différence significative entre le volume ovarien sain, mesuré à 3.4 ± 2.2 cm³, et le volume ovarien cancéreux, mesuré à 368 ± 1176 cm³ [48]. Malgré la grande incertitude sur le volume ovarien cancéreux mesuré, l'ordre de grandeur correspondant à quelques centaines de centimètres cubes est concordant avec les volumes tumoraux initiaux estimés allant de 150 cm³ à 650 cm³.

Ainsi, en conservant les mêmes proportions entre le volume des cellules tumorales (X) et le volume de sang (V_{blood}) utilisés chez la souris [25], étant respectivement de 104 mm^3 pour X (80%) et 26 mm^3 pour V_{blood} (20%) pour atteindre un volume tumoral total (VT) de 130 mm^3 , on peut calculer les volumes tumoraux initiaux pour chacun des stades du cancer de l'ovaire chez l'humain (Tableau 3.2).

Tableau 3.2 Volumes tumoraux initiaux chez l'humain

Stade	Volume de la masse de cellules tumorales (X) (cm^3)	Volume de sang (V_{blood}) (cm^3)	Volume tumoral total (VT) (cm^3)
I	120	30	150
II	196	49	245
III	320	80	400
IV	520	130	650

3.2.7 Taux de croissance tumoral chez l'humain

Le taux de croissance tumoral, noté μ et équivalent à V_{growth} , ainsi que le taux de croissance des capillaires μ_{blood} sont exprimés en h^{-1} . De plus, X est exprimé en L. Les volumes tumoraux X , V_{blood} et VT évoluent selon le système d'équations différentielles présenté aux équations 3.21 à 3.23.

$$\frac{dX}{dt} \left[\frac{\text{L}}{\text{h}} \right] = \mu \left[\frac{1}{\text{h}} \right] * X[\text{L}] \quad (3.21)$$

$$\frac{dV_{blood}}{dt} \left[\frac{\text{L}}{\text{h}} \right] = u_{blood} \left[\frac{1}{\text{h}} \right] * V_{blood}[\text{L}] \quad (3.22)$$

$$\frac{dV_{tumor}}{dt} \left[\frac{\text{L}}{\text{h}} \right] = \frac{dX}{dt} \left[\frac{\text{L}}{\text{h}} \right] + \frac{dV_{blood}}{dt} \left[\frac{\text{L}}{\text{h}} \right] \quad (3.23)$$

La vitesse maximale de la réaction V_{growth} a été calibrée à $0,033 \text{ h}^{-1}$ d'après les données expérimentales de croissance tumorale obtenues chez la souris [25]. Cependant, cette vitesse maximale de réaction n'est pas nécessairement adaptée à une tumeur se développant chez l'humain. Bedia *et al.* (2025) ont mené une analyse exploratoire portant sur des femmes atteintes d'un cancer épithélial invasif de l'ovaire et ont notamment mesuré le temps de doublement tumoral (TDT) pour plusieurs sous-types du cancer. Pour les patientes atteintes du HGSOC, soit un des sous-types les plus agressifs du cancer de l'ovaire, la médiane du TDT était de 2,9 mois [49]. En modélisant une croissance exponentielle de la tumeur, le TDT peut être utilisé pour estimer le taux de croissance tumoral μ chez l'humain (équations 3.24 et 3.25).

$$TDT = 2,9 \text{ mois} * \frac{1 \text{ an}}{12 \text{ mois}} * \frac{365,25 \text{ jours}}{\text{an}} * \frac{24 \text{ heures}}{\text{jour}} \approx 2118 \text{ heures} \quad (3.24)$$

$$\mu \equiv \frac{\ln(2)}{TDT} = \frac{\ln(2)}{2118 \text{ heures}} \approx 0,00033 \text{ h}^{-1} \quad (3.25)$$

3.2.8 Flux sanguin tumoral chez l'humain

Le flux sanguin tumoral F représente le débit de sang circulant dans la tumeur, généralement exprimé en millilitres par minute et par gramme de tissu tumoral ($\text{ml} * \text{min}^{-1} * \text{g}^{-1}$). Dans le modèle métabolique, ce flux est converti et exprimé en litres par heure (L/h). Ce flux joue un rôle clé dans l'apport en nutriments, en oxygène et en médicaments à la tumeur, influençant ainsi sa croissance, son métabolisme et sa réponse au traitement. À notre connaissance, aucune valeur spécifique n'a été reportée pour une tumeur de l'ovaire, bien que plusieurs études aient mesuré le flux sanguin tumoral dans d'autres types de cancer.

En effet, Cao *et al.* (2021) ont rapporté un flux sanguin moyen de $87,68 \pm 22,36 \text{ mL}/(\text{min} * 100 \text{ g})$ dans la tumeur primaire de 62 patients atteints du cancer du nasopharynx [50]. De leur côté, Bakke *et al.* (2020) ont mesuré un flux sanguin de $96 \pm 33 \text{ mL}/(\text{min} * 100 \text{ g})$ chez des patients atteints du cancer du rectum ayant une faible réponse locale au traitement néoadjuvant utilisé [51]. Par ailleurs, Katayama *et al.* (2021) ont observé un flux sanguin de $0,42 \pm 0,42 \text{ mL}/(\text{min} * \text{cm}^3)$ chez des patients atteints du cancer du poumon non à petites cellules [52]. Ainsi, il est plausible d'estimer le flux sanguin tumoral à environ $80 \text{ mL}/(\text{min} * 100 \text{ g})$ (équation 3.26), ce qui correspond à un flux de $7,2 \text{ L/h}$ pour un volume tumoral de 150 cm^3 (équation 3.27).

$$F = 80 \frac{ml}{min * 100g} = 4,8 \frac{L}{h * 100g} \quad (3.26)$$

$$F = 4,8 \frac{L}{h * 100g} * 150 \text{ cm}^3 * \frac{1 \text{ g}}{\text{cm}^3} = 7,2 \left[\frac{L}{h} \right] \quad (3.27)$$

3.2.9 Optimisation du taux de croissance tumoral et du flux sanguin tumoral

Le modèle métabolique étant initialement adapté pour la souris, la modification seule des paramètres macroscopiques identifiés, soient le volume tumoral, le taux de croissance tumoral et le flux sanguin tumoral, ne permet pas une transposition efficace du modèle de la souris vers un modèle humain. Ces paramètres doivent être optimisés pour atteindre une fonction objectif permettant d'évaluer la fidélité du modèle métabolique transposé. Dans ce travail, la fonction objectif choisie est le doublement tumoral en 2,9 mois. Afin de simplifier le processus d'optimisation, le volume tumoral initial est fixé à 150 cm³, correspondant à une tumeur de stade I, et les données transcriptomiques utilisées seront celles d'une patiente classifiée au stade I. Le taux de croissance tumoral ainsi que le flux sanguin tumoral sont posés constants pour tous les stades étudiés. Il s'agit alors de déterminer les combinaisons de ces deux paramètres permettant d'obtenir un doublement tumoral en 2,9 mois. Étant donné qu'il existe une infinité de combinaisons possibles, il est plus pertinent de considérer quelques valeurs de flux sanguin tumoral et d'optimiser la vitesse maximale du taux de croissance tumoral, notée $v_{max,growth}$, pour chacune d'elles. Ainsi, quatre valeurs de flux sanguin ont été envisagées :

1. Un flux sanguin tumoral de $8,663 \times 10^{-8}$ L/h correspondant au flux calibré pour le modèle de la souris.
2. Un flux sanguin de 7,2 L/h correspondant au flux calculé d'après la littérature pour d'autres types de cancers chez l'humain.
3. Une modification d'ordre de grandeur du flux sanguin pour adapter le paramètre calibré pour la souris aux rapports de masse entre la souris et l'humain :
 - Une multiplication par 10^3 du flux sanguin de $8,663 \times 10^{-8}$ L/h
 - Une division par 10^3 du flux sanguin de 7,2 L/h

Les valeurs de $v_{max,growth}$ permettant un doublement de VT en 2218 h, soit en 2,9 mois, ont été calculées pour chacune de ces quatre possibilités (Tableau 3.3). Ces valeurs ont été optimisées au dix-millième près afin de conserver deux chiffres significatifs. De plus, il a été déterminé qu'une variation de 0,0001 sur la vitesse maximale de réaction entraîne une erreur maximale de 4 % sur le TDT, correspondant à un doublement tumoral se produisant jusqu'à quatre jours avant ou après le TDT cible de 2,9 mois.

Tableau 3.3 Optimisation de la vitesse maximale du taux de croissance tumoral selon le flux sanguin tumoral pour obtenir un temps de doublement tumoral de 2,9 mois

Flux sanguin tumoral (L/h)	$v_{max,growth}$ (h ⁻¹)
$8,663 \times 10^{-8}$	$0,0021 \pm 0,0001$
$8,663 \times 10^{-5}$	$0,0020 \pm 0,0001$
$7,2 \times 10^{-3}$	$0,0023 \pm 0,0001$
7,2	$0,0024 \pm 0,0001$

En comparant les fluctuations métaboliques et la croissance tumorale simulées par le modèle métabolique, la combinaison jugée optimale est celle ayant un flux sanguin tumoral de $8,663 \times 10^{-5}$ L/h et un $v_{max,growth}$ de 0,0020 h⁻¹. Le choix de cette combinaison s'est fait notamment selon l'impact des différents flux métaboliques sur la croissance tumorale et sur les métabolites extracellulaires (Figure 3.3).

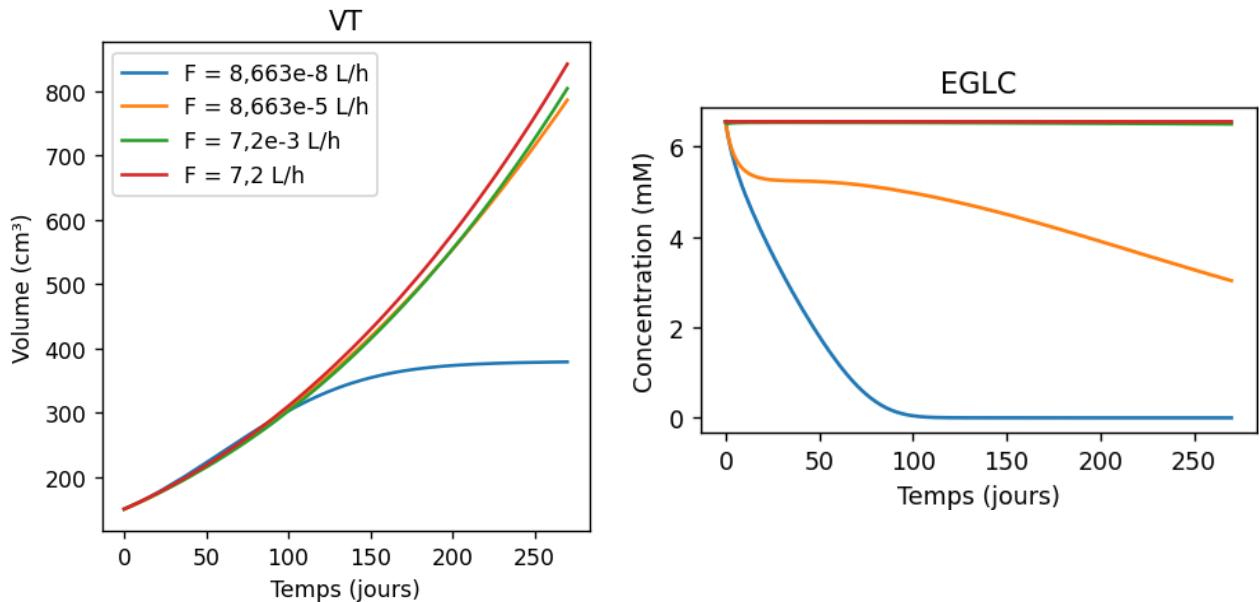


Figure 3.3 Effet du flux sanguin tumoral (F) sur le volume tumoral total (VT) et sur le glucose extracellulaire (EGLC)

Les flux sanguins tumoraux de 7,2 L/h et de $7,2 \times 10^{-3}$ L/h sont trop élevés pour permettre des fluctuations visibles des métabolites extracellulaires, tels que le glucose, qui sont entièrement renouvelés par un débit sanguin élevé. Nous notons qu'il est décrit dans la littérature que la croissance rapide des tumeurs mène à une limitation en nutriments comme le glucose [53]. À l'inverse, le flux sanguin de $8,663 \times 10^{-8}$ L/h semble trop faible, entraînant une consommation trop rapide des métabolites extracellulaires et donc des conditions limitantes irréalistes. Dans ce cas, la croissance tumorale atteint un point d'inflexion vers le 150^e jour, puis plafonne vers le 150^e jour, probablement en raison de l'épuisement de certains métabolites essentiels.

Ainsi, bien que l'ordre de grandeur du flux sanguin tumoral de $8,663 \times 10^{-5}$ L/h soit irréaliste, cette combinaison, associée à un $v_{max,growth}$ de 0,0020 h⁻¹, apparaît comme la plus appropriée pour reproduire simultanément une croissance exponentielle de la tumeur pendant au moins 270 jours et des fluctuations significatives des métabolites extracellulaires, tout en évitant un épuisement complet de ces derniers. Nous sommes conscients du caractère approximatif de ces conditions physiologiques, mais nous avons privilégié cette approche afin d'obtenir des résultats, tels que la croissance tumorale et les fluctuations en métabolites extracellulaires, qui sont dans des

plages réalistes de simulation. Ce choix repose donc sur la nécessité d'équilibrer la représentativité biologique du modèle avec la capacité à observer des variations métaboliques pertinentes.

3.2.10 Résumé des modifications apportées au modèle métabolique

Par rapport au modèle métabolique développé par da Veiga Moreira *et al.* [25], plusieurs modifications ont été apportées afin d'atteindre les objectifs suivants :

- Intégrer les données transcriptomiques dans les cinétiques enzymatiques d'un modèle métabolique.
- Transposer un modèle métabolique murin vers un modèle métabolique humain du cancer de l'ovaire.

Ces objectifs visent à adapter le modèle métabolique développé da Veiga Moreira *et al.* pour qu'il soit représentatif de la croissance tumorale chez l'humain, notamment par l'incorporation de données transcriptomiques et par la modification de certains paramètres macroscopiques tels que le volume tumoral, le taux de croissance et le flux sanguin. Le Tableau 3.4 récapitule les principales modifications réalisées pour répondre aux deux objectifs. Ces modifications affectent principalement les vitesses maximales de réaction ainsi que les valeurs de ces paramètres macroscopiques.

Tableau 3.4 Principales modifications apportées au modèle métabolique

	Modèle métabolique développé par da Veiga Moreira <i>et al.</i> (2019)	Modèle métabolique développé dans le cadre de ce projet
Forme générale des vitesses maximales de réaction	$v_{max} = v_{max,th}$	$v_{max} = v_{max,th} * FC * F(t)$
Vitesse maximale de la réaction V_{growth}	$v_{max,growth} = 0,033 \text{ h}^{-1}$	$v_{max,growth} = 0,002 \text{ h}^{-1} * FC_{growth} * (1 + f_{moy,ATP}(t))$
Flux sanguin tumoral	$F = 8,663 \times 10^{-8} \text{ L/h}$	$F = 8,663 \times 10^{-5} \text{ L/h}$
		Stade I: $X = 120 \text{ cm}^3$; $Vblood = 30 \text{ cm}^3$; $VT = 150 \text{ cm}^3$

Volumes tumoraux initiaux	$X = 104 \text{ mm}^3$	Stade II: $X = 196 \text{ cm}^3; V_{blood} = 49 \text{ cm}^3; VT = 245 \text{ cm}^3$
	$V_{blood} = 26 \text{ mm}^3$	Stade III: $X = 320 \text{ cm}^3; V_{blood} = 80 \text{ cm}^3; VT = 400 \text{ cm}^3$
	$VT = 130 \text{ mm}^3$	Stade IV: $X = 520 \text{ cm}^3; V_{blood} = 130 \text{ cm}^3; VT = 650 \text{ cm}^3$

3.3 Apprentissage supervisé

L'apprentissage supervisé consiste à entraîner des algorithmes à établir des liens entre des variables et des classes de prédiction. Dans le cadre de ce projet, les variables correspondent aux niveaux d'expression génique obtenus par ARNseq, aussi appelés données transcriptomiques, tandis que les classes de prédiction correspondent à la présence ou à l'absence de cancer pour la détection, ainsi qu'aux stades I, II, III et IV, tels que définis par la FIGO, pour la classification du stade.

3.3.1 Métriques d'évaluation des algorithmes d'apprentissage supervisé

Pour évaluer la performance des modèles, cinq métriques d'évaluation sont considérées pour la détection du cancer et trois pour la classification du stade. Ces métriques permettent d'analyser la justesse des prédictions et de prendre en compte les déséquilibres potentiels quant à la richesse d'information disponible entre les classes. Pour une classification binaire, comme celle de la détection du cancer, les vrais positifs (*VP*) représentent les patientes correctement identifiées comme étant atteintes du cancer, tandis que les vrais négatifs (*VN*) sont celles correctement classées comme étant saines. À l'inverse, les faux positifs (*FP*) désignent les patientes incorrectement prédites comme étant cancéreuses, alors qu'elles sont en réalité saines, et les faux négatifs (*FN*) sont celles erronément classées comme étant saines alors qu'elles présentent effectivement un cancer. Les cinq métriques utilisées pour la détection du cancer sont :

1. L'exactitude, représentant le nombre de patientes correctement classées parmi l'ensemble des patientes testées (équation 3.28).

$$Exactitude_{cancer} = \frac{VP + VN}{VP + FP + VN + FN} \quad (3.28)$$

2. L'exactitude balancée, corigeant les déséquilibres entre classes en prenant la moyenne des patientes correctement identifiées dans chaque classe (équation 3.29).

$$Exactitude \text{ } balancée_{cancer} = \frac{1}{2} * \left(\frac{VP}{VP + FN} + \frac{VN}{VN + FP} \right) \quad (3.29)$$

3. La spécificité, mesurant la capacité de l'algorithme à identifier correctement les patientes n'étant pas atteintes du cancer (équation 3.30).

$$Spécificité_{cancer} = \frac{VN}{VN + FP} \quad (3.30)$$

4. La sensibilité, évaluant la capacité de l'algorithme à détecter correctement les patientes atteintes du cancer (équation 3.31).

$$Sensibilité_{cancer} = \frac{VP}{VP + FN} \quad (3.31)$$

5. Le score F1, correspondant à la moyenne harmonique entre la sensibilité et la précision (équation 3.32),

$$Score \text{ } F1_{cancer} = 2 * \frac{Précision * Sensibilité}{Précision + Sensibilité} \quad (3.32)$$

où la précision est définie par la proportion des prédictions positives réellement correctes (équation 3.33).

$$Précision_{cancer} = \frac{VP}{VP + FP} \quad (3.33)$$

Pour la classification du stade, seules l'exactitude, l'exactitude balancée et le score F1 sont retenus, car la distinction entre les stades repose sur quatre classes, rendant la spécificité et la sensibilité moins directement applicables. De plus, les variables VP , VN , FP et FN ne s'appliquent pas directement. Les variables adaptées VP_i , VN_i , FP_i et FN_i sont définies comme suit :

- VP_i : le nombre de fois où une patiente appartenant au stade i est correctement prédite comme étant de ce stade.
- VN_i : le nombre de fois où une patiente n'appartenant pas au stade i est correctement prédite comme appartenant à un autre stade.

- FP_i : le nombre de fois où une patiente n'appartenant pas au stade i est incorrectement prédictée comme appartenant à ce stade.
- FN_i : le nombre de fois où une patiente appartenant au stade i est incorrectement prédictée comme n'appartenant pas à ce stade.

Les trois métriques utilisées pour la classification du stade sont donc :

1. L'exactitude, représentant la proportion des stades correctement prédis par rapport au nombre total de prédictions effectuées (N_{total}) (équation 3.34).

$$Exactitude_{stade} = \frac{1}{N_{total}} \sum_{i=1}^4 VP_i \quad (3.34)$$

2. L'exactitude balancée, corigeant les déséquilibres entre classes en prenant la moyenne des patientes correctement identifiées dans chaque stade (équation 3.35).

$$Exactitude\ balanced_{stade} = \frac{1}{4} * \sum_{i=1}^4 \left(\frac{VP_i}{VP_i + FN_i} \right) \quad (3.35)$$

3. Le score F1, correspondant à la moyenne, pondérée par le nombre de patientes réellement dans chaque stade (N_i), des moyennes harmoniques entre la précision (P_i) et la sensibilité (S_i) de chaque stade (équation 3.36),

$$Score\ F1_{stade} = \frac{1}{N_{total}} \sum_{i=1}^4 \left(N_i * 2 * \frac{P_i * S_i}{P_i + S_i} \right) \quad (3.36)$$

où P_i et S_i sont définies par les équations 3.37 et 3.38.

$$P_i = \frac{VP_i}{VP_i + FP_i} \quad (3.37)$$

$$S_i = \frac{VP_i}{VP_i + FN_i} \quad (3.38)$$

3.3.2 Méthode d'évaluation des algorithmes d'apprentissage supervisé

L'évaluation de la performance des algorithmes d'apprentissage supervisé se doit d'être à la fois fiable et reproductible. À cet effet, la méthode d'évaluation développée utilise la stratification K-Fold. Celle-ci est implémentée via la fonction *StratifiedKFold*, importée depuis la

classe *model_selection* de la bibliothèque *Scikit-learn*, disponible librement pour le langage de programmation Python [54]. Cette méthode de validation croisée divise les données en un nombre fixe de groupes tout en conservant, pour chacun d'eux, la même proportion de patientes de chaque classe. Chaque groupe stratifié sert, à tour de rôle, d'ensemble de test, tandis que l'ensemble des $n - 1$ autres groupes constitue l'ensemble d'entraînement. Ainsi, chaque patiente apparaît exactement une fois comme échantillon de test et $n - 1$ fois dans l'ensemble d'entraînement (Figure 3.4). Les hyperparamètres utilisés de la fonction *StratifiedKFold* sont : $n_splits = 5$, qui définit le nombre de groupes et $random_state = 42$, qui contrôle l'aléa de chaque *fold*. Ces hyperparamètres permettent de conserver en permanence 80% des données pour l'entraînement et de garantir une évaluation reproductible des performances de l'algorithme. De plus, la fonction *StandardScaler* de la classe *preprocessing* de *Scikit-learn* est utilisée pour standardiser les données en supprimant la moyenne et en les ramenant à une variance unitaire [55]. Enfin, un algorithme de sur-échantillonnage est également appliqué sur les données d'entraînement pour rebalancer les classes.

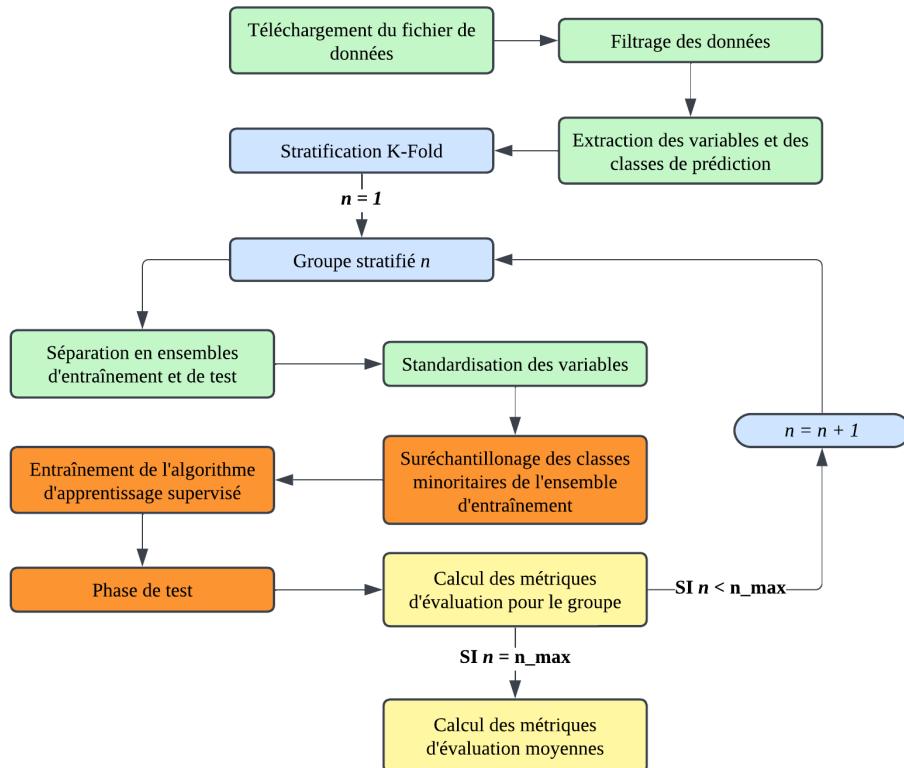


Figure 3.4 Méthode d'évaluation des algorithmes d'apprentissage supervisé

3.3.3 Optimisation de l'algorithme de sur-échantillonnage

Étant donné que, pour la détection du cancer et la classification de son stade, les classes sont déséquilibrées en nombre, l'utilisation d'un algorithme de sur-échantillonnage est à privilégier, car ce déséquilibre peut réduire la performance des algorithmes d'apprentissage supervisé en favorisant la classe majoritaire. Ainsi, l'algorithme utilisé est le *Synthetic Minority Over-sampling Technique* (SMOTE) pouvant être importé depuis la classe *over_sampling* de la bibliothèque libre d'accès *Imbalanced-learn* [56]. Cet algorithme, utilisé seulement pour l'entraînement des algorithmes d'apprentissage supervisé, génère des échantillons synthétiques à partir d'échantillons réels afin d'augmenter le nombre d'échantillons dans les classes minoritaires jusqu'à égaler le nombre d'échantillons dans la classe majoritaire. Par exemple, pour la classification du stade du cancer, la classe majoritaire est celle du stade 3 contenant 436 échantillons. En utilisant l'algorithme SMOTE, chacune des classes correspondant aux stades 1, 2 et 4 générera de nouvelles données jusqu'à obtenir 436 échantillons.

L'algorithme SMOTE prend un hyperparamètre important appelé *k_neighbors*; cet hyperparamètre détermine le nombre de voisins les plus rapprochés dans le voisinage d'échantillons à utiliser pour générer les échantillons synthétiques. Cet hyperparamètre ne peut donc pas être égal ou plus élevé que le nombre d'échantillons disponibles dans la banque d'entraînement. En utilisant les quatre algorithmes d'apprentissage supervisé avec leurs hyperparamètres par défaut, l'hyperparamètre *k_neighbors* de SMOTE est optimisé en fonction des exactitudes balancées moyennes obtenues après 100 simulations pour la détection du cancer et la classification de son stade. Chaque simulation repose sur une répartition aléatoire des échantillons tout en maintenant un ratio de 80% pour l'entraînement et 20 % pour le test afin de faire varier les échantillons contenus dans la banque de test. Les valeurs allant de 1 à 5 seront comparées pour l'hyperparamètre *k_neighbors* puisque, dans le cas de la détection du cancer, la classe minoritaire ne comporte que huit échantillons (huit patientes saines). En réservant 80% des échantillons pour l'entraînement, il y aura au minimum six échantillons de cette classe dans la banque d'entraînement. Ainsi, la valeur maximale que peut prendre *k_neighbors* est 5, car une valeur plus élevée impliquerait un nombre insuffisant de voisins disponibles pour la génération d'échantillons synthétiques.

3.3.4 Optimisation des algorithmes d'apprentissage supervisé

Les algorithmes *Logistic Regression* (LR), *Linear Support Vector Classification* (Linear SVC), *K-Nearest Neighbors* (KNN) et *Random Forest* (RF) ont été sélectionnés pour détecter le cancer de l'ovaire et classifier son stade en raison de leur capacité à effectuer des classifications binaires et multi-classes. Tout d'abord, LR est un modèle linéaire qui prédit la probabilité d'appartenance à une classe en utilisant une fonction logistique. Il est importé depuis la classe *linear_model* de la bibliothèque *Scikit-learn* [57]. De son côté, Linear SVC sépare les classes en traçant une frontière linéaire optimale, tout en maximisant la marge entre les classes. Cet algorithme se retrouve dans la classe *svm* de la bibliothèque *Scikit-learn* [58]. Pour sa part, KNN classe les données en fonction des k plus proches voisins, offrant une méthode simple mais puissante pour les problèmes de classification. Il est disponible depuis la classe *neighbors* de *Scikit-learn* [59]. Enfin, RF, un ensemble d'arbres de décision, combine plusieurs arbres pour améliorer la robustesse et l'exactitude des prédictions. RF peut être importé depuis la classe *ensemble* de *Scikit-learn* [60]. Ainsi, cette combinaison d'algorithmes offre une diversité méthodologique intéressante et reflète les principaux types d'algorithmes utilisés pour la détection du cancer.

Chaque algorithme vient avec ses propres hyperparamètres par défaut. Cependant, les hyperparamètres par défaut peuvent ne pas être les plus appropriés pour le type de classification et les variables utilisées de la base de données. En utilisant l'hyperparamètre *k_neighbors* = 5 de SMOTE, les hyperparamètres les plus importants des quatre algorithmes d'apprentissage supervisé sont optimisés en fonction des exactitudes balancées moyennes obtenues après 20 simulations, séparément pour la détection du cancer et la classification de son stade. Toutes les combinaisons possibles des hyperparamètres présentés (Tableau 3.4) sont donc testées. Pour RF, étant donné que l'algorithme génère des valeurs aléatoires, *random_state* = 42 est intégré dans les hyperparamètres par défaut afin d'assurer la reproductibilité des exactitudes balancées obtenues.

Tableau 3.5 Hyperparamètres testés des algorithmes d'apprentissage supervisé

LR	Linear SVC	KNN	RF
C: [10 ⁻⁵ , 10 ⁻⁴ , 10 ⁻³ , 10 ⁻² , 10 ⁻¹ , 1, 10, 100] penalty: ['l2', 'l1']	C: [10 ⁻⁴ , 10 ⁻³ , 10 ⁻² , 10 ⁻¹ , 1, 10, 100] tol: [10 ⁻⁵ , 10 ⁻⁴ , 10 ⁻³]	n_neighbors: [3, 5, 7, 10] weights: ['uniform', 'distance']	n_estimators: [50, 100, 200]

solver: ['lbfgs', 'saga', 'liblinear']	max_iter: [1000, 2000, 5000]	algorithm: ['auto', 'ball_tree', 'kd_tree', 'brute']	max_depth: [None, 10, 30]
max_iter: [100, 200, 500]	class_weight: ['balanced', None]	p: [1, 2]	min_samples_split: [2, 5]
tol: [10^{-5} , 10^{-4} , 10^{-3}]	dual: [True, False]	leaf_size: [30, 50, 70]	min_samples_leaf: [1, 2]
class_weight: ['balanced', None]	fit_intercept: [True, False]		max_features: ['sqrt', 'log2']
warm_start: [True, False]			class_weight: [None, 'balanced']
			random_state: 42

Comme pour l'optimisation de SMOTE, chaque simulation repose sur une répartition aléatoire des échantillons, avec un ratio de 80% pour l'entraînement et 20% pour le test, garantissant ainsi une variation des échantillons dans l'ensemble de test. Cette première étape d'optimisation permet de filtrer les meilleures combinaisons d'hyperparamètres. Ensuite, une seconde étape d'optimisation sélectionne la meilleure combinaison d'hyperparamètres parmi les dix premières retenues en appliquant une validation croisée grâce à la stratification K-Fold. Les deux étapes d'optimisation sont donc complémentaires; la première réduit plus efficacement l'espace des hyperparamètres, tandis que la seconde garantit une évaluation reproductible et fiable des performances des modèles.

3.4 Ciblage thérapeutique

Dans le cadre de ce projet, le ciblage thérapeutique consiste en un ensemble d'algorithmes développés pour rechercher une solution, caractérisée par un ou plusieurs gènes, chacun ayant son propre facteur de modulation, permettant de réduire la croissance tumorale. Le niveau d'expression de chaque gène (DT) d'une solution trouvée est modulé par un coefficient appelé facteur de modulation (f_m), permettant de simuler la suppression ou la surexpression de ce dernier (équation 3.39). Cette modulation est effectuée au temps $t = 0$ h de la simulation de la croissance tumorale et de l'évolution temporelle des métabolites. Les facteurs de modulation testés sont inclus dans l'intervalle $[0, 2]$, un facteur de modulation de 0 permettant une suppression complète du gène et un facteur de 2 de doubler son expression génique.

$$DT_{modulée} = f_m * DT_{originale} \quad (3.39)$$

Deux algorithmes de recherche ont été développés dans le cadre de ce projet: un algorithme de recherche en profondeur, nommé *Depth Search*, et un algorithme de recherche intelligente, nommé *Intelligent Search*. Le *Depth Search* permet d'explorer toutes les solutions de l'espace des solutions possibles, et présente ainsi l'avantage de retourner une solution optimale. L'*Intelligent Search*, quant à lui, vise à réduire le temps de recherche en explorant en profondeur des voisins contributifs de l'espace des solutions. Bien qu'il n'assure pas l'optimalité de la solution retournée, l'*Intelligent Search* se révèle plus efficient en termes de temps de calcul.

Les hyperparamètres des algorithmes de recherche déterminent à la fois le format des solutions générées et le temps nécessaire à la recherche. Ces hyperparamètres incluent n_{iter} (le nombre d'itérations après lequel la recherche s'arrête, pertinent seulement pour l'*Intelligent Search*), $n_{max,gènes}$ (le nombre maximal de gènes que peut comporter une solution), n_{mod} (le nombre de facteurs de modulation différents pouvant être testés) et $n_{poss,gènes}$ (le nombre de gènes différents pouvant être testés). À partir de ces paramètres, il est possible de calculer le nombre total de solutions différentes (n_{sol}) de l'espace des solutions possibles (équation 3.40).

$$n_{sol} = \sum_{i=1}^{n_{max,gènes}} \left\{ \frac{(n_{poss,gènes})!}{i! * (n_{poss,gènes} - i)!} * (n_{mod})^i \right\} \quad (3.40)$$

Pour l'ensemble des tests, les gènes potentiellement ciblés sont ceux associés au modèle métabolique par les règles GPR. Toutefois, les gènes associés aux réactions *Vgrowth* et *Vresp* du modèle métabolique sont exclus de la recherche. En effet, pour la réaction *Vgrowth*, qui caractérise la croissance de la masse de cellules tumorales, les gènes associés sont des gènes suppresseurs de tumeur et des oncogènes. Bien qu'une mutation d'un de ces gènes puisse initier le développement d'une tumeur ovarienne, leur influence sur la vitesse de croissance est seulement hypothétique. Ainsi, puisque ces gènes sont déjà considérés comme des cibles potentielles pour le diagnostic et le traitement du cancer de l'ovaire (en raison de leur risque de mutation) et que leur impact sur la vitesse de croissance tumorale est théorique, leur utilisation dans un algorithme de ciblage thérapeutique ne présente pas un intérêt majeur. Concernant la réaction *Vresp*, qui caractérise la respiration cellulaire, les gènes associés sont des gènes COX, une famille de gènes codant des sous-unités de l'enzyme *cytochrome c oxidase*, essentielles à la respiration cellulaire. Cependant, la complexité de cette réaction et la difficulté d'inhiber le processus de respiration cellulaire chez une cellule de mammifère rendent le ciblage de ces gènes moins pertinent. Ainsi, en excluant les gènes

associés à *Vgrowth* et à *Vresp*, l'ensemble des 75 gènes restants du modèle métabolique constitue un ensemble raisonnable et pertinent de potentielles cibles thérapeutiques.

Par exemple, si une recherche autorise l'inclusion d'au maximum trois gènes par solution, avec la possibilité de tester quatre facteurs de modulation différents et en considérant l'ensemble des 75 gènes, l'espace des solutions qui en résultera comprendra près de quatre millions de solutions différentes:

$$n_{sol} = 75 * 4 + \frac{(75 * 74)}{2} * (4)^2 + \frac{(75 * 74 * 73)}{6} * (4)^3 \approx 4,37 * 10^6$$

3.4.1 Évaluation d'une solution

Une solution issue du ciblage thérapeutique, caractérisée par un ensemble de gènes et leurs facteurs de modulation respectifs, est évaluée en fonction de la réduction qu'elle induit sur la croissance du volume tumoral par rapport au cas référentiel de la non-modulation (i.e. activités inaltérées) des gènes identifiés de la tumeur, et selon sa praticabilité. Pour déterminer cette praticabilité, deux facteurs sont pris en compte : le nombre de gènes à moduler et la dispersion moyenne de cette modulation autour de 1. Ainsi, une solution comportant un seul gène avec un facteur de modulation de 0,5 sera priorisée par rapport à une solution ayant, pour un impact identique sur la réduction de croissance tumorale, recours à trois gènes tous modulés par un facteur de 0, car cette dernière est jugée plus complexe à mettre en œuvre pratiquement en clinique.

Tout d'abord, l'indice de réduction de la croissance tumorale (*TS*) est défini comme la moyenne des *n* rapports, calculés à partir de la 24^e heure, puis à chaque 48 heures de simulation, entre le volume de la tumeur obtenu avec la solution (*VT_{mod}*) et le volume de la tumeur inaltéré (*VT_{OG}*), après soustraction du volume initial de la tumeur (*VT_{OG,t=0}*), aux volumes évalués (équation 3.41). Le calcul de *TS* permet de quantifier l'effet de la modulation de gènes sur la croissance tumorale de manière constante durant la simulation. De plus, il n'est pas pertinent de calculer un rapport à *t* = 0 h puisque l'effet de la modulation génique sera nul à l'instant initial; c'est pourquoi le premier rapport est donc pris à la fin des premières 24 heures. Par exemple, si la durée de simulation est fixée à 270 jours, 134 rapports (*n_{rapports}*) entre les volumes tumoraux modulé et inaltéré sont moyennés pour déterminer *TS*. Dans ce contexte, *TS* appartient à l'intervalle [0, ∞] puisque selon les équations du modèle métabolique, le volume de la tumeur ne peut décroître et

aucune limite supérieure ne lui est imposée. Ainsi, plus la valeur de TS est faible, plus la croissance tumorale est ralentie par la modulation génique.

$$TS = \frac{1}{n_{rapports}} * \sum_{i=0}^{n_{rapports}-1} \left(\frac{VT_{mod,t=24+48*i} - VT_{OG,t=0}}{VT_{OG,t=24+48*i} - VT_{OG,t=0}} \right) \quad (3.41)$$

Ensuite, l'indice d'impraticabilité lié au nombre de gènes modulés (GV) est proportionnel au nombre de gènes ($n_{gènes}$) constituant la solution, en tenant compte du nombre maximal de gènes qu'une solution peut contenir ($n_{max,gènes}$). Toutefois, si la solution ne comprend qu'un gène, GV est nul, car il s'agit du scénario minimal (il doit y avoir au moins un gène modulé) (équation 3.42). Ainsi, GV varie de 0 à 1, plus sa valeur est faible, plus la solution est considérée comme praticable.

$$GV = \begin{cases} \text{si } n_{gènes} = 1: GV = 0 \\ \text{si } n_{gènes} > 1: GV = \frac{n_{gènes}}{n_{max,gènes}} \end{cases} \quad (3.42)$$

Puis, l'indice d'impraticabilité associé aux facteurs de modulations (MV) quantifie la dispersion des facteurs de modulation de la solution autour de 1, valeur neutre correspondant à un effet nul. MV est également normalisé entre 0 et 1. Ainsi, plus MV est faible, moins les facteurs de modulation s'écartent de 1 et plus la solution est jugée comme viable. En effet, un facteur de modulation fortement différent de 1 implique que le traitement destiné à supprimer ou activer le gène doit être particulièrement performant, ce qui est considéré comme moins réaliste et moins facilement applicable. Les étapes suivantes détaillent le calcul de MV :

1. Pour chaque gène i , on calcule la différence absolue entre son facteur de modulation ($f_{m,i}$) et 1, afin de quantifier son écart par rapport à l'effet neutre.
2. On normalise cette différence en la recentrant par rapport au facteur de modulation le plus proche de 1 autorisé par les hyperparamètres $f_{m,min}$, en soustrayant la valeur absolue de ($f_{m,min} - 1$).
3. On divise cette différence normalisée par la valeur absolue de ($f_{m,max} - 1$), $f_{m,max}$ étant le facteur de modulation le plus éloigné de 1 permis par les hyperparamètres de l'algorithme de recherche, afin d'obtenir une valeur comprise entre 0 et 1.
4. Enfin, on calcule la moyenne de ces valeurs sur l'ensemble des gènes constituant la solution pour obtenir MV (équation 3.43).

$$MV = \frac{1}{n_{gènes}} * \sum_{i=1}^{n_{gènes}} \left(\frac{|f_{m,i} - 1| - |f_{m,min} - 1|}{|f_{m,max} - 1|} \right) \quad (3.43)$$

En calculant une moyenne pondérée entre TS , GV et MV , une valeur notée V_{sol} est attribuée à chaque solution (équation 3.44).

$$V_{sol} = 1 - (0,98 * TS + 0,01 * GV + 0,01 * MV) \quad (3.44)$$

Avec une pondération de 98%, la croissance tumorale, représentée par TS , influence majoritairement V_{sol} . Les pondérations de 1% attribuées aux indices d'impraticabilité GV et MV permettent de distinguer des solutions ayant un effet très similaire sur la croissance tumorale. Ces pondérations résultent d'un processus d'optimisation principalement qualitatif, visant à évaluer la pertinence des solutions retournées par les algorithmes de recherche. Étant donné que TS varie de 0 à l'infini et que GV ainsi que MV varient de 0 à 1, V_{sol} appartient à l'intervalle $[-\infty, 1]$. Plus V_{sol} se rapproche de 1, meilleure est considérée la solution par l'algorithme de recherche.

3.4.2 Recherche en profondeur de solutions (*Depth Search*)

L'algorithme de recherche en profondeur *Depth Search* explore séquentiellement l'ensemble des solutions possibles dans l'espace des solutions. L'algorithme est relativement simple dans sa conception et garantit le retour d'une solution optimale puisqu'il poursuit l'exploration jusqu'à ce que l'intégralité des solutions aient été évaluées. L'ordre d'exploration de *Depth Search* se décompose en trois étapes:

1. Variation du nombre de gènes : il explore successivement des ensembles de gènes de taille croissante, de 1 à jusqu'à $n_{max,gènes}$.
2. Combinaisons de gènes : pour chaque taille d'ensemble, il génère tous les sous-ensembles possibles parmi les 75 gènes testables ($n_{poss,gènes}$).
3. Combinaisons de facteurs de modulation : pour chaque ensemble de gènes sélectionné, il évalue toutes les combinaisons possibles des n facteurs de modulation testables (n_{mod}).

Ces étapes permettent à l'algorithme de couvrir l'ensemble des espaces de solutions.

3.4.3 Recherche intelligente de solutions (*Intelligent Search*)

L'algorithme de recherche intelligente *Intelligent Search* combine une exploration aléatoire et une exploration ciblée de l'espace des solutions. Il fonctionne sur un nombre fixe d'itérations (n_{iter}), chaque itération correspondant à l'évaluation d'une nouvelle solution (Figure 3.5).

L'algorithme commence par tester une solution aléatoire. Ensuite, en fonction de la valeur associée à cette solution, il peut soit explorer son voisinage pendant un nombre fixe d'itérations, soit générer une nouvelle solution aléatoire. Ainsi, *Intelligent Search* sélectionne des solutions au hasard dans l'espace des solutions jusqu'à identifier une solution intéressante. Une solution est considérée comme intéressante si sa valeur (V_{sol}) est strictement positive et représente au moins 60% de la valeur de la meilleure solution trouvée jusqu'alors. Lorsqu'une telle solution est identifiée, l'algorithme consacre 5% des itérations totales à explorer son voisinage.

Exploration du voisinage

Une solution voisine est obtenue en modifiant la solution actuelle de l'une des manières suivantes :

1. Modification des gènes (70% de probabilité) :
 - Si la solution contient moins de gènes que le nombre de gènes maximal alloué ($n_{max,gènes}$), un gène est ajouté. Le gène ajouté reçoit un facteur de modulation aléatoire parmi ceux autorisés.
 - Si la solution contient $n_{max,gènes}$, un gène est retiré.
2. Modification des facteurs de modulation (30% de probabilité) :
 - Chaque gène de la solution actuelle reçoit un nouveau facteur de modulation aléatoire parmi ceux permis.

Si la solution voisine est meilleure que la solution actuelle, elle la remplace et le voisin suivant est généré à partir de cette nouvelle solution. Sinon, la solution actuelle est conservée pour générer le prochain voisin. Après avoir exploré le voisinage pendant 5% des itérations, l'algorithme reprend la génération de nouvelles solutions aléatoires et le cycle se poursuit jusqu'à 90% des itérations n_{iter} .

Phase finale

Les 10% d'itérations restantes sont exclusivement consacrés à l'exploration du voisinage de la meilleure solution trouvée jusqu'à présent, afin d'optimiser le résultat final avant l'arrêt de l'algorithme.

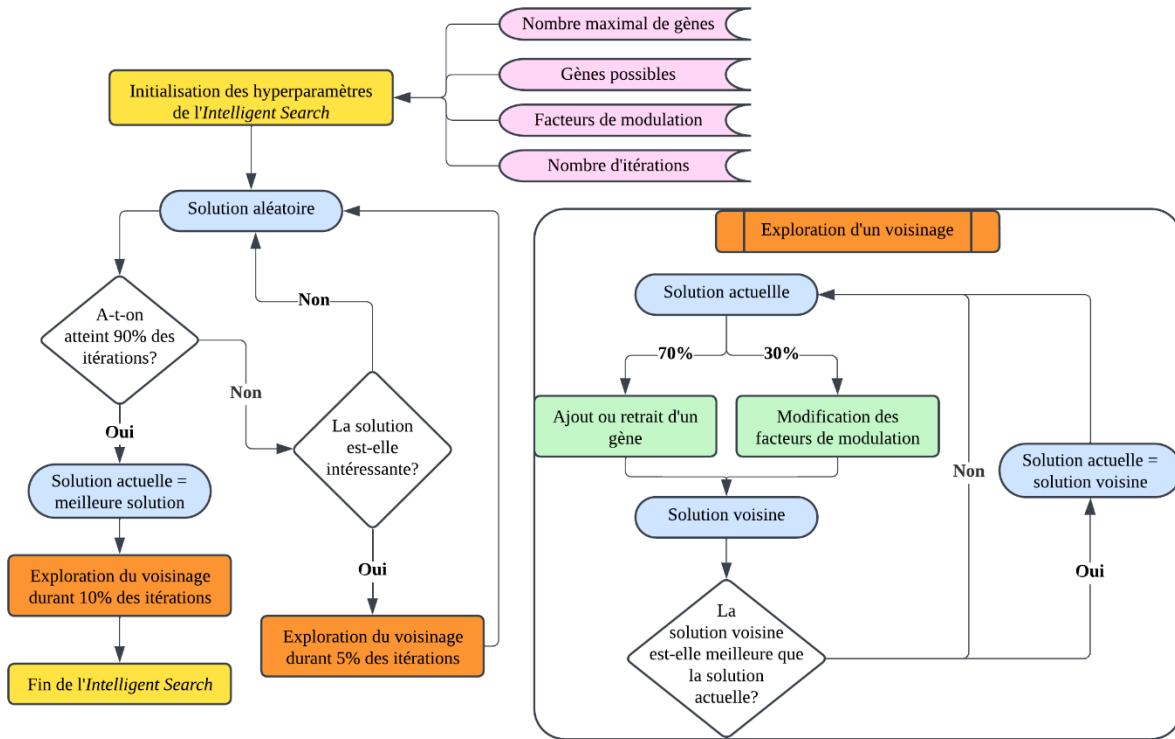


Figure 3.5 Schématisation de l'algorithme de recherche intelligente (*Intelligent Search*)

CHAPITRE 4 RÉSULTATS

Ce chapitre présente les résultats obtenus en appliquant la méthodologie décrite au chapitre précédent. En cas de détection positive du cancer, le stade tumoral est prédit à l'aide des algorithmes d'apprentissage supervisé. Ce stade permet ensuite d'estimer le volume tumoral, qui sert de base à la modélisation métabolique. Enfin, cette modélisation permet l'identification de cibles thérapeutiques spécifiques au profil transcriptomique et à l'évolution du cancer chez la patiente.

4.1 Apprentissage supervisé

Cette section évalue les performances des quatre algorithmes d'apprentissage supervisé sélectionnés pour détecter le cancer de l'ovaire, classifier son stade de développement et identifier la récurrence, en utilisant des niveaux d'expression génique, obtenus par ARNseq, comme variables d'apprentissage. De plus, l'impact de l'optimisation des hyperparamètres des algorithmes utilisés et l'effet des variables d'apprentissages sont également présentés dans cette section.

4.1.1 Hyperparamètre de l'algorithme de sur-échantillonnage

Afin d'évaluer l'efficacité de l'algorithme de sur-échantillonnage SMOTE pour la détection du cancer et la classification de son stade, les quatre algorithmes d'apprentissage supervisé ont été testés, en utilisant leurs hyperparamètres par défaut et en se basant sur les 112 gènes associés au modèle métabolique. Pour chaque algorithme et pour chaque type de classification, l'absence de SMOTE a été comparée à son utilisation avec différentes valeurs de l'hyperparamètre $k_neighbors$ (allant de 1 à 5). Les exactitudes balancées ont été calculées sur la base de 100 simulations par test (Tableau 4.1).

Dans certains cas, l'utilisation de SMOTE améliore considérablement la performance des algorithmes d'apprentissage supervisé, notamment LR pour la détection du cancer (plus de 99% contre 92,9%), KNN pour la classification du stade (plus de 33% contre 26,1%) et RF pour la détection du cancer (plus de 72% contre 51,5%). Concernant le paramètre $k_neighbors$, sa valeur n'influence pas de manière significative les exactitudes balancées de manière générale. Ainsi, puisque la valeur par défaut de $k_neighbors$ est de 5, c'est-à-dire que, sauf indication contraire,

SMOTE considère automatiquement les cinq points les plus proches pour prendre une décision, SMOTE sera utilisé avec $k_neighbors = 5$, tant pour la détection du cancer que pour la classification de son stade, et ce, pour l'ensemble des algorithmes d'apprentissage supervisé.

Tableau 4.1 Optimisation de l'hyperparamètre de SMOTE

Algorithme	Hyperparamètre $k_neighbors$ de SMOTE	Exactitude balancée pour la détection du cancer (%)	Exactitude balancée pour la classification du stade (%)
LR	Sans SMOTE	92,9	28,4
	$k = 1$	99,7	28,4
	$k = 2$	99,8	29,1
	$k = 3$	99,8	29,8
	$k = 4$	99,8	28,6
	$k = 5$	99,8	29,2
Linear SVC	Sans SMOTE	99,8	28,7
	$k = 1$	99,6	29,1
	$k = 2$	99,8	29,0
	$k = 3$	99,8	28,1
	$k = 4$	99,8	28,7
	$k = 5$	99,8	29,6
KNN	Sans SMOTE	99,9	26,1
	$k = 1$	99,3	33,9
	$k = 2$	99,1	34,7
	$k = 3$	99,1	34,3
	$k = 4$	98,9	33,9
	$k = 5$	99,0	34,9
RF	Sans SMOTE	51,5	24,9
	$k = 1$	72,3	25,5
	$k = 2$	75,5	25,8
	$k = 3$	75,3	26,0
	$k = 4$	77,3	26,0
	$k = 5$	75,0	26,2

4.1.2 Hyperparamètres des algorithmes d'apprentissage supervisé

Les hyperparamètres des algorithmes d'apprentissage supervisé influencent également leur performance. En fixant $k_neighbors$ de SMOTE à 5 et en utilisant les 112 gènes associés au modèle métabolique comme variables d'apprentissage, le Tableau 4.2 récapitule les hyperparamètres

optimaux des quatre algorithmes sélectionnés permettant de maximiser l'exactitude balancée dans la détection du cancer et la classification du stade.

Tableau 4.2 Optimisation des hyperparamètres des algorithmes d'apprentissage supervisés

Algorithm	Classification	Hyperparamètres optimisés	Exactitude balancée (%)
LR	Cancer	C: 10, penalty: 'l2', solver: 'lbfgs', max_iter: 500, tol: 10^{-5} , class_weight: None, warm_start: False	99,7
	Stade	C: 10^{-4} , penalty: 'l2', solver: 'saga', max_iter: 200, tol: 10^{-5} , class_weight: None, warm_start: True	37,8
Linear SVC	Cancer	C: 1, tol: 10^{-5} , max_iter: 5000, class_weight: 'balanced', dual: False, fit_intercept: True	99,7
	Stade	C: 10^{-3} , tol: 10^{-3} , max_iter: 2000, class_weight: 'balanced', dual: True, fit_intercept: False	35,6
KNN	Cancer	n_neighbors: 3, weights: 'distance', algorithm: 'ball_tree', p: 2, leaf_size: 50	99,1
	Stade	n_neighbors: 10, weights: 'uniform', algorithm: 'brute', p: 1, leaf_size: 70	34,3
RF	Cancer	n_estimators: 50, max_depth: None, min_samples_split: 2, min_samples_leaf: 1, max_features: 'sqrt', class_weight: 'balanced', random_state: 42	75,0
	Stade	n_estimators: 50, max_depth: 10, min_samples_split: 2, min_samples_leaf: 1, max_features: 'log2', class_weight: None, random_state: 42	28,7

Afin d'analyser l'impact de l'optimisation des hyperparamètres des algorithmes d'apprentissage supervisé, nous avons comparé les exactitudes balancées obtenues en utilisant des hyperparamètres optimisés avec celles obtenues avec des hyperparamètres par défaut (Tableau 4.3).

L'optimisation des hyperparamètres des algorithmes d'apprentissage supervisé permet d'augmenter significativement leurs exactitudes balancées lorsque celles-ci sont faibles. En effet,

pour les quatre algorithmes, les exactitudes balancées pour la classification du stade sont plus élevées avec des hyperparamètres optimisés : 37,8% contre 31,4% pour LR, 35,6% contre 26,4% pour Linear SVC, 34,3% contre 31,8% pour KNN et 28,7% contre 26,3% pour RF. Par ailleurs, pour la détection du cancer avec KNN, l'utilisation de paramètres optimisés conduit à une légère augmentation de l'exactitude balancée, passant de 98,6% à 99,1%. Toutefois, étant donné les exactitudes balancées très élevées pour la détection du cancer, notamment en utilisant LR, Linear SVC et KNN, l'optimisation des hyperparamètres n'apporte pas d'amélioration significative pour la détection du cancer.

Tableau 4.3 Comparaison des exactitudes balancées des algorithmes d'apprentissage supervisé avec les hyperparamètres par défaut et les hyperparamètres optimisés

Algorithmme	Classification	Exactitude balancée (%)	
		Hyperparamètres par défaut	Hyperparamètres optimisés
LR	Cancer	99,7	99,7
	Stade	31,4	37,8
Linear SVC	Cancer	99,7	99,7
	Stade	26,4	35,6
KNN	Cancer	98,6	99,1
	Stade	31,8	34,3
RF	Cancer	75,0	75,0
	Stade	26,3	28,7

4.1.3 Impact des variables d'apprentissage sur la détection du cancer

En ayant utilisé les 112 gènes associés au modèle métabolique comme variables d'apprentissage pour l'optimisation des hyperparamètres, il est pertinent d'évaluer si l'utilisation de l'ensemble des gènes disponibles dans la base de données améliore significativement les performances des algorithmes d'apprentissage supervisé pour la détection du cancer. Pour ce faire,

en appliquant les mêmes hyperparamètres optimisés, les métriques d'évaluation ont été calculées pour deux ensembles de variables d'apprentissage : d'une part, les 112 gènes associés au modèle métabolique, et d'autre part, l'ensemble des 13 110 gènes de la base de données (Tableau 4.4).

Tableau 4.4 Évaluation des algorithmes d'apprentissage supervisé pour la détection du cancer selon les variables d'apprentissage utilisées

Algorithmme	Variables d'apprentissage	Métriques d'évaluation				
		Exactitude balancée (%)	Exactitude (%)	Sensibilité (%)	Spécificité (%)	Score F1
LR	112 gènes du modèle métabolique	99,7	99,3	99,3	100,0	1,00
	13 110 gènes du TCGA	99,9	99,8	99,8	100,0	1,00
Linear SVC	112 gènes du modèle métabolique	99,7	99,5	99,5	100,0	1,00
	13 110 gènes du TCGA	92,5	85,1	84,9	100,0	0,92
KNN	112 gènes du modèle métabolique	99,1	98,3	98,3	100,0	0,99
	13 110 gènes du TCGA	99,9	99,8	99,8	100,0	1,00
RF	112 gènes du modèle métabolique	75,0	99,3	100,0	50,0	1,00
	13 110 gènes du TCGA	93,8	99,8	100,0	87,5	0,92

De plus, les matrices de confusion pour la détection du cancer, obtenues à partir des résultats des cinq groupes stratifiés par la fonction *StratifiedKFold*, sont présentées à la Figure 4.1 pour les 112 gènes associés au modèle métabolique, et à la Figure 4.2 pour l'ensemble des 13 110 gènes de la base de données.

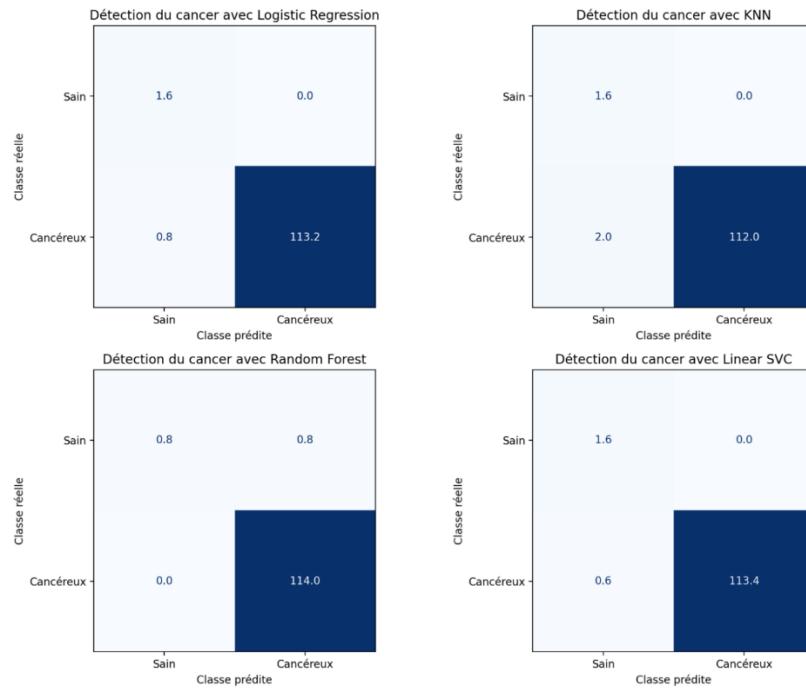


Figure 4.1 Matrices de confusion moyennant cinq K-Folds et utilisant les 112 gènes du modèle métabolique pour la détection du cancer

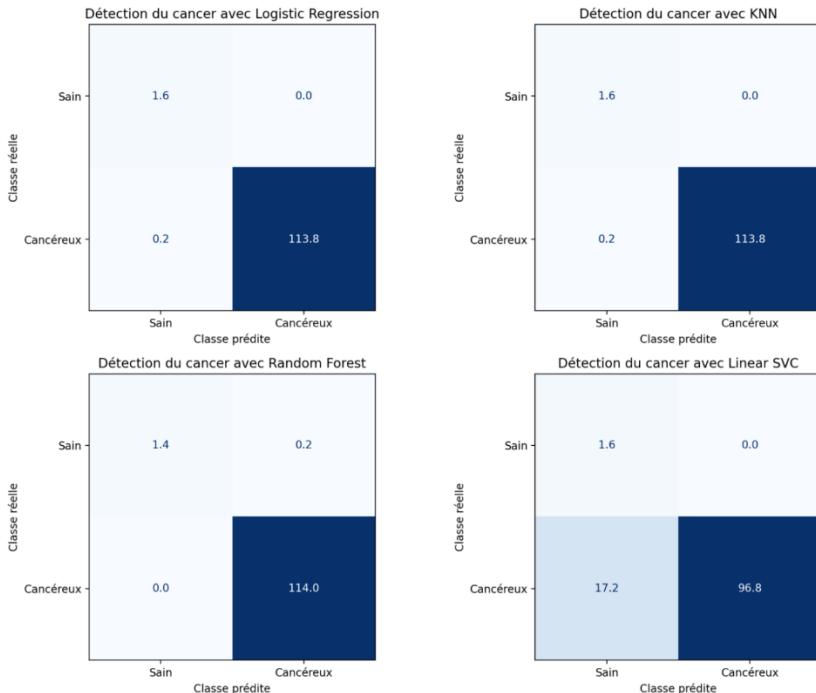


Figure 4.2 Matrices de confusion moyennant cinq K-Folds et utilisant les 13 110 gènes du TCGA pour la détection du cancer

Pour les algorithmes LR et KNN, l'utilisation de l'ensemble complet des gènes du TCGA améliore légèrement certaines métriques d'évaluation pour la détection du cancer. En revanche, pour RF, l'exactitude balancée et la spécificité sont nettement supérieures lorsque l'ensemble des 13 110 gènes sont utilisés (93,8% contre 75,0% pour l'exactitude balancée et 87,5% contre 50,0% pour la spécificité). Toutefois, Linear SVC affiche de meilleures performances avec le plus petit ensemble de variables d'apprentissage, à savoir les 112 gènes associés au modèle métabolique avec 99,7% contre 92,5% pour l'exactitude balancée, 99,5% contre 85,1% pour la sensibilité, 99,5% contre 84,9% pour la spécificité et un score F1 de 1,00 contre 0,92.

4.1.4 Détection précoce du cancer

En ne considérant que les huit patientes saines et les 16 patientes classées au stade I du développement du cancer, les mêmes algorithmes d'apprentissage supervisé ont été employés afin de différencier une patiente saine d'une patiente atteinte du cancer de l'ovaire dès son stade le plus précoce. Les algorithmes ont été utilisés en prenant leurs hyperparamètres optimisés pour la détection du cancer ainsi que les 112 gènes associés au modèle métabolique. Le Tableau 4.5 présente les résultats de la détection précoce du cancer de l'ovaire.

Tableau 4.5 Évaluation des algorithmes d'apprentissage supervisé pour la détection précoce du cancer en ne considérant que les patientes saines et au stade I

Algorithmme	Métriques d'évaluation				
	Exactitude balancée (%)	Exactitude (%)	Sensibilité (%)	Spécificité (%)	Score F1
LR	96,7	96,0	93,8	100,0	0,96
Linear SVC	90,8	87,0	81,3	100,0	0,89
KNN	78,3	72,0	56,3	100,0	0,69
RF	97,5	96,0	93,8	100,0	0,97

Puis, la Figure 4.3 présente les matrices de confusion associées à la détection précoce du cancer de l'ovaire.

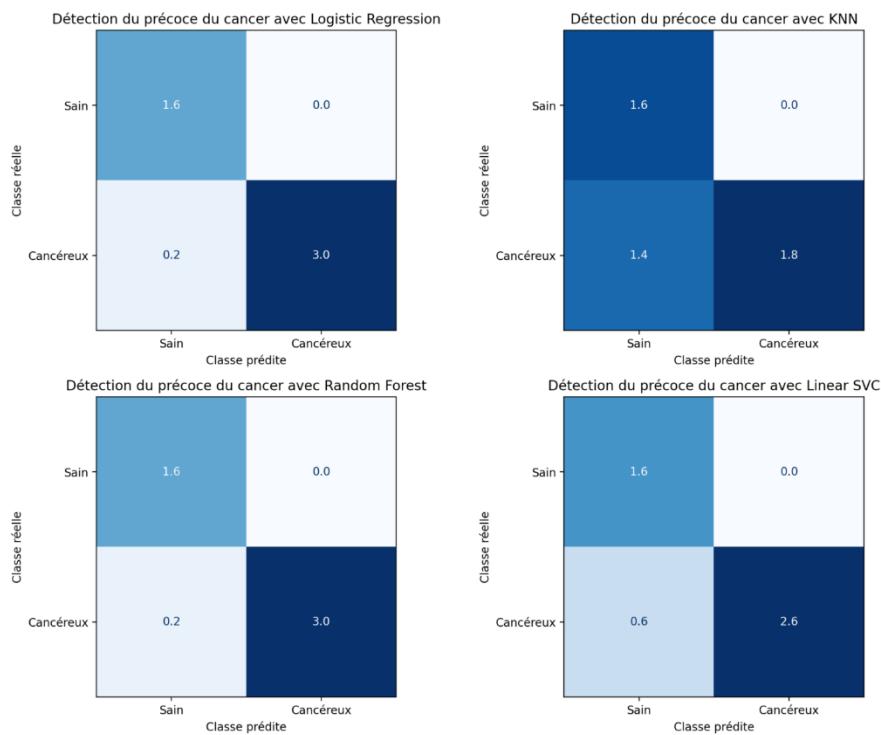


Figure 4.3 Matrices de confusion moyennant cinq K-Folds pour la détection précoce du cancer

L'algorithme RF, qui atteint 97,5% d'exactitude balancée, est celui qui permet le mieux de différencier une patiente présentant un cancer de l'ovaire à un stade précoce d'une patiente saine. Toutefois, il convient de noter que les exactitudes balancées sont généralement plus faibles lorsque l'on ne considère que les patientes saines et celles au stade I. Néanmoins, bien que la détection du cancer de l'ovaire au stade I soit plus difficile, les algorithmes LR, Linear SVC et RF parviennent tous à différencier une patiente saine d'une patiente atteinte d'un cancer précoce avec des exactitudes balancées supérieures à 90%.

4.1.5 Impact des variables d'apprentissage sur la classification du stade

Pour la classification du stade, l'utilisation de l'ensemble complet des 13 110 gènes comme variables d'apprentissage pourrait également s'avérer intéressante. Le Tableau 4.6 présente les

différences entre les métriques d'évaluation en prenant les 112 gènes du modèle métabolique et en prenant l'ensemble des gènes de la base de données.

Tableau 4.6 Évaluation des algorithmes d'apprentissage supervisé pour la classification du stade selon les variables d'apprentissage utilisées

Algorithmme	Variables d'apprentissage	Métriques d'évaluation		
		Exactitude balancée (%)	Exactitude (%)	Score F1
LR	112 gènes du modèle métabolique	37,8	33,8	0,40
	13 110 gènes du TCGA	37,7	51,9	0,56
Linear SVC	112 gènes du modèle métabolique	35,6	25,9	0,30
	13 110 gènes du TCGA	31,2	33,2	0,41
KNN	112 gènes du modèle métabolique	34,3	14,2	0,08
	13 110 gènes du TCGA	32,9	13,9	0,07
RF	112 gènes du modèle métabolique	28,7	74,3	0,70
	13 110 gènes du TCGA	27,1	75,9	0,70

Ensuite, les matrices de confusion pour la classification du stade, obtenues à partir des résultats des cinq groupes stratifiés par la fonction *StratifiedKFold*, sont présentées à la Figure 4.4, en prenant les 112 gènes associés au modèle métabolique, et à la Figure 4.5 en utilisant l'ensemble des 13 110 gènes de la base de données.

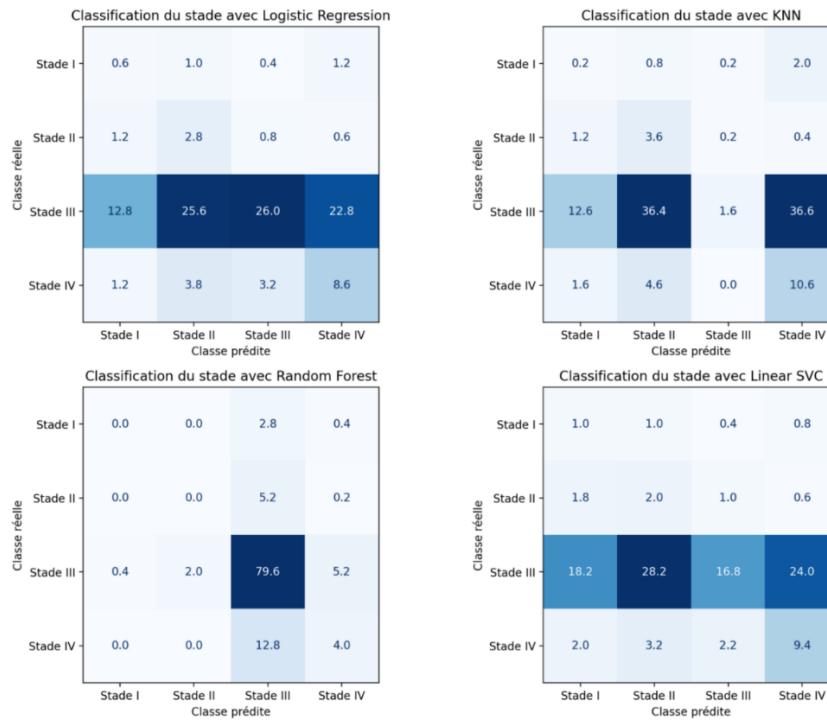


Figure 4.4 Matrices de confusion moyennant cinq K-Folds et utilisant les 112 gènes du modèle métabolique pour la classification du stade

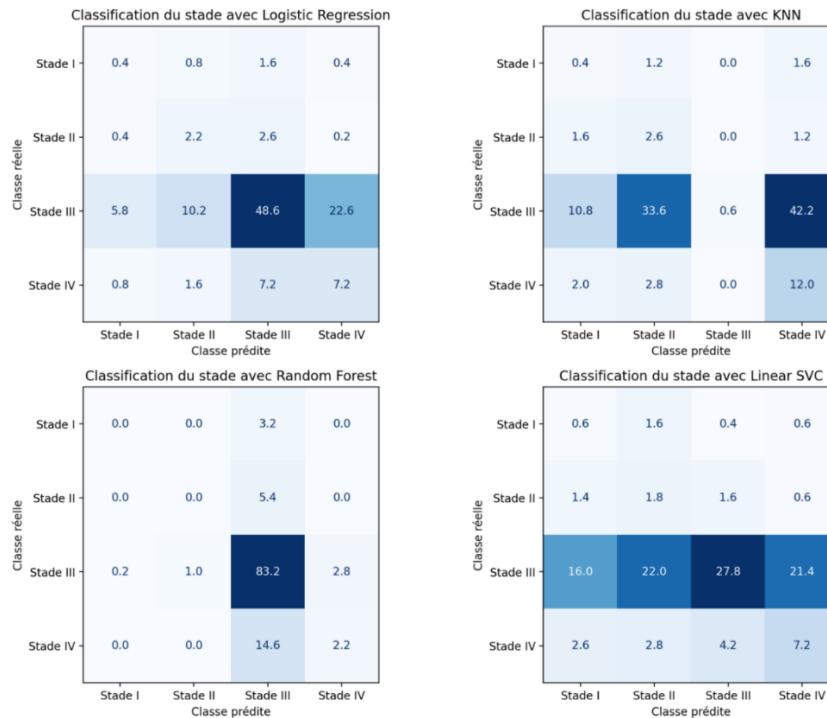


Figure 4.5 Matrices de confusion moyennant cinq K-Folds et utilisant les 13 110 gènes du TCGA pour la classification du stade

Contrairement à la détection du cancer, l'utilisation de l'ensemble des 13 110 gènes pour la classification du stade diminue les exactitudes balancées des algorithmes d'apprentissage supervisé. Cependant, comme les résultats ont été obtenus en utilisant les hyperparamètres optimisés pour les 112 gènes, il serait pertinent d'optimiser à nouveau ces hyperparamètres spécifiquement pour l'ensemble des 13 110 gènes afin d'évaluer leur impact sur la classification du stade.

4.1.6 Impact des variables d'apprentissage sur l'identification de la récurrence

Une autre classe de prédiction à considérer est la récurrence. En effet, les patientes présentant un cancer de l'ovaire récurrent pourraient manifester des altérations intéressantes au niveau de l'expression génique. Le Tableau 4.7 présente les performances des algorithmes d'apprentissage supervisé, en utilisant leurs hyperparamètres par défaut, pour identifier si un cancer de l'ovaire est récurrent ou non.

Tableau 4.7 Évaluation des algorithmes d'apprentissage supervisé pour l'identification de la récurrence du cancer de l'ovaire selon les variables d'apprentissage utilisées

Algorithmme	Variables d'apprentissage	Métriques d'évaluation				
		Exactitude balancée (%)	Exactitude (%)	Sensibilité (%)	Spécificité (%)	Score F1
LR	112 gènes du modèle métabolique	47,4	47,7	53,2	41,7	0,52
	13 110 gènes du TCGA	51,8	52,3	60,9	42,8	0,57
Linear SVC	112 gènes du modèle métabolique	47,8	48,1	52,8	42,8	0,52
	13 110 gènes du TCGA	53,6	53,9	59,5	47,6	0,58
KNN	112 gènes du modèle métabolique	52,5	51,4	31,8	73,1	0,40

	13 110 gènes du TCGA	53,2	51,4	16,7	89,7	0,26
RF	112 gènes du modèle métabolique	48,3	48,6	54,5	42,1	0,52
	13 110 gènes du TCGA	53,2	54,0	61,2	46,1	0,58

Puis, les matrices de confusion associées à la détection de la récurrence sont présentées à la Figure 4.6, en prenant les 112 gènes associés au modèle métabolique, et à la Figure 4.7 pour l'ensemble des 13 110 gènes de la base de données.

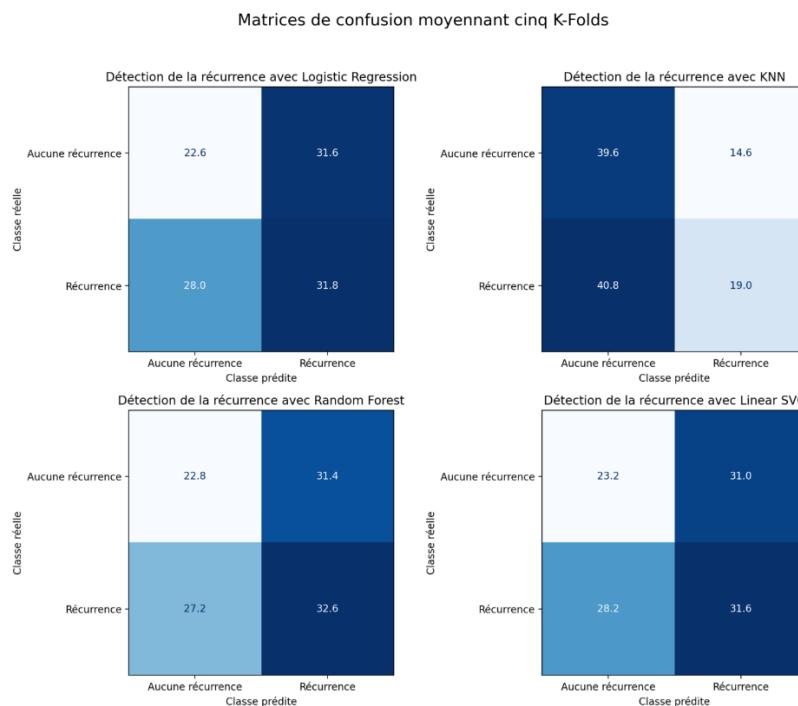


Figure 4.6 Matrices de confusion moyennant cinq K-Folds et utilisant les 112 gènes du modèle métabolique pour l'identification de la récurrence du cancer

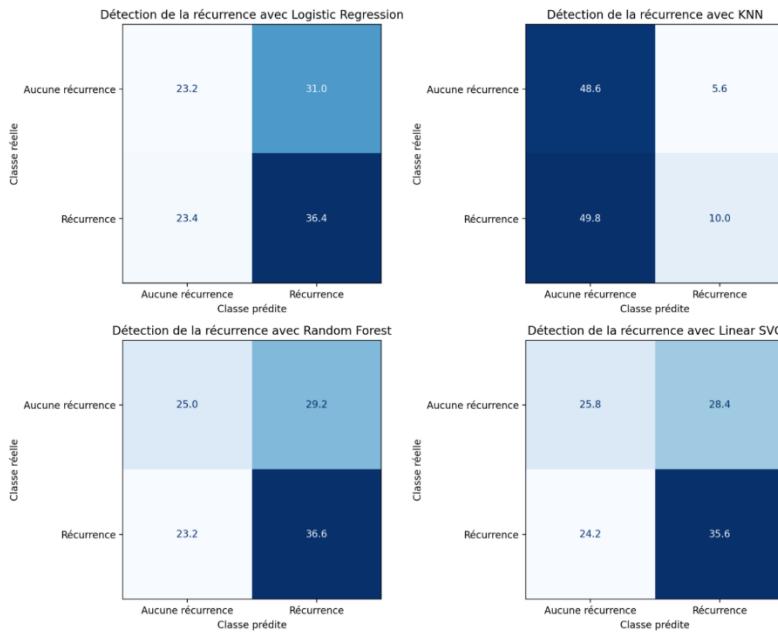


Figure 4.7 Matrices de confusion moyennant cinq K-Folds et utilisant les 13 110 gènes du TCGA pour l’identification de la récurrence du cancer

D’après les exactitudes balancées obtenues, les algorithmes d’apprentissage supervisé avec leurs hyperparamètres par défaut ne permettent pas, à partir de la base de données de patientes utilisée, de détecter efficacement si un cancer de l’ovaire est récurrent. Bien que l’utilisation de l’ensemble des 13 110 gènes comme variables d’apprentissage améliore légèrement les performances, les exactitudes balancées n’excèdent que de très peu celles attendues d’une prédiction aléatoire.

4.2 Modélisation métabolique

Cette section présente les résultats de la modélisation du métabolisme tumoral. Tout d’abord, elle expose la croissance tumorale ainsi que l’évolution temporelle des 35 métabolites impliqués dans le modèle métabolique et des ratios énergétiques. Ensuite, elle examine l’impact de la modification du volume sanguin *Vblood* sur la modélisation métabolique. Une comparaison entre la cinétique enzymatique définie dans la méthodologie et d’autres variantes explorées au cours du

projet est également illustrée. Enfin, une limite importante liée au temps de simulation des équations différentielles est discutée.

4.2.1 Croissance tumorale et fluctuations métaboliques selon le stade du cancer

La simulation du modèle métabolique a été réalisée sur une période de 270 jours pour une patiente représentative de chaque stade du cancer de l'ovaire. À chaque stade, les niveaux d'expression génique utilisés proviennent de la première patiente sans récurrence de la base de données correspondant à ce stade. Ce choix de patientes étant arbitraire, d'autres sélections auraient entraîné de légères variations métaboliques, reflétant l'unicité des profils transcriptomiques de chaque patiente. Un choix devait néanmoins être fait quant à la sélection des patientes représentatives de chaque stade du cancer, et prendre la première patiente sans récurrence à chaque stade a permis d'assurer un critère de sélection clair et objectif. La modélisation permet de suivre l'évolution de la croissance tumorale, et de la concentration en métabolites et en ratios énergétiques au fil du temps. Les volumes tumoraux initiaux sont les seules constantes ajustées en fonction du stade, assurant ainsi une comparaison cohérente entre les différentes simulations. Les Figures 4.8 à 4.12 illustrent les dynamiques métaboliques simulées pour ces patientes.

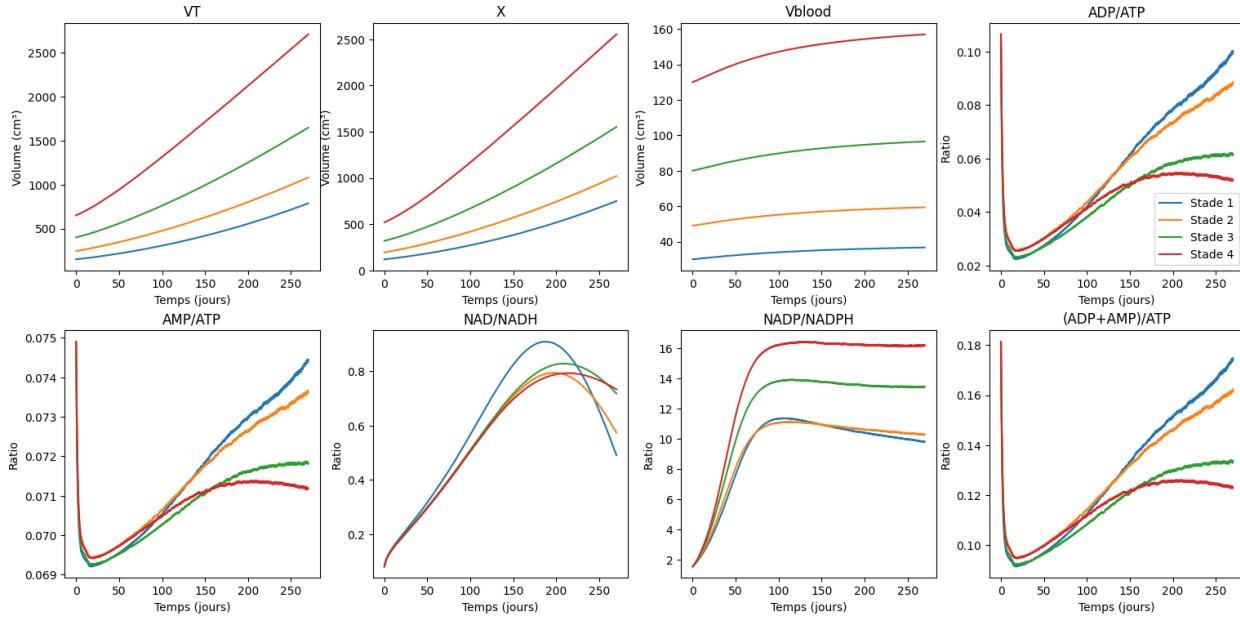


Figure 4.8 Volumes tumoraux et ratios énergétiques selon le stade du cancer de l'ovaire

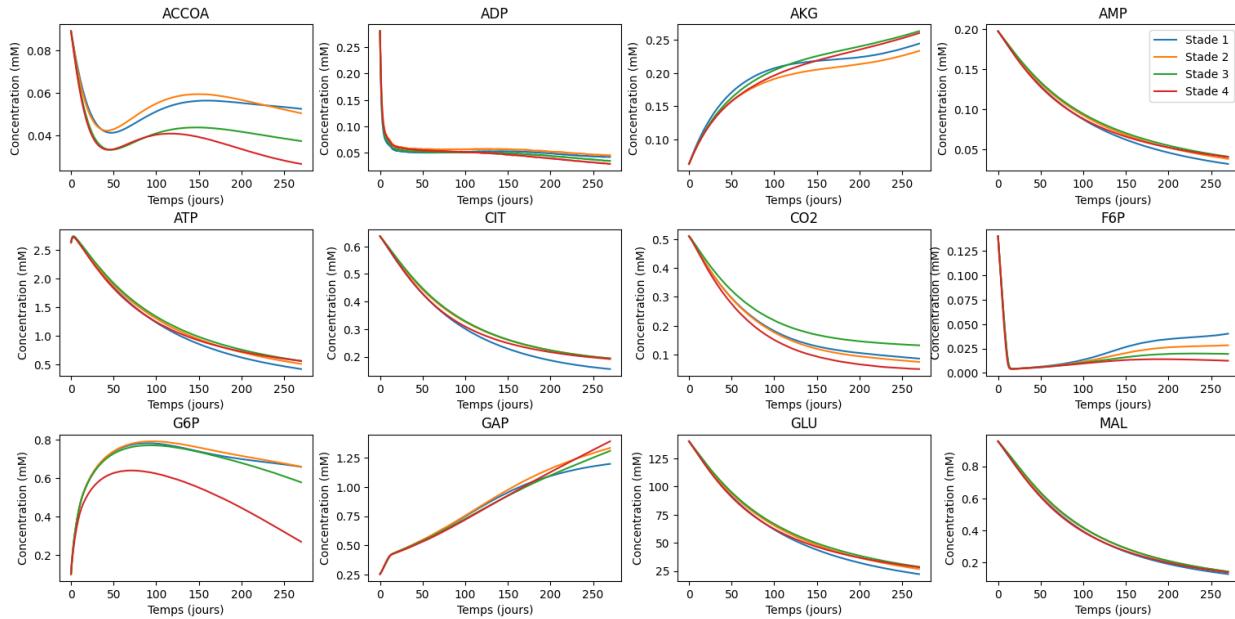


Figure 4.9 Première série de métabolites selon le stade du cancer de l'ovaire

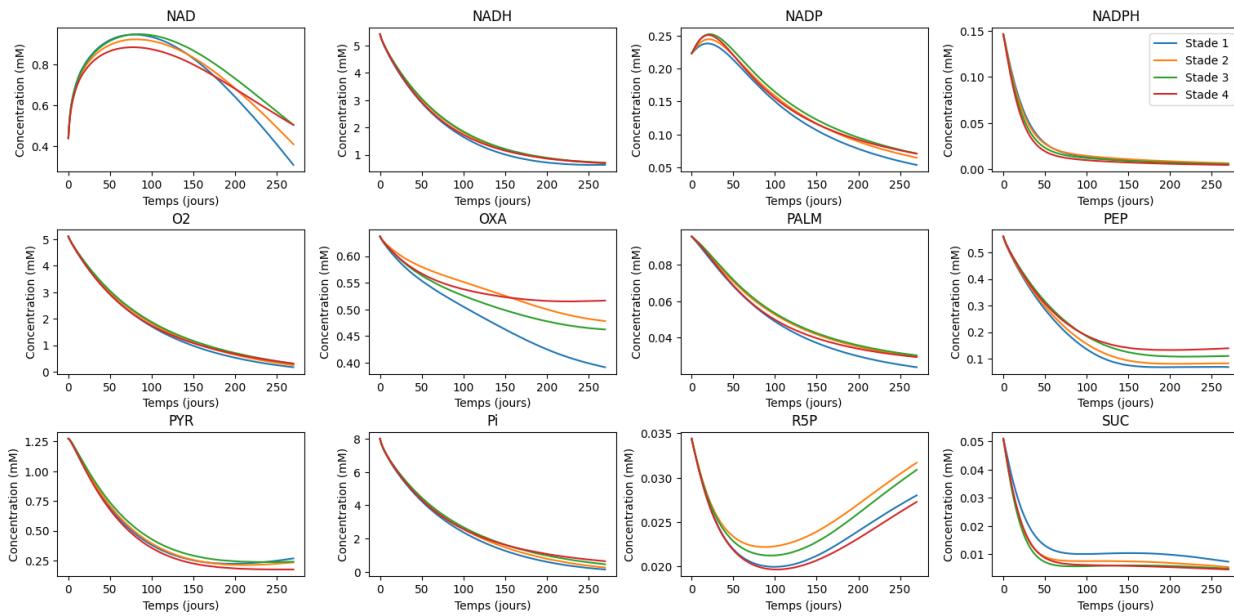


Figure 4.10 Deuxième série de métabolites selon le stade du cancer de l'ovaire

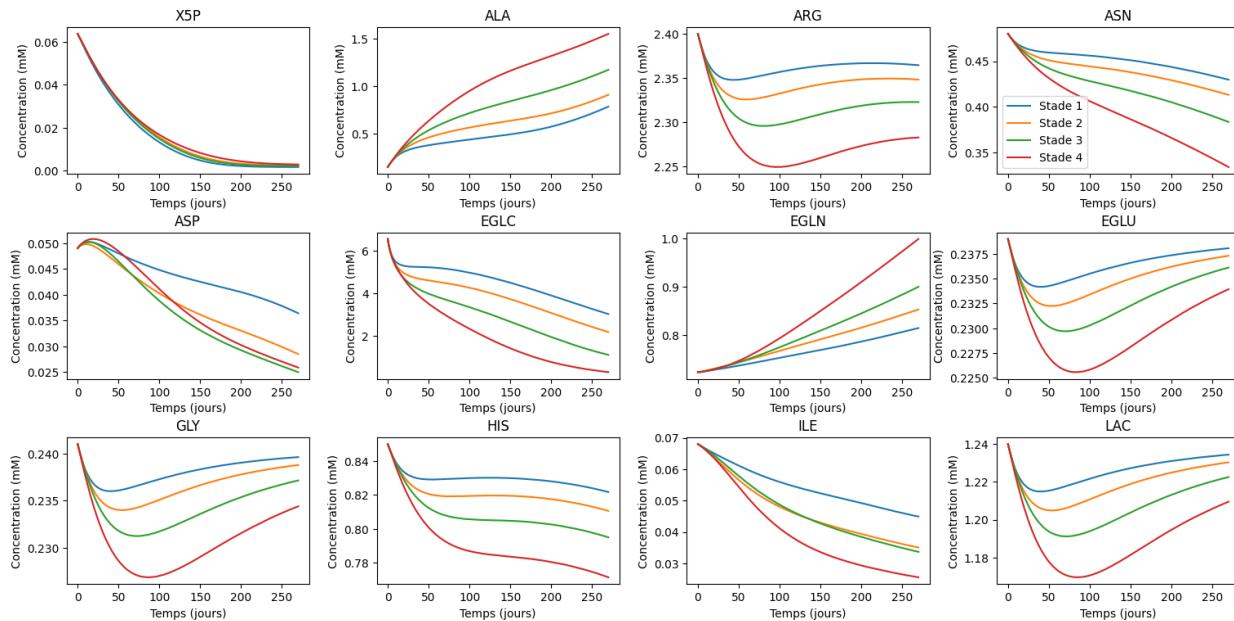


Figure 4.11 Troisième série de métabolites selon le stade du cancer de l'ovaire

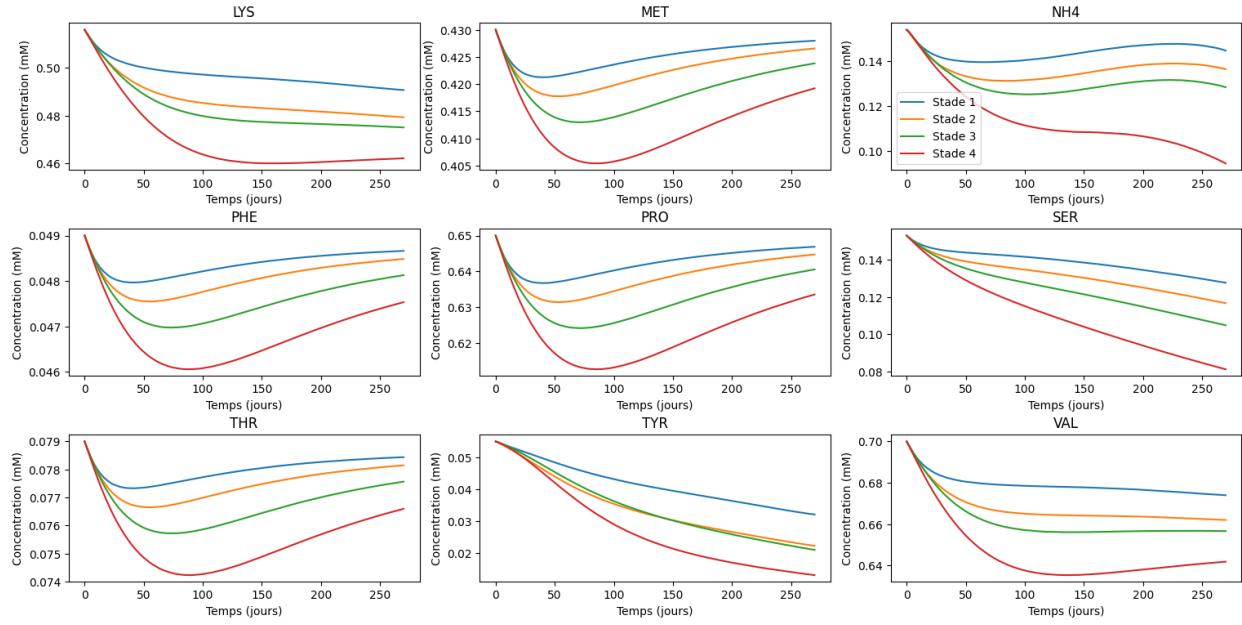


Figure 4.12 Quatrième série de métabolites selon le stade du cancer de l'ovaire

Les volumes tumoraux, à savoir le volume de la masse de cellules cancéreuses (X), le volume de sang (V_{blood}) et le volume tumoral total (VT), ainsi que les concentrations en métabolites extracellulaires, sont les variables qui varient le plus en fonction du stade de développement du cancer. En revanche, les concentrations des métabolites intracellulaires ont tendance à diminuer progressivement et semblent s'approcher de valeurs nulles. Toutefois, un grand nombre de métabolites extracellulaires ne présentent que de faibles fluctuations par rapport à leur concentration initiale. Cela pourrait s'expliquer par un flux sanguin relativement élevé, permettant un apport constant en nutriments aux cellules tumorales et limitant ainsi les variations métaboliques extracellulaires.

4.2.2 Effet de la modification du volume de sang V_{blood}

Le volume sanguin V_{blood} a été modélisé en fonction du taux de croissance des capillaires sanguins μ_{blood} (équation 4.1), lui-même dépendant de X , VT et V_{growth} (Tableau A.7 de l'annexe A).

$$\frac{dV_{blood,original}}{dt} = u_{blood} * V_{blood} \quad (4.1)$$

Cette approche visait à modéliser une augmentation de l'apport sanguin qui devient de moins en moins importante à mesure que le volume du microenvironnement tumoral augmente. Toutefois, elle présente une limitation importante; V_{blood} tend à se stabiliser à une valeur proche de son état initial, ce qui implique qu'il ne croît pas suffisamment pour accompagner l'expansion tumorale. Ainsi, pour un stade donné, V_{blood} ne peut atteindre la condition initiale estimée pour le stade suivant de développement de la tumeur, ce qui entraîne une discontinuité lors de la transition d'un stade à l'autre. Pour remédier à cette discontinuité, nous avons évalué l'hypothèse de maintenir à une valeur constante la proportion volumique de V_{blood} à 20 % de VT , soit 25% de X , afin d'assurer une continuité de modélisation sur l'étalement des quatre phases de développement du cancer de l'ovaire (équation 4.2).

$$\frac{dV_{blood,constant}}{dt} = 0,25 * \frac{dX}{dt} \quad (4.2)$$

Les Figures 4.13, 4.14 et 4.15 montrent l'effet de la modification de V_{blood} pour une patiente au stade I.

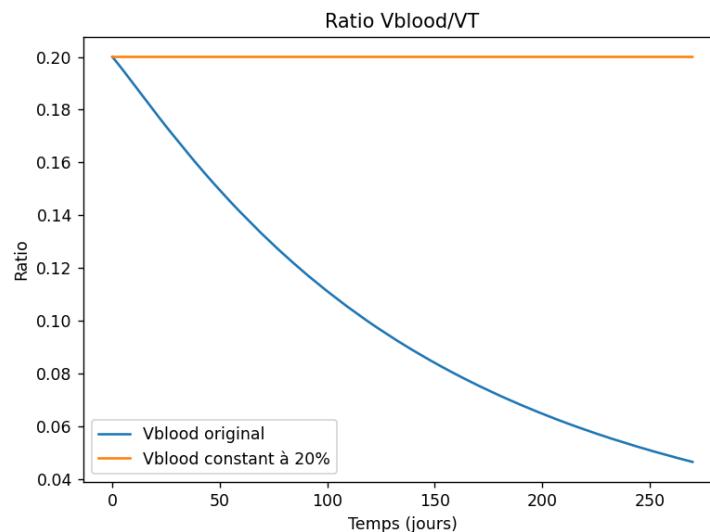


Figure 4.13 Effet de la modification de V_{blood} sur le ratio V_{blood}/VT

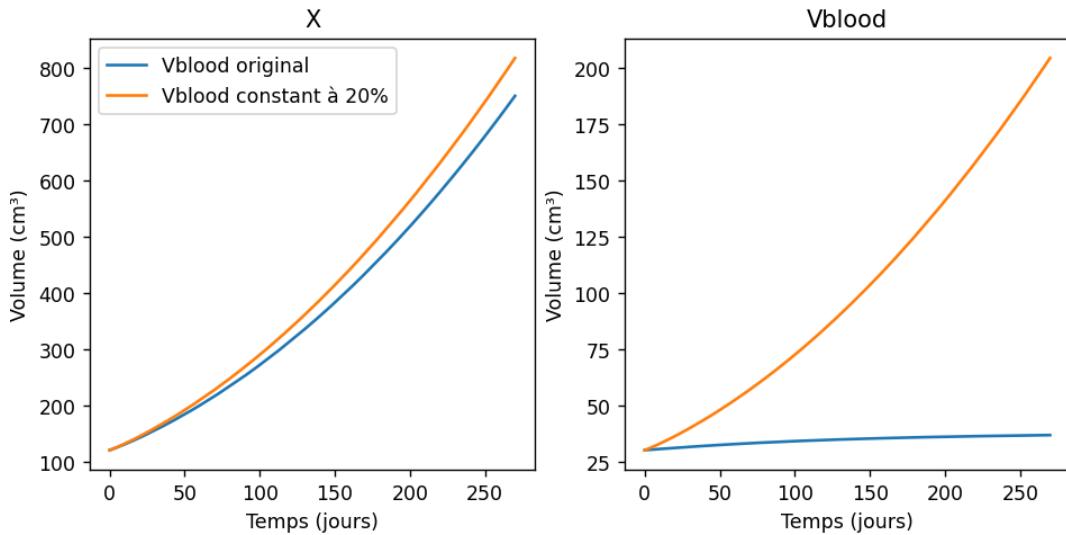


Figure 4.14 Effet de la modification de V_{blood} sur les volumes tumoraux

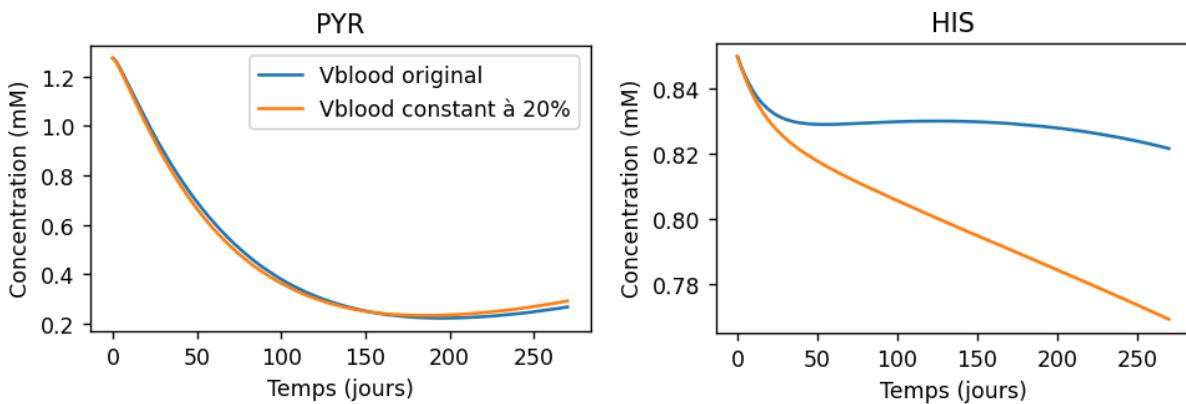


Figure 4.15 Effet de la modification de V_{blood} sur le métabolite intracellulaire PYR et le métabolite extracellulaire HIS

Maintenir V_{blood} constant à 20% du volume tumoral entraîne une augmentation significative de ce dernier et affecte légèrement le volume de la masse de cellules cancéreuses X . La concentration des métabolites intracellulaires, comme le pyruvate, reste globalement inchangée, tandis que celle des métabolites extracellulaires, tels que l'histidine, diminue plus rapidement et de manière plus marquée. Cette diminution s'explique par une dilution des métabolites extracellulaires dans un volume sanguin plus important. Toutefois, comme l'apport en métabolites est régulé par le flux sanguin F et non par V_{blood} , cette modification n'a pas d'impact majeur sur

les quantités de métabolites disponibles. En conséquence, le principal effet du changement de $Vblood$ se manifeste au niveau de VT , qui est défini comme la somme de $Vblood$ et de X .

4.2.3 Comparaison des cinétiques enzymatiques explorées

Durant le projet, plusieurs cinétiques enzymatiques ont été explorées afin d'évaluer leur impact sur la dynamique métabolique spécifique à une épigénétique particulière du cancer. Dans chacun des cas, les vitesses maximales de toutes les réactions du modèle métabolique, notées v_{max} , étaient des variables modifiées (équation 4.3). Ces différentes formulations permettent d'explorer l'impact de la prise en charge des données transcriptomiques sur les flux métaboliques.

$$v_{max} = \begin{cases} cst * FC * f(t): v_{max,th} * FC * F(t) \\ cst: v_{max,th} \\ cst * FC: v_{max,th} * FC \\ cst * \sin(t): v_{max,th} * (1 + 0,2 * \sin(2 * \pi * DT * t)) \end{cases} \quad (4.3)$$

- La première équation, notée $v_{max} = cst * FC * F(t)$, est la cinétique enzymatique proposée dans ce projet, où la vitesse maximale théorique d'une réaction ($v_{max,th}$), telle estimée par da Veiga Moreira *et al.* [25], est multipliée par un coefficient FC et une fonction temporelle $F(t)$.
- La deuxième équation, notée $v_{max,growth} = cst$, est une cinétique enzymatique caractérisée seulement par la vitesse maximale théorique des réactions. Ainsi, cette cinétique enzymatique ne tient pas compte des données transcriptomiques de la patiente.
- La troisième équation, notée $v_{max} = cst * FC$, est une cinétique enzymatique définie par une vitesse maximale de réaction selon le produit entre sa vitesse maximale théorique (i.e. une constante) et son coefficient FC correspondant.
- La quatrième équation, notée $v_{max} = cst * \sin(t)$, vise à intégrer une donnée transcriptomique de la patiente comme étant la fréquence d'une sinusoïde, variant entre 0,8 et 1,2, afin de simuler une variabilité périodique dans l'activité enzymatique. Cette approche permet d'introduire des fluctuations naturelles dans la vitesse maximale d'une réaction, reflétant les variations biologiques de l'expression des enzymes au fil du temps. Le coefficient de 0,2 ajuste ces oscillations dans un intervalle réaliste de modulation de l'activité enzymatique.

Les Figures 4.16 à 4.20 illustrent l'impact de ces différentes cinétiques enzymatiques sur la dynamique métabolique, permettant d'évaluer leur pertinence pour représenter le métabolisme tumoral.

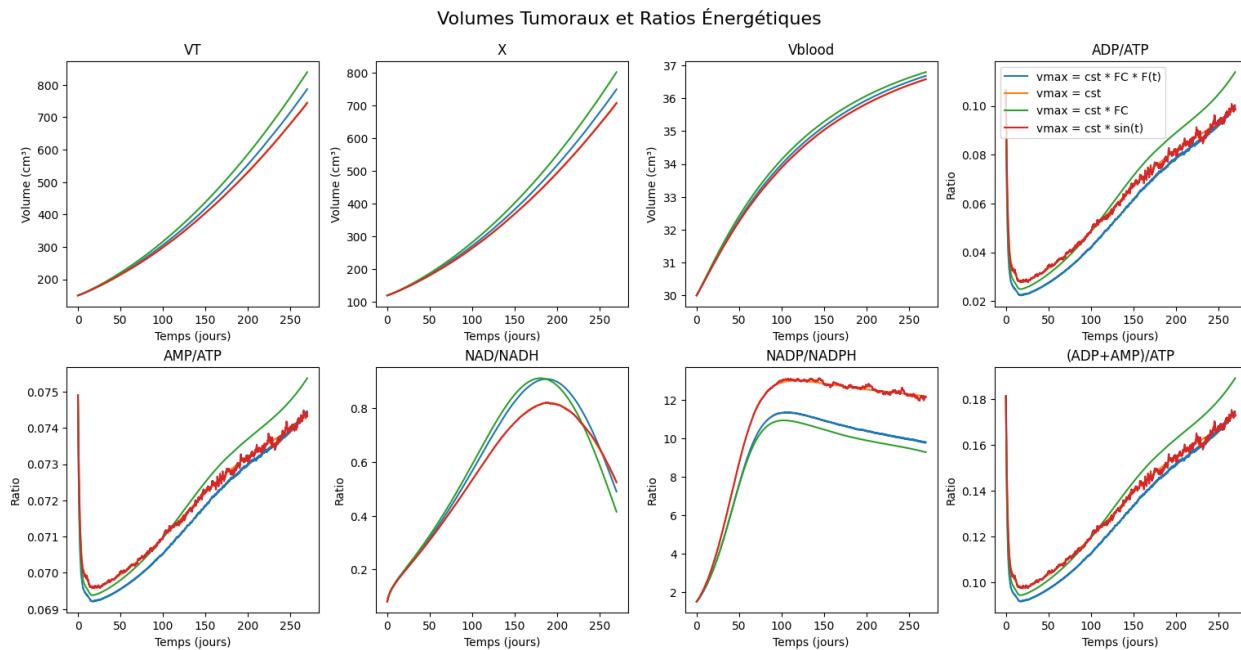


Figure 4.16 Volumes tumoraux et ratios énergétiques selon différentes cinétiques enzymatiques

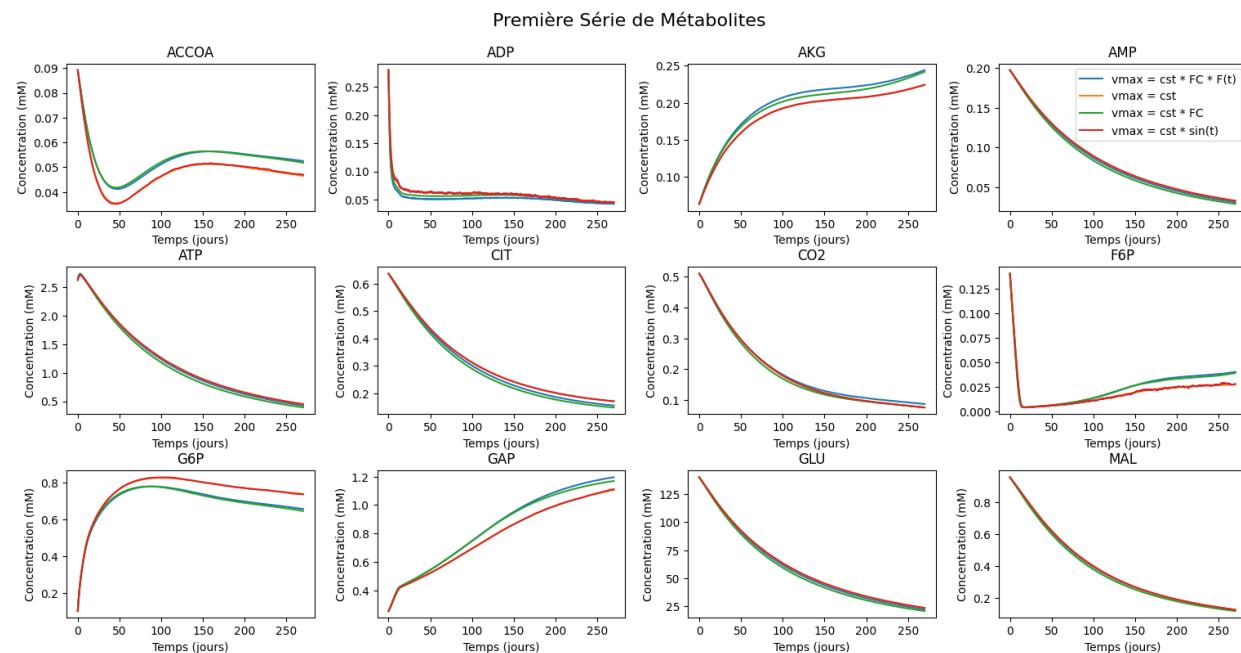


Figure 4.17 Première série de métabolites selon différentes cinétiques enzymatiques

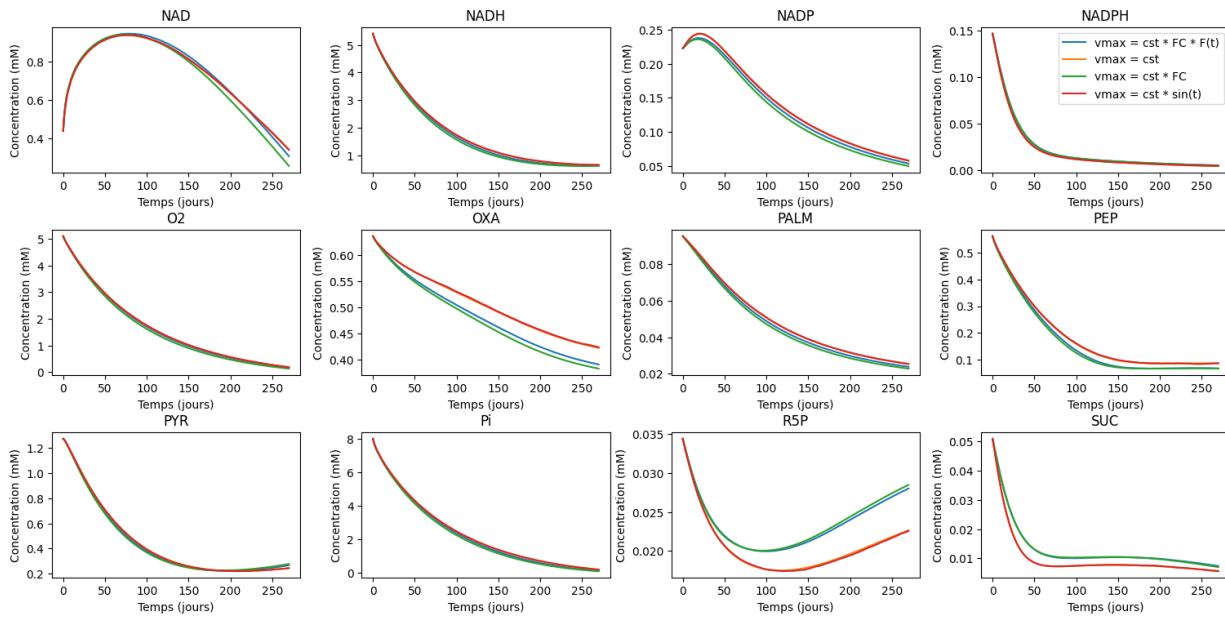


Figure 4.18 Deuxième série de métabolites selon différentes cinétiques enzymatiques

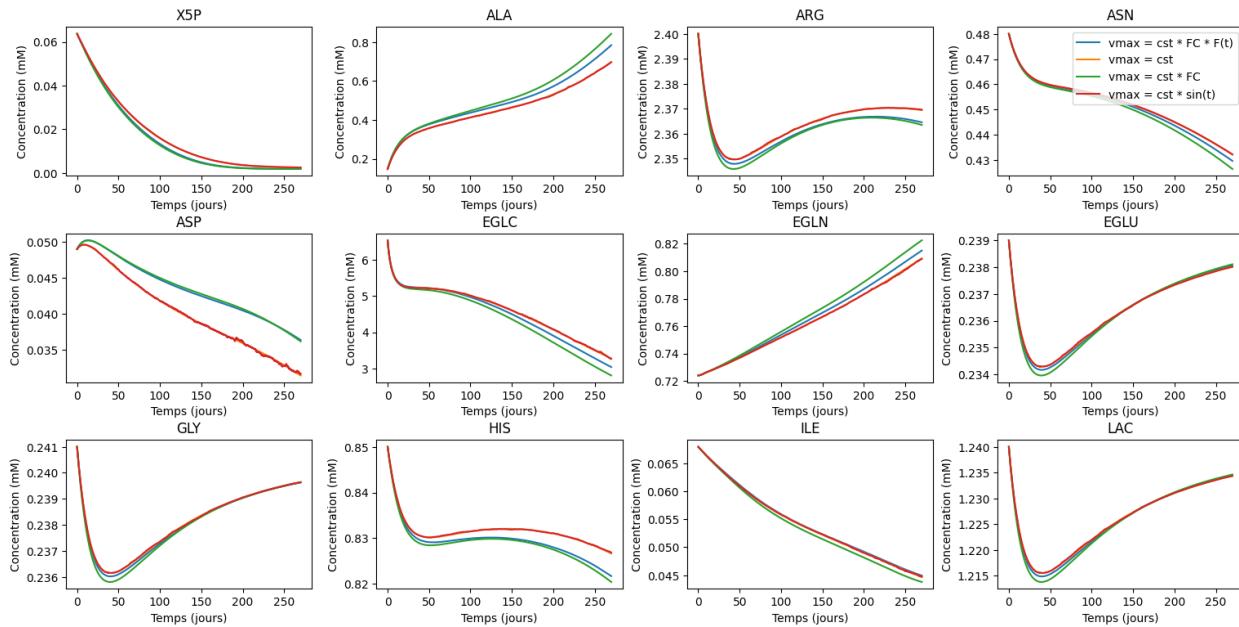


Figure 4.19 Troisième série de métabolites selon différentes cinétiques enzymatiques

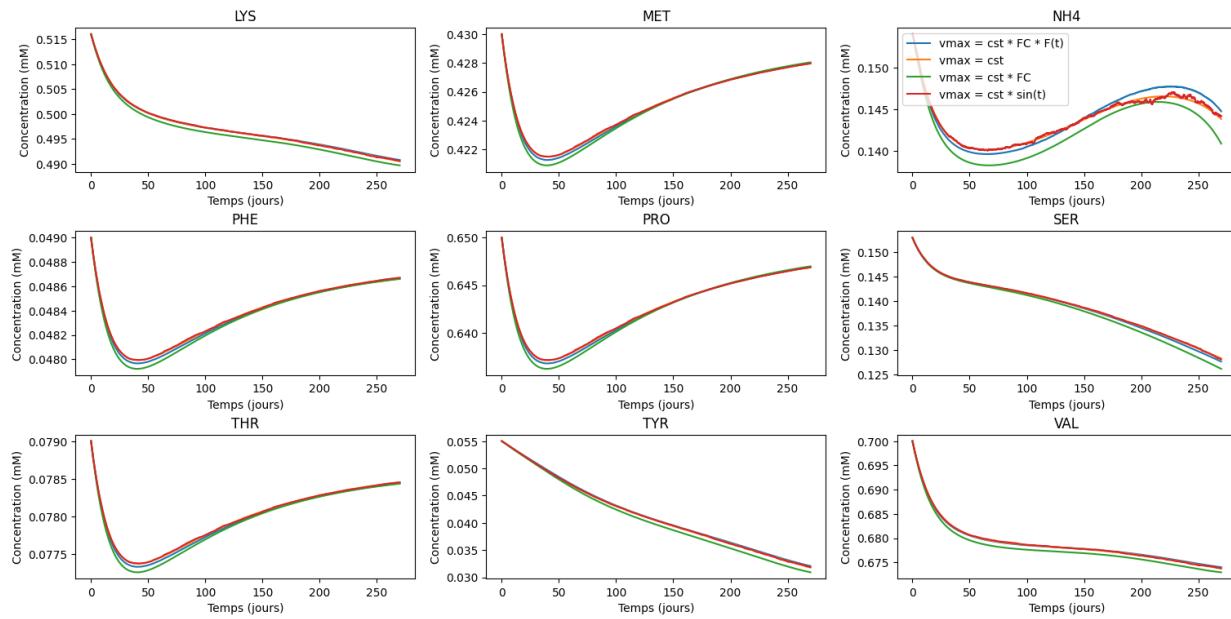


Figure 4.20 Quatrième série de métabolites selon différentes cinétiques enzymatiques

Les variations métaboliques entre les quatre cinétiques enzymatiques ne diffèrent que très peu, voire de manière non significative. En particulier, la cinétique $v_{max} = cst * \sin(t)$ est presque parfaitement superposée à la cinétique $v_{max} = cst$. De plus, les cinétiques $v_{max} = cst * FC * F(t)$ et $v_{max} = cst * FC$ suivent une tendance similaire au niveau métabolique et entraînent une légère augmentation de la croissance tumorale par rapport aux deux autres cinétiques.

4.2.4 Coefficients *Fold Change* (FC)

Les coefficients FC, modulant la vitesse maximale de chaque réaction du modèle métabolique dans la cinétique enzymatique proposée, ont été calculés pour chaque patiente sélectionnée pour représenter les quatre stades de développement du cancer. Le Tableau 4.8 présente les valeurs extrêmes ainsi que la moyenne et l'écart-type des coefficients FC.

Tableau 4.8 Analyse des coefficients FC

Stade de la patiente	Coefficient FC minimal (réaction)	Coefficient FC maximal (réaction)	Moyenne des coefficients FC	Écart-type des coefficients FC
I	0,684 (VME)	1,83 (VNADPHox)	1,01	0,190

II	0,835 (VAK)	1,65 (VNADPHox)	1,05	0,158
III	0,865 (VAAtoSUC)	1,62 (VNADPHox)	1,04	0,127
IV	0,756 (VAK)	1,74 (VNADPHox)	1,03	0,158

Les coefficients FC présentent une moyenne légèrement supérieure à 1, ce qui signifie qu'en moyenne, les vitesses de réaction sont à peine augmentées. Cependant, les coefficients maximaux sont élevés, et sont tous associés à la réaction *VNADPHox*. Les coefficients minimaux, en revanche, restent plus proches de 1 que les coefficients maximaux, la plus faible réduction de la vitesse de réaction étant observée pour la réaction *VME* chez la patiente au stade I. Enfin, les écart-types, d'environ 0,15, indiquent que, bien que la moyenne des coefficients FC soit proche de 1, il existe quand même une variabilité significative autour de cette moyenne.

4.2.5 Limitation sur la durée de la simulation métabolique

En traçant l'évolution temporelle du volume tumoral ainsi que de la concentration des métabolites sur une période de 1095 jours, soit trois ans, avec la cinétique enzymatique proposée et les données d'une patiente au stade I, une limitation apparaît quant à la durée de simulation possible. En effet, le volume tumoral présente plusieurs points d'inflexion brusques, engendrés en raison de discontinuités durant l'intégration des équations différentielles et attribuables à l'épuisement d'un ou de plusieurs métabolites (Figure 4.21).

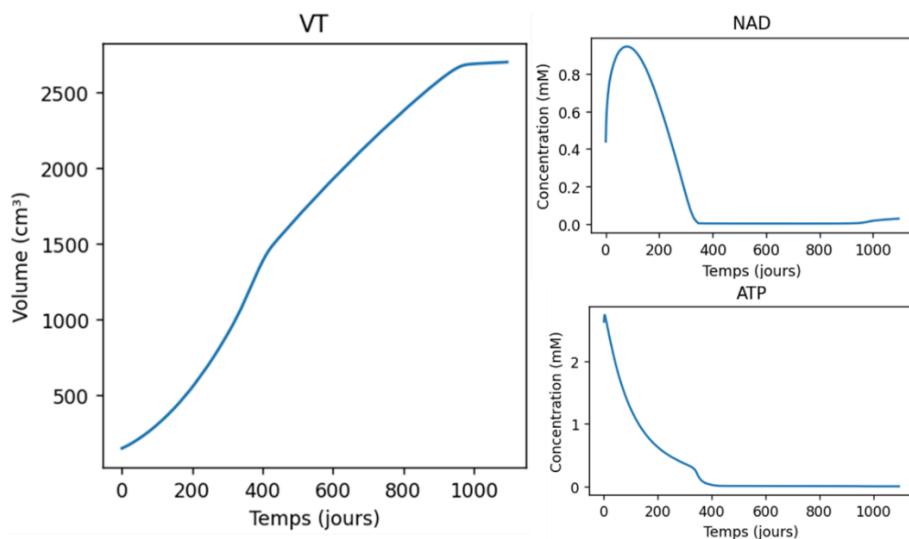


Figure 4.21 Volume tumoral total et concentrations en cofacteurs NAD et ATP sur 1095 jours

Proche du 400^e jour, les concentrations des cofacteurs enzymatiques NAD et ATP chutent à zéro, et un autre point d'infexion de VT observé autour du 1000^e jour suggère qu'au moins un métabolite essentiel est épuisé, limitant ainsi la croissance tumorale (Figure 4.22). Ce métabolite serait le glucose 6-phosphate (G6P), impliqué dans la réaction de croissance tumorale *Vgrowth*, en plus du glycéraldéhyde 3-phosphate (GAP). Leur disparition expliquerait l'arrêt de la croissance tumorale à ce moment précis. Cependant, ces résultats ne sont pas entièrement biologiquement réalisables, car un métabolite ne disparaît généralement pas complètement du métabolisme tumoral. Ainsi, les cellules tumorales sont flexibles et peuvent s'adapter à des conditions de stress, comme la limitation des nutriments, en activant des mécanismes compensatoires. Elles peuvent alterner entre différentes voies métaboliques ou utiliser des processus comme l'autophagie ou la glycolyse aérobie pour maintenir leur croissance même dans des environnements défavorables.

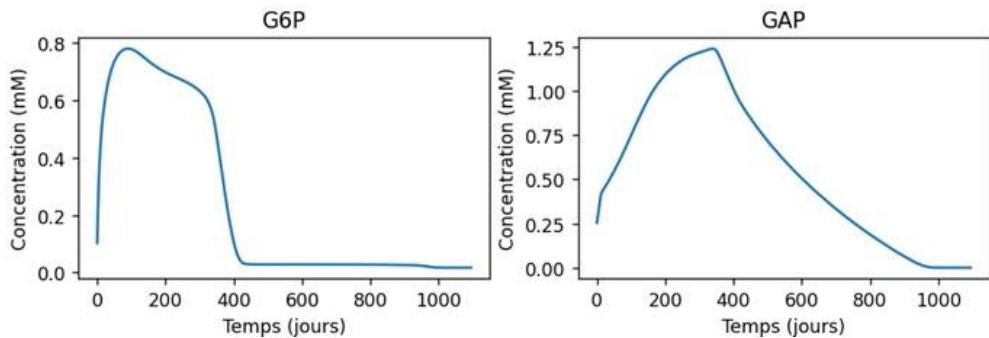


Figure 4.22 Concentrations des métabolites intracellulaires G6P et GAP sur 1095 jours

4.3 Ciblage thérapeutique

Cette section détaille les résultats des algorithmes de recherche de solutions *Depth Search* et *Intelligent Search*. Grâce à ces algorithmes, des gènes cibles sont identifiés comme potentielles cibles thérapeutiques en modulant leurs niveaux d'expression génique.

4.3.1 Suppression et activation d'un seul gène

L'algorithme de recherche en profondeur *Depth Search* a été lancé séparément avec les données de quatre patientes de la base de données (représentant les quatre stades du cancer ovarien) sur un ensemble de 75 gènes du modèle métabolique, excluant ceux associés aux réactions *Vgrowth*

et V_{resp} . Les paramètres utilisés étaient les suivants: un gène par recherche de solution avec un facteur de modulation à 0 (suppression complète). Pour ce test, les mêmes quatre patientes que celles utilisées pour la modélisation métabolique ont été sélectionnées afin de représenter les quatre stades du cancer. Tout comme pour la modélisation métabolique, un choix différent de patientes aurait influencé les résultats, en altérant principalement les pourcentages de réduction de la croissance tumorale, plutôt que les gènes identifiés comme cibles potentielles.

Le gène *glutamic-pyruvic transaminase* (GPT), impliqué dans la réaction réversible *VALaTA*, ressort comme meilleure solution pour les quatre stades selon le pourcentage de réduction moyenne de la croissance tumorale sur 270 jours. Toutefois, l'efficacité de la suppression de GPT diminue significativement avec le stade. La Figure 4.23 illustre la réduction de la croissance tumorale engendrée par la suppression de GPT pour une patiente au stade IV; la suppression de GPT entraîne même une croissance tumorale plus rapide à mesure que le volume tumoral augmente. En effet, sur la Figure 4.23, la courbe orange, représentant l'évolution du volume tumoral en supprimant complètement le gène GPT, s'écarte de la courbe bleue, représentant la croissance tumorale sans modulation génique, lorsque le volume tumoral est proche de sa valeur initiale, puis semble la rejoindre lorsque le volume tumoral est considérablement plus élevé.

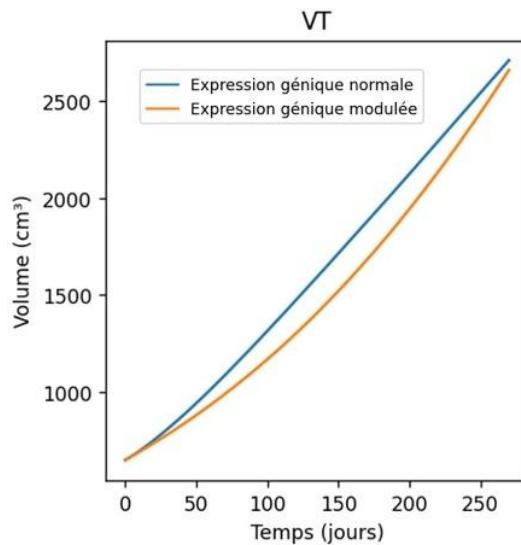


Figure 4.23 Effet de la suppression du gène GPT pour une patiente au stade IV

La suppression des gènes *glucose-6-phosphate dehydrogenase* (G6PD), *phosphogluconolactonase* (PGLS), *phosphogluconate dehydrogenase* (PGD) et *ribose-5-*

phosphate isomerase A (RPIA), tous impliqués dans la réaction *VG6PDH*, montrent également de bons résultats. Étant donné qu'ils sont reliés par l'opération « min » selon les règles GPR, la suppression complète d'un seul de ces gènes suffit à inhiber totalement la réaction *VG6PDH*, ce qui explique les pourcentages de réduction identiques pour chacun de ces gènes (Tableau 4.9).

Tableau 4.9 Meilleures cibles thérapeutiques par la suppression totale ($f_m = 0$) d'un seul gène

Stade de la patiente	Gène	Facteur de modulation	Réduction moyenne de la croissance tumorale (%)	Réduction au jour 270 de la croissance tumorale (%)
I	GPT	0	18,8	20,1
	G6PD	0	4,97	15,0
	PGLS	0	4,97	15,0
	PGD	0	4,97	15,0
	RPIA	0	4,97	15,0
	FASN	0	3,17	8,23
II	GPT	0	16,3	10,6
	G6PD	0	4,85	15,9
	PGLS	0	4,85	15,9
	PGD	0	4,85	15,9
	RPIA	0	4,85	15,9
	HK1	0	3,62	18,1
III	GPT	0	16,9	11,7
	G6PD	0	4,56	14,7
	PGLS	0	4,56	14,7
	PGD	0	4,56	14,7
	RPIA	0	4,56	14,7
	HK1	0	2,88	15,9
IV	GPT	0	11,1	2,44
	G6PD	0	4,18	12,9
	PGLS	0	4,18	12,9
	PGD	0	4,18	12,9
	RPIA	0	4,18	12,9
	HK1	0	3,29	16,3

En ayant lancé le *Depth Search* avec les mêmes paramètres de recherche, à l'exception du facteur de modulation, qui est fixé à 2 pour simuler l'activation d'un gène (doublant ainsi son niveau d'expression), ce sont les gènes *glutamic-oxaloacetic transaminase 1* (GOT1) et *glutamic-oxaloacetic transaminase 2* (GOT2) impliqués dans la réaction *VASTA*, ainsi que les gènes *branched chain amino acid transaminase 1* (BCAT1) et *branched chain amino acid transaminase 2* (BCAT2), impliqués dans la réaction *VAAtoSUC*, qui réduisent le plus la croissance tumorale

lorsqu'ils sont activés. Étonnamment, l'activation du gène GPT réduit également la croissance tumorale, notamment pour un cancer ovarien avancé (Tableau 4.10).

Tableau 4.10 Meilleures cibles thérapeutiques par l'activation ($f_m = 2$) d'un seul gène

Stade de la patiente	Gène	Facteur de modulation	Réduction moyenne de la croissance tumorale (%)	Réduction au jour 270 de la croissance tumorale (%)
I	GOT2	2	7,41	13,7
	GOT1	2	5,95	11,3
	BCAT1	2	4,38	8,02
	BCAT2	2	3,01	5,63
	SDHB	2	1,41	-0,06
	NOX4	2	1,29	1,30
II	GOT2	2	10,1	19,1
	BCAT1	2	6,74	13,8
	GOT1	2	5,67	11,9
	BCAT2	2	2,40	4,66
	GPT	2	1,72	8,69
	PFKL	2	1,10	-0,25
III	GOT2	2	12,0	23,3
	GOT1	2	10,9	20,8
	BCAT2	2	7,61	15,2
	BCAT1	2	3,00	6,45
	GPT	2	2,05	10,2
	SDS	2	0,95	0,82
IV	GOT2	2	13,8	24,1
	BCAT2	2	11,5	21,8
	GOT1	2	10,9	20,1
	GPT	2	5,03	13,2
	BCAT1	2	1,34	2,82
	ATP1A1	2	0,77	0,04

Contrairement à la suppression de GPT qui génère de meilleurs résultats à un stade précoce, l'activation des gènes GOT1, GOT2, BCAT1 et BCAT2 semble réduire davantage la croissance tumorale à des stades avancés (Figure 4.24). En effet, plus le volume tumoral croît, plus la courbe orange, représentant l'évolution du volume tumoral en doublant l'expression génique du gène GOT2, s'écarte de la courbe bleue, représentant la croissance tumorale sans modulation génique.

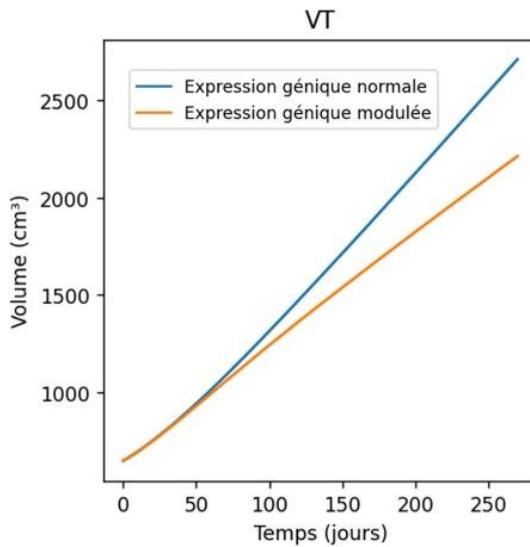


Figure 4.24 Effet de l'activation du gène GOT2 pour une patiente au stade IV

4.3.2 Suppression et activation d'une paire de gènes

L'algorithme de recherche en profondeur (*Depth Search*) a été lancé avec les données du premier échantillon de la base de données, sur l'ensemble des 75 gènes du modèle métabolique (excluant ceux associés aux réactions *Vgrowth* et *Vresp*) et selon les paramètres suivants: deux gènes au maximum par solution et deux facteurs de modulation distincts ($f_m = 0$ et $f_m = 2$). Au total, 11 250 solutions possibles ont été testées, et les trois meilleures solutions pour chaque patiente, sans répétition de gènes, sont représentées au Tableau 4.11.

Tableau 4.11 Meilleures cibles thérapeutiques par la suppression ($f_m = 0$) et/ou l'activation ($f_m = 2$) d'une paire de gènes

Stade de la patiente	Gène(s)	Facteur(s) de modulation	Réduction moyenne de la croissance tumorale (%)	Réduction au jour 270 de la croissance tumorale (%)
I	GPT; FASN	0; 0	21,7	28,0
	PGLS; GOT2	0; 2	10,7	25,0
	PGD; BCAT1	0; 2	8,98	21,1
II	HK1, TKT	0; 0	29,9	82,8
	GLUD1; GPT	2; 0	17,1	10,8
	G6PD; GOT2	0; 2	13,7	29,8

	HK1; TKT	0; 0	35,1	81,5
III	GPT; GOT2	0; 2	19,0	15,8
	GOT1; BCAT2	2; 2	16,3	30,6
	GOT2; BCAT2	2; 2	21,3	38,4
IV	PGLS; GOT1	0; 2	13,8	28,6
	G6PD; TKT	0; 2	6,82	22,6

Pour la patiente au stade I, la suppression simultanée des gènes GPT et *fatty acid synthase* (FASN), FASN étant impliquée dans la réaction *VPALM*, résulte en une réduction moyenne de 21,7% de la croissance tumorale. Cette combinaison de gènes offre une réduction légèrement inférieure à la somme des réductions individuelles de GPT et de FASN, qui s'élève à 22,0% (voir Tableau 4.9). Pour les patientes aux stades II et III, la suppression simultanée des gènes *hexokinase 1* (HK1), impliqué dans la réaction *VHK*, et *transketolase* (TKT), impliqué dans la réaction *VTK*, entraîne une réduction de la croissance tumorale de plus de 80% au jour 270. En supprimant complètement HK1 et TKT, leurs réactions associées sont également inhibées, ce qui conduit à un épuisement complet du métabolite G6P, essentiel à la croissance tumorale via son rôle dans la réaction *Vgrowth*. En effet, seulement deux réactions produisent le G6P: la réaction *VHK* et la réaction *VPGI*. D'une part, la réaction *VHK* est complètement inhibée par la suppression de HK1. D'autre part, la suppression de TKT inhibe la réaction *VTK*, soit la seule réaction, à l'exception de *VPGI*, produisant du fructose 6-phosphate (F6P). Par conséquent, les réserves de F6P ne sont pas renouvelées dans le temps, en plus du fait que la réaction réversible *VPGI* ne produit plus de G6P (Tableau 4.12).

Tableau 4.12 Réactions reliées à la production de G6P

Abréviation	Réaction	Gène(s) associé(s) et règles GPR
VHK	EGLC + ATP → G6P + ADP	HK1
VPGI	G6P ↔ F6P	GPI
VTK	R5P + 2 X5P → 2 F6P + GAP	TKT

Pour les patientes aux stades II et III, un point d'inflexion brusque apparaît au niveau de la croissance tumorale proche du 70^e jour, invalidant l'analyse temporelle subséquente des estimés du volume tumoral et des concentrations en F6P et G6P. La Figure 4.25 présente le cas de la patiente au stade II.

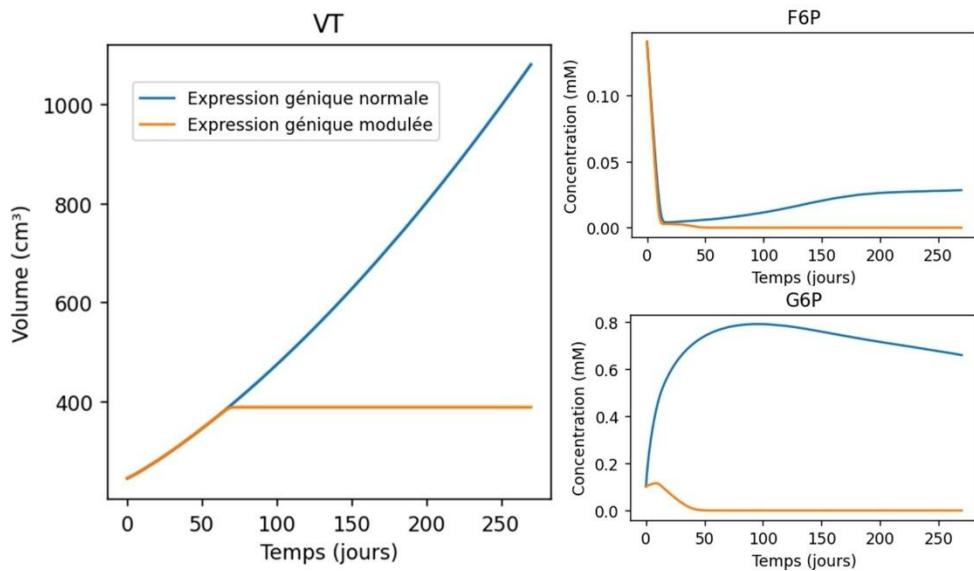


Figure 4.25 Effet de la suppression simultanée des gènes HK1 et TKT pour une patiente au stade II

Enfin, pour la patiente au stade IV, l’activation simultanée des gènes GOT2 et BCAT2 résulte en une réduction moyenne de 21,3% de la croissance tumorale, soit une réduction, encore une fois, inférieure à la somme de 25,% des réductions individuelles, soit en activant séparément GOT2 et BCAT2, obtenues au Tableau 4.10.

4.3.3 Recherche intelligente

L’algorithme de recherche intelligente (*Intelligent Search*) a été lancé sur 400 itérations avec les données du premier échantillon de la base de données sur l’ensemble des 75 gènes du modèle métabolique excluant ceux associés aux réactions *Vgrowth* et *Vresp* et selon les paramètres suivants : trois gènes au maximum par solution et quatre facteurs de modulation différents ($f_m = \{0, 0.5, 1.5, 2\}$). Environ 4 millions de solutions différentes existent avec ces paramètres de recherche. L’algorithme a été en mesure de trouver des combinaisons de gènes comportant les modulations ciblées par les essais avec *Depth Search*. Entre autres, la suppression de GPT et l’activation de GOT1 et GOT2 figurent parmi dans les quatre meilleurs résultats, sans répétition de gène, pour chacune des quatre patientes (Tableau 4.13).

Tableau 4.13 Meilleures cibles thérapeutiques identifiées par l'algorithme *Intelligent Search* en considérant l'activation et la suppression d'un triplet de gènes

Stade de la patiente	Gène(s)	Facteur(s) de modulation	Réduction moyenne de la croissance tumorale (%)	Réduction au jour 270 de la croissance tumorale (%)
I	GPT; PGM1; FASN	0; 2; 0	22,7	27,9
	ACO1; GOT2	0; 2	7,41	13,7
	GOT1	2	5,95	11,3
	BCAT1; PFKP	2; 0	5,01	8,27
II	GPT; PGLS; MT-ND5	0; 0; 0,5	18,2	23,6
	GOT2; PGK1; ME2	2; 0,5; 1,5	10,3	18,3
	AK1; BCAT1	0; 2	6,74	13,8
	GOT1	2	5,67	11,9
III	GPT; GOT2; TKT	0; 2; 2	19,8	16,0
	BCAT2; IDH3G; GOT1	2; 0; 2	16,3	30,6
	RPIA; BCAT1	0; 1,5	4,15	15,1
	PGD	0	4,56	14,7
IV	GOT2; PGD; OGDH	2; 0; 0	17,4	33,6
	PGM3; GPT; HAL	2; 0; 0,5	15,4	2,38
	GOT1; PGLS; RPE	2; 0; 1,5	15,1	31,9
	RPIA; ADCY1	0; 2	4,35	12,9

Pour la patiente au stade I, le meilleur résultat trouvé par *Depth Search* en considérant un maximum de deux gènes par solution, était la suppression simultanée de GPT et FASN, entraînant une réduction moyenne de la croissance tumorale de 21,7%. L'*Intelligent Search*, en autorisant jusqu'à trois gènes par solution, a trouvé une solution incluant cette même suppression de GPT et de FASN, mais avec l'ajout de l'activation du gène *Phosphoglycerate mutase 1* (PGM1), ce qui améliore la réduction moyenne tumorale à 22,7%. Toutefois, les meilleures solutions obtenues par l'*Intelligent Search* pour les patientes au stade II, III et IV ne dépassent pas celles trouvées avec *Depth Search* en limitant les solutions à deux gènes.

CHAPITRE 5 DISCUSSION

La discussion vise à approfondir l'analyse des résultats présentés au chapitre précédent en les comparant avec ceux de la littérature, en abordant les limitations et les défis rencontrés et en soulevant les apports spécifiques de ce projet de recherche.

5.1 Apprentissage supervisé

L'apprentissage supervisé basé sur l'expression génique permet une détection très précise du cancer de l'ovaire. Avec un ensemble de 13 110 gènes, les algorithmes LR et KNN atteignent une exactitude balancée de 99,9%. Même en réduisant ce nombre à 112 gènes métaboliques, Linear SVC maintient une performance élevée avec 99,7% d'exactitude balancée. Dans tous ces cas, la spécificité est de 100%, garantissant l'absence de faux positifs. Ces résultats surpassent ceux obtenus avec des biomarqueurs moléculaires tels que la concentration en CA125, qui affiche une exactitude de 80% avec l'algorithme XGB selon Akazawa et Hashimoto [15] et de 97,4% avec l'algorithme RF selon Lu *et al.* [13]. De plus, les performances observées sont comparables à celles rapportées par Arfiani et Rustam, qui, en utilisant également l'expression génique comme variable d'apprentissage, ont obtenu une exactitude balancée de 98,2% avec RF et de 100% avec la méthode bagging [17]. En ce qui concerne la détection précoce, l'algorithme RF se distingue particulièrement en identifiant avec 97,5% d'exactitude balancée les patientes atteintes d'un cancer au stade I. Ces résultats soulignent ainsi l'intérêt de l'apprentissage supervisé basé sur l'expression génique comme outil complémentaire aux méthodes de diagnostic actuelles pour la détection précoce du cancer de l'ovaire. Toutefois, la principale limitation liée à la détection du cancer de l'ovaire dans ce projet réside dans le déséquilibre marqué, dans la base de données utilisée, entre le nombre de patientes saines et celui des patientes atteintes de cancer. En effet, la base de données utilisée comprend seulement huit patientes saines contre 570 patientes cancéreuses. Bien que les meilleurs algorithmes aient atteint une spécificité de 100% en classant correctement toutes les patientes saines, la taille réduite de cet échantillon limite la fiabilité généralisée des résultats obtenus. Il existe également un risque que la spécificité élevée mesurée provienne d'un surajustement, lors de l'entraînement des algorithmes d'apprentissage supervisé, aux caractéristiques spécifiques de ces huit patientes, plutôt que de la véritable capacité à distinguer efficacement les patientes saines des patientes cancéreuses. Néanmoins, la base de données utilisée

présente, en plus de sa disponibilité, un atout majeur: sa fiabilité. Provenant du TCGA, elle est supervisée par des agences reconnues telles que le NCI et le NHGRI, assurant des données de qualité et renforçant la robustesse des résultats obtenus et leur reproductibilité.

En n'utilisant que des niveaux d'expression génique comme variables d'apprentissage, la classification du stade du cancer de l'ovaire par apprentissage supervisé reste un défi. L'algorithme LR atteint une exactitude balancée de 37,8%, soit une amélioration de 12,8% par rapport à un choix aléatoire, mais les performances globales demeurent limitées. Cette difficulté peut s'expliquer par le fait que la progression tumorale est un processus graduel, tandis que la classification en stades impose une certaine discréétisation. Par exemple, une patiente en stade III avancé peut être erronément classée en stade IV par un algorithme d'apprentissage supervisé, ce qui semble être une erreur significative alors que la différence biologique est minime. Ainsi, pour améliorer la classification du stade, il pourrait être pertinent d'intégrer d'autres types de données. Par exemple, en développant des réseaux de neurones récurrents (RNN) à partir d'images médicales et de données génétiques, Ghoniem *et al.* (2021) ont réussi à classifier le stade du cancer de l'ovaire avec une exactitude moyenne de 98,76% [61]. Il est intéressant de noter que les données génétiques utilisées ne se limitaient pas aux niveaux d'expression génique, mais incluaient également les variations du nombre de copies d'un gène, c'est-à-dire une quantification des altérations où certaines régions du génome sont dupliquées ou supprimées. Néanmoins, la détection du cancer demeure l'étape diagnostique la plus cruciale, tandis que la classification du stade, bien qu'utile pour estimer le volume tumoral nécessaire à la modélisation métabolique, reste secondaire en importance et peut être réalisée par la suite lors d'un examen clinique.

Pour la détection de la récurrence du cancer de l'ovaire, les algorithmes d'apprentissage supervisé montrent des performances limitées, à peine supérieures à une prédiction aléatoire. L'algorithme Linear SVC obtient la meilleure exactitude balancée, avec 53,6%. Ainsi, l'apprentissage supervisé basé uniquement sur l'expression génique ne suffit pas pour prédire la récurrence du cancer de l'ovaire, probablement en raison de l'absence de marqueurs transcriptomiques suffisamment spécifiques pour détecter si un cancer de l'ovaire est récurrent ou non.

Ensuite, l'utilisation de l'algorithme de suréchantillonnage SMOTE améliore généralement les performances des algorithmes d'apprentissage supervisé en présence de classes déséquilibrées.

Son hyperparamètre $k_neighbors$, qui détermine le nombre de voisins considérés lors de la génération d'échantillons synthétiques, n'a pas montré d'influence significative sur l'exactitude balancée dans plusieurs cas. En utilisant SMOTE avec $k_neighbors = 5$, soit la valeur par défaut de cet hyperparamètre lorsqu'il n'est pas spécifié, l'exactitude balancée de LR pour la détection du cancer passe de 92,9% à 99,8%, celle de KNN pour la classification du stade augmente de 26,1% à 34,9% et celle de RF pour la classification du cancer s'accroît de 61,5% jusqu'à 75,0%. L'application de SMOTE sur l'ensemble d'entraînement est donc fortement recommandée lorsque les classes sont déséquilibrées.

Par ailleurs, l'optimisation des hyperparamètres des algorithmes d'apprentissage supervisé joue un rôle essentiel dans l'amélioration de leurs performances, en particulier lorsque les exactitudes balancées initiales sont faibles. Par exemple, l'optimisation des hyperparamètres de Linear SVC pour la classification du stade a permis d'augmenter l'exactitude balancée de 26,4% à 35,6% lors de l'utilisation des niveaux d'expression des 112 gènes associés au modèle métabolique. Cette amélioration met en évidence l'importance d'un ajustement précis des hyperparamètres afin d'optimiser les performances des algorithmes d'apprentissage supervisé.

Enfin, le nombre de variables d'apprentissage influence également les performances des algorithmes. Par exemple, en utilisant 13 110 niveaux d'expression génique, la spécificité de RF pour la détection du cancer est passée de 50%, avec 112 gènes, à 87,5%. En revanche, pour Linear SVC, l'exactitude balancée a diminué de 99,7%, avec 112 gènes, à 92,5% en utilisant l'ensemble des 13 110 gènes. Toutefois, l'effet du nombre de variables d'apprentissage sur les prédictions pourrait être biaisé par les hyperparamètres des algorithmes d'apprentissage supervisé. En effet, les hyperparamètres ont été optimisés pour l'ensemble des 112 gènes associés au modèle métabolique et ont été appliqués tels quels à l'ensemble élargi de 13 110 gènes. Les résultats obtenus avec l'ensemble de 13 110 gènes pourraient donc ne pas refléter pleinement le potentiel d'utiliser un grand nombre de variables d'apprentissage. Ainsi, adapter l'optimisation des hyperparamètres à l'ensemble élargi de variables d'apprentissage pourrait améliorer encore davantage les performances des algorithmes. Néanmoins, l'ensemble réduit de 112 gènes associés au modèle métabolique offre déjà d'excellents résultats pour la détection du cancer, soulignant son efficacité malgré un nombre restreint de variables d'apprentissage.

5.2 Modélisation métabolique

L'adaptation du modèle métabolique, initialement développé pour la souris, à un modèle humain sans recours à des données expérimentales, représente une avancée importante dans la modélisation métabolique du cancer de l'ovaire. Cette transposition a été effectuée en modifiant certains paramètres macroscopiques tels que le volume tumoral, représenté par les variables VT , X et $Vblood$, le taux de croissance tumoral, représenté par la variable $v_{max,growth}$, et le flux sanguin tumoral, représenté par la variable F . Cependant, le flux sanguin tumoral, fixé à $8,663 \times 10^{-5}$ L/h, introduit une incertitude quant à la plausibilité de la modélisation métabolique proposée. En effet, cette valeur n'est pas réaliste, car selon la littérature, le flux sanguin tumoral est estimé à environ 7,2 L/h chez l'humain. Cette constante cinétique est toutefois la seule à avoir été fixée de manière plutôt arbitraire, dans le but de tester la transposition du modèle métabolique actuel à des fins de prédiction diagnostique, plutôt que de décrire fidèlement la réalité humaine. Une tentative d'adaptation de ce flux a quand même été réalisée en l'ayant ajusté d'un ordre de grandeur de 10^3 par rapport à celui utilisé pour modéliser le métabolisme tumoral chez la souris. Il faut toutefois être conscient qu'un flux sanguin tumoral très faible pourrait entraîner des conséquences sur les fluctuations métaboliques. Notamment, les concentrations des métabolites extracellulaires, réapprovisionnés par le sang, pourraient avoir été sous-estimées par le modèle métabolique. Il faut également souligner que, faute de jeux de données adéquats et d'une littérature limitée chez l'humain, notamment pour le cancer ovarien, il a été impossible d'ajuster l'ensemble des paramètres pour adapter précisément le modèle à l'humain. Dans ces conditions, il a été jugé pertinent de conserver les paramètres difficilement transposables, tels que les constantes cinétiques et le flux sanguin tumoral provenant principalement du modèle métabolique de la souris, afin de permettre une simulation fonctionnelle du métabolisme tumoral tout en surmontant les limitations dues à l'absence de données humaines adéquates. En ce sens, le modèle métabolique modifié s'avère utile dans les conditions actuelles et pourrait servir de base pour une transposition future chez l'humain, une fois que des données plus précises et spécifiques seront disponibles pour ajuster les paramètres métaboliques de manière adéquate.

Un travail considérable a également été réalisé sur les cinétiques enzymatiques du modèle métabolique, principalement dans le but d'y intégrer des données transcriptomiques. Les différentes cinétiques enzymatiques testées ont induit de légères modifications sur la croissance tumorale et les fluctuations métaboliques, soulignant l'importance du coefficient FC pour adapter le modèle en

fonction du profil transcriptomique spécifique des patientes. En effet, les coefficients FC calculés présentent des écart-types d'environ 0,15 et des valeurs moyennes proches de 1, indiquant des ajustements modérés mais significatifs des vitesses de réaction. Par ailleurs, deux approches de modulation temporelle des vitesses enzymatiques ont été explorées: l'une basée sur la variation de la concentration des cofacteurs enzymatiques au cours du cycle cellulaire, et l'autre reposant sur une oscillation sinusoïdale de $\pm 20\%$ de la vitesse maximale de réaction. Bien que l'idée d'associer la fréquence de l'oscillation sinusoïdale aux données transcriptomiques ait été envisagée, elle a été écartée en raison du manque de fondement biologique justifiant une telle association. En effet, les vitesses de réaction ne suivent pas nécessairement un rythme parfaitement sinusoïdal, et il n'existe aucune justification biologique démontrant que les données transcriptomiques devraient correspondre à la fréquence d'une telle oscillation. Somme toute, la cinétique enzymatique proposée dans ce projet, intégrant le coefficient FC et une fonction temporelle $F(t)$, semble être une approche biologiquement plausible et bien fondée pour moduler les paramètres enzymatiques et mieux refléter la réalité biologique. De plus, en modulant les vitesses maximales de réaction en fonction du niveau d'expression des gènes d'une patiente, cette cinétique permet de personnaliser la modélisation métabolique du cancer de l'ovaire.

Pour des patientes à des stades différents, la croissance tumorale semble suivre une allure exponentielle plus rapide à des stades avancés, accompagné d'un volume tumoral initial plus élevé. Cette accélération de la croissance tumorale à des stades avancés peut résulter d'une plus grande quantité de cellules tumorales dès le départ, combinée à une intensification de leur métabolisme, favorisant une prolifération plus rapide. De plus, les concentrations des métabolites intracellulaires ne varient que légèrement selon le stade, suggérant qu'elles sont davantage influencées par le profil transcriptomique individuel que par le stade du cancer. Ainsi, le choix des patientes pour représenter chaque stade du cancer a probablement eu un impact important sur les résultats métaboliques obtenus. En revanche, les métabolites extracellulaires présentent des variations plus marquées selon le stade, avec une diminution plus prononcée de leur concentration à mesure que la tumeur progresse, à l'exception de lalanine (ALA), l'aspartate (ASP) et la glutamine extracellulaire (EGLN). Cette disparité entre l'impact du stade sur les métabolites extracellulaires et intracellulaires pourrait être attribué à une consommation accrue des ressources extracellulaires avec le développement tumoral. Parallèlement, les cellules tumorales semblent ajuster leurs

concentrations internes afin de préserver un équilibre métabolique, limitant ainsi les variations intracellulaires malgré la progression du cancer.

5.3 Ciblage thérapeutique

Les algorithmes de recherche de solutions *Depth Search* et *Intelligent Search* ont été utilisés pour identifier les niveaux d'expression génique dont la modulation résulte en la réduction de la croissance tumorale. En exploitant les données transcriptomiques d'une patiente, ces algorithmes permettent de déterminer des cibles spécifiques en fonction du profil d'expression spécifique à une patiente, offrant ainsi une approche personnalisée. Parmi les solutions obtenues, certains gènes se sont démarqués comme potentielles cibles thérapeutiques contre le cancer de l'ovaire (Tableau 5.1).

Tableau 5.1 Cibles thérapeutiques identifiées grâce aux algorithmes de recherche

Gène(s)	Modulation	Réaction(s) altérée(s)	Voie(s) métabolique(s) altérée(s)
GPT	Suppression	VAlaTA	Métabolisme des acides aminés
GOT1 ou GOT2	Activation	VASTA	Métabolisme des acides aminés
BCAT1 ou BCAT2	Activation	VAAtoSUC	Métabolisme des acides aminés
G6PD, PGLS, PGD ou RPIA	Suppression	VG6PDH	Voie des pentoses phosphates
HK1 et TKT	Suppression	VHK et VTK	Glycolyse et voie des pentoses phosphates
FASN	Suppression	VPALM	Lipogenèse

La suppression du gène GPT, impliqué dans le métabolisme des acides aminés, s'est avéré particulièrement efficace dans les premiers stades du cancer de l'ovaire, bien que son efficacité semble diminuer avec la progression de la maladie. De plus, la suppression d'un gène parmi G6PD, PGLS, PGD et RPIA, impliqués dans la voie des pentoses phosphates, entraîne une réduction plus modérée de la croissance tumorale. L'activation de certains gènes s'avère également bénéfique, en particulier pour les stades avancés; l'augmentation de l'expression de GOT1, GOT2, BCAT1 ou BCAT2, également impliqués dans le métabolisme des acides aminés, réduit de manière significative la croissance tumorale. Étant donné que l'efficacité de la modulation des gènes dépend du stade d'avancement du cancer ainsi que du volume tumoral, la modulation séquentielle pourrait offrir une approche thérapeutique plus personnalisée et optimisée en fonction de l'évolution

spécifique à chaque patient du cancer. Par exemple, supprimer le gène GPT à un stade précoce, puis activer les gènes GOT1 et GOT2 lorsque la tumeur progresse, pourrait s'avérer des plus efficace. De plus, en recherchant des combinaisons de gènes, d'autres cibles thérapeutiques peuvent être identifiées. La suppression simultanée de GPT et FASN au stade I génère une réduction moyenne de la croissance tumorale de 21,7%, et dans certains cas particuliers, la suppression simultanée de HK1 et TKT entraîne une réduction drastique de la croissance tumorale en engendrant un épuisement complet du métabolite G6P. Par ailleurs, l'activation simultanée de GOT2 et BCAT2 aux stades avancés améliore l'inhibition de la croissance tumorale, offrant une réduction de 38,4% au 270^e jour pour une patiente au stade IV. Néanmoins, l'effet combiné de la modulation de plusieurs gènes ne dépasse pas toujours la somme des effets individuels.

L'algorithme *Intelligent Search* confirme globalement les tendances obtenues par *Depth Search*, mais il ne parvient pas toujours à identifier les combinaisons de gènes les plus efficaces. De plus, il semble particulièrement apte à sélectionner des gènes ayant un impact individuel sur la croissance tumorale, mais il éprouve plus de difficulté à détecter des interactions synergiques où l'effet ne se manifeste que lorsque plusieurs gènes sont modulés simultanément. Par exemple, la suppression isolée de HK1 ou de TKT n'a qu'un effet limité, alors que leur suppression simultanée peut entraîner une réduction drastique de la croissance tumorale en provoquant un épuisement complet du métabolite G6P. Cet effet n'émerge que par l'action conjointe des deux suppressions, illustrant la notion de « synergie ». Or, l'*Intelligent Search* n'a pas identifié cette combinaison pour les patientes aux stades II et III, bien que toutes ses solutions incluent des gènes ayant un fort impact individuel comme GPT, GOT1, GOT2, BCAT1 et BCAT2. L'algorithme *Intelligent Search* s'avère toutefois bien plus efficace pour explorer un large espace de solutions. Sa performance a été démontrée en identifiant des combinaisons de gènes réduisant significativement la croissance tumorale en seulement 400 itérations, tandis que l'algorithme *Depth Search* aurait nécessité l'évaluation d'environ 4 millions de solutions différentes.

Les gènes identifiés par les algorithmes de recherche rejoignent généralement ceux identifiés dans la littérature. En effet, l'isoforme GOT2 du gène GPT est reconnu dans la littérature comme un biomarqueur de résistance à la chimiothérapie, et la glutaminolyse, catalysée par l'enzyme GPT, est essentielle à la survie des cellules de cancer de l'ovaire résistantes au platine [62]. De plus, l'expression des gènes GOT1 et GOT2 est associée à certains cancers, comme le cancer du pancréas, et leur surexpression peut servir de biomarqueur pour le diagnostic et le

pronostic de divers types de cancer [63]. Cependant, dans le cadre de ce projet, l'activation de GOT1 et GOT2, et non leur inhibition, a conduit à une réduction de la croissance tumorale. Par ailleurs, la surexpression de BCAT2 a été liée à un pronostic défavorable dans plusieurs cancers, mais, dans certains cas, une forte expression de BCAT2 était associée à une meilleure survie, ce qui pourrait indiquer un rôle protecteur ou régulateur de maintien de l'homéostasie tumorale [64]. Enfin, bien que les gènes *solute carrier family 2 member 1* (SLC2A1) et *hexokinase 2* (HK2), bien qu'identifiés dans la littérature comme des cibles thérapeutiques potentielles, n'ont pas été intégrés dans les règles GPR. G6PD et FASN, également identifiés dans la littérature, ont émergé comme cibles thérapeutiques dans les algorithmes de recherche, mais leur suppression n'a pas donné de résultats aussi prometteurs que celle de GPT, ou l'activation de GOT1 et GOT2.

La principale limitation du ciblage thérapeutique dans ce projet réside dans les règles GPR établies, qui constituent une simplification importante des gènes impliqués dans le métabolisme tumoral. En effet, l'utilisation d'opérations prédéfinies, comme le minimum et le maximum, exclut complètement les expressions géniques qui ne sont pas minimales ou maximales dans le calcul de la donnée transcriptomique associée à une réaction. De plus, les réactions du modèle métabolique utilisées dans ce projet ne regroupent qu'un sous-ensemble des réactions biologiques composant le métabolisme de cellules cancéreuses. Toutefois, il est largement considéré que le métabolisme central du carbone compris dans ce sous-ensemble permet d'évaluer les voies essentielles à la croissance cellulaire, soit le trait caractéristique majeur de cellules tumorales. Des gènes identifiés dans la littérature comme des cibles thérapeutiques potentielles, tels que SLC2A1 et HK2, n'ont pas pu être intégrés dans les règles GPR, car ils ne sont pas directement inclus dans le modèle en mains. De plus, bien que 112 gènes aient été identifiés comme étant liés au métabolisme tumoral, ce nombre a dû être réduit à 75 en raison des contraintes de complexité reliées aux réactions *Vresp* et *Vgrowth* tel que décrites dans notre étude. Par ailleurs, le gène *asparaginase* (ASPG), qui code pour l'enzyme portant le même nom et catalysant la réaction VASN, ne figure pas dans la banque de données de 13 110 gènes disponible, rendant son analyse impossible à l'aide des règles GPR. La difficulté à associer des gènes aux réactions du modèle métabolique a donc limité le nombre de gènes étudiés et restreint considérablement les cibles thérapeutiques potentielles.

5.4 Contributions du projet et état de la recherche

Tout d'abord, l'apprentissage supervisé basé sur l'expression génique constitue une approche prometteuse pour la détection du cancer de l'ovaire. En particulier, l'ensemble réduit de 112 gènes métaboliques offre d'excellentes performances, avec une exactitude balancée très élevée, en faisant un candidat pertinent pour une application clinique. L'utilisation de SMOTE s'est également révélée efficace pour améliorer la performance des modèles en présence de classes déséquilibrées. Par ailleurs, la transposition du modèle métabolique de la souris à la physiologie humaine permet d'affiner la compréhension des dynamiques tumorales du cancer de l'ovaire. Les ajustements apportés aux cinétiques enzymatiques, notamment l'introduction du coefficient FC, améliorent la représentation des variations biologiques propres à chaque profil transcriptomique. Enfin, le développement et l'application des algorithmes *Depth Search* et *Intelligent Search*, basés sur des concepts d'intelligence artificielle, marquent une avancée importante dans l'identification de cibles thérapeutiques. En intégrant le profil transcriptomique spécifique des patientes, le développement de ces algorithmes innovants ouvre la voie à une médecine personnalisée, capable d'adapter les interventions thérapeutiques à l'évolution dynamique de la maladie. Par exemple, la suppression du gène GPT aux stades précoces et l'activation des gènes GOT1 et GOT2 aux stades avancés illustrent le potentiel d'une modulation séquentielle ciblée. Par ailleurs, le Tableau B.1 de l'Annexe B résume les résultats obtenus et met en évidence les principales limitations de chaque volet du projet, synthétisant ainsi les pistes d'amélioration pour les travaux futurs. Somme toute, ce projet contribue à la recherche contre le cancer en offrant de nouvelles perspectives pour une prise en charge précoce et personnalisée du cancer de l'ovaire.

CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS

Ce projet de recherche propose une approche innovante exploitant des données transcriptomiques issues d'une biopsie pour optimiser la détection précoce du cancer de l'ovaire et identifier des cibles thérapeutiques pour des stratégies de traitements métaboliques et immunothérapeutiques. Les algorithmes d'apprentissage supervisé optimisés dans ce projet permettent d'identifier avec une grande exactitude les patientes atteintes de ce cancer, même à ses stades précoces, favorisant ainsi une détection hâtive cruciale pour améliorer les chances de survie. L'intégration de ces données dans un modèle métabolique permet également d'analyser le profil métabolique tumoral d'une patiente, en simulant la croissance tumorale et ses fluctuations métaboliques sur une période d'au moins 270 jours. De plus, les algorithmes de recherche de solution développés facilitent l'identification de gènes dont la modulation de l'expression pourrait freiner la croissance tumorale. En combinant détection précoce et stratégies thérapeutiques ciblées, cette approche ouvre la voie à une médecine plus personnalisée, adaptée aux caractéristiques génétiques et métaboliques de chaque patiente.

Bien que les résultats obtenus soient prometteurs, certaines limitations doivent être prises en considération. Tout d'abord, la performance des algorithmes de classification doit encore être validée sur des cohortes avec des classes plus équilibrées et mieux peuplées. Pour garantir que les résultats par rapport à la détection du cancer de l'ovaire soient représentatifs de la réalité, il serait donc nécessaire d'augmenter le nombre de patientes saines dans l'échantillon utilisé pour l'apprentissage supervisé. De plus, le modèle métabolique reste une simplification du métabolisme tumoral, notamment en raison de l'absence de modélisation de la nécrose tumorale, et nécessite des données expérimentales pour garantir sa représentativité de la réalité humaine. L'ajout de métabolites, de voies supplémentaires et de mécanismes de nécrose pourrait donc améliorer la précision et la pertinence du modèle pour le cancer de l'ovaire. Il serait également essentiel d'adapter les paramètres biocinétiques à la réalité humaine, notamment le flux sanguin tumoral, afin d'augmenter la validité du modèle dans un contexte clinique. Enfin, les règles GPR établies dans ce projet ne sont pas exhaustives, limitant ainsi le nombre de cibles thérapeutiques pouvant être identifiées. Afin de surmonter cette limitation, il serait pertinent de développer des règles GPR plus complexes, en ajustant les opérations strictes comme le minimum ou le maximum, et d'intégrer un plus grand nombre de gènes, possiblement en ajoutant des réactions enzymatiques supplémentaires expressément liées au métabolisme tumoral. En ce qui concerne la capacité à agir sur

des gènes spécifiquement liés au métabolisme tumoral, l'objectif est de potentiellement limiter le risque d'effets secondaires, ce qui permettrait ainsi de surmonter les problèmes associés à la chimiothérapie.

Les résultats de ce projet ouvrent de nombreuses perspectives pour l'application des algorithmes développés, notamment leur utilisation dans la détection précoce, nous l'espérons, ainsi que le traitement d'autres types de cancers. De plus, l'intégration de données complémentaires, telles que celles issues de l'imagerie médicale, et l'adoption de modèles d'intelligence artificielle plus avancés, comme les réseaux de neurones profonds, pourraient renforcer la précision des analyses et affiner la personnalisation des traitements. Par ailleurs, l'adaptation de cette approche à un suivi longitudinal des patientes, en récoltant plusieurs séries de données transcriptomiques, permettrait d'évaluer en temps réel l'évolution tumorale et d'ajuster de manière dynamique les stratégies thérapeutiques.

En somme, ce projet établit des bases solides pour le développement de nouvelles approches dans la détection et le traitement du cancer de l'ovaire. Bien que les résultats nécessitent encore des validations, ils montrent des perspectives prometteuses pour améliorer la prise en charge de ce cancer et, potentiellement, celle d'autres cancers.

RÉFÉRENCES

- [1] C. C. S. / S. canadienne du cancer, « Statistiques canadiennes sur le cancer », Société canadienne du cancer. Consulté le: 7 juin 2024. [En ligne]. Disponible à: <https://cancer.ca/fr/research/cancer-statistics/canadian-cancer-statistics>
- [2] Y. Xiao, M. Bi, H. Guo, et M. Li, « Multi-omics approaches for biomarker discovery in early ovarian cancer diagnosis », *eBioMedicine*, vol. 79, p. 104001, mai 2022, doi: 10.1016/j.ebiom.2022.104001.
- [3] L. A. Torre *et al.*, « Ovarian Cancer Statistics, 2018 », *CA. Cancer J. Clin.*, vol. 68, n° 4, p. 284-296, juill. 2018, doi: 10.3322/caac.21456.
- [4] R. Lowe, N. Shirley, M. Bleackley, S. Dolan, et T. Shafee, « Transcriptomics technologies », *PLoS Comput. Biol.*, vol. 13, n° 5, p. e1005457, mai 2017, doi: 10.1371/journal.pcbi.1005457.
- [5] N. Pashayan et P. D. P. Pharoah, « The challenge of early detection in cancer », *Science*, vol. 368, n° 6491, p. 589-590, mai 2020, doi: 10.1126/science.aaz2078.
- [6] R. Lawrence, M. Watters, C. R. Davies, K. Pantel, et Y.-J. Lu, « Circulating tumour cells for early detection of clinically relevant cancer », *Nat. Rev. Clin. Oncol.*, vol. 20, n° 7, p. 487-500, juill. 2023, doi: 10.1038/s41571-023-00781-y.
- [7] B. Roskams-Hieter *et al.*, « Plasma cell-free RNA profiling distinguishes cancers from premalignant conditions in solid and hematologic malignancies », *Npj Precis. Oncol.*, vol. 6, n° 1, p. 1-11, avr. 2022, doi: 10.1038/s41698-022-00270-y.
- [8] D. R. Schmidt, R. Patel, D. G. Kirsch, C. A. Lewis, M. G. Vander Heiden, et J. W. Locasale, « Metabolomics in Cancer Research and Emerging Applications in Clinical Oncology », *CA. Cancer J. Clin.*, vol. 71, n° 4, p. 333-358, juill. 2021, doi: 10.3322/caac.21670.
- [9] K. M. Elias, J. Guo, et R. C. Bast, « Early Detection of Ovarian Cancer », *Hematol. Oncol. Clin. North Am.*, vol. 32, n° 6, p. 903-914, déc. 2018, doi: 10.1016/j.hoc.2018.07.003.
- [10] D. Fischerova et A. Burgetova, « Imaging techniques for the evaluation of ovarian cancer », *Best Pract. Res. Clin. Obstet. Gynaecol.*, vol. 28, n° 5, p. 697-720, juill. 2014, doi: 10.1016/j.bpobgyn.2014.04.006.
- [11] Y. Gao *et al.*, « Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: a retrospective, multicentre, diagnostic study », *Lancet Digit. Health*, vol. 4, n° 3, p. e179-e187, mars 2022, doi: 10.1016/S2589-7500(21)00278-8.
- [12] R. Zhang, M. K. Y. Siu, H. Y. S. Ngan, et K. K. L. Chan, « Molecular Biomarkers for the Early Detection of Ovarian Cancer », *Int. J. Mol. Sci.*, vol. 23, n° 19, Art. n° 19, janv. 2022, doi: 10.3390/ijms231912041.
- [13] M. Lu *et al.*, « Using machine learning to predict ovarian cancer », *Int. J. Med. Inf.*, vol. 141, p. 104195, sept. 2020, doi: 10.1016/j.ijmedinf.2020.104195.
- [14] S. J. Skates, « OCS: Development of the Risk of Ovarian Cancer Algorithm (ROCA) and ROCA screening trials », *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.*, vol. 22, n° Suppl 1, p. S24-S26, mai 2012, doi: 10.1097/IGC.0b013e318256488a.

- [15] M. Akazawa et K. Hashimoto, « Artificial Intelligence in Ovarian Cancer Diagnosis », *Anticancer Res.*, vol. 40, n° 8, p. 4795-4800, août 2020, doi: 10.21873/anticanres.14482.
- [16] J. Liñares-Blanco, A. Pazos, et C. Fernandez-Lozano, « Machine learning analysis of TCGA cancer data », *PeerJ Comput. Sci.*, vol. 7, p. e584, juill. 2021, doi: 10.7717/peerj-cs.584.
- [17] A. Arfiani et Z. Rustam, « Ovarian cancer data classification using bagging and random forest », *AIP Conf. Proc.*, vol. 2168, n° 1, p. 020046, nov. 2019, doi: 10.1063/1.5132473.
- [18] « What is Ovarian Cancer | Ovarian Tumors and Cysts ». Consulté le: 12 février 2025. [En ligne]. Disponible à: <https://www.cancer.org/cancer/types/ovarian-cancer/about/what-is-ovarian-cancer.html>
- [19] « Types & Stages », ovarian.org. Consulté le: 12 février 2025. [En ligne]. Disponible à: <https://ovarian.org/about-ovarian-cancer/types-and-stages/>
- [20] M. Koshiyama, N. Matsumura, et I. Konishi, « Subtypes of Ovarian Cancer and Ovarian Cancer Screening », *Diagnostics*, vol. 7, n° 1, Art. n° 1, mars 2017, doi: 10.3390/diagnostics7010012.
- [21] Y. Wang, A. J. Duval, M. Adli, et D. Matei, « Biology-driven therapy advances in high-grade serous ovarian cancer », *J. Clin. Invest.*, vol. 134, n° 1, janv. 2024, doi: 10.1172/JCI174013.
- [22] L. Yang, H.-J. Xie, Y.-Y. Li, X. Wang, X.-X. Liu, et J. Mai, « Molecular mechanisms of platinum-based chemotherapy resistance in ovarian cancer (Review) », *Oncol. Rep.*, vol. 47, n° 4, p. 1-11, avr. 2022, doi: 10.3892/or.2022.8293.
- [23] D. Hanahan et R. A. Weinberg, « Hallmarks of Cancer: The Next Generation », *Cell*, vol. 144, n° 5, p. 646-674, mars 2011, doi: 10.1016/j.cell.2011.02.013.
- [24] M. Vanoni, P. Palumbo, S. Busti, et L. Alberghina, « A critical review of multiscale modeling for predictive understanding of cancer cell metabolism », *Curr. Opin. Syst. Biol.*, vol. 39, p. 100531, déc. 2024, doi: 10.1016/j.coisb.2024.100531.
- [25] J. da Veiga Moreira, M. Hamraz, M. Abolhassani, L. Schwartz, M. Jolicœur, et S. Peres, « Metabolic therapies inhibit tumor growth in vivo and in silico », *Sci. Rep.*, vol. 9, n° 1, p. 3153, févr. 2019, doi: 10.1038/s41598-019-39109-1.
- [26] K. Tondo-Steele et K. McLean, « The “Sweet Spot” of Targeting Tumor Metabolism in Ovarian Cancers », *Cancers*, vol. 14, n° 19, Art. n° 19, janv. 2022, doi: 10.3390/cancers14194696.
- [27] S. Kumari, S. Gupta, A. Jamil, D. Tabatabaei, et S. Karakashev, « Exploring Metabolic Approaches for Epithelial Ovarian Cancer Therapy », *J. Cell. Physiol.*, vol. 240, n° 1, p. e31495, 2025, doi: 10.1002/jcp.31495.
- [28] H. Sajjad, S. Imtiaz, T. Noor, Y. H. Siddiqui, A. Sajjad, et M. Zia, « Cancer models in preclinical research: A chronicle review of advancement in effective cancer research », *Anim. Models Exp. Med.*, vol. 4, n° 2, p. 87-103, mars 2021, doi: 10.1002/ame2.12165.
- [29] J. D. Orth, I. Thiele, et B. Ø. Palsson, « What is flux balance analysis? », *Nat. Biotechnol.*, vol. 28, n° 3, p. 245-248, mars 2010, doi: 10.1038/nbt.1614.

- [30] K. E. Meeson et J.-M. Schwartz, « Constraint-based modelling predicts metabolic signatures of low and high-grade serous ovarian cancer », *Npj Syst. Biol. Appl.*, vol. 10, n° 1, p. 1-12, août 2024, doi: 10.1038/s41540-024-00418-5.
- [31] L. Bezjak *et al.*, « Incorporating RNA-Seq transcriptomics into glycosylation-integrating metabolic network modelling kinetics: Multiomic Chinese hamster ovary (CHO) cell bioreactors », *Biotechnol. Bioeng.*, vol. 118, n° 4, p. 1476-90, 2021, doi: 10.1002/bit.27660.
- [32] B. F. Ganzfried *et al.*, « curatedOvarianData: clinically annotated data for the ovarian cancer transcriptome », *Database J. Biol. Databases Curation*, vol. 2013, p. bat013, 2013, doi: 10.1093/database/bat013.
- [33] D. Bell *et al.*, « Integrated genomic analyses of ovarian carcinoma », *Nature*, vol. 474, n° 7353, p. 609-615, juin 2011, doi: 10.1038/nature10166.
- [34] J. Robitaille, J. Chen, et M. Jolicoeur, « A Single Dynamic Metabolic Model Can Describe mAb Producing CHO Cell Batch and Fed-Batch Cultures on Different Culture Media », *PLOS ONE*, vol. 10, n° 9, p. e0136815, sept. 2015, doi: 10.1371/journal.pone.0136815.
- [35] A. Ghorbaniaghdam, O. Henry, et M. Jolicoeur, « A kinetic-metabolic model based on cell energetic state: study of CHO cell behavior under Na-butyrate stimulation », *Bioprocess Biosyst. Eng.*, vol. 36, n° 4, p. 469-487, avr. 2013, doi: 10.1007/s00449-012-0804-3.
- [36] A. Ghorbaniaghdam, J. Chen, O. Henry, et M. Jolicoeur, « Analyzing Clonal Variation of Monoclonal Antibody-Producing CHO Cell Lines Using an In Silico Metabolomic Platform », *PLOS ONE*, vol. 9, n° 3, p. e90832, mars 2014, doi: 10.1371/journal.pone.0090832.
- [37] A. Ghorbaniaghdam, O. Henry, et M. Jolicoeur, « An *in-silico* study of the regulation of CHO cells glycolysis », *J. Theor. Biol.*, vol. 357, p. 112-122, sept. 2014, doi: 10.1016/j.jtbi.2014.04.035.
- [38] G. Goffaux, I. Hammami, et M. Jolicoeur, « A Dynamic Metabolic Flux Analysis of Myeloid-Derived Suppressor Cells Confirms Immunosuppression-Related Metabolic Plasticity », *Sci. Rep.*, vol. 7, n° 1, p. 9850, août 2017, doi: 10.1038/s41598-017-10464-1.
- [39] « (PDF) Minimal cut sets in biochemical reaction networks », *ResearchGate*, oct. 2024, doi: 10.1093/bioinformatics/btg395.
- [40] « Duration of cell cycle phases for a typical c - Human Homo sapiens - BNID 112260 ». Consulté le: 3 février 2025. [En ligne]. Disponible à: https://bionumbers.hms.harvard.edu/bionumber.aspx?id=112260&s=n&v=3&utm_source=c_hatgpt.com
- [41] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, et P. Walter, « An Overview of the Cell Cycle », dans *Molecular Biology of the Cell. 4th edition*, Garland Science, 2002. Consulté le: 3 février 2025. [En ligne]. Disponible à: <https://www.ncbi.nlm.nih.gov/books/NBK26869/>
- [42] J. da Veiga Moreira *et al.*, « Cell cycle progression is regulated by intertwined redox oscillators », *Theor. Biol. Med. Model.*, vol. 12, n° 1, p. 10, mai 2015, doi: 10.1186/s12976-015-0005-2.

- [43] J. Hellerstein, « An oscillating reaction network with an exact closed form solution in the time domain », *BMC Bioinformatics*, vol. 24, n° 1, p. 466, déc. 2023, doi: 10.1186/s12859-023-05600-w.
- [44] « Interpolation (scipy.interpolate) — SciPy v1.15.1 Manual ». Consulté le: 4 février 2025. [En ligne]. Disponible à: <https://docs.scipy.org/doc/scipy/reference/interpolate.html>
- [45] « Ovarian Cancer Stages | Staging for Ovarian Cancer ». Consulté le: 13 janvier 2025. [En ligne]. Disponible à: <https://www.cancer.org/cancer/types/ovarian-cancer/detection-diagnosis-staging/staging.html>
- [46] J. S. Berek, M. Renz, S. Kehoe, L. Kumar, et M. Friedlander, « Cancer of the ovary, fallopian tube, and peritoneum: 2021 update », *Int. J. Gynecol. Obstet.*, vol. 155, n° S1, p. 61-85, 2021, doi: 10.1002/ijgo.13878.
- [47] E. Petru, C. Huber, E. Sampl, et J. Haas, « Comparison of Primary Tumor Size in Stage I and III Epithelial Ovarian Cancer », *Anticancer Res.*, vol. 38, n° 11, p. 6507-6511, nov. 2018, doi: 10.21873/anticanres.13015.
- [48] R. Aviram, G. Gassner, O. Markovitch, I. Cohen, A. Fishman, et R. Tepper, « Volumes of normal ovaries, ovaries with benign lesions, and ovaries with cancer in menopausal women: Is there an optimal cut-off value to predict malignancy? », *J. Clin. Ultrasound*, vol. 36, n° 1, p. 1-5, janv. 2008, doi: 10.1002/jcu.20416.
- [49] J. S. Bedia *et al.*, « Estimating the ovarian cancer CA-125 preclinical detectable phase, in-vivo tumour doubling time, and window for detection in early stage: an exploratory analysis of UKCTOCS », *eBioMedicine*, vol. 112, févr. 2025, doi: 10.1016/j.ebiom.2024.105554.
- [50] X. Cao *et al.*, « Tumor Blood Flow Is a Predictor of Radiotherapy Response in Patients With Nasopharyngeal Carcinoma », *Front. Oncol.*, vol. 11, août 2021, doi: 10.3389/fonc.2021.567954.
- [51] K. M. Bakke *et al.*, « Sex Differences and Tumor Blood Flow from Dynamic Susceptibility Contrast MRI Are Associated with Treatment Response after Chemoradiation and Long-term Survival in Rectal Cancer », *Radiology*, vol. 297, n° 2, p. 352-360, nov. 2020, doi: 10.1148/radiol.2020200287.
- [52] D. Katayama *et al.*, « Greater reductions in blood flow after anti-angiogenic treatment in non-small cell lung cancer patients are associated with shorter progression-free survival », *Sci. Rep.*, vol. 11, n° 1, p. 6805, mars 2021, doi: 10.1038/s41598-021-86405-w.
- [53] N. Ahmadiankia, « In vitro and in vivo studies of cancer cell behavior under nutrient deprivation », *Cell Biol. Int.*, vol. 44, n° 8, p. 1588-1597, 2020, doi: 10.1002/cbin.11368.
- [54] « StratifiedKFold », scikit-learn. Consulté le: 4 décembre 2024. [En ligne]. Disponible à: https://scikit-learn/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html
- [55] « StandardScaler », scikit-learn. Consulté le: 17 février 2025. [En ligne]. Disponible à: <https://scikit-learn/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>
- [56] « SMOTE — Version 0.13.0.dev0 ». Consulté le: 4 décembre 2024. [En ligne]. Disponible à: https://imbalanced-learn.org/dev/references/generated/imblearn.over_sampling.SMOTE.html

- [57] « LogisticRegression », scikit-learn. Consulté le: 4 décembre 2024. [En ligne]. Disponible à: https://scikit-learn/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
- [58] « LinearSVC », scikit-learn. Consulté le: 4 décembre 2024. [En ligne]. Disponible à: <https://scikit-learn/stable/modules/generated/sklearn.svm.LinearSVC.html>
- [59] « KNeighborsClassifier », scikit-learn. Consulté le: 4 décembre 2024. [En ligne]. Disponible à: <https://scikit-learn/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>
- [60] « RandomForestClassifier », scikit-learn. Consulté le: 4 décembre 2024. [En ligne]. Disponible à: <https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>
- [61] R. M. Ghoniem, A. D. Algarni, B. Refky, et A. A. Ewees, « Multi-Modal Evolutionary Deep Learning Model for Ovarian Cancer Diagnosis », *Symmetry*, vol. 13, n° 4, Art. n° 4, avr. 2021, doi: 10.3390/sym13040643.
- [62] A. L. Ponton-Almodovar, J. Bernard, et S. Horibata, « Abstract B011: GPT2 as a novel driver of platinum-resistant ovarian cancer », *Cancer Res.*, vol. 84, n° 5_Supplement_2, p. B011, mars 2024, doi: 10.1158/1538-7445.OVARIAN23-B011.
- [63] Z. Song *et al.*, « Glutamic oxaloacetic transaminase 1 as a potential target in human cancer », *Eur. J. Pharmacol.*, vol. 917, p. 174754, févr. 2022, doi: 10.1016/j.ejphar.2022.174754.
- [64] Q. Cao, J. Fan, J. Zou, et W. Wang, « Multi-omics analysis identifies BCAT2 as a potential pan-cancer biomarker for tumor progression and immune microenvironment modulation », *Sci. Rep.*, vol. 14, n° 1, p. 23371, oct. 2024, doi: 10.1038/s41598-024-74441-1.
- [65] Z. Wang, D. D. Rao, N. Senzer, et J. Nemunaitis, « RNA Interference and Cancer Therapy », *Pharm. Res.*, vol. 28, n° 12, p. 2983-2995, déc. 2011, doi: 10.1007/s11095-011-0604-5.
- [66] « Tissue-Specific Codon/Codon Pair Usage Tables (TissueCoCoPUTs) ». Consulté le: 20 février 2025. [En ligne]. Disponible à: https://dnahive.fda.gov/dna.cgi?cmd=codon_usage&id=537&mode=tisspec
- [67] K. Zahn, « Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth », *J. Bacteriol.*, vol. 178, n° 10, p. 2926-2933, mai 1996, doi: 10.1128/jb.178.10.2926-2933.1996.
- [68] W. Qian, J.-R. Yang, N. M. Pearson, C. Maclean, et J. Zhang, « Balanced Codon Usage Optimizes Eukaryotic Translational Efficiency », *PLOS Genet.*, vol. 8, n° 3, p. e1002603, mars 2012, doi: 10.1371/journal.pgen.1002603.
- [69] K. Bentele, P. Saffert, R. Rauscher, Z. Ignatova, et N. Blüthgen, « Efficient translation initiation dictates codon usage at gene start », *Mol. Syst. Biol.*, vol. 9, n° 1, p. 675, janv. 2013, doi: 10.1038/msb.2013.32.
- [70] V. Bali et Z. Bebok, « Decoding mechanisms by which silent codon changes influence protein biogenesis and function », *Int. J. Biochem. Cell Biol.*, vol. 64, p. 58-74, juill. 2015, doi: 10.1016/j.biocel.2015.03.011.
- [71] « spearmanr — SciPy v1.15.2 Manual ». Consulté le: 20 février 2025. [En ligne]. Disponible à: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.spearmanr.html#scipy.stats.spearmanr>

ANNEXE A MODÈLE MÉTABOLIQUE

Tableau A.1 Métabolites et composés impliqués dans le modèle métabolique

Abréviation	Nom	Condition initiale
ACCOA	Acétyl-CoA	$8,92 \times 10^{-2}$ mM
ADP	Adénosine diphosphate	$2,80 \times 10^{-1}$ mM
AKG	Alpha-cétoglutarate	$6,37 \times 10^{-2}$ mM
ALA	Alanine	$1,48 \times 10^{-1}$ mM
AMP	Adénosine monophosphate	$1,97 \times 10^{-1}$ mM
ARG	Arginine	2,40 mM
ASN	Asparagine	$4,80 \times 10^{-1}$ mM
ASP	Aspartate	$4,90 \times 10^{-2}$ mM
ATP	Adénosine triphosphate	2,63 mM
CIT	Citrate	$6,37 \times 10^{-1}$ mM
CO2	Dioxyde de carbone	$5,10 \times 10^{-1}$ mM
EGLC	Glucose extracellulaire	6,55 mM
EGLN	Glutamine extracellulaire	$7,24 \times 10^{-1}$ mM
EGLU	Glutamate extracellulaire	$2,39 \times 10^{-1}$ mM
F6P	Fructose 6-phosphate	$1,40 \times 10^{-1}$ mM
G6P	Glucose 6-phosphate	$1,02 \times 10^{-1}$ mM
GAP	Glycéraldéhyde 3-phosphate	$2,55 \times 10^{-1}$ mM
GLU	Glutamate	$1,40 \times 10^2$ mM
GLY	Glycine	$2,41 \times 10^{-1}$ mM
HIS	Histidine	$8,50 \times 10^{-1}$ mM
ILE	Isoleucine	$6,80 \times 10^{-2}$ mM
LAC	Lactate	1,24 mM
LYS	Lysine	$5,16 \times 10^{-1}$ mM
MAL	Malate	$9,55 \times 10^{-1}$ mM
MET	Méthionine	$4,30 \times 10^{-1}$ mM
NAD	Nicotinamide adénine dinucléotide (forme oxydée)	$4,39 \times 10^{-1}$ mM
NADH	Nicotinamide adénine dinucléotide (forme réduite)	5,41 mM
NADP	Nicotinamide adénine dinucléotide phosphate (forme oxydée)	$2,33 \times 10^{-1}$ mM
NADPH	Nicotinamide adénine dinucléotide phosphate (forme réduite)	$1,46 \times 10^{-1}$ mM
NH4	Ammonium	$1,54 \times 10^{-1}$ mM

O2	Dioxygène	5,10 mM
OXA	Oxaloacétate	$6,37 \times 10^{-1}$ mM
PALM	Palminate	$9,55 \times 10^{-2}$ mM
PEP	Phosphoénolpyruvate	$5,61 \times 10^{-1}$ mM
Pi	Phosphate inorganique	8,00 mM
PHE	Phénylalanine	$4,90 \times 10^{-2}$ mM
PRO	Proline	$6,50 \times 10^{-1}$ mM
PYR	Pyruvate	1,27 mM
R5P	Ribose 5-phosphate	$3,44 \times 10^{-2}$ mM
SER	Sérine	$1,53 \times 10^{-1}$ mM
SUC	Succinate	$5,10 \times 10^{-2}$ mM
THR	Thréonine	$7,90 \times 10^{-2}$ mM
TYR	Tyrosine	$5,50 \times 10^{-2}$ mM
VAL	Valine	$7,00 \times 10^{-1}$ mM
Vblood	Volume du sang alimentant la tumeur	Stade I : 30 cm^3
		Stade II : 49 cm^3
		Stade III : 80 cm^3
		Stade IV : 130 cm^3
VT	Volume total de la tumeur	Stade I : 150 cm^3
		Stade II : 245 cm^3
		Stade III : 400 cm^3
		Stade IV : 650 cm^3
X	Volume de la masse de cellules tumorales	Stade I : 120 cm^3
		Stade II : 196 cm^3
		Stade III : 320 cm^3
		Stade IV : 520 cm^3
X5P	Xylulose 5-phosphate	$6,37 \times 10^{-2}$ mM

Tableau A.2 Réactions et règles GPR du modèle métabolique

No	Abréviation	Réaction	Gène(s) associé(s) et règles GPR
1	VHK	$\text{EGLC} + \text{ATP} \rightarrow \text{G6P} + \text{ADP}$	HK1
2	VPGI	$\text{G6P} \leftrightarrow \text{F6P}$	GPI

3	VPFK	$F6P + ATP \rightarrow 2 GAP + ADP$	$\min(\max(PFKM, PFKL, PFKP), \max(ALDOA, ALDOB, ALDOC))$
4	VPGK	$GAP + ADP + NAD \rightarrow PEP + ATP + NADH$	$\min(GAPDH, PGK1, \max(PGM1, PGM3, PGM5), \max(ENO1, ENO2, ENO3))$
5	VPK	$PEP + ADP \rightarrow PYR + ATP$	$\max(PKM, PKLR)$
6	VLDH	$PYR + NADH \leftrightarrow LAC + NAD$	$\min(LDHA, LDHB)$
7	VG6PDH	$G6P + 2 NADP \rightarrow R5P + 2 NADPH + CO2$	$\min(G6PD, PGLS, PGD, RPIA)$
8	VEP	$R5P \rightarrow X5P$	RPE
9	VTK	$R5P + 2 X5P \rightarrow 2 F6P + GAP$	TKT
10	VPDH	$PYR + NAD \rightarrow ACCOA + NADH + CO2$	$\max(\min(PDHA1, PDHA2), PDHB)$
11	VCS	$ACCOA + OXA \rightarrow CIT$	CS
12	VCITS	$CIT + NAD \rightarrow AKG + NADH + CO2$	$\min(\max(ACO1, ACO2), \max(IDH1, IDH2, \min(IDH3A, IDH3B, IDH3G)))$
13	VACL	$CIT + ATP \rightarrow OXA + ACCOA + ADP$	ACLY
14	VAKGDH	$AKG + NAD + ADP + Pi \rightarrow SUC + NADH + ATP + CO2$	$\min(OGDH, SUCLA2)$
15	VSDH	$SUC + 0.66 NAD \rightarrow MAL + 0.66 NADH$	$\min(SDHB, SDHC, SDHD)$
16	VMLD	$MAL + NAD \rightarrow OXA + NADH$	$\max(MDH1, MDH2)$
17	VPC	$PYR + ATP + CO2 \rightarrow OXA + ADP + Pi$	PC
18	VME	$MAL + NADP \rightarrow PYR + NADPH$	$\max(ME1, ME2)$
19	VGlnT	$EGLN + ADP + Pi \leftrightarrow GLU + ATP + NH4$	GLUL
20	VGLDH	$GLU + NAD \leftrightarrow AKG + NADH + NH4$	$\max(GLUD1, GLUD2)$
21	VAlaTA	$GLU + PYR \leftrightarrow AKG + ALA$	GPT
22	VASN	$ASN \rightarrow ASP + NH4$	
23	VASTA	$ASP + AKG \rightarrow OXA + GLU$	$\max(GOT1, GOT2)$
24	VAAtoSUC	$LYS + ILE + VAL + TYR + 7 AKG + ATP + 9 NAD + 2 NADP \rightarrow 4 GLU + 3 SUC + MAL + 8 ACCOA + ADP + 9 NADH + 2 NADPH + CO2$	$\max(BCAT1, BCAT2)$
25	VHISARGTA	$HIS + ARG + AKG \rightarrow 4 GLU + NH4 + CO2$	$\min(HAL, \max(ARG1, ARG2))$
26	VGluT	$GLU \leftrightarrow EGLU$	$\max(SLC1A1, SLC1A2, SLC1A3)$
27	VSDHH	$SER \rightarrow PYR + NH4$	SDS
28	VATPase	$ATP \rightarrow ADP + Pi$	ATP1A1
29	VNADPHox	$NADPH \rightarrow NADP$	$\max(NOX1, NOX3, NOX4, NOX5)$

30	Vresp	2 NADH + 4.2 ADP + O ₂ + 4.2 Pi → 2 NAD + 4.2 ATP	min(COX10, COX11, COX15, COX16, COX17P1//COX17, COX4I1, COX5A, COX5B, COX6A1P2//COX6A1, COX6A2, COX6B1, COX6C, COX7A1, COX7A2, COX7A2L, COX7B, COX7C, COX8A)
31	Vleak	2 NADH → 2 NAD	MT-ND5//MTND5P11
32	VAK	ATP + AMP ↔ 2 ADP	max(AK1, AK2, AK4, AK5)
33	VPPRibP	2 EGLN + 0.6 R5P + 2 ASP + GLY + 2 ATP → 2 GLU + 2 MAL + AMP + 2 ADP	ADCY1
34	VPALM	8 ACCOA + 7 ATP + 14 NADPH → PALM + 7 ADP + 14 NADP	FASN
35	Vgrowth	0.06 ALA + 0.04 ARG + 0.04 ASN + 0.03 ASP + 0.02 EGLN + 0.04 EGLU + 0.06 GLY + 0.02 HIS + 0.09 ILE + 0.06 LYS + 0.01 MET + 0.02 PHE + 0.03 PRO + 0.05 SER + 0.04 THR + 0.02 TYR + 0.04 VAL + 3.78 ATP + 0.03 G6P + 0.03 R5P + 0.09 PALM → X + 3.78 ADP	moy(HRAS, KRAS, NRAS, ERBB2, BRAF, PIK3CA, EGFR, CCND1) / moy(TP53, RB1, BRCA1, BRCA2, PTEN//PTENP1, APC, NF1, VHL, WT1, CDKN2A, SMAD4)

Tableau A.3 Noms complets des gènes associés au modèle métabolique

Symbol	Nom complet	Symbol	Nom complet
ACO1	Aconitase 1	HRAS	HRas proto-oncogene, GTPase
ACO2	Aconitase 2	IDH1	Isocitrate dehydrogenase (NADP(+)) 1
ACLY	ATP citrate lyase	IDH2	Isocitrate dehydrogenase (NADP(+)) 2
ADCY1	Adenylate cyclase 1	IDH3A	Isocitrate dehydrogenase (NAD(+)) 3 catalytic subunit alpha
AK1	Adenylate kinase 1	IDH3B	Isocitrate dehydrogenase (NAD(+)) 3 non-catalytic subunit beta
AK2	Adenylate kinase 2	IDH3G	Isocitrate dehydrogenase (NAD(+)) 3 non-catalytic subunit gamma
AK4	Adenylate kinase 4	KRAS	KRas proto-oncogene, GTPase
AK5	Adenylate kinase 5	LDHA	Lactate dehydrogenase A

ALDOA	Aldolase A, fructose-bisphosphate	LDHB	Lactate dehydrogenase B
ALDOB	Aldolase B, fructose-bisphosphate	MDH1	Malate dehydrogenase 1
ALDOC	Aldolase C, fructose-bisphosphate	MDH2	Malate dehydrogenase 2
APC	APC regulator of WNT signaling pathway	ME1	Malic enzyme 1
ARG1	Arginase 1	ME2	Malic enzyme 2
ARG2	Arginase 2	MT-ND5	Mitochondrially encoded NADH dehydrogenase 5
ATP1A1	ATPase Na+/K+ transporting subunit alpha 1	MTND5P11	Mitochondrially encoded NADH dehydrogenase 5 pseudogene 11
BCAT1	Branched chain amino acid transaminase 1	NF1	Neurofibromin 1
BCAT2	Branched chain amino acid transaminase 2	NOX1	NADPH oxidase 1
BRAF	B-Raf proto-oncogene, serine/threonine kinase	NOX3	NADPH oxidase 3
BRCA1	BRCA1 DNA repair associated	NOX4	NADPH oxidase 4
BRCA2	BRCA2 DNA repair associated	NOX5	NADPH oxidase 5
CCND1	Cyclin D1	NRAS	NRAS proto-oncogene, GTPase
CDKN2A	Cyclin-dependent kinase inhibitor 2A	OGDH	2-oxoglutarate dehydrogenase
COX10	Cytochrome c oxidase assembly protein COX10	PC	Pyruvate carboxylase
COX11	Cytochrome c oxidase assembly protein COX11	PFKL	Phosphofructokinase, liver type
COX15	Cytochrome c oxidase assembly protein COX15	PFKM	Phosphofructokinase, muscle type
COX16	Cytochrome c oxidase assembly protein COX16	PFKP	Phosphofructokinase, platelet type
COX17	Cytochrome c oxidase copper chaperone COX17	PGD	Phosphogluconate dehydrogenase
COX4I1	Cytochrome c oxidase subunit 4 isoform 1	PGK1	Phosphoglycerate kinase 1
COX5A	Cytochrome c oxidase subunit 5A	PGLS	Phosphogluconolactonase
COX5B	Cytochrome c oxidase subunit 5B	PGM1	Phosphoglycerate mutase 1
COX6A1	Cytochrome c oxidase subunit 6A1	PGM3	Phosphoglycerate mutase 3
COX6A2	Cytochrome c oxidase subunit 6A2	PGM5	Phosphoglycerate mutase 5

COX6B1	Cytochrome c oxidase subunit 6B1	PIK3CA	Phosphoinositide-3-kinase catalytic subunit alpha
COX6C	Cytochrome c oxidase subunit 6C	PKLR	Pyruvate kinase, liver and red blood cell
COX7A1	Cytochrome c oxidase subunit 7A1	PKM	Pyruvate kinase, muscle
COX7A2	Cytochrome c oxidase subunit 7A2	PTEN	Phosphatase and tensin homolog
COX7A2L	Cytochrome c oxidase subunit 7A2 like	PTENP1	Phosphatase and tensin homolog pseudogene 1
COX7B	Cytochrome c oxidase subunit 7B	PDHA1	Pyruvate dehydrogenase alpha 1
COX7C	Cytochrome c oxidase subunit 7C	PDHA2	Pyruvate dehydrogenase alpha 2
COX8A	Cytochrome c oxidase subunit 8A	PDHB	Pyruvate dehydrogenase beta
CS	Citrate synthase	RPE	Retinal pigment epithelium-specific protein
EGFR	Epidermal growth factor receptor	RPIA	Ribose-5-phosphate isomerase A
ENO1	Enolase 1	RB1	Retinoblastoma 1
ENO2	Enolase 2	SDS	Sucrose-isomaltase
ENO3	Enolase 3	SLC1A1	Solute carrier family 1 member 1
FASN	Fatty acid synthase	SLC1A2	Solute carrier family 1 member 2
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	SLC1A3	Solute carrier family 1 member 3
GLUD1	Glutamate dehydrogenase 1	SMAD4	SMAD family member 4
GLUD2	Glutamate dehydrogenase 2	SDHB	Succinate dehydrogenase complex iron sulfur subunit B
GLUL	Glutamate-ammonia ligase	SDHC	Succinate dehydrogenase complex subunit C
GPI	Glucose-6-phosphate isomerase	SDHD	Succinate dehydrogenase complex subunit D
GPT	Glutamic-pyruvic transaminase	SUCLA2	Succinate-CoA ligase ADP-forming subunit beta
GOT1	Glutamate oxaloacetate transaminase 1	TKT	Transketolase
GOT2	Glutamate oxaloacetate transaminase 2	TP53	Tumor protein p53
G6PD	Glucose-6-phosphate dehydrogenase	VHL	Von Hippel-Lindau tumor suppressor
HAL	Histidine ammonia-lyase	WT1	WT1 transcription factor
HK1	Hexokinase 1		

Tableau A.4 Cinétiques enzymatiques des vitesses de réactions

No	Équation
1	$v_{HK} = v_{max,HK} * \frac{[ATP]}{K_m,ADP + \frac{[ATP]}{[ADP]}} * \frac{[EGLC] \left(1 + \frac{\beta_{AMP} \frac{[AMP]}{[ATP]}}{\alpha_{AMP} K_{A,ATP}} \right)}{K_m,EGLC \left(1 + \frac{[AMP]}{K_{A,ATP}} \right) + [EGLC] \left(1 + \frac{[AMP]}{K_{A,ATP}} \right)} * \frac{K_{i,G6P}}{K_{i,G6P} + [G6P]}$
2	$v_{PGI} = v_{max,PGI} * \frac{[G6P]}{K_m,G6P + [G6P]} * \frac{K_{i,PEP}}{K_{i,PEP} + [PEP]} - v_{max,r,PGI} * \frac{[F6P]}{K_m,F6P + [F6P]}$
3	$v_{PFK} = v_{max,PFK} * \frac{[ATP]}{K_m,ADP + \frac{[ATP]}{[ADP]}} * \frac{[F6P] \left(1 + \frac{\beta_{AMP} \frac{[AMP]}{[ATP]}}{\alpha_{AMP} K_{A,ATP}} \right)}{K_m,F6P \left(1 + \frac{[AMP]}{K_{A,ATP}} \right) + [F6P] \left(1 + \frac{[AMP]}{K_{A,ATP}} \right)} * \frac{K_{i,CIT}}{K_{i,CIT} + [CIT]} * \frac{K_{i,LAC}}{K_{i,LAC} + [LAC]}$
4	$v_{PGK} = v_{max,PGK} * \frac{[ADP]}{K_m,ADP + \frac{[ADP]}{[ATP]}} * \frac{[GAP]}{K_m,GAP + [GAP]} * \frac{[NAD]}{K_m,NAD + \frac{[NAD]}{[NADH]}}$
5	$v_{PK} = v_{max,PK} * \frac{[ADP]}{K_m,ADP + \frac{[ADP]}{[ATP]}} * \frac{[PEP] \left(1 + \frac{\beta_{F6P} [F6P]}{\alpha_{F6P} K_{A,F6P}} \right)}{K_m,PEP \left(1 + \frac{[F6P]}{K_{A,F6P}} \right) + [PEP] \left(1 + \frac{[F6P]}{K_{A,F6P}} \right)} * \frac{K_{i,ALA}}{K_{i,ALA} + [ALA]}$
6	$v_{LDH} = v_{max,LDH} * \frac{[NADH]}{K_m,NADH + \frac{[NADH]}{[NAD]}} * \frac{[PYR] \left(1 + \frac{\beta_{AMP} \frac{[AMP]}{[ATP]}}{\alpha_{AMP} K_{A,ATP}} \right)}{K_m,PYR \left(1 + \frac{[AMP]}{K_{A,ATP}} \right) + [PYR] \left(1 + \frac{[AMP]}{K_{A,ATP}} \right)} - v_{max,r,LDH} * \frac{[NAD]}{K_m,NAD + \frac{[NAD]}{[NADH]}} * \frac{[LAC]}{K_m,LAC + [LAC]} * \frac{K_{i,PYR}}{K_{i,PYR} + [PYR]}$
7	$v_{G6PDH} = v_{max,G6PDH} * \frac{[G6P]}{K_m,G6P + [G6P]} * \frac{[NADP]}{K_m,NADP + \frac{[NADP]}{[NADPH]}}$
8	$v_{EP} = v_{max,EP} * \frac{[R5P]}{K_m,R5P + [R5P]}$
9	$v_{TK} = v_{max,TK} * \frac{[R5P]}{K_m,R5P + [R5P]} * \frac{[X5P]}{K_m,X5P + [X5P]}$
10	$v_{PDH} = v_{max,PDH} * \frac{[NAD]}{K_m,NAD + \frac{[NAD]}{[NADH]}} * \frac{[PYR]}{K_m,PYR + [PYR]}$
11	$v_{CS} = v_{max,CS} * \frac{[ACCOA]}{K_m,ACCOA + [ACCOA]} * \frac{[OXA]}{K_m,OXA + [OXA]}$

12	$v_{CITS} = v_{max,CITS} * \frac{[CIT]}{K_{m,CIT} + [CIT]} * \frac{\frac{[NAD]}{[NADH]}}{K_{m,\frac{NAD}{NADH}} + \frac{[NAD]}{[NADH]}}$
13	$v_{ACL} = v_{max,ACL} * \frac{\frac{[ATP]}{[ADP]}}{K_{m,\frac{ATP}{ADP}} + \frac{[ATP]}{[ADP]}} * \frac{[CIT]}{K_{m,CIT} + [CIT]}$
14	$v_{AKGDH} = v_{max,AKGDH} * \frac{\frac{[ADP]}{[ATP]}}{K_{m,\frac{ADP}{ATP}} + \frac{[ADP]}{[ATP]}} * \frac{[AKG]}{K_{m,AKG} + [AKG]} * \frac{\frac{[NAD]}{[NADH]}}{K_{m,\frac{NAD}{NADH}} + \frac{[NAD]}{[NADH]}} * \frac{[Pi]}{K_{m,Pi} + [Pi]}$
15	$v_{SDH} = v_{max,SDH} * \frac{\frac{[NAD]}{[NADH]}}{K_{m,\frac{NAD}{NADH}} + \frac{[NAD]}{[NADH]}} * \frac{[SUC]}{K_{m,SUC} + [SUC]}$
16	$v_{MLD} = v_{max,MLD} * \frac{[MAL]}{K_{m,MAL} + [MAL]} * \frac{\frac{[NAD]}{[NADH]}}{K_{m,\frac{NAD}{NADH}} + \frac{[NAD]}{[NADH]}}$
17	$v_{PC} = v_{max,PC} * \frac{\frac{[ATP]}{[ADP]}}{K_{m,\frac{ATP}{ADP}} + \frac{[ATP]}{[ADP]}} * \frac{[CO2]}{K_{m,CO2} + [CO2]} * \frac{[PYR]}{K_{m,PYR} + [PYR]}$
18	$v_{ME} = v_{max,ME} * \frac{[MAL]}{K_{m,MAL} + [MAL]} * \frac{\frac{[NADP]}{[NADPH]}}{K_{m,\frac{NADP}{NADPH}} + \frac{[NADP]}{[NADPH]}}$
19	$v_{GlnT} = v_{max,GlnT} * \frac{\frac{[ADP]}{[ATP]}}{K_{m,\frac{ADP}{ATP}} + \frac{[ADP]}{[ATP]}} * \frac{[Pi]}{K_{m,Pi} + [Pi]} * \frac{[EGLN]}{K_{m,EGLN} + [EGLN]} - v_{max,r,GlnT} * \frac{\frac{[ATP]}{[ADP]}}{K_{m,\frac{ATP}{ADP}} + \frac{[ATP]}{[ADP]}} * \frac{[GLU]}{K_{m,GLU} + [GLU]} * \frac{[NH4]}{K_{m,NH4} + [NH4]}$
20	$v_{GLDH} = v_{max,GLDH} * \frac{[GLU]}{K_{m,GLU} + [GLU]} * \frac{\frac{[NAD]}{[NADH]}}{K_{m,\frac{NAD}{NADH}} + \frac{[NAD]}{[NADH]}} - v_{max,r,GLDH} * \frac{[AKG]}{K_{m,AKG} + [AKG]} * \frac{\frac{[NADH]}{[NAD]}}{K_{m,\frac{NADH}{NAD}} + \frac{[NADH]}{[NAD]}} * \frac{[NH4]}{K_{m,NH4} + [NH4]}$
21	$v_{AlaTA} = v_{max,AlaTA} * \frac{[GLU]}{K_{m,GLU} + [GLU]} * \frac{[PYR]}{K_{m,PYR} + [PYR]} - v_{max,r,AlaTA} * \frac{[AKG]}{K_{m,AKG} + [AKG]} * \frac{[ALA]}{K_{m,ALA} + [ALA]} * \frac{K_{i,EGLN}}{K_{i,EGLN} + [EGLN]}$
22	$v_{ASN} = v_{max,ASN} * \frac{[ASN]}{K_{m,ASN} + [ASN]}$
23	$v_{ASTA} = v_{max,ASTA} * \frac{[AKG]}{K_{m,AKG} + [AKG]} * \frac{[ASP]}{K_{m,ASP} + [ASP]}$
24	$v_{AAtoSUC} = v_{max,AAtoSUC} * \frac{[AKG]}{K_{m,AKG} + [AKG]} * \frac{\frac{[ATP]}{[ADP]}}{K_{m,\frac{ATP}{ADP}} + \frac{[ATP]}{[ADP]}} * \frac{\frac{[NAD]}{[NADH]}}{K_{m,\frac{NAD}{NADH}} + \frac{[NAD]}{[NADH]}} * \frac{\frac{[NADP]}{[NADPH]}}{K_{m,\frac{NADP}{NADPH}} + \frac{[NADP]}{[NADPH]}} * \frac{[ILE]}{K_{m,ILE} + [ILE]} * \frac{[LYS]}{K_{m,LYS} + [LYS]} * \frac{[TYR]}{K_{m,TYR} + [TYR]} * \frac{[VAL]}{K_{m,VAL} + [VAL]}$
25	$v_{HISARGTA} = v_{max,HISARGTA} * \frac{[AKG]}{K_{m,AKG} + [AKG]} * \frac{[ARG]}{K_{m,ARG} + [ARG]} * \frac{[HIS]}{K_{m,HIS} + [HIS]}$

26	$v_{GluT} = v_{max,GluT} * \frac{[GLU]}{K_{m,GLU} + [GLU]} - v_{max,r,GluT} * \frac{[EGLU]}{K_{m,EGLU} + [EGLU]}$
27	$v_{SDHH} = v_{max,SDHH} * \frac{[SER]}{K_{m,SER} + [SER]}$
28	$v_{ATPase} = v_{max,ATPase} * \frac{[ATP]}{K_{m,ATP} + [ATP]}$
29	$v_{NADPHox} = v_{max,NADPHox} * \frac{[NADPH]}{K_{m,NADPH} + [NADPH]}$
30	$v_{resp} = v_{max,resp} * \frac{\frac{[ADP]}{[ATP]}}{K_{m,ATP} + \frac{[ADP]}{[ATP]}} * \frac{\frac{[NADH]}{[NAD]}}{K_{m,NADH} + \frac{[NADH]}{[NAD]}} * \frac{[O2]}{K_{m,O2} + [O2]} * \frac{[Pi]}{K_{m,Pi} + [Pi]}$
31	$v_{leak} = v_{max,leak} * \frac{[NADH]}{K_{m,NADH} + [NADH]}$
32	$v_{AK} = v_{max,AK} * \frac{[AMP]}{K_{m,AMP} + [AMP]} * \frac{[ATP]}{K_{m,ATP} + [ATP]} - v_{max,r,AK} * \frac{[ADP]}{K_{m,ADP} + [ADP]}$
33	$v_{PPRibP} = v_{max,PPRibP} * \frac{[ATP]}{K_{m,ATP} + [ATP]} * \frac{[R5P]}{K_{m,R5P} + [R5P]} * \frac{[ASP]}{K_{m,ASP} + [ASP]} * \frac{[EGLN]}{K_{m,EGLN} + [EGLN]}$ $* \frac{[GLY]}{K_{m,GLY} + [GLY]}$
34	$v_{PALM} = v_{max,PALM} * \frac{[ACCOA]}{K_{m,ACCOA} + [ACCOA]} * \frac{\frac{[ATP]}{[ADP]}}{K_{m,ADP} + \frac{[ATP]}{[ADP]}} * \frac{\frac{[NADPH]}{[NADP]}}{K_{m,NADP} + \frac{[NADPH]}{[NADP]}}$
35	$v_{growth} = v_{max,growth} * \frac{[ATP]}{K_{m,g,ATP} + [ATP]} * \frac{[G6P]}{K_{m,g,G6P} + [G6P]} * \frac{[PALM]}{K_{m,g,PALM} + [PALM]} * \frac{[R5P]}{K_{m,g,R5P} + [R5P]}$ $* \frac{[ALA]}{K_{m,g,ALA} + [ALA]} * \frac{[ARG]}{K_{m,g,ARG} + [ARG]} * \frac{[ASN]}{K_{m,g,ASN} + [ASN]} * \frac{[ASP]}{K_{m,g,ASP} + [ASP]}$ $* \frac{[EGLN]}{K_{m,g,EGLN} + [EGLN]} * \frac{[EGLU]}{K_{m,g,EGLU} + [EGLU]} * \frac{[GLY]}{K_{m,g,GLY} + [GLY]} * \frac{[HIS]}{K_{m,g,HIS} + [HIS]}$ $* \frac{[ILE]}{K_{m,g,ILE} + [ILE]} * \frac{[LYS]}{K_{m,g,LYS} + [LYS]} * \frac{[MET]}{K_{m,g,MET} + [MET]} * \frac{[PHE]}{K_{m,g,PHE} + [PHE]}$ $* \frac{[PRO]}{K_{m,g,PRO} + [PRO]} * \frac{[SER]}{K_{m,g,SER} + [SER]} * \frac{[THR]}{K_{m,g,THR} + [THR]} * \frac{[TYR]}{K_{m,g,TYR} + [TYR]}$ $* \frac{[VAL]}{K_{m,g,VAL} + [VAL]} * \frac{K_{i,LAC}}{K_{i,LAC} + [LAC]} * \frac{K_{i,NH4}}{K_{i,NH4} + [NH4]}$

Tableau A.5 Vitesses maximales de réaction

No	Vitesse maximale	Unité
1	$v_{max,HK} = 1 * 10^{-2} * FC_{HK} * (1 + f_{moy,ATP}(t))$	mM/h
2a	$v_{max,PGI} = 7,6 * 10^{-4} * FC_{PGI}$	mM/h
2b	$v_{max,r,PGI} = 3,5 * 10^{-4} * FC_{PGI}$	mM/h
3	$v_{max,PFK} = 7,5 * 10^{-5} * FC_{PFK} * (1 + f_{moy,ATP}(t))$	mM/h

4	$v_{max,PGK} = 1,07 * 10^{-3} * FC_{PGK} * (1 + f_{moy,ADP}(t) + f_{moy,NAD}(t))$	mM/h
5	$v_{max,PK} = 1,38 * 10^{-3} * FC_{PK} * (1 + f_{moy,ADP}(t))$	mM/h
6a	$v_{max,LDH} = 6 * 10^{-9} * FC_{LDH} * (1 + f_{moy,NADH}(t))$	mM/h
6b	$v_{max,r,LDH} = 1,9998 * 10^{-9} * FC_{LDH} * (1 + f_{moy,NAD}(t))$	mM/h
7	$v_{max,G6PDH} = 1,5 * 10^{-5} * FC_{G6PDH} * (1 + f_{moy,NADP}(t))$	mM/h
8	$v_{max,EP} = 1,3 * 10^{-5} * FC_{EP}$	mM/h
9	$v_{max,TK} = 2,3 * 10^{-5} * FC_{TK}$	mM/h
10	$v_{max,PDH} = 7,1 * 10^{-5} * FC_{PDH} * (1 + f_{moy,NAD}(t))$	mM/h
11	$v_{max,CS} = 8,8 * 10^{-5} * FC_{CS}$	mM/h
12	$v_{max,CITS} = 3,9 * 10^{-5} * FC_{CITS} * (1 + f_{moy,NAD}(t))$	mM/h
13	$v_{max,ACL} = 2,4 * 10^{-5} * FC_{ACL} * (1 + f_{moy,ATP}(t))$	mM/h
14	$v_{max,AKGDH} = 2,6 * 10^{-4} * FC_{AKGDH} * (1 + f_{moy,ADP}(t) + f_{moy,NAD}(t))$	mM/h
15	$v_{max,SDH} = 4 * 10^{-4} * FC_{SDH} * (1 + f_{moy,NAD}(t))$	mM/h
16	$v_{max,MLD} = 9,5 * 10^{-5} * FC_{MLD} * (1 + f_{moy,NAD}(t))$	mM/h
17	$v_{max,PC} = 2,1 * 10^{-4} * FC_{PC} * (1 + f_{moy,ATP}(t))$	mM/h
18	$v_{max,ME} = 1,3 * 10^{-5} * FC_{ME} * (1 + f_{moy,NADP}(t))$	mM/h
19a	$v_{max,GlnT} = 2,54 * 10^{-6} * FC_{GlnT} * (1 + f_{moy,ADP}(t))$	mM/h
19b	$v_{max,r,GlnT} = 1,9 * 10^{-5} * FC_{GlnT} * (1 + f_{moy,ATP}(t))$	mM/h
20a	$v_{max,GLDH} = 1,2 * 10^{-6} * FC_{GLDH} * (1 + f_{moy,NAD}(t))$	mM/h
20b	$v_{max,r,GLDH} = 6,3 * 10^{-5} * FC_{GLDH} * (1 + f_{moy,NADH}(t))$	mM/h
21a	$v_{max,AlaTA} = 1,7 * 10^{-4} * FC_{AlaTA}$	mM/h
21b	$v_{max,r,AlaTA} = 2,4 * 10^{-4} * FC_{AlaTA}$	mM/h
22	$v_{max,ASN} = 6,2 * 10^{-6}$	mM/h
23	$v_{max,ASTA} = 7,8 * 10^{-5} * FC_{ASTA}$	mM/h
24	$v_{max,AAtoSUC} = 1,3 * 10^{-4} * FC_{AAtoSUC} * (1 + f_{moy,NADP}(t) + f_{moy,NAD}(t) + f_{moy,ATP}(t))$	mM/h
25	$v_{max,HISARGTA} = 1,9 * 10^{-5} * FC_{HISARGTA}$	mM/h
26a	$v_{max,GluT} = 1,8 * 10^{-6} * FC_{GluT}$	mM/h
26b	$v_{max,r,GluT} = 2,5 * 10^{-6} * FC_{GluT}$	mM/h
27	$v_{max,SDHH} = 8,5 * 10^{-6} * FC_{SDHH}$	mM/h
28	$v_{max,ATPase} = 9,14 * 10^{-4} * FC_{ATPase} * (1 + f_{moy,ATP}(t))$	mM/h

29	$v_{max,NADPHox} = 1,4 * 10^{-5} * FC_{NADPHox} * (1 + f_{moy,NADPH}(t))$	mM/h
30	$v_{max,resp} = 8,9 * 10^{-3} * FC_{resp} * (1 + f_{moy,ADP}(t) + f_{moy,NADH}(t))$	mM/h
31	$v_{max,leak} = 2,9 * 10^{-5} * FC_{leak} * (1 + f_{moy,NADH}(t))$	mM/h
32a	$v_{max,AK} = 1,8 * 10^{-8} * FC_{AK} * (1 + f_{moy,ATP}(t))$	mM/h
32b	$v_{max,r,AK} = 1,2 * 10^{-8} * FC_{AK} * (1 + f_{moy,ADP}(t))$	mM/h
33	$v_{max,PPRibP} = 2,6 * 10^{-9} * FC_{PPribP} * (1 + f_{moy,ATP}(t))$	mM/h
34	$v_{max,PALM} = 2,95 * 10^{-5} * FC_{PALM} * (1 + f_{moy,ATP}(t) + f_{moy,NADPH}(t))$	mM/h
35	$v_{max,growth} = 2 * 10^{-3} * FC_{growth} * (1 + f_{moy,ATP}(t))$	1 / h

Tableau A.6 Constantes cinétiques du modèle métabolique

Constante cinétique	Valeur	Constante cinétique	Valeur
K _{m,EGLC}	3 mM	K _{i,NH4}	4,44 mM
K _{m,EGLN}	1 x 10 ⁻¹ mM	K _{A,AMP/ATP}	9,03776 x 10 ⁻²
K _{m,LAC}	1 x 10 ⁻¹ mM	K _{A,F6P}	3,12102 mM
K _{m,NH4}	1 x 10 ⁻² mM	α _{AMP/ATP}	4,712 x 10 ⁻¹
K _{m,ALA}	1 x 10 ⁻² mM	β _{AMP/ATP}	10,4557
K _{m,ARG}	7,3 x 10 ⁻² mM	α _{F6P}	4,11128
K _{m,ASP}	4 x 10 ⁻² mM	β _{F6P}	1,74503
K _{m,ASN}	1 x 10 ⁻² mM	K _{m,g,ASP}	1,03225 x 10 ⁻² mM
K _{m,HIS}	5,8 x 10 ⁻¹ mM	K _{m,g,ASN}	4,2 x 10 ⁻³ mM
K _{m,ILE}	3,8 x 10 ⁻² mM	K _{m,g,HIS}	1,68556 x 10 ⁻² mM
K _{m,LYS}	9,8 x 10 ⁻² mM	K _{m,g,ARG}	1,46417 x 10 ⁻² mM
K _{m,SER}	2,2 x 10 ⁻¹ mM	K _{m,g,LYS}	2,74893 x 10 ⁻² mM
K _{m,TYR}	4,8 x 10 ⁻² mM	K _{m,g,ILE}	1,01028 x 10 ⁻² mM
K _{m,GLY}	5 x 10 ⁻² mM	K _{m,g,VAL}	1,33139 x 10 ⁻² mM
K _{m,VAL}	1 x 10 ⁻² mM	K _{m,g,TYR}	1,15531 x 10 ⁻² mM
K _{m,G6P}	1,82548 x 10 ⁻² mM	K _{m,g,MET}	3,51315 x 10 ⁻² mM
K _{m,F6P}	1 x 10 ⁻² mM	K _{m,g,SER}	1,07979 x 10 ⁻³ mM
K _{m,GAP}	1 x 10 ⁻¹ mM	K _{m,g,THR}	9,60848 x 10 ⁻³ mM
K _{m,PEP}	5 x 10 ⁻² mM	K _{m,g,PRO}	1,01713 x 10 ⁻² mM
K _{m,R5P}	4,4586 x 10 ⁻² mM	K _{m,g,PHE}	1,63977 x 10 ⁻² mM
K _{m,X5P}	7,00637 x 10 ⁻³ mM	K _{m,g,GLY}	1 x 10 ⁻² mM
K _{m,PYR}	1 x 10 ⁻¹ mM	K _{m,g,R5P}	7,17994 x 10 ⁻⁵ mM
K _{m,ACCOA}	1 x 10 ⁻² mM	K _{m,g,G6P}	6,30901 x 10 ⁻⁷ mM
K _{m,OXA}	1,40127 x 10 ⁻¹ mM	K _{m,g,ATP}	9,15637 x 10 ⁻³ mM
K _{m,CIT}	1,01911 x 10 ⁻¹ mM	K _{m,g,ALA}	7,56376 x 10 ⁻² mM
K _{m,AKG}	7,00637 x 10 ⁻¹ mM	K _{m,g,EGLN}	1,39129 x 10 ⁻³ mM
K _{m,SUC}	1,46497 x 10 ⁻¹ mM	K _{m,g,EGLU}	1,68472 x 10 ⁻⁶ mM
K _{m,MAL}	5 x 10 ⁻¹ mM	K _{m,g,PALM}	2,54777 x 10 ⁻² mM
K _{m,EGLU}	1 x 10 ⁻¹ mM	V _{g,R5P}	7,35 x 10 ⁻⁵ mM
K _{m,GLU}	1 x 10 ⁻² mM	V _{g,G6P}	8,79 x 10 ⁻⁵ mM
K _{m,CO2}	1 x 10 ⁻¹ mM	V _{g,PALM}	7,35 x 10 ⁻⁵ mM

$K_{m,O2}$	$1 \times 10^{-2} \text{ mM}$	$V_g,EGLU$	$1,22 \times 10^{-4} \text{ mM}$
$K_{m,Pi}$	$6,36943 \times 10^{-1} \text{ mM}$	$V_g,EGLN$	$1,01 \times 10^{-4} \text{ mM}$
$K_{m,ATP}$	1 mM	V_g,ALA	$1,89 \times 10^{-4} \text{ mM}$
$K_{m,ADP}$	$2 \times 10^{-1} \text{ mM}$	V_g,ARG	$1,19 \times 10^{-4} \text{ mM}$
$K_{m,AMP}$	$1,52866 \times 10^{-3} \text{ mM}$	V_g,ASN	$9,07 \times 10^{-5} \text{ mM}$
$K_{m,NADH}$	$1 \times 10^{-1} \text{ mM}$	V_g,ASP	$1,13 \times 10^{-4} \text{ mM}$
$K_{m,NADPH}$	$5,35032 \times 10^{-2} \text{ mM}$	V_g,GLY	$1,69 \times 10^{-4} \text{ mM}$
$K_{m,ATP/ADP}$	5	V_g,HIS	$4,5 \times 10^{-5} \text{ mM}$
$K_{m,ADP/ATP}$	2×10^{-1}	V_g,ILE	$1,02 \times 10^{-4} \text{ mM}$
$K_{m,NADH/NAD}$	$8,3$	V_g,LYS	$1,8 \times 10^{-4} \text{ mM}$
$K_{m,NAD/NADH}$	$4,8 \times 10^{-2}$	V_g,MET	$4,35 \times 10^{-5} \text{ mM}$
$K_{m,NADP/NADPH}$	2	V_g,PHE	$6,9 \times 10^{-5} \text{ mM}$
$K_{m,NADPH/NADP}$	5×10^{-1}	V_g,PRO	$9,86 \times 10^{-5} \text{ mM}$
$K_{i,G6P}$	$1,08971 \times 10^{-2} \text{ mM}$	V_g,SER	$1,65 \times 10^{-4} \text{ mM}$
$K_{i,PEP}$	$2,35669 \times 10^{-1} \text{ mM}$	V_g,THR	$1,22 \times 10^{-4} \text{ mM}$
$K_{i,PYR}$	$2,86624 \times 10^{-1} \text{ mM}$	V_g,TYR	$5,37 \times 10^{-5} \text{ mM}$
$K_{i,CIT}$	$471,338 \text{ mM}$	V_g,VAL	$1,31 \times 10^{-4} \text{ mM}$
$K_{i,EGLN}$	$3,5 \text{ mM}$	V_g,ATP	$1,19 \times 10^{-2} \text{ mM}$
$K_{i,ALA}$	$22,7872 \text{ mM}$	V_g,ADP	$1,19 \times 10^{-2} \text{ mM}$
$K_{i,LAC}$	170 mM	F	$8,663 \times 10^{-5} \text{ L/h}$

Tableau A.7 Équations différentielles du modèle métabolique

No	Équation différentielle
0*	$\mu_{blood} = \left(1 - \frac{X}{VT}\right) * v_{growth}$
1	$\frac{d[ACCOA]}{dt} = v_{PDH} - v_{CS} + v_{ACL} + (8 * v_{AAtoSUC}) - (8 * v_{PALM}) - (v_{growth} * [ACCOA])$
2	$\frac{d[ADP]}{dt} = v_{HK} + v_{PFK} - v_{PGK} - v_{PK} + v_{ACL} - v_{AKGDH} + v_{PC} - v_{GlnT} + v_{AAtoSUC} + v_{ATPase} - (4,2 * v_{resp}) + (2 * v_{AK}) + (2 * v_{PPRiBP}) + (7 * v_{PALM}) - (v_{growth} * [ADP]) + (v_{g,ADP} * v_{growth})$
3	$\frac{d[AKG]}{dt} = v_{CITS} - v_{AKGDH} + v_{GLDH} + v_{AlaTA} - v_{ASTA} - v_{HISARGTA} - (7 * v_{AAtoSUC}) - (v_{growth} * [AKG])$
4	$\frac{d[AMP]}{dt} = -v_{AK} + v_{PPRiBP} - (v_{growth} * [AMP])$
5	$\frac{d[ATP]}{dt} = -v_{HK} - v_{PFK} + v_{PGK} + v_{PK} - v_{ACL} + v_{AKGDH} - v_{PC} + v_{GlnT} - v_{AAtoSUC} - v_{ATPase} + (4,2 * v_{resp}) - v_{AK} - (2 * v_{PPRiBP}) - (7 * v_{PALM}) - (v_{growth} * [ATP]) - (v_{g,ATP} * v_{growth})$
6	$\frac{d[CIT]}{dt} = v_{CS} - v_{CITS} - v_{ACL} - (v_{growth} * [CIT])$
7	$\frac{d[CO2]}{dt} = v_{G6PDH} + v_{PDH} + v_{CITS} - v_{PC} + v_{AAtoSUC} + v_{HISARGTA} - (v_{growth} * [CO2])$
8	$\frac{d[F6P]}{dt} = v_{PGI} - v_{PFK} + (2 * v_{TK}) - (v_{growth} * [F6P])$
9	$\frac{d[G6P]}{dt} = v_{HK} - v_{PGI} - v_{G6PDH} - (v_{growth} * [G6P]) - (v_{g,G6P} * v_{growth})$

10	$\frac{d[GAP]}{dt} = (2 * v_{PFK}) - v_{PGK} + v_{TK} - (v_{growth} * [GAP])$
11	$\frac{d[GLU]}{dt} = v_{GlnT} - v_{GLDH} - v_{AlaTA} + v_{ASTA} + (4 * v_{AAtoSUC}) + (4 * v_{HISARGTA}) - v_{Glut} + (2 * v_{PPRibP}) - (v_{growth} * [GLU])$
12	$\frac{d[MAL]}{dt} = v_{SDH} - v_{MLD} - v_{ME} + v_{AAtoSUC} + (2 * v_{PPRibP}) - (v_{growth} * [MAL])$
13	$\frac{d[NAD]}{dt} = -v_{PGK} + v_{LDH} - v_{PDH} - v_{CITS} - v_{AKGDH} - (0,66 * v_{SDH}) - v_{MLD} - v_{GLDH} - (9 * v_{AAtoSUC}) + (2 * v_{resp}) + (2 * v_{leak}) - (v_{growth} * [NAD])$
14	$\frac{d[NADH]}{dt} = v_{PGK} - v_{LDH} + v_{PDH} + v_{CITS} + v_{AKGDH} + (0,66 * v_{SDH}) + v_{MLD} + v_{GLDH} + (9 * v_{AAtoSUC}) - (2 * v_{resp}) - (2 * v_{leak}) - (v_{growth} * [NADH])$
15	$\frac{d[NADP]}{dt} = -(2 * v_{G6PDH}) - v_{ME} - (2 * v_{AAtoSUC}) + v_{NADPHox} + (14 * v_{PALM}) - (v_{growth} * [NADP])$
16	$\frac{d[NADPH]}{dt} = (2 * v_{G6PDH}) + v_{ME} + (2 * v_{AAtoSUC}) - v_{NADPHox} - (14 * v_{PALM}) - (v_{growth} * [NADPH])$
17	$\frac{d[O2]}{dt} = -v_{resp} - (v_{growth} * [O2])$
18	$\frac{d[OXA]}{dt} = -v_{CS} + v_{ACL} + v_{MLD} + v_{PC} + v_{ASTA} - (v_{growth} * [OXA])$
19	$\frac{d[PALM]}{dt} = v_{PALM} - (v_{growth} * [PALM]) - (v_{g,PALM} * v_{growth})$
20	$\frac{d[PEP]}{dt} = v_{PGK} - v_{PK} - (v_{growth} * [PEP])$
21	$\frac{d[PYR]}{dt} = v_{PK} - v_{LDH} - v_{PDH} - v_{PC} + v_{ME} - v_{AlaTA} + v_{SDHH} - (v_{growth} * [PYR])$
22	$\frac{d[Pi]}{dt} = -v_{AKGDH} + v_{PC} - v_{GlnT} + v_{ATPase} - (4,2 * v_{resp}) - (v_{growth} * [Pi])$
23	$\frac{d[R5P]}{dt} = v_{G6PDH} - v_{EP} - v_{TK} - (0,6 * v_{PPRibP}) - (v_{growth} * [R5P]) - (v_{g,R5P} * v_{growth})$
24	$\frac{d[SUC]}{dt} = v_{AKGDH} - v_{SDH} + (3 * v_{AAtoSUC}) - (v_{growth} * [SUC])$
25	$\frac{d[X5P]}{dt} = v_{EP} - (2 * v_{TK}) - (v_{growth} * [X5P])$
26	$\frac{d[ALA]}{dt} = \frac{(F * [ALA_0] - F * [ALA])}{V_{blood}} + \left(v_{AlaTA} * \frac{X}{V_{blood}} \right) - \left(v_{g,ALA} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [ALA])$
27	$\frac{d[ARG]}{dt} = \frac{(F * [ARG_0] - F * [ARG])}{V_{blood}} - \left(v_{HISARGTA} * \frac{X}{V_{blood}} \right) - \left(v_{g,ARG} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [ARG])$
28	$\frac{d[ASN]}{dt} = \frac{(F * [ASN_0] - F * [ASN])}{V_{blood}} - \left(v_{ASN} * \frac{X}{V_{blood}} \right) - \left(v_{g,ASN} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [ASN])$
29	$\frac{d[ASP]}{dt} = \frac{(F * [ASP_0] - F * [ASP])}{V_{blood}} + \left(v_{ASN} * \frac{X}{V_{blood}} \right) - \left(v_{ASTA} * \frac{X}{V_{blood}} \right) - \left(2 * v_{PPRibP} * \frac{X}{V_{blood}} \right) - \left(v_{g,ASP} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [ASP])$

30	$\frac{d[EGLC]}{dt} = \frac{(F * [EGLC_0] - F * [EGLC])}{V_{blood}} - \left(v_{HK} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [EGLC])$
31	$\frac{d[EGLN]}{dt} = \frac{(F * [EGLN_0] - F * [EGLN])}{V_{blood}} - \left(v_{GlnT} * \frac{X}{V_{blood}} \right) - \left(2 * v_{PPRibP} * \frac{X}{V_{blood}} \right) - \left(v_{g,EGLN} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [EGLN])$
32	$\frac{d[EGLU]}{dt} = \frac{(F * [EGLU_0] - F * [EGLU])}{V_{blood}} + \left(v_{GluT} * \frac{X}{V_{blood}} \right) - \left(v_{g,EGLU} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [EGLU])$
33	$\frac{d[GLY]}{dt} = \frac{(F * [GLY_0] - F * [GLY])}{V_{blood}} - \left(v_{PPRibP} * \frac{X}{V_{blood}} \right) - \left(v_{g,GLY} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [GLY])$
34	$\frac{d[HIS]}{dt} = \frac{(F * [HIS_0] - F * [HIS])}{V_{blood}} - \left(v_{HISARGTA} * \frac{X}{V_{blood}} \right) - \left(v_{g,HIS} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [HIS])$
35	$\frac{d[ILE]}{dt} = \frac{(F * [ILE_0] - F * [ILE])}{V_{blood}} - \left(v_{AAtoSUC} * \frac{X}{V_{blood}} \right) - \left(v_{g,ILE} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [ILE])$
36	$\frac{d[LAC]}{dt} = \frac{(F * [LAC_0] - F * [LAC])}{V_{blood}} + \left(v_{LDH} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [LAC])$
37	$\frac{d[LYS]}{dt} = \frac{(F * [LYS_0] - F * [LYS])}{V_{blood}} - \left(v_{AAtoSUC} * \frac{X}{V_{blood}} \right) - \left(v_{g,LYS} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [LYS])$
38	$\frac{d[MET]}{dt} = \frac{(F * [MET_0] - F * [MET])}{V_{blood}} - \left(v_{g,MET} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [MET])$
39	$\frac{d[NH4]}{dt} = \frac{(F * [NH4_0] - F * [NH4])}{V_{blood}} + \left((v_{GlnT} + v_{GLDH} + v_{ASN} + v_{HISARGTA} + v_{SDHH}) * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [NH4])$
40	$\frac{d[PHE]}{dt} = \frac{(F * [PHE_0] - F * [PHE])}{V_{blood}} - \left(v_{g,PHE} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [PHE])$
41	$\frac{d[PRO]}{dt} = \frac{(F * [PRO_0] - F * [PRO])}{V_{blood}} - \left(v_{g,PRO} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [PRO])$
42	$\frac{d[SER]}{dt} = \frac{(F * [SER_0] - F * [SER])}{V_{blood}} - \left(v_{SDHH} * \frac{X}{V_{blood}} \right) - \left(v_{g,SER} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [SER])$
43	$\frac{d[THR]}{dt} = \frac{(F * [THR_0] - F * [THR])}{V_{blood}} - \left(v_{g,THR} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [THR])$
44	$\frac{d[TYR]}{dt} = \frac{(F * [TYR_0] - F * [TYR])}{V_{blood}} - \left(v_{AAtoSUC} * \frac{X}{V_{blood}} \right) - \left(v_{g,TYR} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [TYR])$
45	$\frac{d[VAL]}{dt} = \frac{(F * [VAL_0] - F * [VAL])}{V_{blood}} - \left(v_{AAtoSUC} * \frac{X}{V_{blood}} \right) - \left(v_{g,VAL} * v_{growth} * \frac{X}{V_{blood}} \right) - (\mu_{blood} * [VAL])$
46	$\frac{dX}{dt} = v_{growth} * X$
47	$\frac{dV_{blood}}{dt} = \mu_{blood} * V_{blood}$
48	$\frac{dVT}{dt} = \frac{dX}{dt} + \frac{dV_{blood}}{dt}$

ANNEXE B RÉSUMÉ DES RÉSULTATS ET LIMITATIONS DU PROJET

Tableau B.1 Résumé des résultats et limitations des volets de la détection du cancer de l'ovaire, de la modélisation métabolique et du ciblage thérapeutique

	Principaux résultats	Limitations
Détection du cancer de l'ovaire	<ul style="list-style-type: none"> - Les niveaux d'expression génique permettent une excellente détection du cancer (99,91% avec LR), une faible classification du stade (37,77% avec LR) et une détection quasi nulle de la récurrence (53,57% avec Linear SVC). - L'algorithme de suréchantillonnage SMOTE améliore les performances de l'apprentissage supervisé lors d'un débordement entre classes. - Les hyperparamètres des algorithmes et le nombre de variables d'apprentissage influencent significativement les performances de l'apprentissage supervisé. 	<ul style="list-style-type: none"> - Le débordement entre les classes de la détection du cancer diminue la fiabilité des résultats. - L'impact du nombre de variables d'apprentissage peut être biaisé par l'absence de l'optimisation des hyperparamètres avec l'ensemble des 13 110 gènes. - Les données utilisées ne contiennent qu'un type de variables.
Modélisation métabolique	<ul style="list-style-type: none"> - La croissance tumorale, les concentrations de 35 métabolites et cinq ratios énergétiques sont modélisables sur une période d'au moins 270 jours. - En modifiant les volumes tumoraux initiaux et certains paramètres cinétiques macroscopiques, le modèle murin initial peut être transposé vers un modèle humain du cancer de l'ovaire. - Les niveaux d'expression génique influencent la croissance tumorale et les concentrations métaboliques, sans pour autant modifier leur allure générale. - Le stade d'avancement du cancer de l'ovaire influence notamment sa croissance ainsi que les concentrations des métabolites extracellulaires. 	<ul style="list-style-type: none"> - Le modèle métabolique est un modèle réduit et simplifié du métabolisme tumoral, et ne modélise pas la nécrose tumorale. - Le flux sanguin tumoral de $8,663 \times 10^{-5}$ L/h réduit la plausibilité du modèle métabolique. - La croissance tumorale et les fluctuations métaboliques ressorties par le modèle ne sont pas validées expérimentalement. - Les paramètres du coefficient FC et de la fonction temporelle F(t) de la cinétique enzymatique proposée restent à être validés et affinés. - Le volume sanguin <i>Vblood</i> présente des discontinuités entre les stades du cancer. - L'épuisement complet d'un ou de plusieurs métabolites extracellulaires limite la durée de simulation du modèle métabolique.

Ciblage thérapeutique	<ul style="list-style-type: none"> - Les algorithmes <i>Depth Search</i> et <i>Intelligent Search</i> permettent d'identifier des gènes cibles en fonction des données transcriptomiques d'une patiente. <i>Depth Search</i> assure une solution optimale tandis que <i>Intelligent Search</i> permet une recherche plus efficiente. - La suppression individuelle des gènes GPT, G6PD, PGLS, PGD, RPIA et FASN entraîne une réduction de la croissance tumorale, tout comme l'activation individuelle des gènes GOT1, GOT2, BCAT1 et BCAT2. - La suppression simultanée des gènes HK1 et TKT peut, dans certains cas, entraîner un épuisement complet du métabolite G6P, arrêtant la croissance tumorale. - Le volume tumoral influence l'effet de la modulation génique sur la croissance tumorale. - L'algorithme <i>Intelligent Search</i> identifie facilement des gènes ayant un impact individuel sur la croissance tumorale. 	<ul style="list-style-type: none"> - Les règles GPR formulées ne sont pas exhaustives. Elles utilisent des opérations strictes et ne considèrent qu'un petit échantillon de gènes comme potentielles cibles thérapeutiques. - La complexité computationnelle devient rapidement un obstacle pour l'algorithme <i>Depth Search</i>. - L'algorithme <i>Intelligent Search</i> éprouve de la difficulté à détecter des interactions synergiques entre les gènes. - L'effet simulé sur la croissance tumorale des cibles thérapeutiques reste à être validé expérimentalement.
------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

ANNEXE C DESIGN D'UN ARN INTERFÉRENT

Un ARN interférent a comme fonction d'inhiber la traduction d'un ARNm, permettant ainsi de réduire la quantité de sa protéine associée, et donc son niveau global d'activité. Plusieurs types d'ARN interférents existent: l'ARN interférent court (ARNic), l'ARN en épingle à cheveux courts (ARNsh) et le micro-ARN (miARN).

Les ARNic s'intègrent dans le complexe *RNA-Induced Silencing Complex* (RISC), un complexe protéique chargé de réguler l'expression des gènes. Leur brin guide s'apparie à un ARNm cible, ce qui induit sa coupure et sa dégradation. Toutefois, leur sensibilité aux nucléases, des enzymes capables de briser le lien entre deux nucléotides, et leur courte demi-vie *in vivo* nécessitent des modifications chimiques pour améliorer leur stabilité [65]. Les ARNsh, quant à eux, représentent une alternative aux ARNic et permettent un silençage génique durable. Transcrits à partir d'un vecteur d'expression, ils adoptent une structure en boucle d'épingle à cheveux. Contrairement aux ARNic, qui se dégradent rapidement, les ARNsh sont continuellement synthétisés, assurant ainsi un silençage plus stable, et leur production en grande quantité est moins coûteuse. Néanmoins, ils sont plus difficiles à modifier, et leur efficacité peut être compromise par des retards dans le traitement cellulaire du vecteur d'expression [65]. Enfin, les miARN sont des petits ARN non codants qui régulent les fonctions cellulaires. Ils se lient de manière partiellement complémentaire aux ARNm cibles, bloquant ainsi le processus de traduction. Les miARN peuvent interagir avec différentes régions de l'ARNm et, dans certains cas, même stimuler la traduction ou réguler l'expression génique en se liant à l'ADN. Contrairement aux ARNic et ARNsh, ils n'exigent pas une correspondance parfaite pour exercer leur effet, ce qui permet à un seul miARN d'affecter l'expression de centaines de gènes différents, souvent de manière imprévisible [65].

En tant que volet exploratoire du projet, une nouvelle approche visant à interférer avec le processus de traduction a été conceptualisée. Cette approche vise à réduire l'expression d'un gène cible, et par conséquent la croissance tumorale. Elle se base sur l'épuisement des acides aminés correspondants aux codons ayant une fréquence relative élevée dans le gène cible, par rapport à leur fréquence d'apparition dans le tissu ovarien. La fréquence relative d'un codon du gène ciblé reflète ainsi dans quelle mesure son acide aminé correspondant est requis pour la synthèse de la protéine codée par le gène. En complément, l'approche proposée cherche également à perturber l'expression de gènes associés au gène cible afin d'amplifier l'effet d'interférence par la

modulation des voies biologiques connexes et de renforcer l'inhibition tumorale. L'objectif ultime est de produire en grande quantité des peptides sans fonction biologique, synthétisés à partir d'un ARNm et composés des acides aminés les plus demandés pour la traduction du gène cible et de ses gènes associés. Cette synthèse aurait ainsi pour effet d'épuiser localement, soit dans le microenvironnement tumoral, la concentration en certains acides aminés, de ralentir la traduction des gènes concernés et, par conséquent, de réduire leur expression dans le but de freiner la croissance du cancer de l'ovaire.

Le design pour concevoir l'ARN interférent proposé suivrait les étapes suivantes:

1. Identification des gènes associés au gène cible.
2. Calculs des fréquences d'apparition des codons.
3. Sélection des codons à intégrer dans la séquence nucléotidique de l'ARN interférent.
4. Design de la séquence nucléotidique de l'ARN interférent.

Identification des gènes associés au gène cible

L'approche proposée pour interférer avec la traduction d'un gène cible viserait également à interférer avec des gènes associés au gène cible. Pour ce faire, la corrélation par rang de Spearman pourrait être utilisée pour identifier les gènes les plus influencés par ce dernier. Cette fonction attribue un coefficient de corrélation, variant de -1 à 1, aux paires de gènes, chaque paire étant composée du gène cible et d'un autre gène étant impliqué dans le modèle métabolique. Une corrélation négative indique que lorsque l'expression d'un gène augmente, celle de l'autre diminue, tandis qu'une corrélation positive signifie que les deux gènes augmentent ou diminuent leur expression simultanément. Plus la valeur absolue du coefficient de Spearman est élevée, plus il est probable que les deux gènes soient considérés comme dépendants l'un de l'autre. Les gènes ayant les valeurs absolues des coefficients de corrélation de Spearman les plus élevées avec le gène cible pourraient donc être considérés comme étant « associés » au gène cible.

Calcul des fréquences d'apparition des codons

Quatre types de fréquences d'apparition des codons, soient la fréquence locale, la fréquence globale, la fréquence relative et la fréquence relative moyenne, devraient être calculées dans le but de sélectionner les codons à intégrer dans la séquence nucléotidique de l'ARN interférent. Tout d'abord, la fréquence locale d'un codon dans un gène est définie comme le nombre de fois où il apparaît dans ce gène, normalisé pour 1000 codons. Puis, la fréquence globale d'un codon

correspond à sa proportion, normalisée sur 1000 codons, parmi l'ensemble des codons des gènes exprimés dans le tissu ovarien. Les fréquences globales des 64 codons dans le tissu ovarien peuvent être obtenues à partir des tables d'usage des paires de codon, appelées *TissueCoCoPUTs*, de la *Food & Drug Administration* (FDA) [66]. Ensuite, la fréquence relative d'un codon est déterminée par le rapport entre sa fréquence locale et sa fréquence globale. Enfin, la fréquence relative moyenne est calculée à l'aide d'une moyenne pondérée des fréquences relatives observées dans le gène cible et dans ses gènes associés.

Sélection des codons

Afin de synthétiser le peptide interférant avec la traduction du gène cible et de ses gènes associés, une séquence de nucléotides prenant la forme d'un ARNm pourrait être conçue en sélectionnant les codons présentant les fréquences relatives moyennes les plus élevées. Cette approche permettrait donc de privilégier les codons les plus fréquemment utilisés dans le gène cible, par rapport à l'usage global dans le tissu ovarien, tout en tenant compte de ceux des gènes associés.

Design de la séquence nucléotidique de l'ARN interférent

Afin d'éviter que le peptide traduit n'interagisse de manière imprévisible avec d'autres biomolécules, il est essentiel qu'il soit inactif. Une stratégie efficace pour y parvenir consisterait à favoriser un mauvais repliement post-traductionnel. Par ailleurs, il est crucial de comprendre l'impact de l'ordre des codons dans une séquence de nucléotides afin d'optimiser la conception d'un ARNm interférent performant. Selon la littérature, la surexpression d'un ARNm enrichi en codons rares inhibe la synthèse des protéines et freine la croissance cellulaire [67]. L'ordre des codons influe à la fois sur le repliement de la protéine et sur la vitesse de traduction [68], et la structure de l'ARNm au début de traduction est un facteur déterminant du rendement protéique [69]. De plus, la présence de codons rares, particulièrement dans les 50 premiers acides aminés, ralentit l'allongement de la protéine et atténue les embouteillages au niveau des ribosomes [70]. Enfin, dans la majorité des organismes, les gènes les plus exprimés contiennent un pourcentage élevé de codons correspondant aux ARN de transfert (ARNt) les plus abondants, tandis que les gènes les moins exprimés utilisent davantage des codons rares [70].

En tenant compte de ces aspects du processus de traduction, la conception de l'ARNm interférent devrait s'appuyer sur les deux principes suivants :

1. Exploiter des codons rares : sélectionner des codons pour lesquels les ARNt sont faiblement disponibles permettrait d'épuiser les ressources en ARNt. Cela contribuerait à réduire plus efficacement l'expression du gène cible et de ses gènes associés.
2. Induire un mauvais repliement protéique : favoriser l'utilisation de codons rares entraînerait un mauvais repliement de la protéine, la rendant potentiellement inactive. Une approche possible consisterait à organiser la séquence de sorte que les codons les plus fréquents soient utilisés dans la première moitié de l'ARNm, suivis par les codons rares dans la seconde moitié. Cet agencement induirait une synthèse initiale rapide, susceptible de provoquer des embouteillages au niveau des ribosomes, suivie d'une synthèse ralentie, favorisant ainsi la malformation de la structure moléculaire.

Résultats préliminaires obtenus

La fonction *spearmanr* [71] du module *stats* de la librairie *SciPy*, en libre d'accès dans le langage de programmation Python, a été employée pour analyser la monotonie de la relation entre certains gènes cibles, identifiés par les algorithmes de ciblage thérapeutique, et les autres gènes du modèle métabolique. Afin de calculer les coefficients de corrélation, cette fonction prend en paramètre les niveaux d'expression génique issus de la base de données du TCGA de toutes les patientes atteintes du cancer de l'ovaire. Puis, les fréquences relatives moyennes d'apparition des codons ont été calculées afin de sélectionner, pour chaque gène cible, cinq codons pouvant être intégrés dans les séquences nucléotidiques des ARN interférents (Tableau C.1).

Tableau C.1 Gènes associés, coefficients de Spearman, codons sélectionnés et fréquences relatives moyennes pour le design de l'ARN interférent de certains gènes cibles

Gène cible	Gènes associés et coefficients de Spearman	Codons sélectionnés et fréquences relatives moyennes
GPT	ACLY: -0,287; NF1: -0,258; MDH2: 0,248; BRCA1: -0,247; KRAS: -0,233; ADCY1: 0,223; IDH3G: 0,214; SDHD: 0,203; CS: -0,195; NRAS: -0,192	GCG: 2,39; GUG: 1,62; CGU: 1,48; CUG: 1,44; UAC: 1,41
G6PD	IDH3G: 0,325; OGDH: 0,193; COX7A2L: -0,179; SMAD4: -0,177; IDH2: 0,165; SUCLA2: -0,165; SDHC: -0,158; HK1: 0,158; SLC1A2: 0,153; PDHA1: 0,153	AUC: 1,72; AAC: 1,57; UUC: 1,54; GUG: 1,46; UAC: 1,42
PGLS	COX4I1: 0,251; COX7A2L: 0,235; MDH1: 0,202; NOX1: 0,199; PIK3CA: -0,199; COX7C: 0,198; COX5B: 0,189; IDH3G: 0,186; ACO1: -0,168; TKT: 0,168	CUA: 2,46; GCG: 2,43; UCG: 2,33; CCG: 2,15; GCA: 1,65

PGD	ENO1: 0,520; SDHB: 0,467; GAPDH: 0,264; NRAS: 0,244; LDHB: 0,242; SMAD4: 0,240; GPI: 0,237; MDH1: 0,232; KRAS: 0,216; CS: 0,216	GGA: 1,58; AUU: 1,49; GUA: 1,47; AUC: 1,45; CUA: 1,42
RPIA	COX7A2L: 0,347; MDH1: 0,229; COX5B: 0,190; COX6C: -0,179; ACO2: -0,173; SDS: -0,164; RB1: 0,160; ADCY1: -0,153; PTEN//PTENP1: -0,147; NF1: -0,145	GCU: 1,76; GGG: 1,56; AUC: 1,54; CGA: 1,42; GUG: 1,40
HK1	PFKL: 0,267; EGFR: 0,233; COX7A2: -0,222; FASN: 0,207; PGM1: 0,197; MT-ND5//MTND5P11: 0,196; ATP1A1: 0,196; PDHA1: 0,190; BRCA2: -0,188; LDHA: 0,187	ACG: 1,76; AUC: 1,53; AUA: 1,40; AUG: 1,34; AGG: 1,33
TKT	PDHB: 0,310; COX5A: 0,224; COX4I1: 0,217; ATP1A1: 0,208; PKM: 0,192; PGM3: -0,188; PTEN//PTENP1: 0,183; FASN: 0,182; PDHA1: 0,181; NOX4: -0,175	AUC: 1,90; GCC: 1,59; GCU: 1,33; GUA: 1,32; GAC: 1,28
FASN	ACLY: 0,302; SDHD: -0,242; HK1: 0,207; COX7C: -0,196; ADCY1: -0,192; TKT: 0,182; CS: 0,176; GOT2: 0,171; PKM: 0,170; IDH2: 0,166	CUG: 1,65; GCC: 1,63; GUG: 1,54; GGC: 1,54; UCG: 1,50

Les coefficients de corrélation de Spearman entre les gènes cibles et leurs gènes associés ne sont pas particulièrement élevés, se situant autour de 0,3 en valeur absolue pour les meilleurs gènes. Cependant, certains gènes présentent des corrélations plus significatives, tels que le gène PGD, qui affiche un coefficient de 0,520 avec le gène ENO1 et de 0,467 avec le gène SDHB, suggérant une relation plus forte entre ces gènes. Par ailleurs, certains codons sélectionnés présentent des fréquences relatives remarquablement élevées, comme CUA (2,46) et GCG (2,43) pour le gène PGLS, ainsi que GCG (2,39) pour le gène GPT. L'épuisement des acides aminés correspondants en surchargeant leurs ARNt pourrait donc avoir un effet bénéfique, puisque ces codons sont utilisés plus de deux fois plus fréquemment dans les gènes cibles et leurs gènes associés par rapport à leur usage global dans le tissu ovarien. Toutefois, la spécificité de l'ARN interférent conçu peut être un enjeu majeur. En ciblant des codons fréquents, il existe un risque de perturber involontairement l'expression d'autres gènes non ciblés qui partagent ces mêmes codons. En effet, avec seulement 64 codons disponibles, il est probable que les codons sélectionnés pour le gène cible soient également présents en abondance dans d'autres gènes. Utiliser un système de délivrance ciblée afin d'acheminer l'ARN interférent directement vers la masse tumorale est donc essentiel afin de minimiser les effets involontaires sur les cellules saines.