Computational and mathematical approaches to understand glioma cells metabolism : characterization from transcriptomic data and focus on HIF pathways

Kévin Spinicci

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES
et de SWANSEA UNIVERSITY

École doctorale : ISCE - Ingénierie pour la Santé la Cognition et l'Environnement
Spécialité : MBS - Modèles, méthodes et algorithmes en biologie, santé et environnement
Unité de recherche : Translational Innovation in Medicine and Complexity

Approches computationnelles et mathématiques pour comprendre le métabolisme des cellules de gliomes : caractérisation à partir de données transcriptomiques et focus sur les voies HIF

Computational and mathematical approaches to understand glioma cells metabolism : characterization from transcriptomic data and focus on HIF pathways

Présentée par :

Kévin SPINICCI

Direction de thèse :

Angélique STÉPHANOU Chargée de recherche CNRS	Directrice de thèse
GiBIN POWATHIL Professor, Swansea University	Co-Directeur de thèse

Rapporteurs :

Mark CHAPLAIN
Professor, University of St Andrews
Tommaso LORENZI
Associate Professor, Politecnico di Torino

Soutenance de thèse publique, prévue le **20 décembre 2023**, devant le jury composé de :

Mark CHAPLAIN Professor, University of St Andrews	Rapporteur
Adeline LECLERCQ SAMSON Professeure, Université Grenoble Alpes	Président
Tommaso LORENZI Associate Professor, Politecnico di Torino	Rapporteur
Biagio LUCINI Professor, Swansea University	Examinateur

Kévin SPINICCI: *Computational and mathematical approaches to understand glioma cells metabolism : characterization from transcriptomic data and focus on HIF pathways,*

Supervised by:
Angélique STÉPHANOU
Gibin POWATHIL

Laboratory:
Translational Innovation in Medicine and Complexity (TIMC)
University:
Université Grenoble Alpes - Swansea University
Team and department:
Équipe BCM - TIMC
Department of Mathematics - Swansea University

"Yhorm, Old Friend.
I, Siegward Of The Knights Of Catarina, Have Come To Fulfill My Promise.
Let The Sun Shine Upon This Lord Of Cinder!"
— Siegward of Catarina - Dark Souls 3

Declarations

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed
Date: 23/10/2023

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed
Date: 23/10/2023

I hereby give consent for my thesis, if accepted, to be available for electronic sharing

Signed
Date: 23/10/2023

The University's ethical procedures have been followed and, where appropriate, that ethical approval has been granted.

Signed
Date: 23/10/2023

Abstract

This Thesis questions the effect of hypoxia on cell metabolism, how different oxygen levels within the tumour lead to heterogeneity, and what is the impact on tumour growth. The study focuses on the main protein orchestrating the cellular adaptation to hypoxia: the Hypoxia Inducible Factor (HIF). Despite being the topic of many studies, HIF seems to be only characterized through its effect on genes observed during biological experiments. Mathematical modelling of HIF does not integrate the protein in models simulating large population of cells. This work will try to unravel some of the mechanisms by which HIF impact the tumour growth with a focus on the metabolim and the cell invasion. In this regard, an agent-based model (ABM) has been developed and the investigation paths for the model have been determined from a statistical analysis.

The mathematical model developed describes a network of genes regulated by HIF which includes the two important metabolic genes lactate dehydrogenase (LDH) and pyruvate dehydrogenase (PDH) with a focus on the appearance of the Warburg Effect, a mechanism often considered as a tumour cell characteristic. An individual-based approach has been chosen to better represent cellular heterogeneity as each cell has its own set of parameters and experiences a different local environment. With that model, we could test different environmental conditions and genetic regulation. We saw that both rapid variations of extracellular oxygen and increased levels of HIF are enough to induce a Warburg phenotype, which therefore appears as influenced by both a contextual effect and a genetic effect.

A statistical analysis of glioblastoma transcriptomic data has been performed to unravel the deregulated mechanisms into two datasets: one available on the The Cancer Genome Atlas (TCGA) platform and a Patient Derived Cell Lines (PDCL) dataset provided by the ICM (Paris). The workflow of analysis begins with a Differential Expression (DE) analysis of the transcriptomic data to find the deregulated genes, which are submitted to pathway enrichment tools to determine what are the corresponding pathways. Among the results, collagen biosynthesis was often deregulated in both datasets and cholesterol metabolism was often deregulated in PDCL data. Owing to its link to the current thematic, collagen biosynthesis has been selected as the new candidate for mathematical description in the current model.

Thus, following the results of the statistical analysis, the collagen biosynthesis have been implemeted in the model of metabolism. Collagen and matrix remodelling have a noticeable impact on cellular migration and thus the invasion of distant tissues by cancer cells. The three genes P4HA1, MT1-MMP and LOX have been documented to be upregulated by HIF in the literature and to impact the secretion, degradation and cross-linking of collagen. Genetic parameters were fitted to transcriptomic data to define the genetic parameters of these new genes as well as for the ones already included. The results obtained with the new version of the model indicated that collagen density has a substantial effect on the proliferation of the tumour and its shape. Like collagen, oxygen alone could impact the tumour growth. Moreover, results showed that the oxygen and collagen seems to have a dual effect on tumour proliferation. Increased P4HA1 and MT1-MMP quantities, respectively decreased and increased cellular migration speed while LOX seemed to have little effect. Here, no cells adopted a Warburg phenotype in any of the simulations. The similarity of the production of H^+ in this model observed when PDH sensitivity to HIF was reduced in the previous model indicates that this may be due to the new parameters values. Consequently, even if the microenvironment favours a Warburg phenotype, genetic regulation may dictate the cell ability to induce it.

This model made it possible to show how HIF can impact the tumour growth and cellular invasion through few regulations. It suggests that both fast changes of oxygen levels and increased HIF stabilization in normoxia can induce a Warburg Effect. However, the ability of the cell to reduce the oxygen consumption through adaptation to hypoxia seemed to be a limiting factor preventing the adoption of a Warburg phenotype.

Résumé

Cette Thèse s'interroge sur l'effet de l'hypoxie sur le métabolisme des cellules, comment différents niveaux d'oxygène vont se répercuter sur l'hétérogénéité, et quel est l'impact sur la croissance tumorale. L'étude se centre sur la principale protéine qui orchestre l'adaptation à l'hypoxie : le Facteur Induit de l'Hypoxie, en anglais "Hypoxia Inducible Factor" (HIF). Bien que celle-ci fasse l'objet de nombreuses études, HIF semblent n'être charactérisé seulement via son effet sur des gènes observés durant des expériences biologiques. La modélisation mathématique de HIF n'intégre bien souvent pas la protéine au sein de modèle simulant la croissance d'une large population de cellules. Ce travail va donc s'atteler à la compréhension de certains des mécanisms par lesquels la protéine HIF va impacter la croissance de la tumeur en se concentrant sur le métabolisme et l'invasion des cellules cancéreuses. A cet égard, un modèle multi-agent a été développé et le choix des voies d'investigation ont été déterminées par une analyse statistique.

Le modèle mathématique décrit un réseau de gènes régulés par HIF qui inclut deux gènes métaboliques important que sont les gènes du lactate deshydrogènase (LDH) et de la pyruvate deshydrogènase (PDH). Le but étant de refléter l'effet de HIF sur le métabolisme en se centrant sur l'apparition de l'effet Warburg, un mécanisme souvent considéré comme une caractéristique de la cellule tumorale. Une approche dîte individu-centrée a été choisie pour représenter au mieux l'hétérogénéité cellulaire car chaque cellule possède son propre jeux de paramètres et est exposée un environnement différent. Grâce à ce modèle, nous avons pu tester différentes conditions environnementales et régulations génétiques. Nous avons pu observer que des variations rapides des niveaux d'oxygènes ainsi que des niveaux accrûs de HIF ont déclenché l'adoption du phénotype Warburg, qui par conséquent apparaît comme influencé à la fois par le contexte environmental et génétique.

Une analyse statistique de données de transcriptomique provenant de glioblastomes a été réalisée pour déterminer les mécanismes dérégulés dans deux jeux de données : l'un provenant de la plateforme The Cancer Genome Atlas, l'autre comprenant une série de lignées issues de patient (PDCL, fournies par l'ICM). La méthode d'analyse débute par une analyse d'expression différentielle des données de transcriptomique pour trouver les gènes dérégulés dans nos jeux de données. Ces gènes sont ensuite soumis à des outils d'enrichissement de pathways pour en extraire la liste des pathways correspondants. Parmis les résultats, la biosynthèse du collagène remonte comme souvent dérégulée dans les deux jeux de données, et le métabolisme du cholésterol est trouvé souvent dérégulé dans les données de PDCL. A cause de son lien avec la thématique courante, la biosynthèse du collagène a été retenue comme candidate pour être inclute dans le modèle.

De ce fait, suivant les résultats de l'analyse statistique, la biosynthèse du collagène a été implémentée dans le modèle du métabolisme. Le collagène et le remodellage de la matrice ont un impact notable sur la migration cellulaire et de ce fait l'invasion des cellules cancéreuses dans les tissus distants. Les trois gènes P4HA1, MT1-MMP et LOX sont sur-régulés par la protéine HIF selon la littérature et impactent la sécrétion, la dégradation et la réticulation du collagène. Des données de transcriptomique ont été utlisées pour définir les paramètres génétiques de ces nouveaux gènes ainsi que ceux déjà présents. Les résultats obtenus avec la nouvelle version du modèle indiquent que la densité de collagène génère un effet sur la prolifération et la forme de la tumeur. Comme le collagène, l'oxygène peut impacter la croissance de la tumeur à lui seul. De plus, les résultats montrent que l'oxygène et le collagène semblent avoir un effet double sur la prolifération de la tumeur. Des quantités accrûes de P4HA1 et MT1-MMP vont, respectivement, réduire et augmenter la vitesse de la migration cellulaire dans le modèle alors que LOX semble n'avoir que peu d'effet. Dans ce modèle, l'effet Warburg n'a été observé dans aucune simulation. La similarité de la production de H^+ dans ce nouveau modèle avec les résultats obtenus dans le modèle précédent lorsque la sensibilité de PDH à HIF était réduite indique que cela est dû aux nouveaux paramètres pour les gènes. En conséquence, si l'environnement de la cellule pourrait favoriser l'apparition d'un phénotype Warburg, les régulations génétiques pourraient contrôler la capacité de la cellule à l'adopter.

Ce modèle permet de montrer comment HIF peut impacter la croissance de la tumeur et l'invasion cellulaire à travers quelques régulations. Il suggère que des variations rapides des niveaux d'oxygène ainsi qu'une stabilisation accrûe de HIF peuvent induire un effet Warburg. Cependant, la capacité de la

cellule à réduire la consommation d'oxygène au travers de l'adaptation à l'hypoxie semble être un facteur limitant pouvant prévenir l'adoption d'un phénotype Warburg.

Acknowledgments

If we think that doing science is hard, writing and expressing its gratitude is not less complicated. This exercice is even more complex in a co-tutelle as someone meet with even more people with whom we have the pleasure to share a few moments.

The first persons to receive my acknowledgments are my thesis supervisors Angélique STÉPHANOU and Gibin POWATHIL, for being by my side during the Thesis, for their advices, for our scientific interactions and our less scientific discussion.

A big thank to Nicolas GLADE, without whom I would not have known about the opportunity to do this PhD and for his pre-PhD advivcs that prepared me. A big thank to Pierre JACQUET for investigating PhysiCell beforehand, for the help during my work, for sharing an office and lunch breaks.

A big thank to Samuel OLIVER and Kira PUGH for all the moments we had in Wales that made my travel heartwarming, for the laugh, for teaching me how to prononce Worcester and Llanfairpwllgwyngyll-gogerychwyrndrobwllllantysiliogogogoch.

Finally, I would like to thank every persons from the TIMC laboratory and Computational Foundry that I could not cite. Thank you for sharing these moments with me and making this PhD a sweet experience.

Remerciements

Si l'on pense que faire des sciences est dur, écrire et formuler sa gratitude n'en est pas moins compliqué. Cet exercice est encore plus complexe dans le cas d'une co-tutelle car l'on rencontre d'autant plus de personne avec qui nous avons eu le plaisir de partager certains moments.

Les premières personnes à recevoir mes remerciements sont mes directeurs de thèse Angélique STÉPHANOU et Gibin POWATHIL, pour m'avoir accompagné au cours du doctorat, pour leur conseils, pour nos échanges scientifiques ainsi que nos discussion qui l'étaient moins.

Un grand merci à Nicolas GLADE, sans qui je n'aurais jamais eu connaissance de cette thèse et pour les conseils pré-thèse qui auront su me préparer. Un grand merci à Pierre JACQUET d'avoir essuyer les plâtres de PhysiCell, pour l'aide apportée dans mon travail, pour avoir partager un bureau et nos pauses déjeuner.

Un grand merci à Samuel OLIVER et Kira PUGH pour les moments passés au Pays de Galle qui auront rendu mon voyage chaleureux, pour les rires, pour m'avoir appris comment prononcer Worcester et Llanfairpwllgwyngyllgogerychwyrndrobwllllantysiliogogogoch.

Enfin, j'aimerai remercier toutes les personnes du Laboratoire TIMC et du bâtiments Computational Foundry que je n'ai pu citer. Merci d'avoir partagé ces moments avec moi et d'avoir fait de cette thèse une expérience agréable.

Contents

List of Figures

List of Tables

Acronyms

ABM agent-based model
ADAM a disintegrin and metalloproteinases
α-KG α-ketoglutarate
ATP Adenosine Triphosphate
BBB Blood-Brain Barrier
BF boolean function
bHLH basic-loop-helix
BN boolean network
BRAF B-Raf proto-oncogene serine/threonine kinase
CA carbonic anhydrase
CAF cancer associated fibroblast
CDK cyclin-dependant kinase
CNS Central Nervous System
CPM cellular potts model
C-TAD C-terminal
DE Differential Expression
ECM Extra-Cellular Matrix
EGFR Epidermal Growth Factor Receptor
EMT epithelial to mesenchymal transition
ES Enrichment Score
FACIT fibril-associated collagens with interrupted triples helice
FBA flux balance analysis
FCS Functional Class Scoring
FDR False Discovery Rate
FIH Factor-Inhibiting HIF
FPKM fragments per kilobase of exon per million mapped reads
GBM glioblastoma
GEO Gene Expression Omnibus
GLM generalized linear model
GMT gene matrix transposed
GSEA gene set enrichment analysis
HD-SFRT high-dose single-fraction hypofractionated radiation
HIF Hypoxia Inducible Factor
hiPSC human induced Pluripotent Stem Cells
HPC high performance computing
HRE Hypoxia Response Elements
ICM Institut du Cerveau et de la Moelle épinière
IDH Isocitrate Dehydrogenase
KEGG Kyoto Encyclopedia of Genes and Genomes

LDH lactate dehydrogenase
LOX lysyl oxidase
ltNES Long-Term Neural Epithelial Stem Cell Culture
MCT monocarboxylate transporter
MET mesenchymal to epithelial transition
MGMT O-6-methylguanine-DNA methyltransferase
MMP matrix metalloproteinase
MRI magnetic resonance imaging
mTOR mechanistic target of rapamycin kinase
N-TAD N-terminal
ODD oxygen-dependent degradation
ODE ordinary differential equation
ORA Over-Representation Analysis
OXPHOS oxidative phosphorylation
P4H prolyl 4-hydroxylase
p53 tumour protein 53
PAS Per-Arnt-Sim
PCA Principal Component Analysis
PDCL Patient Derived Cell Lines
PDE partial differential equation
PDH pyruvate dehydrogenase
PDK pyruvate dehydrogenase kinase
PenDA Personalized Differential Analysis
PHD Prolyl Hydroxylases
PI3K phosphoinositide 3-kinase
PLOD procollagen-lysine,2-oxoglutarate 5-dioxygenase
PTEN Phosphatase and Tensin Homolog
PT Pathway Topology
RB retinoblastoma
RNA-Seq RNA-Sequencing
ROS Reactive Oxygen Species
TCA tricarboxylic acid
TCGA The Cancer Genome Atlas
TERT Telomerase Reverse Transcriptase
TGF Transforming Growth Factor
TIMP tissue inhibitors of MMP
TMZ temozolomide
TPM transcripts per million
VEGF Vascular Endothelial Growth Factor
VHL von Hippel Lindau
WHO World Health Organization

Introduction

Cancer is a major medical concern as this is the leading cause of death worldwide. Cancer caused the death of 157,400 persons in France in 2023 [1] and 167,142 persons in the United Kingdom between 2017 and 2019 [2]. During Antiquity, Hippocrate compared the tumour to a crab due to its extensions similar to the legs of that animal and thus, first called it “karkinos” and “karkinoma” (greek for crab) [3]. Today, treatment and patient outcome has improved and many cases can be cured. Yet, several challenges still need to be addressed as some types of tumours are very lethal and survival very low.

One of the reasons for treatment failure is the remarkable heterogeneity displayed by tumours at several layers: inter-individual, inter-tumoural and even intra-tumoural. Cancer takes so many forms that it may be more appropriate to talk about *cancers*. Consequently, cancer exhibits different characteristics depending on its primary site and type with an associated level of aggressiveness. Nonetheless, Hanahan and Weinberg identified common features of cancers, labelled as “Hallmarks of Cancer”, in their publication in 2011 [4]. A few of these hallmarks include resistance to cell death, sustained proliferative signals, induced angiogenesis, deregulated cellular genetics, invasion and metastasis.

A common environmental feature of tumours that has been the subject of many studies in the past is the hypoxia created by the tumour as a result of the high cellular density and the abnormal vasculature due to angiogenetic factor secretion. Hypoxic conditions affect the cell behaviour that adapts by secreting angiogenetic factors and relying more on anaerobic processes to sustain these harsh conditions. Studies have linked hypoxia to cancer cell aggressiveness, poor clinical outcome, increased metastasis and higher resistance to radiotherapy. Naturally, the potential to use hypoxic conditions to improve treatment outcomes has caught the interest of the scientific community. The metabolism of cancer cells itself is an emerging topic that received more attention recently owing to its therapeutic potential. Otto Warburg observed in 1927 that cancer cells produced more lactate and consumed more glucose. While regular cells show an increased consumption of glucose and secretion of lactate in hypoxia as they adapt to the low amount of oxygen, the specificity of cancer cells is to exhibit this behaviour even in normoxia when oxygen should be sufficient. This phenomenon, now termed the Warburg Effect, has been addressed in the literature through many studies in the hope of uncovering the survival advantage it confers to the cell. It is easy to see a link between the Warburg Effect and hypoxic conditions.

This Thesis aims to understand the interplay between hypoxia and cell metabolism: what is the impact on the tumour growth, how different oxygen conditions within the tumour affect its growth and what are the consequences on the cell metabolism. Since hypoxia is the main topic of this study, this work will focus on the main actor of the cellular response to hypoxia: the Hypoxia Inducible Factor (HIF). HIF is a heterodimer that acts as a transcription factor to increase the transcriptional activity of a target gene called Hypoxia Response Elements (HRE). Through its action, HIF initiates transcriptional changes so the cell can sustain the hypoxic conditions. The transcription factor will increase the usage of anaerobic pathways to produce energy as oxygen becomes unavailable, induces the arrest of the cell cycle, triggers the secretion of angiogenetic factors and initiates the migration of the cell. The present work will focus on the effect of HIF on the metabolism and the migration of the cell.

Mathematical modelling is a powerful tool that has shown its efficiency in handling the complexity of cancer to clarify some of its concepts. It is highly suited to test hypotheses, guide research studies and make progress in understanding the disease. In this context, biological systems are a target of choice due to their inherent complexity. Mathematical modelling has been applied: to describe the dynamic of a drug in pharmacokinetics studies, to model gene-regulation networks, predict population growth or to predict treatment efficiency and tumour growth in cancer. Apart from mathematical modelling, statistical analyses are well suited to find patterns or differences in biological data. They have been successfully applied to classify cancer into different sub-types depending on their genetic and histologic characteristics, find frequent mutation common or specific to certain types of tumours, and understand the genetic regulation changes in different conditions. In this Thesis, both a mathematical modelling approach and a statistical analysis approach will be used to study the effect of HIF on the tumour. The results of the statistical analysis will define the processes that will be implemented in the model.

Firstly, I will describe in chapters 1 and 2 the biological background and the mathematical modelling landscape. I will start by explaining what is the current knowledge on HIF, the metabolism, their interaction and the cellular migration. This part gives the knowledge necessary to understand the concerns and the features of the mathematical model. Then, I will detail some of the mathematical approaches that have been documented in the literature: how metabolism has been described mathematically, what are the different computational models used and how continuous models were combined with discrete methods to simulate tumour growth. In this chapter, I will also talk about the mathematical approach used in the Thesis. I will justify the choices of modelling taken: the type of model, 2D simulations, the experiments replicated, programming choices, etc.

A first model of the impact of HIF on the metabolism is described in chapter 3. The aim is here to understand the impact of HIF regulation on the metabolism and how it affects the appearance of the Warburg Effect. The model will include a small network of metabolic genes including the two key genes lactate dehydrogenase (LDH) and pyruvate dehydrogenase (PDH). Here, metabolism will dictate the fate of the cell: the cell will be allowed to proliferate if the energy produced is sufficient. I will show how environmental conditions and HIF upregulation can trigger the Warburg Effect in the results.

The analysis of glioblastoma transcriptomic data will be described in chapter 4. The aim is to assess the deregulated pathways in that type of tumour to find relevant processes to include in the model. The workflow is mainly composed of a Differential Expression analysis followed by a Pathway Enrichment analysis. A Differential Expression analysis finds deregulated genes, genes whose expression has changed, between two conditions (here normal and glioblastoma). The genes found deregulated are then submitted for Pathway Enrichment analysis, a method that finds deregulated pathways in a list of genes. The results will be presented at the end of the chapter with a conclusion explaining the candidates selected for modelling and a link to the next work.

In the last chapter 5, the processes identified and selected during the statistical analyses will be implemented in the model of metabolism to understand the impact of HIF on cellular migration. At the beginning of the chapter, the interaction between the selected processes and the protein HIF will be explained. I will show how collagen was included in the model and how the cell modifies the collagen content (state and quantity) depending on the genes included that are regulated by HIF. The impact that collagen density has on the cellular migration speed will be detailed. The last part of this chapter will investigate the results obtained with the model to show how different collagen and oxygen conditions can impact the growth of the tumour, and what is the role of the different genes in our model.

1 Biological Background

1.1 Glioblastoma

Glioblastoma (GBM) (figure 1.1) is the most frequent and deadly tumour of the Central Nervous System (CNS) [5] with 3-7 cases for 100,000 persons per year [6, 7]. A pathological feature of GBM is the necrotic foci with surrounding cellular pseudopalisades and microvascular hyperplasia (an exacerbated form of angiogenesis) [8].

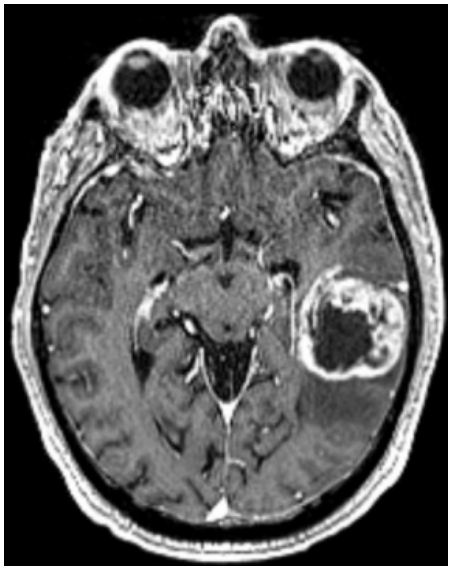

Figure 1.1: Magnetic resonance imaging (MRI) scan of grade IV glioblastoma.
Source: https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Glioblastoma-Multiforme

1.1.1 Classification of Glioblastoma

Glioblastomas belong to the group of *malignant glioma* which is a group of heterogeneous tumours derived from glial cells [6]. Primary GBM arise *de novo* and constitute the majority of cases (90%) [6–9] while secondary GBM arise from a previous tumour such as glioma [6, 7, 9, 10], an oligodendroglioma and their anaplastic variants [6] or from a lower-grade astrocytoma [8]. Glioblastoma in older population mainly develop *de novo* [7]. The World Health Organization (WHO) updated the classification of CNS tumours using molecular parameters, based on the IDH gene mutation status, in addition to histological parameters. Glioblastomas now belong to the IDH-Wild Type group or to the IDH-mutant group [8]. The IDH1 gene mutation status is considered a reliable and objective marker of secondary GBM.

Glioblastomas have been classified into three different sub-types: classical, mesenchymal and proneural (figure 1.2) [5–8, 11]. Several sub-types can be present inside the same tumour in different regions, studies have shown that sub-types can vary following the time and the therapy [11]. They are defined by different genetic mutations together with different gene expression patterns. For example, the mesenchymal sub-type is defined by a high mutation frequency of tumour suppressor genes (such as NF1 [5, 7, 8], PTEN [5] and p53 [5]) and mesenchymal markers (such as CHI3L and MET [8]). The neural sub-type has also been mentioned in the literature although to a lesser extent than the three sub-types discussed above [5, 6, 8].

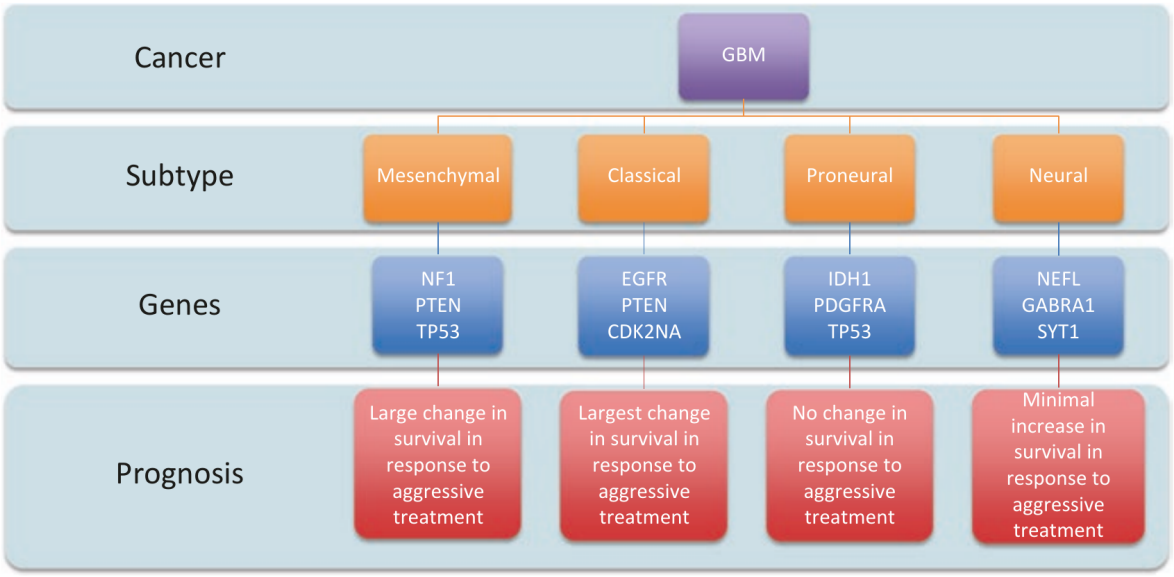

Figure 1.2: The different subtypes of glioblastomas with their corresponding identified mutations.
Source : Quinones et *al* [5]

1.1.2 Heterogeneity

Heterogeneity in glioblastoma is known to be one of the main reasons for therapy failure. GBM exhibit a remarkable inter- and intra-patients heterogeneity at the genomic and gene expression levels that can be driven by therapy or mutation.

Studies on tumour heterogeneity have shown that 60-80 % of the mutations are shared between the bulk tumour and its margin. Gene expression, genetic aberrations and copy number variations differ from the bulk tumour and its periphery, between hypoxic and non-hypoxic regions and between different zones in the same tumour [12].

Neftel et *al* have identified four different cellular states for glioblastoma cells genes linked to specific states: neural-progenitor-like, oligodendrocyte-progenitor-like, astrocyte-progenitor-like and mesenchymal-progenitor-like. Each tumour in their dataset displayed at least two of the four cellular states with the majority of tumours displaying every cellular state. The frequency of each state varied between tumours and within the same tumour depending on the region [11].

Clinical evidence has shown that GBM can spread further depending on the type; for example, the mesenchymal sub-type shows an increased invasive potential compared to the three other subtypes [13].

Primary and recurrent tumours differ by their molecular trait: the number of mutations shared between a primary and recurrent tumour can vary from 11 to 98 % with distant recurrent tumours sharing 25 % while local recurrent tumours share on average 70 %. Not only does almost all glioblastomas recur, but recurrent tumours are more resistant to therapy and usually too deep in the brain to be accessible for a second surgical resection. In fact, only 20-30 % of recurrent GBM are accessible with surgery [12].

1.1.3 Survival and Therapy

Migration and invasion are important aspects of glioblastoma as this type of tumour often recurs with the recurrent tumour becoming harder to treat. Metastasis originating from other tumours (mainly from lung, breast or melanoma) that form in the brain are more frequent than brain tumours [7].

Population-based studies report a median survival of 42.4 % at six months, 17.7 % after one year and a 5 years survival rate of 5 %. Glioblastoma therapy must face several difficulties: the infiltrative nature of the tumour that prevents it from being surgically removed, the rapid proliferation rate of GBM malignant cells, the appearance of treatment-resistant clones after initial therapy, the Blood-Brain Barrier (BBB) preventing drugs to reach the brain, the activation of multiples transduction signalling pathways, specific genetic mutations inside the tumour and the sensitivity of certain brain areas to radiation therapy. The clinical and molecular factors associated with long-term survival are still unknown. Yet, methylation of the O-6-methylguanine-DNA methyltransferase (MGMT) promoter has been confirmed as a prognostic

marker to predict an increased survival and predict the response to chemotherapy. Furthermore, patients carrying the IDH1 mutation show better survival. The long-term survival group seems over-represented with female patients displaying hyper-methylation of the MGMT promoter and good initial performance. Combination of chemotherapy with temozolomide (TMZ) increases the mean survival rate from 10.4 % to 26.5 % at 14.6 months and at 2 years [6]. The standard of care consists of surgical resection of the tumour followed by a treatment composed of radiotherapy and chemotherapy combined with TMZ (up to 6 cycles) [12, 14].

Most of the therapeutic agents developed to target GBM are first tested on recurrent GBM in clinical trials [12]. Not only genetic alterations can impact the treatment efficiency, such as sensitivity to TMZ can be predicted by the methylation status of the MGMT promoter [8, 9, 14], but the therapy will impact the genetic properties of the cell as well. A study reported a 10-fold increased mutation rate in lower-grade gliomas treated with temozolomide. Here TMZ induced a hyper-mutator phenotype [12]. Similarly to the effect of temozolomide, areas in recurrent tumours treated by radiation therapy display an increased absence of MGMT methylation mutation compared to untreated region and changes linked to the VEGF pathway have been associated with the bevacizumab agent [12], showing that therapy can influence tumour heterogeneity.

Recently, the field of drug repurposing has gained interest and is an emerging field in anti-cancer therapy. The repurposing of a drug approved for the treatment of other diseases than cancer [14]. This field brings hope to reduce drug development costs and time as patient data are already available for the drugs being investigated. Multiple approved drugs are already under investigation for their potential use in cancer therapy such as thioridazine [15] and metformin [14].

The majority of clinical trials focuses on the intrinsic targets of GBM addressing the oncogenic signalling pathways through tyrosine receptor kinases, control of the cell cycle and the ability to induce apoptosis [14].

1.1.4 Main genetic alterations in Glioblastoma

In this section, I will present a few of the genetic alterations commonly found in glioblastoma to show the diversity of mutations and the heterogeneity of this tumour. As seen earlier, the methylation status of the MGMT promoter, an enzyme involved in DNA repair by removing alkyl groups from guanine residues, is associated with increased sensitivity of GBM to alkylating agents (such as TMZ). The three pathways that are the most often deregulated in glioblastoma are: the RTK/RAS/PI3K signalling (88%); the tumour protein 53 (p53) signalling (87%) and the retinoblastoma (RB) signaling (78%) (see figure 1.3).

There is a significant probability that a gene in each of these three pathways is deregulated: it has been reported that 74 % of the samples show a genetic alteration in each pathway which seems to indicate that they are required for the pathogenesis of GBM. However, there is a statistical tendency toward the mutual exclusivity of the component altered in each pathway [9]. Among the most frequent genetic alterations is found: inactivation of NF1 (18 %), activation of Epidermal Growth Factor Receptor (EGFR) (45 %), missense activating mutations on the PI3KCA (PI3K) protein (15 %), inactivation of the tumour suppressor gene Phosphatase and Tensin Homolog (PTEN) (36 %) [9, 10] and an homozygous deletion on CDKN2A/B (45 %) [10]. IDH wild-type lower-grade glioma might be a precursor to wild-type glioblastoma as they share similar alterations on certain loci despite GBM prevalence to mutations is higher than lower-grade gliomas. EGFR and FGR3 activating mutation are predicted to be restricted to IDH wild-type glioma with similar frequencies to the one seen in GBM [10]. The RAF kinase member, B-Raf proto-oncogene serine/threonine kinase (BRAF), is involved in the MAPK/ERK signalling pathway which promotes proliferation. BRAF activating mutations are observed in several different types of tumours and have been confirmed as a potential target for therapy, especially in melanoma [14]. The RB pathway, controlling the cell cycle, is altered in the majority of Isocitrate Dehydrogenase (IDH) wild-type due to a homozygous CDKN2A/B deletion or an amplification on CDK4/CDK6 or an alteration on the gene RB1 [14]. Alteration of the Telomerase Reverse Transcriptase (TERT) promoter is among the most common in IDH wild-type GBM and increases TERT transcription which in turn supports GBM cells immortalization [14]. TGF-β1/2 is the major actor of the Transforming Growth Factor (TGF) Beta family members, it was considered for decades as a key molecule responsible for glioblastoma mediated immunosuppression [14].

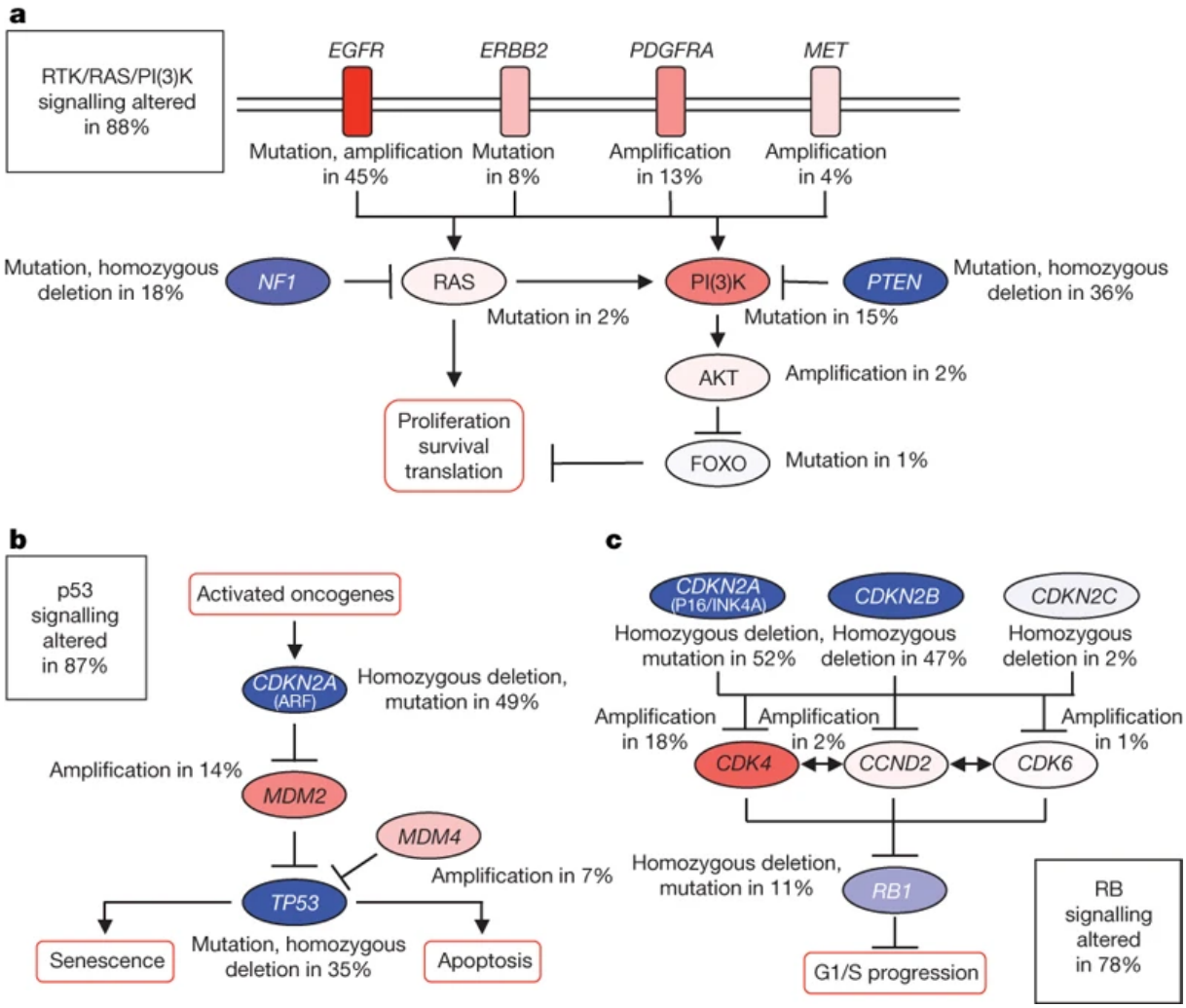

Figure 1.3: Frequently altered pathways in glioblastoma as reported by a study on gene mutations conducted by The Cancer Genome Atlas initiative for the (a) RTK/RAS/PI3K pathway, (b) the p53 pathway and (c) the RB pathway. Activating genetic alterations are shown in red while blue indicates inactivating alterations. The stronger the shade, the more frequently the gene is altered.
Source : McLendon et *al* [9]

1.2 Implication of hypoxia in cancer

1.2.1 Motivation to study hypoxia

While most of the *in vitro* studies are conducted at around 20-21 % O_2 , 5 % O_2 would be a better approximation of *in vivo* tissue oxygenation as oxygen pressure drops in the tissue. McKeown et *al* define 5 % O_2 as “physoxia”. However, median oxygen levels measured in tumours are around 2 %, 1.7 % in brain tumours (3.4 % in normal conditions) [16].

Blood vessels in GBM are disorganized, tortuous, very permeable and characterized by abnormalities in the endothelial wall due to lack of coverage by pericytes. Their diameter is also significantly larger and their basement membrane is thicker. This inconsistent vascular oxygenation brings hypoxia, acidity and necrosis [8]. Long-term hypoxia selects the cells the most tolerant to low oxygen conditions, resistant to stress, the most malignant and cells with defects in apoptosis. Acute hypoxia induces DNA strand breaks that lead to, if not repaired, mutations in cell culture and animal models. There is also evidence indicating that DNA reparation processes in tumours are affected by hypoxia and are linked to an increased genomic instability [16]. Hypoxic niches are associated with a higher mutation rate as well as alteration of oncogenes and tumour repressor genes such as TP52, PTEN and MYC [17].

Hypoxic cells are significantly more resistant to radiotherapy with a marked drop in sensitivity at 2.6 % O_2 and even more at 0.5 % O_2 [16]. The usage of High-dose single-fraction hypofractionated

radiation (HD-SFRT) in anti-cancer therapy has shown promising results. A study has shown that hypoxia increases the cell's resistance to HD-SFRT therapy in pre-clinical models which was associated with HIF1-α : the main actor in the cell response to hypoxia [18]. It was shown that fractionated radiotherapy enhances HIF1-α expression in glioblastoma cells even in normoxia. The therapy affects PHD-2 and VHL which makes them unstable and leads to HIF1-α stabilization [19].

Glioblastoma expresses high levels of Vascular Endothelial Growth Factor (VEGF), especially in necrotic and hypoxic areas, as it is one of the most vascularized cancer. Hence, anti-angiogenic therapies may be more efficient when treating this type of tumour. Angiogenesis-inhibiting drugs display clinical and radiological benefits, yet in 40 % to 60 % of cases, the initial response is then followed by a dramatic progression of the disease. Bevacizumab, a monoclonal antibody targeting the Vascular Endothelial Growth Factor (VEGF), has a beneficial effect on progression-free survival for patients with recurrent glioblastoma when chemotherapy and radiotherapy have failed. After an initial response showing a decrease in tumoural effect, the tumour develops a resistance to the therapy and displays a deeper infiltration in tissues adjacent to the brain [20]. Anti-angiogenic therapies are known to elicit malignant progression as a consequence of the hypoxic conditions induced by the therapy and have been correlated to the early formation of metastasis [7].

1.2.2 The Hypoxia Inducible Factor: the main actor in the response to hypoxia

53 % of malignant tumours including prostate cancer, lung cancer, breast cancer, colon cancer, pancreas cancer, brain cancer, stomach cancer, kidney cancer and melanomas show higher expression levels of the Hypoxia Inducible Factor (HIF). HIF has been detected even in non-hypoxic cancer cell lines. Studies have shown that HIF1-α activation in aerobic conditions can be caused by mutations like VHL loss of function due to PTEN inactivating mutations or activation of mTOR by PI3K [21].

HIF has been the subject of numerous studies as hypoxia is a common effect observed in cancer and due to the mechanism causing the normoxic upregulation of HIF in cancer cells. This part will be dedicated to the description of the HIF protein: its structure, role and mechanism, regulation and finally the impact it has on several cellular processes.

The Hypoxia Inducible Factor (HIF) protein was discovered in 1991 during a study on the erythropoietin gene by Semenza et *al* . They discovered that a DNA sequence was essential for the transcriptional activation of the gene in hypoxia. This DNA sequence was named Hypoxia Response Elements (HRE) and later linked to the HIF protein [22].

HIF is an heterodimeric complex composed of two subunits : an α subunit regulated by oxygen and a β subunit consitutively expressed [8, 22, 23]. Several isoforms of the α subunit were identified namely HIF1-α , HIF2-α and HIF3-α . The structure of HIF1-α can be seen in figure 1.4.

HIF1-α and HIF2-α are the most studied proteins and are considered to be the main actors in the cellular response to hypoxia [8]. The HIF1-α isoform is ubiquitous [8] while HIF2-α would be expressed only in endothelial cells, in the heart, the lung, the placenta and the kidney [24]. Both isoforms share target genes but have distinct targets as well [8, 23, 24], and some genes might be more effectively regulated by one of the two isoforms than the other [23]. These isoforms are regulated by different physiological and pathological conditions. For example, they may play a different role in the tumorigenesis following the microenvironment [8].

Little is known about the HIF3-α subunit, the existence of multiple variants makes the study of its functions harder. Some variants are expressed in specific tissues and are regulated in different ways by hypoxia. The complete HIF3-α protein acts as a transcription factor that triggers a unique transcriptional program in response to hypoxia, but some variants act as negative regulators of HIF1-α and HIF2-α [8]. HIF3-α might mainly be expressed in epithelial cells of the kidney and the lung [24].

HIF subunits contain several domains that are involved in its mechanism of action and its regulation. The units HIF1-α , HIF2-α and HIF-β contain a basic-loop-helix (bHLH) domain, a Per-Arnt-Sim (PAS) and a C-terminal (C-TAD) domain. The α units also contain an oxygen-dependent degradation (ODD) domain and a N-terminal (N-TAD) domain [8].

1.2.3 Role and mechanism of action of HIF

HIF is a master regulator in hypoxic conditions acting like a transcription factor that will bind to HRE to increase the transcriptional activity of the targeted gene. At least 98 target genes involved in 20 pathways have been identified [25], which includes cell proliferation, survival, apoptosis, erythropoiesis and angiogenesis [26]. HIF will impact the metabolism as well: it will cause the cell to favour the use

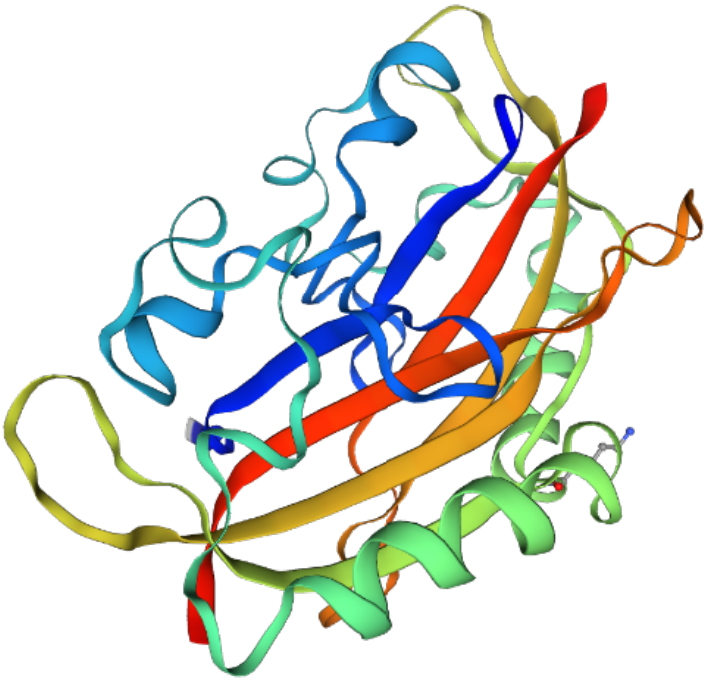

Figure 1.4: 3D Structure of HIF1-α from protein databases.

of anaerobic (glycolysis) pathways rather than aerobic ones (glycolysis and oxidative phosphorylation) to produce Adenosine Triphosphate (ATP) [27]. The table 1.1 summarizes a few HRE identified in the literature.

The binding of HIF to the HRE requires the formation of a complex between the HIF1-α and HIF-β subunit coupled with the CBP/p300 co-factor. The bHLH and PAS domain are involved in the binding between the α and the β subunits while the C-TAD and the N-TAD domains are responsible for the interaction between HIF and the coactivators [8, 22]. Lastly, the ODD domain, only present on α subunits, is important to mediate the oxygen-dependent degradation of HIF [22].

1.2.4 Regulation of HIF by oxygen dependant and independent mechanisms

HIF is mainly regulated at the protein level by oxygen-dependant mechanisms, yet oxygen-independant mechanisms regulating HIF also exist. HIF1-α degradation in normoxia is very fast, the protein has a half-life of roughly 5 minutes [22].

Oxygen-depend degradation of HIF relies mainly on the action of two enzymes: Prolyl Hydroxylases (PHD) and Factor-Inhibiting HIF (FIH). PHD will catalyze the hydroxylation of two prolyl residues on the HIF1-α protein facilitating its binding with the von Hippel Lindau (VHL) protein. VHL then recruit E3 ubiquitin ligases that target HIF1-α for proteasomal degradation [8, 22, 23]. FIH catalyzes the hydroxylation of an asparagin residue on the HIF1-α preventing its binding with the CBP/p300 cofactor and reducing its transcriptional activity [8, 22, 28]. The hydroxylation activity of both enzymes requires α-ketoglutarate (α-KG) and O_2 , they are considered to be oxygen sensors [29].

Mechanisms that do not rely on oxygen to regulate HIF levels have been documented in the literature. Mutation on tumour suppressor genes and oncogenes can cause HIF1-α stabilization [30]. Among them, pathways with oncogenic activation such as PI3K/Akt/mTOR and RAS/RAF/MEK/ERK have been identified to induce HIF accumulation [22]. Figure 1.5 summarizes the regulation mechanisms of HIF.

Oxygen-independent mechanisms can increase the level of HIF in normoxia or decrease it in hypoxia. These includes oncogenic genes that are frequently deregulated in tumours: PTEN [8, 22], mechanistic target of rapamycin kinase (mTOR) [22], EGFR [8], the RAS/RAF/MEK/ERK signalling pathways [22] and IDH1/IDH2 in GBM [29]. It has been found that HIF is hydroxylated even in hypoxia when the production of Reactive Oxygen Species (ROS) is low [28, 29]. Generation of lactate and pyruvate by glycolysis can induce HIF stabilization in normoxia. It has been reported that lactate and pyruvate compete with PHD

Function	Genes
Cell proliferation	Cyclin G2, IGF2, IGF-BP1, IGF-BP-2, IGF-BP-3, WAF-1, TGF-α, TGF-β3
Cell survival	ADM, EPO, IGF2, IGF-BP1, IGF-BP-2, IGF-BP-3, NOS2, TGF-α, VEGF
Apoptosis	NIP3, NIX, RTP801
Motility	ANF/GPI, c-MET, LRP1, TGF-α
Cytoskeletal structure	KRT14, KRT18, KRT19, VIM
Cell adhesion	MIC2
Erythropoiesis	EPO
Angiogenesis	EG-VEGF, ENG, LEP, LRP1, TGF-β3, VEGF
Vascular tone	α 1B -adrenergic receptor, ADM, ET1, Haem oxygenase-1, NOS2
Transcriptional regulation	DEC1, DEC2, ETS-1, NUR77
pH regulation	Carbonic anhydrase 9
Regulation of HIF-1 activity	P35srj
Epithelial homeostasis	Intestinal trefoil factor
Drug resistance	MDR1
Nucleotide metabolism	Adenylate kinase 3, Ecto-5'-nucleotidase
Iron metabolism	Ceruloplasmin, Transferrin, Transeferrin receptor
Glucose metabolism	HK1, HK2, AMF/GPI, ENO1, GLUT1, GAPDH, LDHA, PFKBF3, PFKL, PGK1, PKM, TPI, ALDA, ALDC
Extracellular-matrix metabolism	CATHD, Collagen type V (α1), FN1, MMP2, PAI1, Prolyl-4-hydroxylase α (1), UPAR
Energy metabolism	LEP
Amino-Acid metabolism	Transglutaminase 2

ADM, adrenomedullin; ALDA, aldolase A; ALDC, aldolase C; AMF, autocrine motility factor; CATHD, cathepsin D; EG-VEGF, endocrine-gland-derived VEGF; ENG, endoglin; ET1, endothelin-1; ENO1, enolase 1; EPO, erythropoietin; FN1, fibronectin 1; GLUT1, glucose transporter1; GLUT3, glucose transporter 3; GAPDH, glyceraldehyde-3-P-dehydrogenase; HK1, hexokinase 1; HK2, hexokinase 2; IGF2, insulin-like growth-factor 2; IGF-BP1, IGF-factor-binding-protein 1; IGF-BP2, IGF-factor-binding-protein 2; IGF-BP3, IGF-factor-binding-protein 3; KRT14, keratin 14;KRT18, keratin 18; KRT19, keratin 19; LDHA, lactate dehydrogenase A; LEP, leptin; LRP1, LDL-receptor-related protein 1; MDR1, multidrugresistance 1; MMP2, matrix metalloproteinase 2; NOS2, nitric oxide synthase 2; PFKBF3, 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase-3;PFKL, phosphofructokinase L; PGK 1, phosphoglycerate kinase 1; PAI1, plasminogen-activator inhibitor 1; PKM, pyruvate kinase M; TGF-α, transfor-ming growth factor-α; TGF-β3, transforming growth factor-β3; TPI, triosephosphate isomerase; VEGF, vascular endothelial growth factor; UPAR,urokinase plasminogen activator receptor; VEGFR2, VEGF receptor-2; VIM, vimentin.

Table 1.1: HIF targets genes in different biological functions.
Source: Lee et *al* [26]

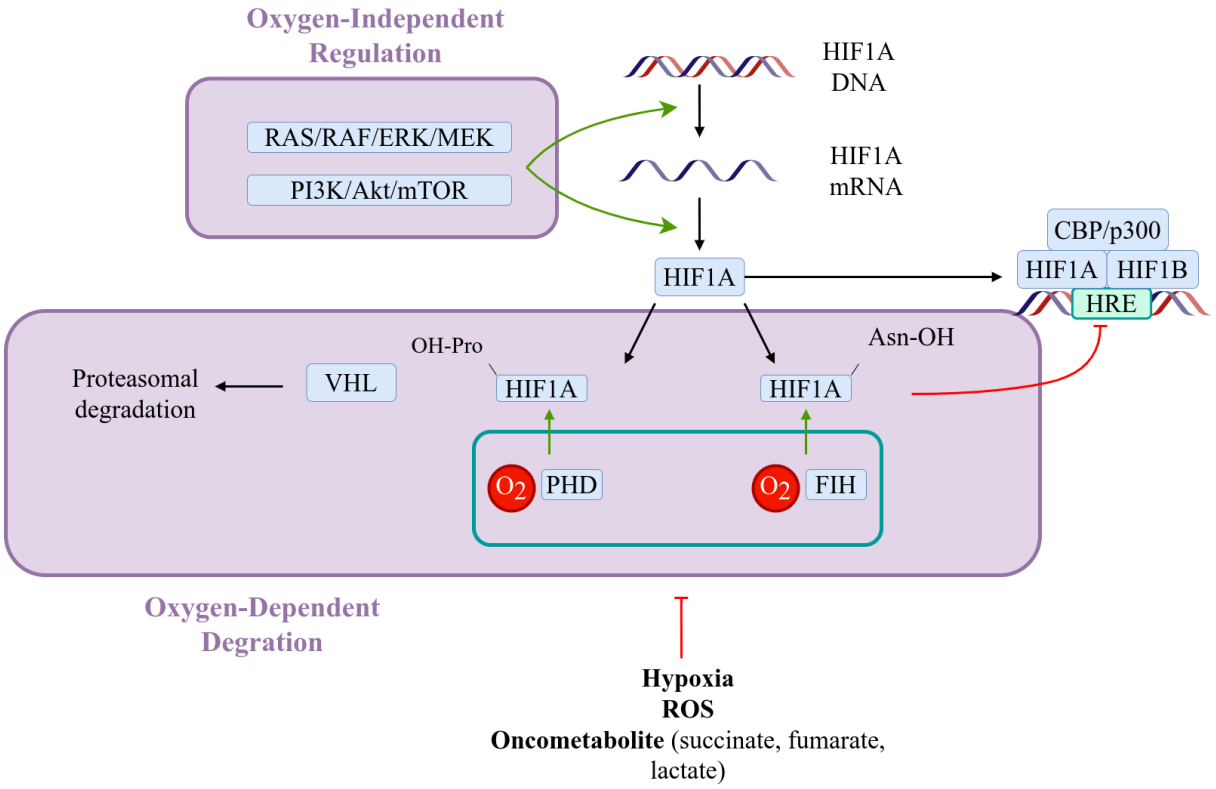

Figure 1.5: Regulation of Hypoxia Inducible Factor by oxygen-dependent and oxygen-independent mechanisms. Source : Spinicci et *al* [31]

for binding to α-KG which inhibits PHD activity [29]. Interaction between HIF and the tumour suppressor protein p53 has been reported, though its exact nature is quite controversial [32].

1.2.5 Adaptation of the cell to hypoxia

In this part, I will describe briefly the impact that HIF has upon the cell cycle, the metabolism and the epithelial to mesenchymal transition (EMT). The models I have developed include the regulation of the cell cycle through the quantity of ATP produced and, in a later version, the migration of the cell. I will not go into deep details on the mechanism used by HIF to regulate these processes as they are not modelled. The impact of HIF on metabolism will be introduced here but will be presented in more details in the section about metabolism. Despite the major role of HIF in angiogenesis, I will not talk about it in this thesis since it is not part of this study.

Literature indicates that HIF can prevent the G1/S transition by regulating cyclin-dependant kinase (CDK) inhibitors (p21, p27) and cyclins proteins (cyclin G2, cyclin E) [33]. Still, the activation of p27 in hypoxia and its role in the cell cycle arrest is discussed [27]. Cyclin E downregulation is mediated through the direct inhibition of cyclin D by HIF [34, 35], causing a slowing down or arrest of the cell cycle in the G1 phase and promoting the entry into quiescence, which can be a mechanism to escape chemotherapy [36]. Cyclin G2 is considered as a tumour suppressor in glioma due to its role in suppressing cell proliferation, initiating cell apoptosis and negatively regulating LDHA a critical enzyme of anaerobic pathways [37]. Interestingly, a study has reported a doubled rate of proliferation of several GBM cell lines at 1 % O_2 that can be attenuated when inhibiting HIF1-α and HIF2-α [20].

Certain transcriptional effects induced by HIF1-α cause a phenotype similar to the Warburg Effect, which I will present in the section on metabolism, making HIF a potential regulator of this behaviour. HIF1-α activates the transcriptional activity of the gene SLC2A1 and SLC2A3 coding respectively for the GLUT1 and GLUT3 transporters: two transporters that facilitate the entry of glucose in the cell. In addition, HIF1-α induced the expression of PDK1 an inhibitor of PDH, a key enzyme for the use of TCA cycle [21].

Evidences supports the idea that detachment of cancer cells from the tumour is initiated by an embryonic developmental program called EMT where cells lose their cell-cell interaction and acquire

a more motile and invasive phenotype called *mesenchymal* [8]. Hypoxia has been reported to induce mesenchymal changes in GBM, which can be reduced by inhibiting HIF with no differences between HIF1-α and HIF2-α . Exposure to 0.2 % or 1 % O_2 elevates the levels of EMT associated genes such as Slug, Snail, TWIST and MMP [20]. HIF1-α has been recognised as a mastor regulator of the EMT, invasion and metastasis in breast cancer [24]. It has been suggested that hypoxic conditions created by therapy such as bevacizumab can trigger EMT. An *in vivo* experiment that two tumours both treated with radiotherapy, chemotherapy and temozolomide, but with one only treated with bevacizumab, revealed that cells migrated 5 cm from the tumour when treated with bevacizumab [20].

1.3 Metabolism

1.3.1 The energetic metabolism of the cell

The cell produces Adenosine Triphosphate (ATP) through aerobic or anaerobic processes depending on the availability of oxygen in the environment, yet, glycolysis is involved in both processes. Glycolysis is a set of reactions producing energy by turning glucose into pyruvate with a yield of two ATP molecules [38, 39] (the steps of glycolysis are shown in figure 1.6).

Overall equation of glycolysis:

$$Glucose + 2ADP + 2P_i + 2NAD^+ \rightarrow 2Pyruvate + 2ATP + 2NADH + 2H^+$$

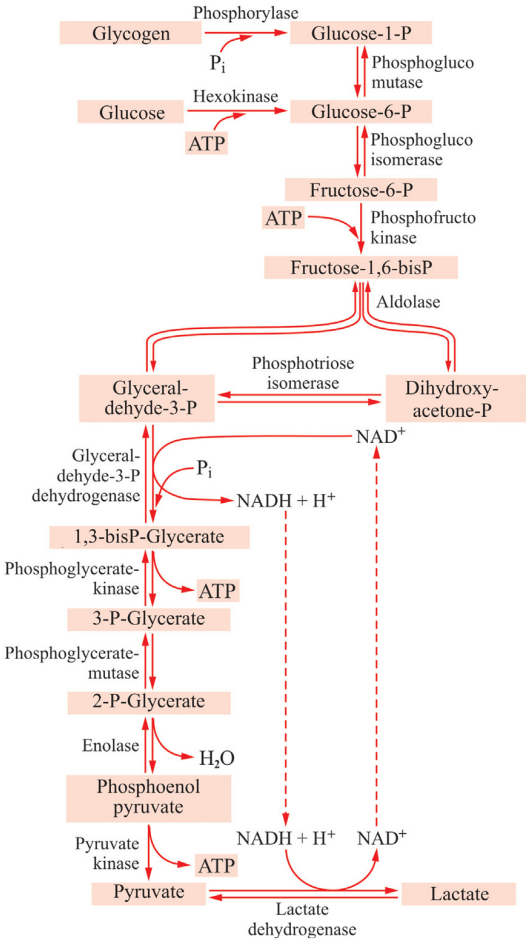

Figure 1.6: The different reactions of the glycolysis.
Source: A. Blanco et G. Blanco, « Chapter 14 - Carbohydrate Metabolism », in Medical Biochemistry [40]

The fate of pyruvate is dictated by the availability of oxygen in the environment. When oxygen is not available for the subsequent aerobic pathways, pyruvate is turned into lactate by the action of the lactate

dehydrogenase (LDH) enzyme to regenerate the enzymes required to carry on the glycolysis [40]. The lactate produced is then secreted in the environment by the monocarboxylate transporters (MCTs).

Conversion of pyruvate to lactate (when oxygen is low):

$$Pyruvate + NADH + H^+ \rightarrow Lactate + NAD^+$$

When oxygen is available, pyruvate is turned into Acetyl CoenzymeA by the action of the pyruvate dehydrogenase (PDH) enzyme to be used in the tricarboxylic acid (TCA) cycle (see figure 1.7 for the step of the TCA cycle).

Overall reaction of oxidative phosphorylation:

$$Pyruvate + 5O_2 + 38ADP + 38P_i \rightarrow 38ATP$$

The theoretical yield of ATP is 38 but experimental measurements have shown that the actual yield is lower due to losses [41].

The TCA cycle generates NADH and $FADH_2$ that are used by the different members of the electron chain during the oxidative phosphorylation (OXPHOS). Oxidation of NADH and $FADH_2$ by the electron chain creates a high proton concentration in the intermembrane space of the mitochondrial matrix, giving energy to the ATP synthase to produce ATP when oxygen is used by the last member of the chain. This process is generally called more efficient than glycolysis as it generates a theoretical yield of 38 ATP molecules [40]. Figure 1.7 shows the metabolism of glucose when oxygen is available.

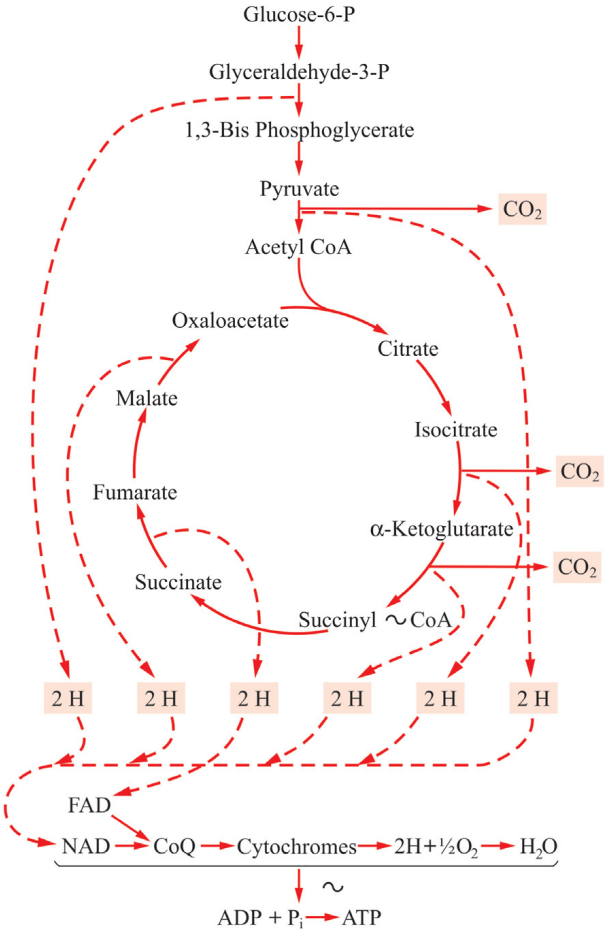

Figure 1.7: Oxidation of glucose under aerobic pathways.
Source: A. Blanco et G. Blanco, « Chapter 14 - Carbohydrate Metabolism », in Medical Biochemistry [40]

Glycolysis is considered to be an ancient way of producing energy as it is commonly accepted that cells rose in an atmosphere deprived of oxygen [21]. Constrained-based modelling such as flux balance analysis (FBA) have been the most used models to simulate cancer cell metabolism [42].

In glioblastoma, the combination of genetic alterations induces multiple metabolic phenotypes. For example, mTOR targeting therapies limit cellular proliferation, glucose uptake and lactate production but fail to promote cell death [5]. The IDH enzyme, responsible for the carboxylation of isocitrates in α-KG, producing NADPH, has an activity decreased by 50 % in GBM. Combined with impaired mitochondrial function, IDH mutant GBM produces less energy and has a better patient prognosis [8].

HIF is able to modulate the energetic metabolism by inducing transcriptional changes favouring the use of glycolysis without coupling with the TCA cycle. Firstly, HIF downregulate the PDH enzyme by enhancing the synthesis of pyruvate dehydrogenase kinase (PDK) [27]. Secondly, it promotes a glycolytic phenotype by stimulating the production of specific isoforms of the glycolytic enzymes: HKI/HKII, HPI, PFK-L, ALD-A/ALD-C, TPI, GAPDH, PGK1, PGAM-B, ENO-α, PYK-M2, LDH-A [26] (figure 1.8).

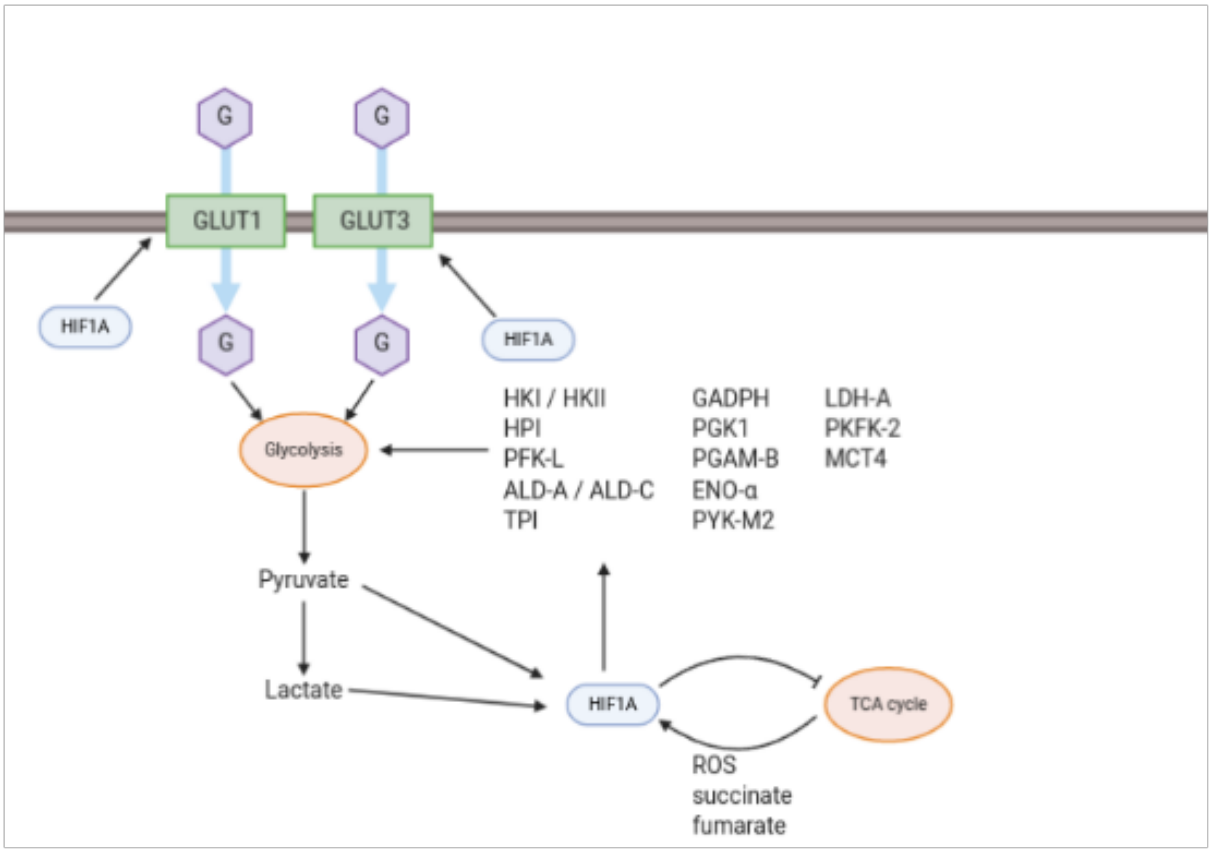

Figure 1.8: Summary of the impact of HIF1-α on the energetic metabolism. HIF will increase the expression of specific isoforms of the glycolytic enzymes while reducing the usage of TCA.

1.3.2 The acidic environment of cancer cells

Tumour microenvironment is characterized by a lower pH compared to what is seen in normal tissues. Furthermore, it is widely acknowledged that tumour cells survive at lower pH than normal cells. The first model developed during the thesis questions the implication of HIF in the metabolism and the consequences with a focus on the Warburg effect. Otto Warburg conducted in 1927 an experiment to measure the consumption of glucose, oxygen and the production of lactate by the tumour. He measured the quantity of each molecule in the veins before and after passage through the tumour. He observed that the tumour consumed more glucose and produced more lactate than healthy tissue with 66 % of the consumed glucose being used for fermentation [43]. This “over-production” of lactate has been defined as the “Warburg Effect”. Warburg hypothesized that the mitochondria in cancer cells was dysfunctional and that the cell used fermentation to make up for the energy need [44]. We know that this is rarely the case and most of the time the mitochondria is functional in cancer cells. There is no clear definition for the

Warburg effect since the meaning of the term seems to have changed over time. Nowadays, this term is mainly associated with aerobic glycolysis [45], that is the use of anaerobic processes even when oxygen is available. The usage of fermentation (glycolysis followed by lactate secretion) when oxygen is not available is normal for the cell. Hence this definition highlights that what makes the use of fermentation "abnormal" is the presence of oxygen for aerobic pathways. Scientists studied the ability of HIF to induce a Warburg effect owing to its role in the glycolytic phenotype. It was shown that increased activity of HIF is enough, but not compulsory, to trigger the Warburg effect [46].

The acidic microenvironment is a driver of cancer that can facilitate tumour growth. An acidic pH can cause breaks on the DNA strands and delay DNA repairing systems in non-cancerous cells. *In vitro* and *in vivo* data support the hypothesis that pH can influence epigenetic and RNA processing in cancerous cells. Exposure to an acidic environment can favour genomic instability and accumulation of chromosomal aberrations. Epigenetic changes may contribute to an early transition to a cancerous state. Furthermore, acidic pH can lead to the stabilization of HIF with the accumulation of a derivated of α-KG caused by a mutation on the IDH gene present in glioblastoma. Lactate can upregulate VEGF in a hypoxia-independent manner and trigger angiogenesis [47]. Acidity is also a mechanism to evade the immune system as it disturbs the metabolism of T cells [48]. In addition to its effect on the immune system, lower pH is seen as a mechanism of invasion as it increases apoptosis among the healthy cells [49]. Acidity stimulates HIF functions and it has been shown that cells in low pH were more invasive both *in vitro* and *in vivo* [13, 50]. Tumour cells preferentially invade healthy tissues in lower pH while little to no invasion is observed in alkaline pH. Similarly, type I and IV collagen degradation was increased at pH 6.4 compared to pH 7.4 [50]. Moreover, comparable acidity was found in tumours with no lactate accumulation generated by glycolysis. A proposed mechanism was accounted to the carbonic anhydrase (CA)-9, a protein upregulated by HIF, which is involved in pH regulation and acidification of the environment [51].

1.4 Invasion of distant tissues by cancer cells

1.4.1 The Extra-Cellular Matrix: a complex structure

The Extra-Cellular Matrix (ECM) is a complex structure containing 300 proteins, water and polysaccharides. Its composition can greatly vary depending on the tissue (figure 1.9).

The proteins of the ECM regulate tissue homeostasis, organ development, inflammation and diseases. Most of them are fibrous proteins such as collagen, elastin, fibronectin and laminin. The ECM contains proteoglycans such as chondroitin sulphate, heparan sulphate, keratan sulphate and hyaluronic acid [53, 54]. ECM in the brain has a different composition, it contains hyaluronan, proteoglycans, tenascin-C and does not have rigid structures such as fibrous collagen [13].

The stiffness of the matrix is the resistance of the matrix to deformation in response to the forces applied. This characteristic of the matrix can influence the cell migration direction and its migration velocity. The stiffness can be influenced by the architecture of the tissue in 3D (figure 1.10).

The ECM has an important role as it dictates the behaviour of the cell: it influences its proliferation, growth, migration and apoptosis [54]. The ECM manages the cellular organisation by interacting with receptors on the cell surface to promote the activation of signalling pathways that trigger a variety of physiological responses [56]. The formation of the mechanical microenvironment depends heavily on the intracellular component (vimentin, actin and neurofilaments), the extracellular components (collagen and fibrin), inter-cellular signals (integrins) and stromal cells (fibroblast) [17].

The stiffness of the matrix can influence angiogenesis: increases of stiffness inhibit angiogenesis [57], and compaction of cells increases the expression of angiogenic factor [58].

Compared to healthy brain, the matrix in glioblastoma has a higher content of collagen, laminin and fibronectin mainly in the basement membrane of blood vessels [54]. In progressive glioma, the brain matrix becomes more disorganized and stiffer with the growing malignancy of the tumour [7]. Increased collagen deposition is associated with breast cancer development and progression [59]. In cancer, the enhanced deposition of collagen is the most recognized alteration of tumorous tissues. Cancer cells recruit cancer associated fibroblast (CAF) to stimulate the synthesis of ECM proteins. Hypoxic areas in the tumour have increased deposition of collagen [53].

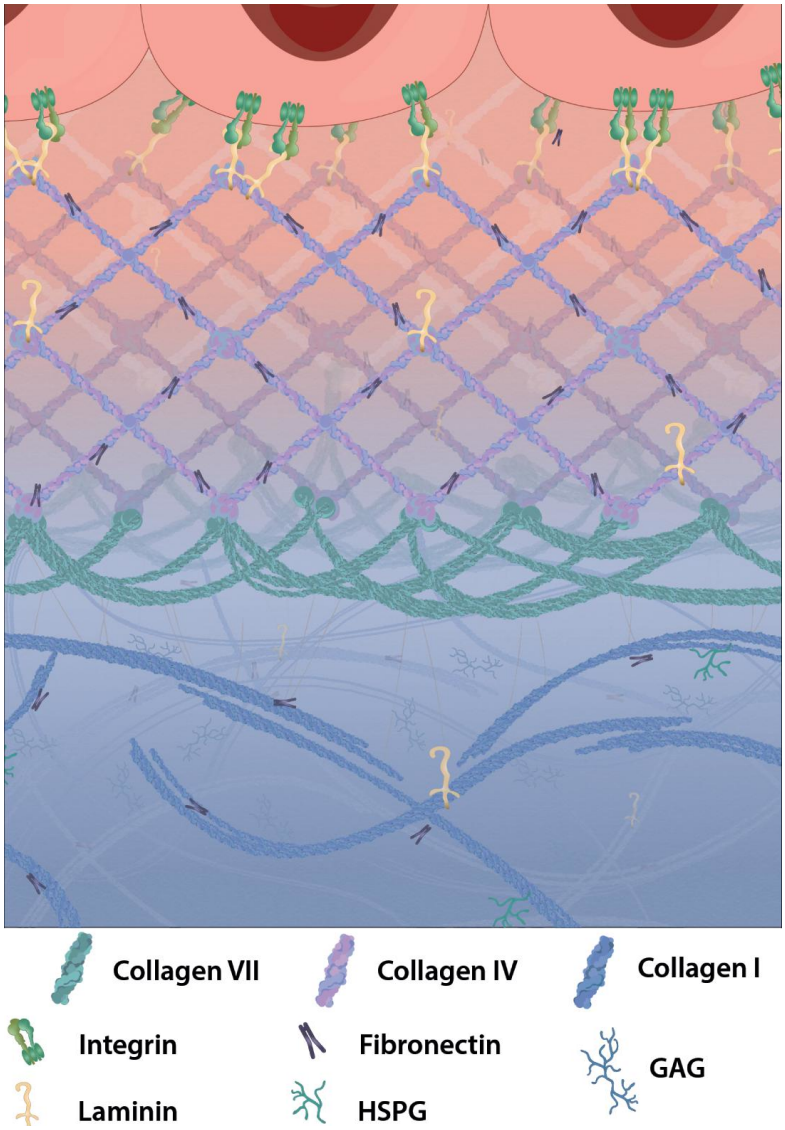

Figure 1.9: Composition of the Extra-Cellular Matrix (ECM) in the mammalian skin. Adapted from Pfisterer et *al* [52]

1.4.2 Collagen: the main constituent of the ECM

Collagen is the most abundant fibrous protein constituting up to 30 % of the total proteins in humans. Fibroblasts are the main source of collagen, they are responsible for the organization and alignment of collagen fibres [60]. Yet, it has been shown that glioblastoma cells are able not only to produce collagen-like fibroblasts [61, 62] but fibronectin and laminin as well [62]. Various pathologies such as osteoporosis or arterial aneurysm have been associated with collagen mutations. Tumourigenesis affects both the structure and biosynthesis of collagen. The collagen family is large with 28 different collagen types classified into 7 categories. Only the fibril-associated collagens with interrupted triples helices (FACITs) will be covered in this thesis for their role in the ECM composition and the cell migration. Fibrillar collagen includes the types I, II, III, V, XI, XXIV and XXVII. It functions as a tensile element for the tissues and can affect cell migration, adhesion, angiogenesis, tissue development and repair by sending signals to the cells [60]. Usually, one type of collagen is predominant in one structure. Type I collagen is commonly found in skin, bone, teeth, tendons, ligaments, vascular ligatures and organs. Type II collagen is found in cartilage. Type III collagen is found in the skin, muscle and blood vessels. Type IV collagen is found in the basement membrane (in the epithelium-secreted layer) and in the basal lamina. Type V collagen is found at the cell surface and in the placenta [63].

The biosynthesis of collagen is a complex process that includes intra-cellular post-translational modification and assembly of pro-collagen chains (figure 1.11).

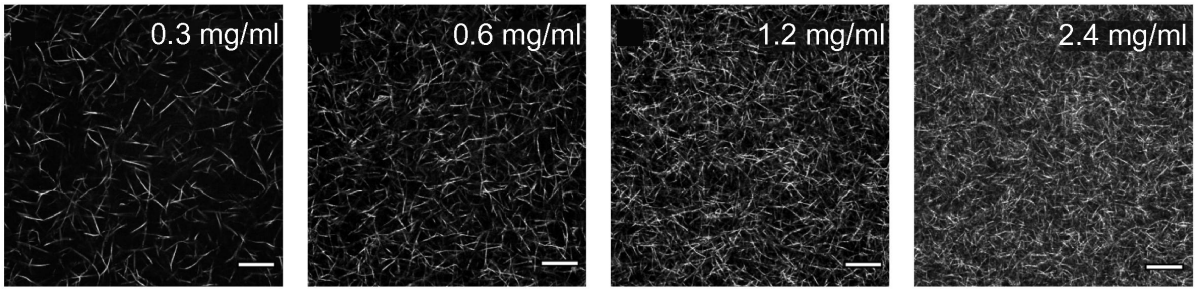

Figure 1.10: Optical section of collagen gel with different concentrations of collagen (top-right).
Source: Lang et *al* [55]

Collagen is composed of three α chains assembled in a triple helical structure formed by a repetition of the motif *GXY* where *G* is a glycine, *X* is often a proline or a lysine and *Y* is often a hydroxyproline. C-terminal and N-terminal peptides from pro-collagen are removed by procollagen C-proteinases and procollagen N-proteinases resulting in the formation of tropocollagen that can self-assemble into fibrils. Hydroxylation of proline residues by prolyl 4-hydroxylase (P4H) is essential for the proper folding of collagen chains into the triple helical structure. Hydroxylation of lysine by procollagen-lysine,2-oxoglutarate 5-dioxygenase (PLOD) drives the stability of collagen cross-linking giving collagen its tensile strength and mechanical stability [56, 64]. Collagen fibrils are further stabilized by lysyl oxidase (LOX) cross-linking. LOX is a copper-dependent enzyme that deaminates the ε-amino groups of specific lysine forming highly reactive aldehyde leading to the formation of intra- or inter-molecular cross-links [65, 66]. Four lysyl oxidase-like proteins have been identified: LOXL, LOXL2, LOXL3 and LOXL4. Recently, LOX has been associated with the functions of tumour suppression, cellular senescence developmental control and chemotaxis [65].

1.4.3 Collagen degradation by matrix metalloproteinases

Degradation and remodelling of the matrix in localized regions facilitate the growth and the spreading of cancer due to the available spaces created by degraded tissue [8, 67, 68]. Collagen is tightly regulated by cells as it participates in the bone and cartilage structure. The proteolysis of collagen is considered to be a contributing factor to several pathologies such as tumour cell spreading, arthritis, tissue ulcerations, cardiovascular diseases and neurodegenerative diseases. An interesting property of collagen is its ability to regulate its own levels in the pericellular environment through a dynamic feedback mechanism that not only regulates its expression but also the activity of the proteinases involved in its degradation. Collagen can be degraded by different molecules such as the proteinases family “a disintegrin and metalloproteinases (ADAM)” and the cathepsins, proteases involved in protein turnover [68–71].

However, most of the collagen degradation activity seems to be attributed to matrix metalloproteinases (MMPs), a family of Zn^{2+} dependent endopeptidases. They are inhibited by tissue inhibitors of MMP (TIMP) which comprises a family of four protease inhibitors namely TIMP-1, TIMP-2, TIMP-3 and TIMP-4. The collagenolytic activity of several MMPs is redundant and compensatory. Interstitial collagen is cleaved at specific sites by all collagenolytic MMPs, 3/4 away from the N-terminus. Cleavage initiates the denaturation of the triple helical collagen into gelatin which is sensitive to many different proteinases. Interstitial collagen can be processed by MMP1, MMP2, MMP8, MMP9, MMP13 and MT1-MMP. MMP activity is required, at least partially, for the stromal invasion by epithelial cells stimulated by growth factor in organ-cultured skin [69, 71–74].

Further to the assembly of collagen fibrils into fibres, collagen cross-linking seems to form fibres resistant to proteolysis [69, 71]. A study has shown that corneal collagen cross-linked by RFUVA was resistant to degradation by MMP1, MMP2, MMP9 or MMP13. The author emitted the hypothesis that structural changes may prevent the access of MMPs to their cleavage site.

Extracellular pH can impact the activity of different enzymes that are involved in the ECM remodelling. It has been shown that acidic regions coincide with higher levels of CAIX (CA9), LDHA, MMP9 and MT1-MMP in *in vivo* murine tumour model. MMPs and cathepsins are secreted by cancer cells in an acidic environment to catalyze ECM degradation [47]. CAFs secrete MMPs (such as MMP2, MMP3 andMMP9) or activate proteins to promote ECM degradation, matrix remodelling, EMT and cancer cell stemness [17].

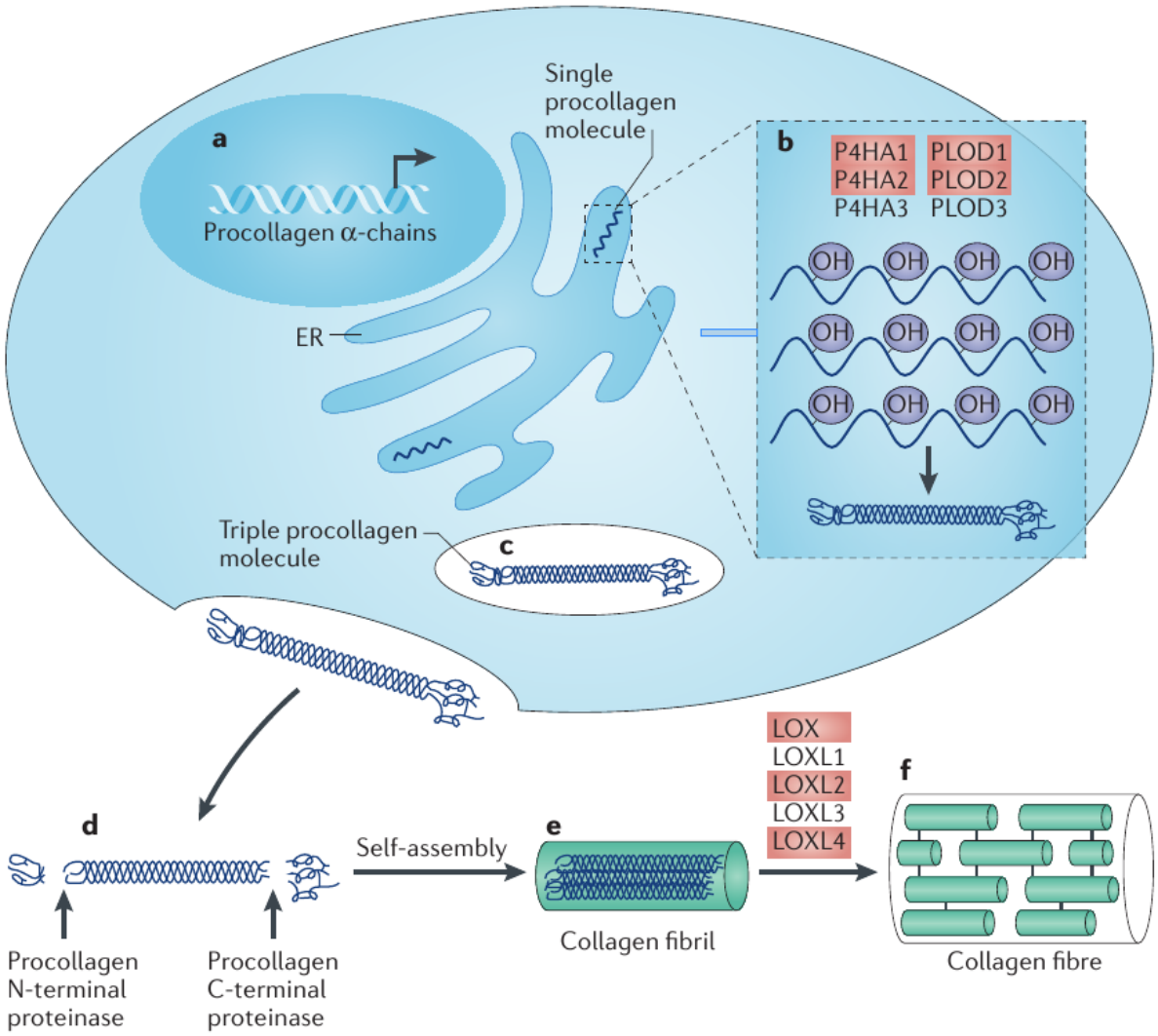

Figure 1.11: Biosynthesis of collagen. Procollagen α-chains are modified by the action of P4Hs and PLODs enzymes to allow collagen chains to form fibrils and to increase the stability of cross-linkings. Secreted procollagen molecules are cleaved at the N-terminal and C-terminal by proteinases to allow for the self-assembly into collagen fibrils. LOX reaction leads to collagen cross-linking and increased stiffness of the ECM.
Source: Gilkes et *al* [53]

1.4.4 The cellular migration of cancer cells

Four essential hallmarks of metastatic cells have been proposed: motility and invasion, the ability to modulate the metastatic micro-environment, high and reversible cellular plasticity of metastatic cells, the ability to proliferate and colonize other tissues. Generally, cancer cells spread through the bloodstream but they can also spread through lymphatic vessels or coelomic cavities (although both are connected to the bloodstream) [7]. The balance between proliferation and migration can be seen as a go-or-grow dichotomy as biological evidence suggests that some signalling pathways are shared by proliferation and migratory processes. Hypoxia influences not only the regulation of angiogenesis but ECM production and degradation as well, thus it is a factor controlling this balance [49]. Transition to an invasive phenotype is driven by changes in adhesion independent from the tissue of origin. These changes include loss of cadherin-dependent junction and loss of hemidesmosomes (if they were present). E-cadherin is important for epithelial cell-cell adhesion, strong evidence suggests that loss of E-cadherin mediated adhesion is essential for the adoption of a malignant phenotype [68].

The key process involved in the metastasis is the epithelial to mesenchymal transition (EMT). During EMT, tumour cells upregulate mesenchymal markers, stop expressing surface and cystoskeleton proteins, lose their polarity and cell-cell adhesion, adopt *spindle-like* instead of an epithelial shape [7, 13]. In glioma, a transition to a mesenchymal sub-type is often observed after recurrence and is associated with

lower survival. Fast changes can be induced by drug or radiation therapy which results in loss of effect of therapies targeting one sub-type [13]. Cells which have undergone EMT can transition back through a process called mesenchymal to epithelial transition (MET) [7], and form a new tumour [13].

Two migration modes, based on the motion type, have been identified in the literature: *mesenchymal* or *amoeboid* (figure 1.12).

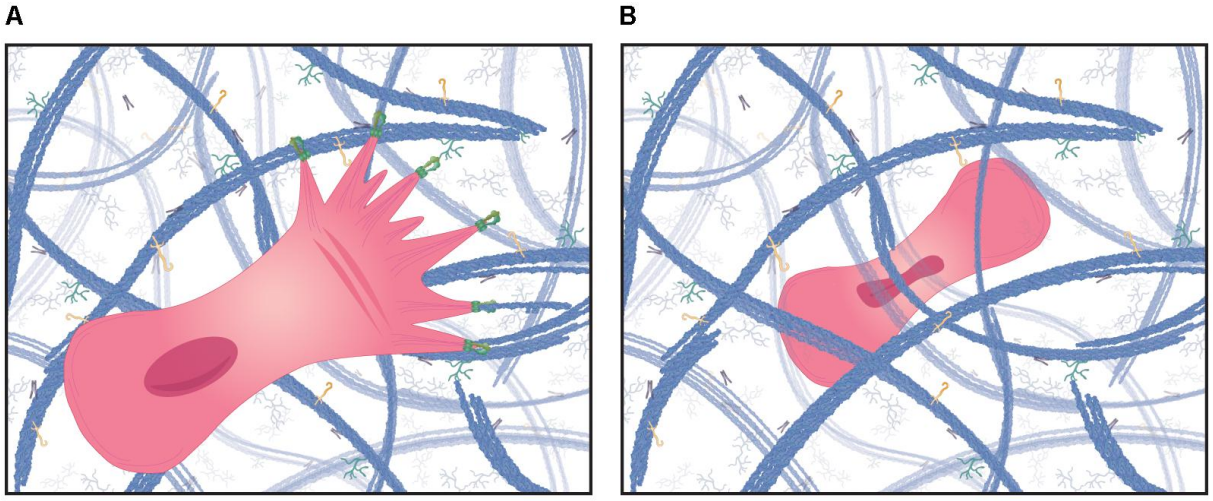

Figure 1.12: Cell migration modes: (a) mesenchymal and (b) amoeboid.
Adapted from Pfistere et *al* [52]

In the *amoeboid* mode, the cell changes its shape by developing protrusions and retracting them to move forward. In that case, cells can squeeze through the matrix pores with minimal attachment and do not rely on adhesion molecules (such as integrins). In the mesenchymal mode, cells have a highly polarized shape and develop protrusions (called lamellipodia or filopodia) that form focal adhesion points with the fibres. Here, the cell can retract their protrusions to generate traction forces while breaking the bonds at the rear [75, 76]. A major component for cell migration is the cell-fiber interaction: cells are slower in a stiff matrix but tend to migrate to stiff areas of the matrix [77].

2 Mathematical and Computational approaches

The Thesis aims to investigate cell metabolism in the context of a growing tumour. To do that, a hybrid multiscale model integrating cell and tissue scales has been developed. It addresses the metabolism at the cell level and integrates this into a tumour at the tissue level taking into account the substrates at the environmental level. This chapter will present the mathematical and computational approaches to describe the metabolism, biological networks and tumour growth. I will explain what is the current state of Hypoxia Inducible Factor (HIF) modelling and what are the features implemented in the model. Here, I will mostly focus on the computational method and the numerical implementation while mathematical description will be explained in the related chapters.

2.1 Landscape of modelling techniques

2.1.1 Modelling of the metabolism

Flux balance analysis (FBA) is a method to study a network of biochemical reactions that can be large up to the genome-scale. FBA represents metabolite reactions as a stochiometric matrix **S** of size $m \times n$ where m is the number of metabolite in the network and n is the number of reactions. Each value in the matrix **S** is the stochiometric coefficient of the metabolite in one reaction. The coefficient is positive when a metabolite is produced, negative when it is consumed or null when it does not participate in the reaction. Fluxes in FBA are represented by a vector **v** of size n with a set of upper and lower bounds for each v_i. This method assumes that the system studied is (1) at a steady-state, and (2) optimized for a phenotype. For example, we may want to study which phenotype, or fluxes value, maximizes growth or ATP production. The phenotype maximizing growth is mathematically described by an objective-function of the form $Z = C^T v$ where C is a vector of weights representing how much each reaction contributes to the phenotype of interest. The system $Sv = 0$ is then solved with linear programming. FBA does not require any kinetic parameters and has a low computational cost making them attractive to study perturbation applied to biological systems like genetic manipulations. This method does have limitations though. They are not suitable for metabolite concentration predictions. Moreover, no regulatory effects (like gene regulations) are considered as the system is predicted at a static state. Hence, this method cannot be used to represent a temporal dynamic [41, 78].

Description of metabolite quantity over time has typically been achieved through modelling using ordinary differential equations (ODEs). A plethora of mathematical frameworks are available to the modeller, yet the common ones rely mainly on the mass-action law introduced over 150 years ago by two Swedish scientists, the mathematician Gulberg and the chemist Waage. The formulation of the Mass-Action law describes a simple system allowing to keep the equations very simple [41, 79, 80].

Mass-Action Law
In mass-action law, the reaction:

$$\eta_1 X_1 + \eta_2 X_2 + \cdots + \eta_N X_N \rightarrow Y,$$

is represented by the equation:

$$\frac{dY}{dt} = k \prod_{i=1}^{N} X_i^{\eta_i},$$

where k is a constant, η_i is the stochiometric coefficient of the compound X_i and Y the product of the reaction.

The power-law, a new mathematical formalism based on the mass-action law, was proposed by Savageau et *al* in 1987 to structure the description of biochemical reactions using a simple mathematical formulation. A noticeable inconvenient of the power-law is that it does not saturate.

Power-Law

$$\frac{dY}{dt} = \alpha \prod_{i=1}^{N} X_i^{k_i} - \beta \prod_{i=1}^{N} X_i^{k_i}, \tag{2.1}$$

where α and β are the rate constant of the increase and decrease of the metabolite Y, k_i is the kinetic order of the metabolite X_i. $k_i > 0$ represents an activation, while $k_i < 0$ represents an inhibition.

One of the most used functions in biochemical systems is the Michaelis-Menten equation. It was originally introduced to represent enzyme kinetic reactions under the two assumptions of (1) equilibrium (the total concentration of enzyme does not vary significantly over time) and (2) steady-state (the concentration of the enzyme-substrate complex is approximately constant). Compared to the power-law formalism, the formulation of the Michaelis-Menten equation does saturate when the quantity of substrate is far higher than the concentration of enzyme [41, 81]. The impact of the different parameters of the Michaelis-Menten equation can be seen in figure 2.1.

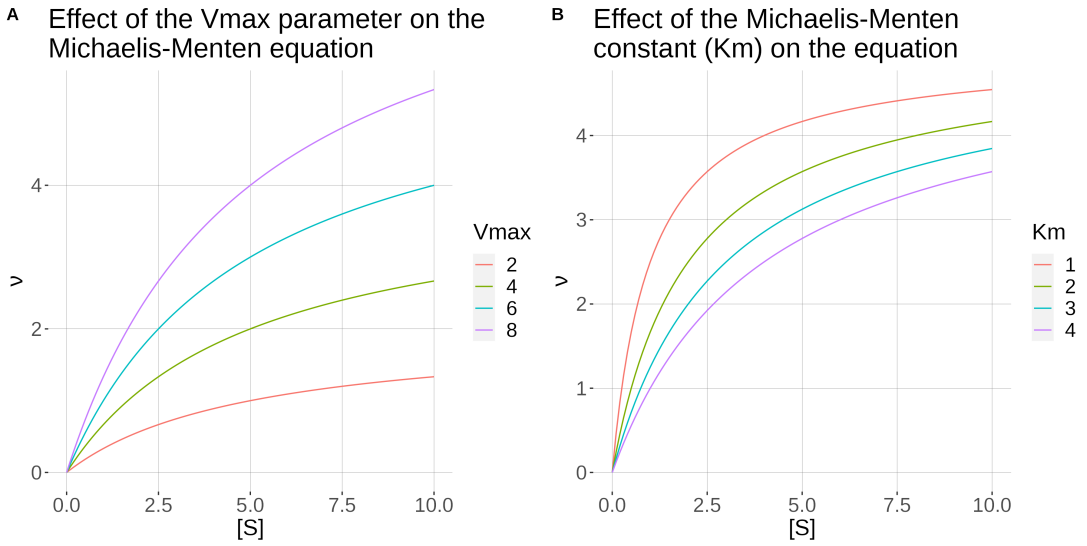

Figure 2.1: Effect of (A) the V_{max} parameter and (B) the K_m parameter on the Michaelis-Menten equation. $[S]$ is the substrate concentration and ν is the speed of the reaction.

Michaelis-Menten equation:

$$\nu = V_{max} \frac{[S]}{K_m + [S]}. \tag{2.2}$$

Where ν is the speed of the the reaction, V_{max} is the maximal speed, K_m is the Michaelis-Menten constant and $[S]$ is the substrate concentration. The Michaelis-Menten constant is the substrate concentration at which the speed of the reaction is half-maximal.

The formulation of the Hill function is similar to the Michaelis-Menten one but further includes cooperativity between the enzyme and its substrate. Some enzymes change their structures when a ligand bounds to one of their receptor, increasing its affinity to that same ligand. For example, when oxygen binds to the haemoglobin, its structural conformation changes to increase its affinity and carry more molecule of oxygens. This phenomenon captured by the Hill function can be quantified using the hill

coefficient term. Hill function is commonly used in pharmacodynamics to study drug-dose response [82]. The impact of the different parameters of the Hill equation can be seen in figure 2.2.

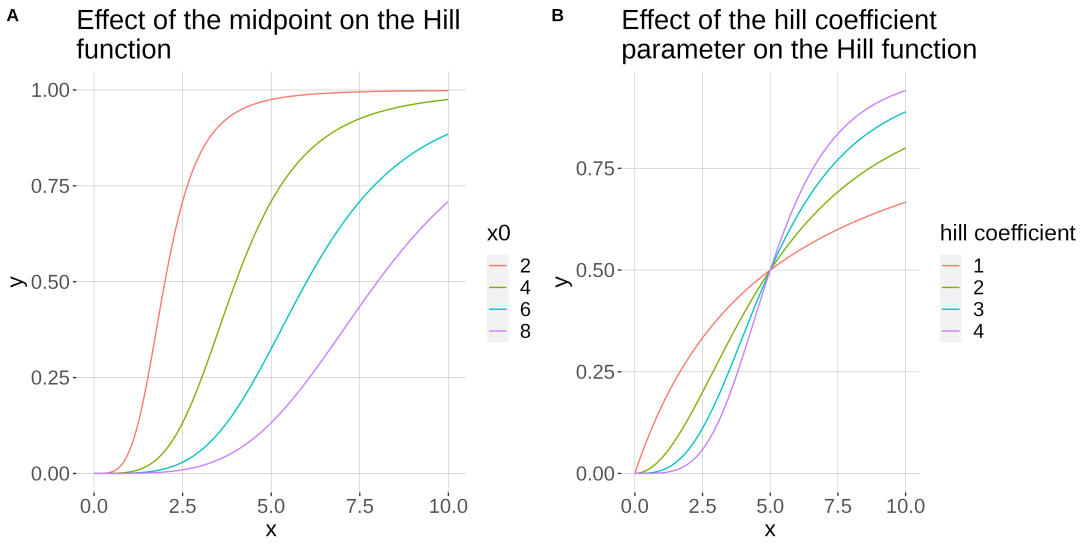

Figure 2.2: Effect of (A) the midpoint x_0 and (B) the hill coefficient n on the Hill equation.

Hill-function:

$$B = \frac{L^n}{K_D^n + L^n}. \tag{2.3}$$

The Hill function was developed to describe the fraction B of a protein bound to a ligand. L is the concentration of ligand, K_D is the dissociation constant and n the Hill coefficient. The Hill coefficient is a way to quantify the interaction between a ligand and the binding sites of the protein studied.

Although it is less used for biochemical reactions, the logistic function is a simple way to represent regulating mechanisms. Like the Michaelis-Menten and the Hill function, the logistic function does saturate but unlike them its curve is symmetrical around the midpoint. The impact of the different parameters of the logistic function can be seen in figure 2.3.

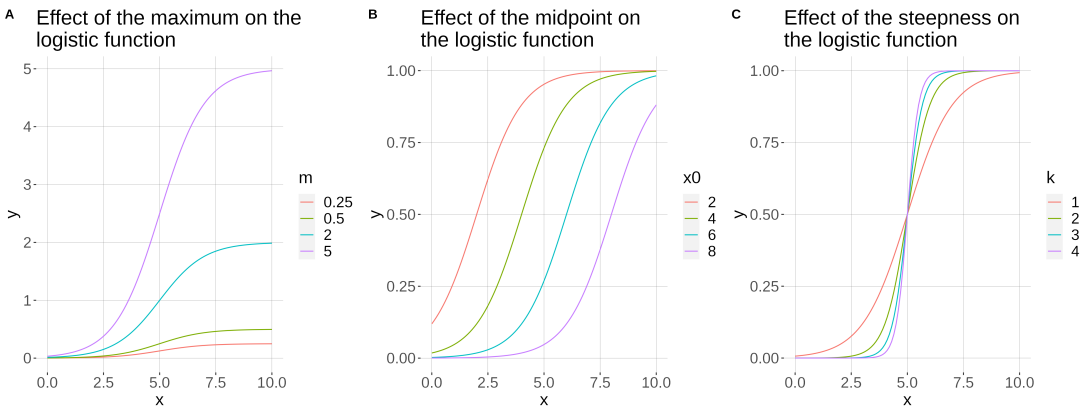

Figure 2.3: Effect of (A) the maximum m, (B) the midpoint x_0 and (C) the steepness k on the logistic function.

Logistic function:

$$\frac{dY}{dt} = \frac{m}{1 + e^{-k(x-x_0)}},$$

where m is the maximum of the function, x_0 is the mid-point and k is the steepness of the curve.

Most of these equations have been derived to include more phenomenon. For example, the Michaelis-Menten function has been modified to account for bi-directional reaction or inhibiting effect.

2.1.2 Modelling complex networks: the case of Boolean Networks

Kinetic functions described in 2.1.1 can be integrated into a complex model to simulate more exhaustive networks of interacting metabolites and reactions representing a biological pathway, usually using ODEs. An emerging method that has been successfully applied in biological system modelling is boolean networks (BNs). In BN, the system is represented using a boolean formalism where variables (nodes) have binary values: one (ON) or zero (OFF). Variables are connected by boolean function (BF) (edges) used to update the state of each variable depending on an input. Graphically, BN can be seen as a network graph with nodes connected by edges (figure 2.4). The choice of variables and interactions depends on the system studied. It impacts the topology of the network and is usually built from processes documented in the literature. For example, variables can represent time-series gene expression inferred from data that were binarized to represent the ON/OFF state of a gene in each timestep. Three basics BF are: AND, OR and NOT; yet other BF like the canalysing function (a function that depicts a hierarchical relationship between multiple inputs) or the probability distribution exists as well. Different analyses can be performed to understand the properties of the system. The properties of the modelled system can be understood through attractor and perturbation analysis. An attractor analysis finds stable behaviour, called attractors, corresponding to biological endpoints. Through a perturbation analysis, the topological robustness of a network is examined by changing the state of a variable to assess what are the resulting attractors. It simulates a perturbation like a drug would disrupt the network. The main advantage of boolean networks is that they do not require kinetic parameters which makes them easier to build. Yet their qualitative nature makes them unsuitable for quantitative prediction. Nonetheless, they can be combined with quantitative methods (such as ODEs) to overcome this limit. Boolean networks have been applied to various problems of medicine like phenotype prediction, drug target prediction and cellular crosstalk prediction and cancer research [84].

2.1.3 Modelling tumour growth

A large variety of approaches has been documented in the literature to study biological systems. Population growth has been simulated through ODEs with systems such as the Lokta-Volterra system to study the growth of two populations with a prey-predator relationship (implemented in some models of tumour growth to account for cell competition) or the logistic growth applied to cell cultures. In 2021, Raatz et *al*

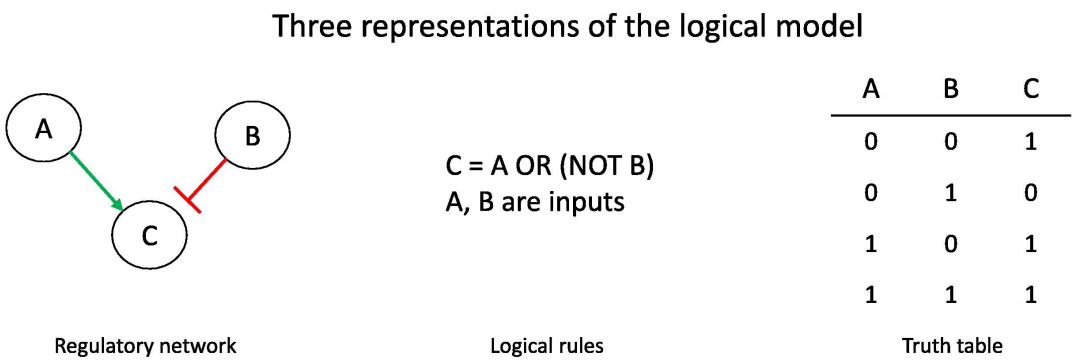

Figure 2.4: Representation of the logical model of boolean network (BN): the regulatory network representing the different variables and interactions in the model, the logical rules defining the interactions and the resulting truth table. Adapted from: Calzone et *al* [83].

developed a mathematical model based on ODEs to study the impact of tumour cell heterogeneity on treatment and relapse dynamic [85]. In their model, each variable is a different cell population with an associated growth-rate and treatment sensitivity. Cells can transit from one population to another trading a faster growth rate for a higher sensitivity to drugs.

Tumour growth can be further characterized in a spatio-temporal manner using partial differential equation (PDE) based systems. Not only these systems can accurately describe nutrient concentration at any point of the domain but they can depict cell density as well. Similarly to a diffusible, cell movement can be modelled using a diffusion coefficient term [86]. PDE models have proposed to examine glioma density and spread with respect to surrounding tissue, chemical signals, vasculature and pH. Heterogeneity was included with the dynamic of different populations of cells (proliferating/necrotic/hypoxic) [49].

However, PDE models still fail to account for a per-cell heterogeneity, a problem that is tackled by discrete individual-based approaches (figure 2.5). Individual-based methods are often combined with continuous models (ODEs and PDE) to represent more complex systems.

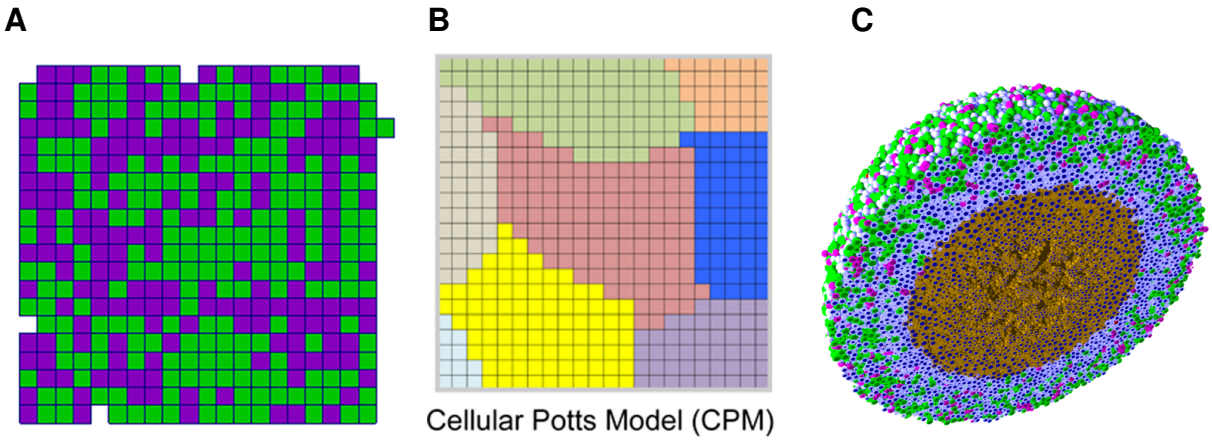

Figure 2.5: Representation of three different individual-based approaches: (A) Cellular automaton, (B) Cellular potts model and (C) Agent-based model.
Adapted from Hirashima et *al* [87], Osborne et *al* [88] and Ghaffarizadeh et *al* [89].

Cellular automata are modelled on a square grid where the cell is the smallest unit. Each cell has a state that can take its value among a finite set of possibilities. The state of each cell is updated every timestep [90]. They permit to describe different cell states an individual can adopt and transition from one state to another controlled by algorithmic rules (such as the transition from proliferative to quiescent, alive to necrotic, etc). Cellular automata have become increasingly important due to their versatility and were applied to tumour growth [91].

Cellular potts models (CPMs) are very popular for morphogenesis studies due to their ability to simulate complex and realistic shapes. They can be traced back to q-Potts models. In CPM, the biological cell is represented as a set of elements on a square lattice. It was developed to test the differential adhesion hypothesis: motile cells organize in a way that minimises energy. Cell behaviour is updated following the forces applied on the system (cell adhesion, interaction with the Extra-Cellular Matrix (ECM)) represented by a Hamiltonian function and solved by a Monte-Carlo method [87, 92, 93].

While cellular automata and CPM are lattice-based methods, agent-based model (ABM) does not bind the agent to a square grid. Traditionally, the agent (here a biological cell) is a point representing its centre with a spherical shape of variable diameter. Because the cell is not bound to a lattice it can move freely in the environment. Interactions between agents are described by adhesion and repulsion forces that can allow the overlap of spheres depending on the parameters applied [41, 94]. The ABM technique allows the description of very realistic tumour morphologies and cellular densities matching quantitative experiments (comparison of the model results with histological slices, etc).

2.2 Choice of mathematical approach for the Thesis

2.2.1 State of HIF mathematical modelling

HIF mathematical modelling is not new, the variety of processes impacted by HIF makes it an interesting target for testing hypotheses through mathematical and computational techniques. Although, HIF role

on cellular functions such as proliferation is usually simplified to represent an effect of hypoxia. The impact of the protein is simplified through a simple rule like “if oxygen is below a threshold, the cell becomes quiescent”. Not only HIF could impact tumour growth and proliferation due to its effect on both the metabolism and cell cycle arrest, but we saw in section 1.2.4 that a few mechanisms like onco-pathway activation can induce the upregulation of HIF in normoxia. Models simulating the growth of a tumour could thus benefit from modelling HIF levels explicitly. Yet its effect is often simplified to the level of O_2 . Even in models of angiogenesis, one of the most known roles of HIF, the protein variations are not explicitly described. A few examples of model studying the impact of heterogeneity, metabolic phenotype, acidity or angiogenesis on tumour growth can be found in the literature (see [91, 95–98]).

On the opposite, models explicitly describing the level of HIF tend to detail a specific aspect. The four main topics that have mathematically characterized HIF were:

1. The nuclear accumulation and transcriptional activity of HIF by oxygen-mediated processes.
2. The sensitivity of HIF to the molecular microenvironment.
3. The Factor-Inhibiting HIF (FIH) driven activity and stability of HIF.
4. The dynamic over time of the HIF response to hypoxia.

In addition, continuous methods relying on ODEs are mostly used to describe biochemical kinetics defining HIF-related networks [99]. ODE systems were also developed to understand the importance of compartmentalization and to unravel the “switch-like” behaviour modulating cell cycle arrest [35, 100]. In line with metabolism, HIF-metabolic interactions has been investigated through a network centered on the HIF-AMPK relationship [42].
Several conclusions can be drawn from the approaches mentioned here:

- Most of the studies modelling mathematically HIF focus on one specific aspect.
- Model of tumour growth usually does not explain HIF variations explicitly.
- Most of the HIF-based systems rely on ODEs which cannot efficiently depict heterogeneity, while most studies on tumour heterogeneity implement a computational model (usually as cellular automaton) that does not account for HIF.

The pattern in these statements clearly indicates what position the work in this Thesis is taking: describing a model of tumour growth to represent heterogeneity with respect to HIF. I hope to answer or find some leads to the following questions: How HIF will impact tumour growth? Can it drive the Warburg Effect? Are there other ways HIF can facilitate tumour proliferation or aggressiveness? This section will mainly focus on the general base of the computational method but not on the mathematical equations describing the cell behaviour. Indeed, the model developed was extended with the information gathered from a statistical analysis of transcriptomic data. Hence, the specific equations of the models present in chapters 3 and 5 will be described in those chapters.

2.2.2 Choice of an agent-based model

It has been decided that the model will reproduce the growth of tumour spheroids grown *in vitro*. These biological models are able to capture more accurately the interactions between cells due to a 3D environment compared to 2D cultures. Furthermore, the cellular density and oxygen profiles are closer to the conditions of a tumour that would grow *in vivo*. Our team uses spheroid culture to unravel the adaptation of cancer cells to acidic conditions and study the kinetic of temozolomide (TMZ) [101].

The tumour dynamic is a complex spatiotemporal phenomenon. As the tumour grows over time, it influences its environment (for example oxygen conditions through angiogenesis) and in turn, the environment influences the behaviour of the cells. We previously have seen that tumours are characterized by a high heterogeneity driven by many factors such as mutations and the microenvironment. The goal of the model is to depict the heterogeneity caused by the different levels of oxygen experienced by the cells inside the tumours. With that in mind, the main criteria are:

1. The ability to simulate a large number of cells to represent heterogeneity.
2. Oxygen diffusion is a major aspect of the phenomenon studied and should depend on the local cellular density.

3. We focus on the impact of HIF on the energetic metabolism, hence angiogenesis will not be implemented despite its effect on oxygenation.

4. Cell division can introduce new mutations further increasing the genetic heterogeneity over time. To keep the model deterministic, children cells will retain the same genetic parameters the parent cell has.

The agent-based model (ABM) approach is the one that fits the best the requirements stated. In ABMs, cells are modelled as different individuals called *agent*. Each of them will progress in an environment that varies from cell to cell in the simulation and each cell can have its own parameters. Yet compared to other techniques like cellular automaton, they are not bound to a square grid and move freely through the simulation. This allows for a more accurate description of cellular density which is crucial for oxygen diffusion as the core of the tumour is usually hypoxic due to a lowered diffusion.

Many software designed to run simulation of cell proliferation are available (Chaste [102], Biocellion [103], CompuCell3D [104], Timothy [105], CellSys [106]). Physicell [107], a C++ ABM framework created by Paul Macklin (figure 2.6), was selected for its ease of use while offering good customisation.

Figure 2.6: Picture of Paul Macklin, the creator of PhysiCell.

It is parallelized with the OpenMP C++ library to ensure good performance on desktop and single-node high performance computing (HPC) servers. PhysiCell provides the user with an XML configuration file to specify most, if not all, the simulation parameters. Thus, the software does not need to be compiled for every run and only the value in the XML can be changed. The software handles out-of-the-box substrate diffusion through PDEs, progression through the cell cycle, cell division and death, cell secretion and uptake, cell-cell adhesion and repulsion, and can run both 2D and 3D simulations. Each agent represents a cell with a phenotype and used-defined associated data. Cells are modelled as spheres that cannot deform, their volume and radius can change over time (for example before the division).

PhysiCell integrates three different scales for biological processes:

- Environmental scale: It describes the concentration of the substrates, the diffusive processes, cell uptake and secretion that may influence the extracellular quantities.

- Cell mechanical scale: Describe the adhesion and repulsion forces exerted on a cell by its neighbourhood and simulate cellular migration.

- Phenotypic changes: Represent the changes in phenotype, whether the cell changes its behaviour, or switch states (proliferative/quiescent/necrotic).

These processes have different speed and it would increase the computational cost with no benefit on accuracy to evaluate all of them at the same time. Therefore they are evaluated at three different time steps (figure 2.7): $dt_{diffusion}$ for diffusive processes (environmental scale), $dt_{mechanic}$ for cell mechanics and migration and $dt_{pehnotype}$ for the changes in cell phenotype.

Cells can move freely in the environment, as they are not bound to the domain grid, with adhesion and repulsion forces described by a potential function. As shown on figure 2.8, environmental substrate concentrations in PhysiCell are discretized on a square grid to solve the PDE, each grid element is called a voxel. The size of the domain and voxels can be defined in the XML configuration file. Voxel substrate quantities can be sampled to influence the behaviour of the system depending on substrate concentration. Consequently, the smaller the size of each voxel will be, the more precise the description of extracellular quantity is. Dirichlet-boundary conditions are applied on the border of the domain for diffusible molecules. When this condition is imposed over the boundary of an ODE or a PDE, the value taken at the boundary of the domain is specified. This means that the value at the borders of the simulation is fixed.

Substrate diffusion is modelled by PDEs built-in PhysiCell, they include parameters accounting for cell impact on extracellular quantities such as cell secretions and uptakes. Diffusion and decay are

Figure 2.7: Graphical description of the different timesteps in PhysiCell.

solved using the finite volume method, each dimension (x-, y- and z-) is split using the locally one-dimensional method and solved by the Thomas algorithm [108]. The PDE solver is parallelized for increased performance and reduced computation time.

PDE **describing extracellular quantities for a substance ρ in PhysiCell** (as described in the software documentation):

$$\frac{\partial \rho}{\partial t} = \vec{D}\nabla^2\rho - \lambda \cdot \rho + \sum_i \delta\left(\vec{x} - \vec{x}_i\right)\left[V_i\vec{S}_i \cdot \left(\rho_i^* - \rho\right) - V_i\vec{U}_i \cdot \rho + \vec{E}_i\right],$$

where

Symbol	Meaning	Dimensions
$\vec{D}$	vector of diffusion coefficients	length^2/time
λ	vector of decay rates	1/time
V_i	volume of cell i	volume
$\vec{x}_i$	cell i's position (center)	length
$\vec{S}_i$	vector of cell i's secretion rates	1/time
ρ_i^*	vector of cell i's secretion saturations	substance/volume
$\vec{U}_i$	vector of cell i's uptake rates	1/time
$\vec{E}_i$	vector of cell i's net export rates	substance/time

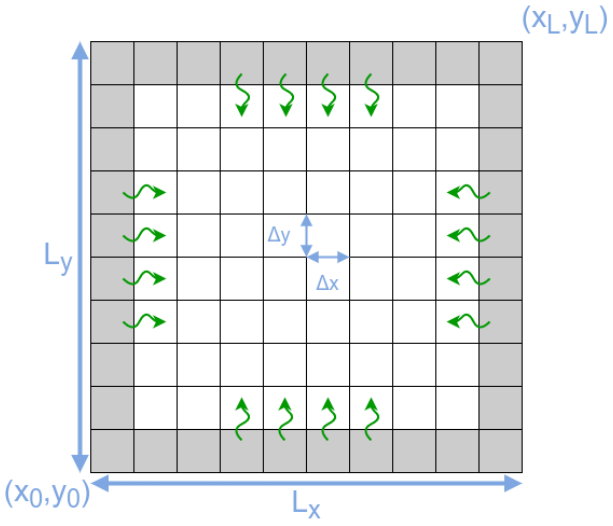

Figure 2.8: Discretized domain of length $L_x \times L_y$ in PhysiCell. Each square is a voxel of width Δx and height Δy. Grey squares represent the Dirichlet Boundary conditions and green arrows shows the substrate diffusion from the border of the domain.

While naked PhysiCell uses an isotropic diffusion, an anisotropic diffusion was implemented by Pierre Jacquet during his Thesis when he worked on a tumour growth model using the same software [41].

Anisotropic diffusion:

$$\frac{\partial X}{\partial t} = \nabla \cdot (D(\rho)\nabla X),$$
$$D(\rho) = D_0(1 - \alpha\rho),$$

where X is a diffusible molecule, ρ the local cellular density, D_0 is the diffusion coefficient in interstitial liquid, α is a term tuning the impact of the cellular density ρ.

An isotropic diffusion would imply that no matter the tissue considered, the diffusion coefficient is the same across the domain (the substrate moves at the same speed at each point of the domain). Anisotropic diffusion would reflect the impact of cellular density on substrate diffusion. With the current focus on modelling heterogeneity caused by different oxygen conditions, an anisotropic diffusion would lead to a better depiction of spatial heterogeneity and impact of HIF. The computational cost is, however, higher and the simulation takes longer. The parameter α was left to 0.25 following the conclusion of Jacquet. The cellular density is updated every time PhysiCell updates the mechanical properties of each cell.

The local cellular density is computed using the neighbourhood of a cell. The neighbourhood of each cell is composed of the cells in the same voxel and cells in the second order Moore-Neighbourhood. Then, a sampling sphere is drawn to collect the neighbour cells present in it (the sphere has a diameter of 45 micrometres which is roughly 5 cells). The total volume (or partial for cells near the edge of the sphere) is summed and then divided by the volume of the sphere to calculate the percentage of density. This gives a value between 0 and 1 (1 being 100 % of coverage). Because cells are modelled as non-deformable spheres, the only way to obtain a density of 100 % is by overlapping cells. By adjusting the mechanical parameters and the initial tissue, it is possible to replicate the cellular density observed in experiments.

PhysiCell models the cell cycle and cellular death as a network of connected phases where the cell transits from one phase to another at a defined rate. Several built-in cell cycle models are already included and the one implemented will be described later. In the death model, cells first start to swell and then shrink before removal from the environment. Removal from the environment was disabled as we are modelling similar conditions to *in vitro cultures*.

3 Modelling the impact of HIF on LDH and PDH genes and its consequences on the Warburg Effect

In this chapter, I will present the first mathematical model developed to study the impact of hypoxia on the cellular metabolism. The main focus is to study the interaction of the Hypoxia Inducible Factor (HIF) protein on the metabolism and the possible consequences for the Warburg Effect. The model describes the impact HIF has on the genetic regulation of selected genes on the metabolism in a 2D multicellular setting. The chapter is divided into three main sections:

1. The biological motivation and the mathematical approach used to describe it.
2. The numerical implementation of the model.
3. The impact of microenvironmental and genetic regulations on the result.

3.1 Biological Assumptions of the model

We already described how the energetic metabolism functions in the cell to generate energy in the form of Adenosine Triphosphate (ATP) in section 1.3.1. But what advantages does the usage of anaerobic pathways confer to the cancer cell? Is there a factor that drives the Warburg Effect? Many studies have brought more clarification or hypotheses but no clear answer has been found yet. We saw that HIF plays an important role in promoting the use of anaerobic pathways (glycolysis followed by the generation of lactate) instead of aerobic pathways (glycolysis followed by the transformation of pyruvate in Acetyl-CoenzymeA). In addition, HIF is found to be often present in higher quantity in cancerous tissues compared to normal tissues as (1) median O_2 in tumours is hypoxic and (2) several mutations including ones commonly found in glioblastomas can promote HIF1-α stabilization (sometimes even in normoxia). One issue that arises is: Does HIF play a role in the Warburg Effect? To recall, the usage of the concept *Warburg Effect* varied over time and sometimes led to contradiction with the original observations made by Otto Warburg [41, 45]. For that reason, it seems a good practice to explicitly write the biology that drives the modelling approach and the question to investigate:

- Lactate is produced in both normoxia and hypoxia.
- Lactate production is increased in hypoxic conditions due to the increased anaerobic pathway (fermentation) to make up for the ATP demand.
- In normoxia, lactate secretion should be relatively low compared to hypoxia.

In the next sections, I will use the terms “Warburg Effect” or “Warburg Phenotype” to refer to *an increased production of lactate in normoxic conditions*. The main goal of this first modelling approach is to investigate if HIF can increase the production of lactate even when oxygen levels are sufficient for the cell to produce ATP through oxidative phosphorylation, in another word if HIF can trigger the appearance of a Warburg Phenotype. Because HIF acts as a transcription factor targeting Hypoxia Response Elements (HRE) genes, the model includes genetic regulations. Metabolism has been studied through many different approaches to describe both intracellular metabolites and extracellular diffusibles. Naturally, metabolism is included in the model using ordinary differential equations (ODEs) to account for intracellular molecules, proteins and ATP. Because metabolism interaction with microenvironmental conditions act as a double feedback, extracellular concentrations of the main diffusible substrates O_2 , glucose and lactate will be determined using partial differential equations (PDEs).

3.1.1 Genetic/Protein regulation

As we have seen before, the fate of pyruvate is driven by the two enzymes lactate dehydrogenase (LDH) and pyruvate dehydrogenase (PDH) which can be considered key players in the balance between aerobic/anaerobic pathways. Since they are both regulated by HIF (directly or through an intermediate

molecule), it is evident that they should be included in the model to reflect the effect of HIF on the aerobic/anaerobic balance. It is important to note that the model aims to describe the **qualitative** variations of the proteins studied and does not intend to quantitatively represent their intracellular levels. I based the mathematical formalism of genetic regulations following the one described by Li et *al* in their study of the metabolic landscape of cancer cells [109]. Figure 3.1 represents a diagram of the network they designed. It combines gene-metabolic interactions to identify the underlying metabolic steady-states and investigate how the metabolic landscape is shaped by the parameters of their model. The network is

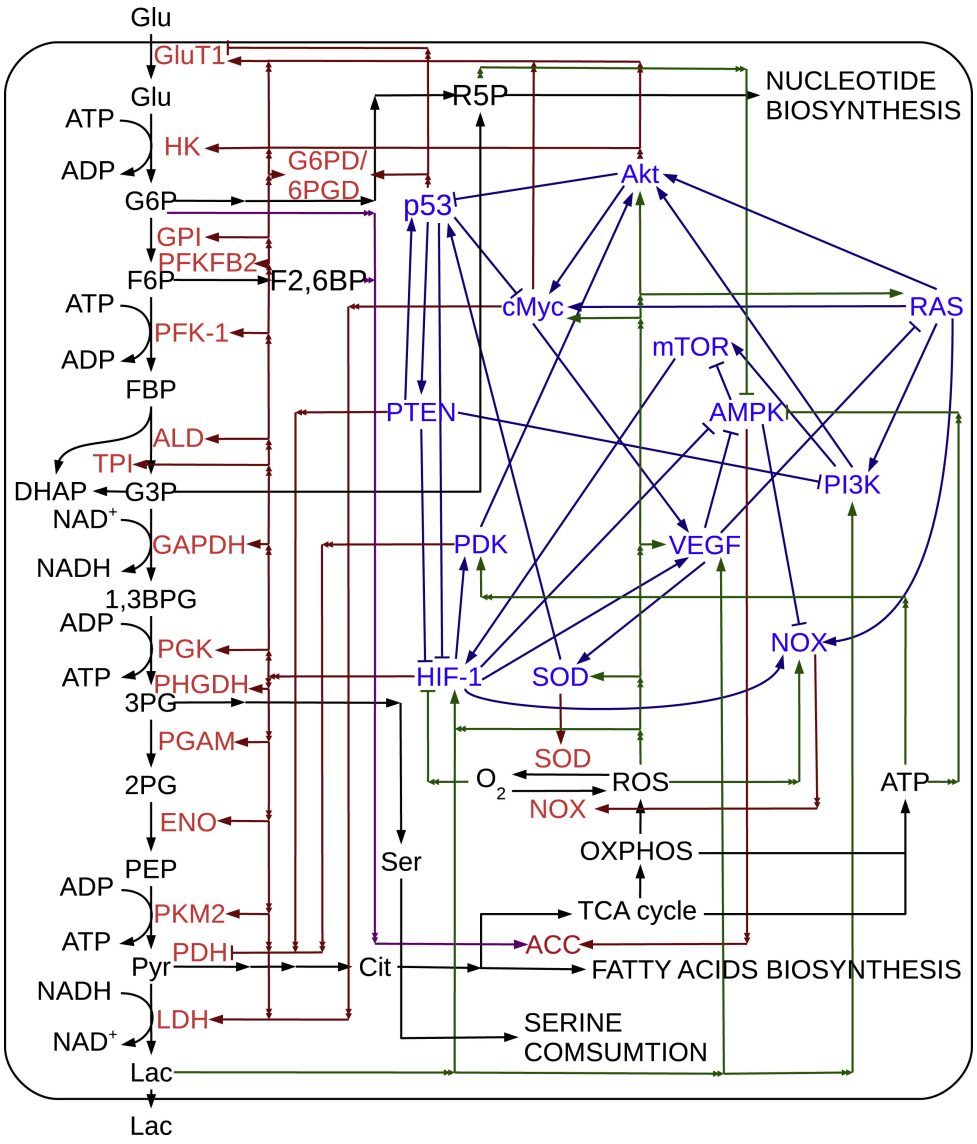

Figure 3.1: The model used to design the equations for genetic/protein regulations. The model represents a gene-metabolism regulatory network to identify the underlying landscape and probability flux. It includes a total of 53 substances: 30 proteins and 23 metabolites. The authors found four different metabolic steady-states: a normal state, an oxidative state, a glycolytic state and an intermediate state.
Source: Li et *al* [109]

quite complex as it describes 53 nodes composed of 30 proteins and 23 metabolites of the metabolism. The authors found four different attractors that correspond to four different metabolic states (see figure 3.2), namely: a normal state, an oxidative state, a glycolytic state and an intermediate state. Gene-gene network interactions and gene-metabolites interactions were determined using ODEs. Despite oxygen, glucose and lactate being included, there is no PDEs to describe the extracellular quantities of these metabolites. Genes and proteins are modelled as a single variable to avoid redundancy. While the authors describe their network as a *gene-gene regulatory network*, it seems more appropriate to use the term *protein-protein regulatory network*. Especially because there is no such thing as a “gene level” but rather *protein level*.

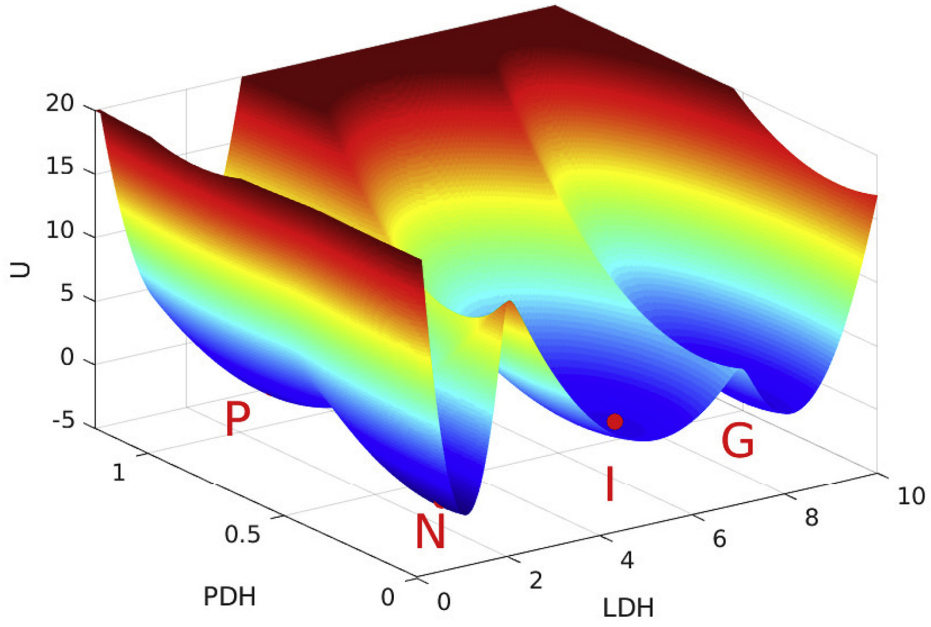

Figure 3.2: Metabolic landscape of cancer cells metabolism. Four basins can be seen on the figure representing steady-states or attractors: a normal state (N), a glycolytic state (G), an oxidative state (P) and an intermediate state (I). Source: Li et *al* [109]

Mathematical formulation of the gene-gene network and gene-metabolite network by Li et *al* [109]:

$$\frac{dX_i}{dt} = A_i \prod_{j=1}^{N_i} H_{ij} - D_i X_i, \tag{3.1}$$

$$H_{ij} = \frac{S_{ji}^n}{S_{ji}^n + X_{ji}^n} + \gamma_{ij} \frac{X_{ji}^n}{S_{ji}^n + X_{ji}^n}, \tag{3.2}$$

$$\frac{dY_i}{dt} = \sum_{j=1}^{N_i} X_j r_j. \tag{3.3}$$

Here, $\frac{dX_i}{dt}$ represent the level of the X_i protein; A_i and D_i are the production rate and the degradation rate of the X_i protein. H_ij is a non-linear term, a Shifted-Hill function, to reflect the regulation of a protein i by a protein j; S_{ji}^n is the quantity of protein at which the production of the regulated protein is half-maximal; n is the hill-coefficient; X_{ji}^n is the regulator and γ_{ij} is a positive number representing the strength of regulation. A value of $\gamma_{ij} > 1$ indicates an upregulation while a value of $\gamma_{ij} < 1$ indicates an inhibition. $\frac{dY_i}{dt}$ represent the concentration of metabolite which is driven by the summation of the enzyme kinetic speed r_j multiplied by the related level of the enzyme X_j.

Li et *al* mathematically described protein-protein interactions (upregulation/downregulation) using a non-linear function called the Shifted-Hill function (equation 3.4).

Shifted-Hill function:

$$S(Y, s, \gamma) = \frac{s^n}{s^n + Y^n} + \gamma \frac{Y^n}{s^n + Y^n}, \tag{3.4}$$

where n is the Hill coefficient, s is the level of the regulating protein Y at which the regulation if half-maximal, γ is the regulating strength of the gene Y on X. γ is a positive parameter that represents both an upregulating effect, when $\gamma > 1$, and a downregulating effect, when $\gamma < 1$. The Shifted-Hill function consists mainly of the summation of two Hill terms.

The Shifted-Hill function was first used by Lu et *al* [110] to depict the implication of transcription factors and microRNA in a network associated with metastasis. Figure 3.3 shows how the different parameters shape the function.

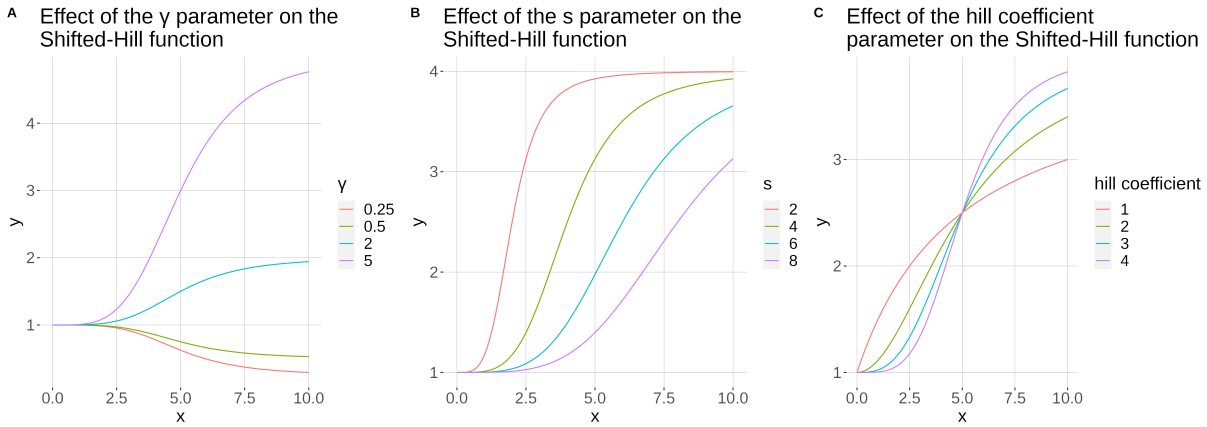

Figure 3.3: Effect of the different parameters of the Shifted Hill on its shape. (A) shows the impact of the γ parameter, $s = 5$ and $n = 4$. (B) shows the impact of the s parameter (midpoint), $\gamma = 4$ and $n = 4$. (C) shows the impact of the hill coefficient, $s = 5$ and $\gamma = 4$.

The advantages of the Shifted-Hill function are (1) up- and downregulation are simulated using a single term; (2) the midpoint can introduce a delay when tuning this parameter to understand how the model behaves. As said, the original model depicts a complete network of interacting proteins and metabolites combining both protein regulation and kinetic reactions. However, only a subset of genes is relevant to the current issue being studied and keeping the model as is would increase the complexity without bringing actual benefit, or even complicate the interpretation of the result as numerous players would be involved. Related to the issues being studied, only the proteins: HIF, LDH, PDK and PDH are included. I selected LDH and PDH due to their role in the fate of pyruvate. In fact, Li et *al* studied the balance between aerobic/anaerobic pathways using these two genes as markers. pyruvate dehydrogenase kinase (PDK) is included as it is an inhibitor by which HIF can downregulate PDH. The final protein-protein network is shown in figure 3.4.

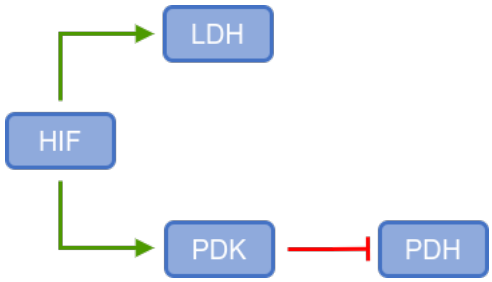

Figure 3.4: Diagram of the gene network modelled in the metabolic model.

Equations 3.5 to 3.8 describe the different interactions in this network.

Parameter	Value	Dimension	Parameter	Value	Dimension
α_h	0.005	1/min	d_h	0.005	1/min
α_l	0.005	1/min	d_l	0.005	1/min
α_k	0.005	1/min	d_k	0.005	1/min
α_q	0.005	1/min	d_q	0.005	1/min
$s_{O\to h}$	0.02085	mmol/L	$s_{h\to l}$	4.48	-
$s_{h\to k}$	5.0	-	$s_{k\to q}$	2.2	-
$\gamma_{O\to h}$	10.0	-	$\gamma_{h\to l}$	3.61	-
$\gamma_{h\to k}$	6.97	-	$\gamma_{k\to q}$	0.14	-
β_h	10	-			

Table 3.1: Parameters for genetic regulations. The symbol "-" stands for dimensionless.

$$\frac{dh}{dt} = \alpha_h\ \beta_h - d_h\ S(O, s_{O\to h}, \gamma_{O\to h})\ h, \tag{3.5}$$

$$\frac{dl}{dt} = \alpha_l\ S(h, s_{h\to l}, \gamma_{h\to l}) - d_l\ l, \tag{3.6}$$

$$\frac{dk}{dt} = \alpha_k\ S(h, s_{h\to k}, \gamma_{h\to k}) - d_k\ k, \tag{3.7}$$

$$\frac{dq}{dt} = \alpha_q\ S(k, s_{k\to q}, \gamma_{k\to q}) - d_q\ q. \tag{3.8}$$

h, l, k and q describe, respectively, the level of HIF, LDH, PDK and PDH. α and d are the production and degradation rates of each protein in the model. $S(Y, s, \gamma)$ account for the interaction between proteins. $s_{x\to y}$ represent the mid-point of the Shifted-Hill function while $\gamma_{x\to y}$ is the strength and type of regulation (see function 3.4). Since the interactions between proteins in the network mainly increase or decrease their production, $S(Y, s, \gamma)$ only modulate α. One exception is HIF itself as we saw earlier that HIF is mainly regulated at the protein level through oxygen-dependant mechanisms. Still, it can be upregulated even in normoxia due to different factors such as mutations on HIF-inhibitors genes. Thus β_h is a parameter used to tune the production of HIF proteins so that: (1) when oxygen is not present in the environment, HIF production is high; and (2) mutations stabilizing increase the production of HIF independently on other conditions.

Parameters used for genetic regulations are shown in table 3.1.

3.1.2 Energetic metabolism and genetic regulation

It is evident that metabolism must be implemented in the model. But compared to the implementation of metabolism by Li et *al* , here it should be as simple as possible. With that in mind, it was decided to only model three diffusible molecules: oxygen (O), glucose (G) and pH with the production of protons (H). It is not the aim of the model to compute the intracellular quantities of these nutrients, hence only their extracellular quantities will be described using PDEs. Robertson-Tessi et *al* [95] described a hybrid cellular automaton model to study the impact of metabolic heterogeneity on tumour growth and treatment. Their model includes functions to describe the consumptions and secretions of the cell. Interestingly, they included a parameter to account for the increased reliance of cancer cells on the glycolytic pathway i. e.the Warburg Effect. That parameter, along with a parameter for resistance to low pH, is allowed to vary when the cell divides to reproduce phenotypic variations during tumour growth. They showed that during tumour growth, cancer cells tend to acquire acid-resistant and a glycolytic phenotype as a result of the development of hypoxic regions. Figure 3.5 shows one of their simulations.

Original Model from Robertson-Tessi et *al* [95]:
The concentration of a diffusible molecule C:

$$\frac{\partial C}{\partial t} = D\nabla^2 C + f(C,p). \tag{3.9}$$

Where D is a diffusion constant and $f(C,p)$ is a function to describe the production and consumption of the molecule. The diffusible molecules included in the model are oxygen (O), glucose (G) and protons (H). Oxygen consumption is determined by a Michaelis-Menten dynamic:

$$f_O = -V_O \frac{O}{O + k_O}. \tag{3.10}$$

V_O is the maximal O_2 consumption rate and k_O is the oxygen level at which the consumption is half-maximal aka the Michaelis-Menten constant of oxygen. Similarly, glucose consumption follows a Michaelis-Menten dynamic with a modified speed term:

$$f_G = -\left(\frac{p_G + A_0}{2} + \frac{27 f_O}{10}\right) \frac{G}{O + k_G}. \tag{3.11}$$

p_G is a parameter representing the altered glucose metabolism of cancer cells (Warburg Effect), A_O is the target ATP level the cell needs to produce. Healthy cells have values of $p_G = 1$ while cancer cells have values of $p_G > 1$ to depict the increased glucose consumption of cancer cells. Consumption of glucose is driven by the need of the cell to meet ATP demand: when aerobic pathways do not produce sufficient ATP the cell consumes more glucose to make up for the ATP need. ATP production is simply calculated from the sum of nutrients consumed multiplied by their stoichiometric coefficient:

$$f_A = -\left(2 f_G + \frac{27 f_O}{5}\right). \tag{3.12}$$

Proton production is linked to the amount of glucose consumed that does end up in aerobic pathways.

$$f_H = k_H \left(\frac{29\,(p_G V_O + f_O)}{5}\right). \tag{3.13}$$

k_H is a proton buffering term. Functions f_O, f_G and f_H are used in 3.9 while ATP production (f_A) is used to update the cell behaviour.

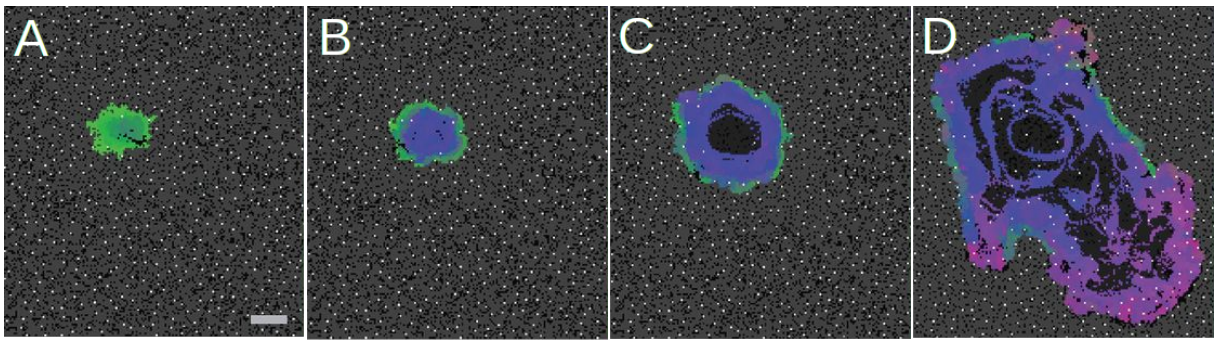

Figure 3.5: Result of tumour growth simulation at (A) 1,270 days, (B) 1,392 days, (C) 1,610 days and (D) 1,912 days. White squares correspond to vasculature, black to empty spaces, dark grey to necrosis, and medium grey to normal tissue. Other colours represent cancer cells depending on their phenotype as shown in the phenotype flow diagram in the original paper (not shown here).
Adapted from Robertson-Tessi et *al* [95].

I selected this model for the following reasons:

- The functions for metabolism are simple and easy to implement in another framework.
- The parameter for the increased glycolysis of cancer cells can be easily redirected to account for the impact of the LDH gene.
- It includes all the variables relevant to the current work without adding unnecessary complexity.

Nutrient consumption is described using Michaelis-Menten dynamics. Oxygen consumption (equation 3.14) is represented by:

$$f_O = \Psi_O V_O \frac{O}{K_O + O}, \tag{3.14}$$

$$\Psi_O = \frac{\Phi_O - \phi_O}{1 + \exp\big(-\lambda_q (q - q_0)\big)} + \phi_O. \tag{3.15}$$

V_O is the maximal consumption rate of oxygen and K_O is the concentration of oxygen at which the consumption is half-maximal (Michaelis-Menten constant of oxygen). Ψ_O is a logistic function to tune the consumption of oxygen depending on the level of the gene PDH. Φ_O and ϕ_O represent the maximum and minimum values, λ_q is the steepness of the curve, q is the level of PDH and q_0 is the level at which the effect of PDH is half-maximal (midpoint of the logistic curve, similar to the Michaelis-Menten constant). Like oxygen, glucose consumption (equation 3.16) is described by a Michalis-Menten dynamic with a modified term for the maximal speed:

$$f_G = \left(\frac{\Psi_G A_0}{2} - \frac{29 f_O}{10}\right) \frac{G}{K_G + G}, \tag{3.16}$$

$$\Psi_G = \frac{\Phi_G - \phi_G}{1 + \exp\big(-\lambda_l (l - l_0)\big)} + \phi_G. \tag{3.17}$$

A_0 is the target ATP demand the cell needs to meet, K_G is the concentration of glucose at which the consumption is half-maximal. Similarly to oxygen, glucose consumption is influenced through the term Ψ_G which represents the effect of the LDH gene on glucose consumption. Φ_G and ϕ_G represent the maximum and minimum values, λ_l is the steepness of the curve, l is the level of LDH and l_0 is the level at which the effect of LDH is half-maximal. It is possible to see that glucose consumption is driven by two factors: the need to meet ATP requirement and the influence of LDH. If oxygen is lacking in the environment, the cell will naturally increase its consumption of glucose to produce the energy required. But, the most noticeable impact will be due to the LDH genes which drive the consumption of glucose "independently" from the oxygen conditions and the ATP produced through aerobic conditions. ATP is computed by adding the consumption of oxygen and glucose multiplied by the stoichiometric coefficient of the reactions giving ATP:

$$f_A = 2 f_G + \frac{29}{5} f_O. \tag{3.18}$$

In the original model H^+ was calculated by the amount of glucose consumed that was not used in aerobic pathways. I preferred to simplify the model by multiplying the glucose consumed by the amount of lactate produced:

$$f_H = 2 f_G \, k_H. \tag{3.19}$$

k_H is a proton-buffering term.
The PDEs 3.20 to 3.22 describe the extracellular quantity of the diffusible implemented in the model:

$$\frac{\partial O}{\partial t} = D_O \nabla^2 O - \sum_{N}^{i=1} f_O^i, \tag{3.20}$$

$$\frac{\partial G}{\partial t} = D_G \nabla^2 G - \sum_{N}^{i=1} f_G^i, \tag{3.21}$$

$$\frac{\partial H}{\partial t} = D_H \nabla^2 H + \sum_{N}^{i=1} f_H^i. \tag{3.22}$$

With D_O, D_G and D_H the diffusion coefficient of the diffrent diffusibles. f_O and f_G are functions describing the consumption of oxygen and glucose, respectively, while f_H determine the secretion of protons. Note that all f functions are summed in $\sum_N^{i=1} f^i$. Since the simulating domain is a square domain composed of voxels (more details in section 3.2), this summation represents the consumption and production of all the cells present in the voxel. In other word, $\sum_N^{i=1} f^i$ is the local effect of the cells. It is assumed that cells only consume glucose and oxygen. The same goes for protons although it is known that monocarboxylate transporter (MCT) are bi-reversible and can uptake lactate leading to decreased acidity. Here, we assume that cells only secrete lactate (produce protons) and have no intake. Like Robertson-Tessi et *al* , the ATP produced dictates the behaviour of the cell. It is not an intracellular quantity but rather the ATP produced at each timestep in the simulation. It is computed by the following equation:

$$\frac{dA}{dt} = f_A. \tag{3.23}$$

f_A is the function accounting for the production of ATP based on nutrient consumption.
Parameters used in the modelled are summarised in Table 3.2.

3.2 Implementation of the model

PhysiCell PDEs were configured to correspond to the equations presented in section 3.1.2 (equation 3.20 - 3.22).
The following cell cycle model was implemented (figure 3.6A): 5 hours in G1, 8 hours in S, 4 hours in G2 and 1 hours in M [111] and a supplementary node for G0 is included. The G0 (quiescence) and G1 phases interact in a bi-directional way, the cell can either enter and leave quiescence depending on the ATP generated. The rate of transition between G0/G1 or G1/G0 is changed to reflect deterministic quiescence/proliferation. The other phases are linked to the next one in the cell cycle and the transition rate does not change. Cancer cells are known to be resistant to apoptosis thus it was disabled.

There are no conditions on the neighbourhood of a cell, it can divide even if it is surrounded by others as long as nutrients are available. Consequently, certain regions of the tumour will exhibit a higher cell density reducing the diffusion of nutrients. The cell state is updated at each $dt_{phentoype}$ according to the ATP generated and the extracellular pH following the scheme (figure 3.6B):

- If the quantity of ATP generated is less than a threshold A_q, the cell enters quiescence and is then prevented from completing the G1 phase.
- If the quantity of ATP generated is less than a threshold A_d or if the pH is less than a threshold pH_d, the cell dies.

Therefore, they can proliferate and divide only if they are able to generate enough ATP and if extracellular pH is higher than the acid resistance of the cell. Intracellular ODE and signalling was not implemented in the PhysiCell version used. Thus they were implemented manually and simulated at each timestep $dt_{genetic}$. Diffusion is evaluated every 0.01 minutes, cell mechanic and cell genetics are evaluated every 0.1 minutes and phenotype update is evaluated every 6 minutes. This allows to avoid increasing the computational cost without sacrificing the accuracy. Simulations were run in 2D to simulate 14 days in a $1,000 \times 1,000 \mu m^2$ domain split into $20 \times 20 \mu m^2$ voxels.

Parameter	Value	Unit
Oxygen (3.14,3.15)		
V_O	0.01875	$mmol/L/min$
K_O	0.0075	$mmol/L$
Φ_O	1	-
ϕ_O	0	-
λ_q	15	-
q_0	0.575	-
Glucose (3.16,3.17)		
K_G	0.04	$mmol/L$
A_0	0.10875	$mmol/L/min$
Φ_G	50	-
ϕ_G	1	-
λ_l	4	-
l_0	2.35	-
H^+ (3.19)		
K_H	$2.5 \cdot 10^{-4}$	-
Diffusion (3.20 - 3.22)		
D_O	$109,200$	$\mu m^2/min$
D_G	$30,000$	$\mu m^2/min$
D_{H+}	$27,0000$	$\mu m^2/min$

Table 3.2: Parameters for metabolism. The symbol "-" stands for dimensionless.

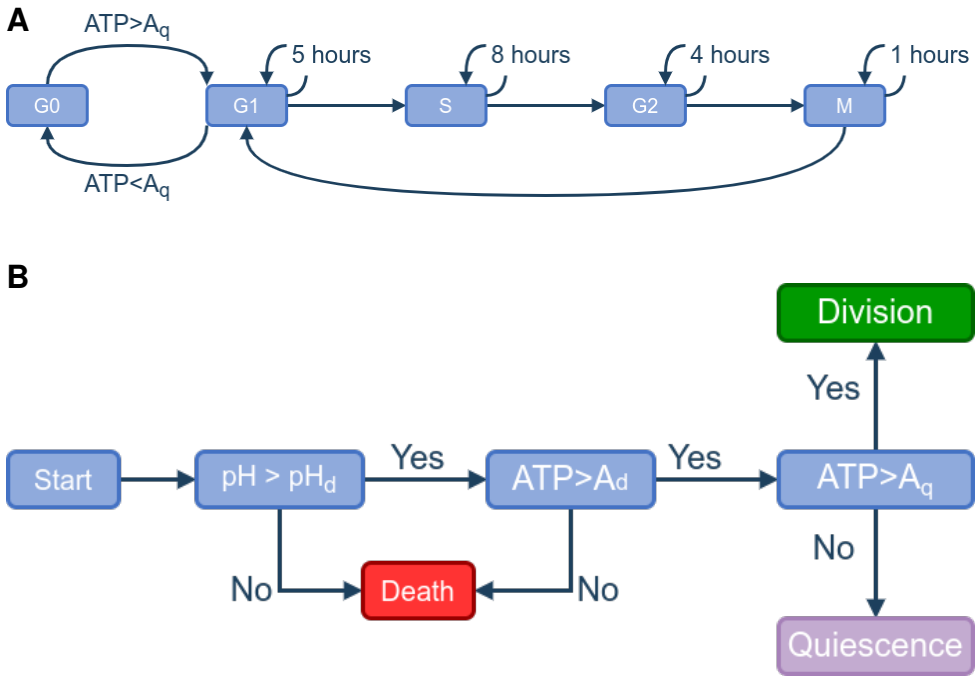

Figure 3.6: (A) Cell cycle graph model implemented in PhysiCell. (B) Decision process to update the state of the cell.

We used the following initial values:

- Oxygen (equation 3.20): $O(x, y, t_0) = 0.056$ mmol/L in normoxia and 0.012 mmol/L in hypoxia [16].
- Glucose (equation 3.21): $G(x, y, t_0) = 5.0$ mmol/L [91].
- H^+ (equation 3.22): $H^+(x, y, t_0) = 3.98 \times 10^{-5}$ mmol/L (pH 7.4).
- HIF, LDH, PDK and PDH: 1 (default level).

Assuming x_0 and y_0 the lower boundary of the domain in x and y, x_L and y_L the upper boundary in x and y. We set up the following Dirichlet-Boundary conditions:

- Oxygen (equation 3.20): $O(x_0, y, t) = O(x_L, y, t) = O(x, y_0, t) = O(x, y_L, t) =$ 0.056 mmol/L in normoxia and 0.012 mmol/L in hypoxia.
- Glucose (equation 3.21): $G(x_0, y, t) = G(x_L, y, t) = G(x, y_0, t) = G(x, y_L, t) = 5.0$ mmol/L.
- H^+ (equation 3.22): $H^+(x_0, y, t) = H^+(x_L, y, t) = H^+(x, y_0, t) = H^+(x, y_L, t) = 3.98 \times 10^{-5}$ mmol/L (pH 7.4).

3.3 Exploration of the impact of the microenvironment and the cell genetic properties on the Warburg Phenotype

I first attempted to assess the impact that genetic regulations can have on the Warburg Effect without considering the temporal changes of the environment nor the effect the growing tumour has on it. ODEs for genetic regulations and metabolism were implemented in a python script that ran the model at different oxygen and glucose concentrations to calculate the production of H^+ to evaluate whether we can already see higher levels of H^+ secretion in normoxia. To reflect the impact of genetic regulation, the γ parameter was modified to represent the up- or down-regulation of the different genes in the model. To recall, γ controls the strength and type of regulation. Thus a map of the rate of production of protons could be created for a few different genetic regulations (3.7). In normal conditions, the proton production ranges

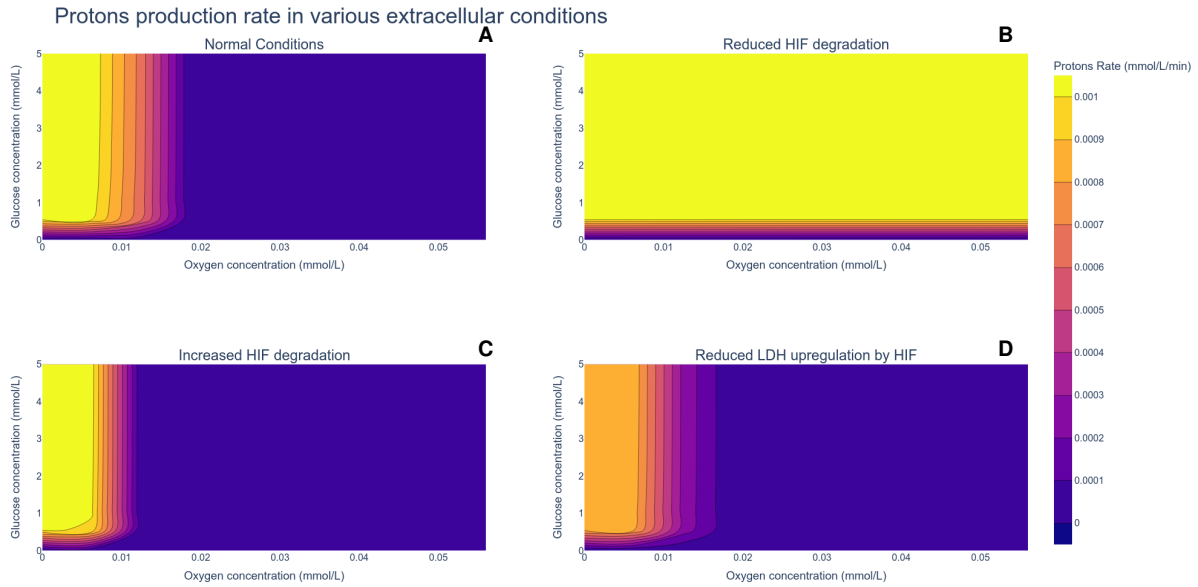

Figure 3.7: Influence of genetic upregulation or inhibition on the production rate of protons at different glucose and oxygen concentrations. (A) Result with no genetic deregulation. ($\gamma_{O\to h} = 10.0$, $\gamma_{h\to l} = 3.61$, $\gamma_{h\to w} = 6.97$, $\gamma_{k\to q} = 0.14$) (B) Result with inhibition of the oxygen-dependant degradation of HIF. ($\gamma_{O\to h} = 1.0$, $\gamma_{h\to l} = 3.61$, $\gamma_{h\to k} = 6.97$, $\gamma_{k\to q} = 0.14$) (C) Result with over-degradation of HIF by oxygen. ($\gamma_{O\to h} = 40.0$, $\gamma_{h\to l} = 3.61$, $\gamma_{h\to k} = 6.97$, $\gamma_{k\to q} = 0.14$) (D) Result with loss of upregulation of LDH by HIF. ($\gamma_{O\to h} = 10.0$, $\gamma_{h\to l} = 3.0$, $\gamma_{h\to k} = 6.97$, $\gamma_{k\to q} = 0.14$)

from 0.0001 mmol/L/min to 0.001 mmol/L/min with the default γ values. H^+ secretion increase from 0.0001 to 0.001 mmol/L/min at around 0.01 mmol/L of O_2 (1%). When HIF is not degraded by oxygen

anymore, H^+ secretion is only dependent on the glucose concentration. Lactic acid secretion of 0.001 mmol/L/min is reached even in normal oxygen pressure, corresponding to the Warburg effect. When oxygen-dependent degradation of HIF is increased, the oxygen threshold at which the cell's H^+ secretion rises is reduced. In that case, stronger hypoxic conditions are needed to achieve increased acidification of the environment. When degradation of HIF by oxygen-dependent mechanisms behave normally H^+ production starts to increase at 0.08 mmol/L and reaches a maximum at 0.019 mmol/L. This span is reduced when degradation of HIF is reduced. Consequently, oxygen-dependent degradation of HIF controls the range at which H^+ secretion increases. It is similar to the finding of a modelling paper which found that a lower degradation rate of HIF increases the probability the cell use glycolysis [42]. Lastly, inhibiting the sensitivity of LDH to HIF upregulation causes the maximum H^+ production rate to fall to 0.0008 mol/L/min compared to the 0.001 mmol/L/min in normal conditions. Increasing the sensitivity of LDH to HIF does not have the opposite effect. Interfering with the regulation of PDK by HIF or the regulation of PDH by PDK does not affect the production of H^+ in the model. From those results, we can hypothesize that HIF degradation will impact the appearance of a Warburg phenotype in the simulations.

I then wanted to assess the impact of environmental oxygen on the tumour growth, the production of H^+ and the levels of the LDH/PDH genes in the cell population. Three simulations were run to that end:

- A simulation in normoxia (5 % O_2 or 0.056 mmol/L) as the reference: Tumour growth was initiated in normoxia and Dirichlet boundary conditions were set up to normoxia as well.
- A simulation in pathological hypoxia (1 % O_2 or 0.01112 mmol/L): Tumour growth was initiated in hypoxia with hypoxic Dirichlet boundary conditions.
- A simulation with oscillating oxygen conditions: Tumour growth is initiated in normoxia with normoxic Dirichlet boundary conditions. The oxygen diffusing at the boundary condition is repeatedly decreased from normoxia to pathological hypoxia over a period of 6h and is then increased back to normoxic levels at the same rate until the end of the simulation. 6h were chosen as kinetics measurements show that HIF1-α peaks after 6 hours but reaches an equilibrium after 24-48h [35]. This replicates stressful conditions for cells as we do not allow HIF1-α values to reach an equilibrium but simulate a high hypoxic reaction.

The effect these different conditions have on tumour growth is shown in figure 3.8. In all three different simulations, the core of the tumour is necrotic which is a famous characteristic of tumours. As expected, tumour growth is higher in normoxia than in hypoxia. Oscillating conditions do not seem to affect much the proliferation of the tumour itself although compared to normoxia a thicker ring of *necrotic swelling* cells is visible. It seems like the death of cells is more synchronized in oscillating conditions than in normoxia. In normoxic and oscillating conditions, but not in hypoxia, some cells survived in the necrotic core and continued to divide. These cells probably entered quiescence when nutrients started lacking but as neighbour cells were dying more nutrients became available to quiescent cells since the dead ones do not consume nutrients and their shrinkage created more space for the substrate to diffuse. It might be a mechanism by which quiescent cells can regrow a tumour. Previous modelling work also found that quiescence may constitute a mechanism to escape drug effect [36].

But to assess in which conditions cells adopt the Warburg phenotype, we need to investigate the secretion of H^+ depending on the extracellular oxygen. Figure 3.9 represents the secretion of H^+ in the different oxygen conditions stated above. While cells can reach pathological hypoxia (1 % O_2), the threshold of hypoxia chosen is 2 % O_2 (0.025085 mmol/L) which is the level at which HIF response is half-maximal [16]. The area which corresponds to the Warburg phenotype is added to the figure for better clarity. Boundaries for the Warburg phenotype were chosen to show cells above the hypoxic threshold and with a H^+ secretion rate higher than cells in the normoxic simulation. The figure shows that environmental oxygen greatly influences the production of H^+ . In the normoxic simulation, cells start with a moderately low H^+ production rate as oxygen is available. Some areas of the tumour reach hypoxia after 2 days of simulated time. It is possible to see that acidification of the environment starts to increase a little before cells reach hypoxic threshold. At the end of the simulation, all cells are in hypoxic conditions and the production of lactate is only dependent on the extracellular glucose. Compared to normoxia, hypoxia directly selects cells with increased glycolytic activity in the model. In that case, they rely mainly on glucose and do not consume oxygen. Lastly, cells in oscillating conditions generally have higher secretion of H^+ despite having access to more oxygen than the ones in normoxia. It suggests that environmental stress is enough to trigger the appearance of the Warburg Effect in the simulations.

But why do cells growing in normoxic conditions generally have access to lower oxygen and reach more severe hypoxia than cells with their growth started in hypoxic conditions? To answer that question,

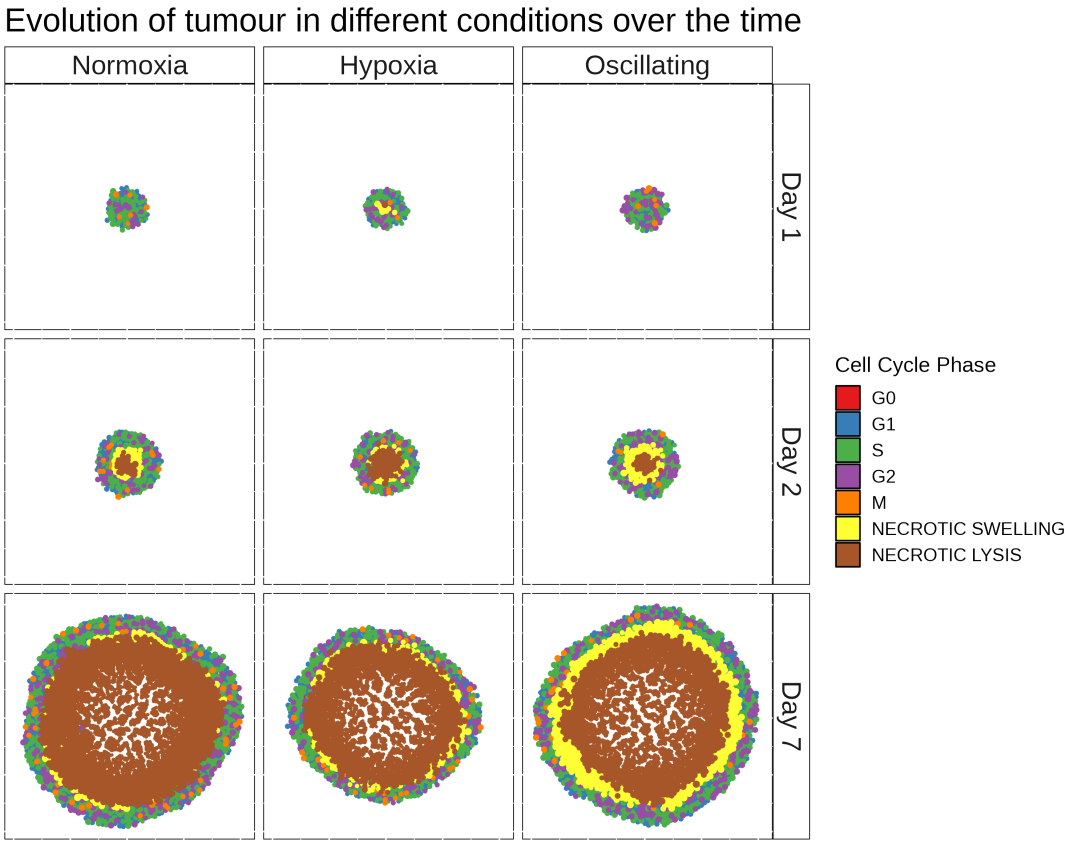

Figure 3.8: Evolution of tumour growth at different times in different conditions: normoxia (5 % O_2 or 0.056 mmol/L), pathological hypoxia (1 % O_2 or 0.01112 mmol/L) and oscillating. In oscillating conditions, the oxygen concentration is slowly decreased from normoxia to pathological hypoxia during 6 hours, then cells are slowly put back in normoxia at the same rate. This process is repeated until the end of the simulation. Some cells survived in the necrotic core of the tumour in a quiescent state and may be able to reform a tumour if conditions are good enough.

we need to investigate the map of LDH/PDH gene levels in figure 3.10. On that figure, cells tend to travel between two genetic states that look similar to the result of Li et *al* [109]: an OXPHOS state with a high level of PDH and a low level of LDH, and a glycolytic state with a low level of PDH and a high level of LDH. But compared to the original genetic network it does not seem that an intermediate stable state exists. Instead, it looks like levels of the two genes slowly change from one state to another following two routes whether the cell transit from the OXPHOS to the glycolytic state or from the glycolytic state to OXPHOS one. The appearance of cellular states is more visible in hypoxia where the conditions force the cell to only adopt the ratio high PDH/low LDH. Consequently, the period when the cell is adapting is similar to the intermediate state found by Li et *al* as, in that case, cells rely on both oxygen and glucose. Recalling that oxygen consumption is driven by the level of PDH and that these levels are higher in normoxic conditions than the others, we may find an explanation for the observation stated earlier. Cells may rely more on oxygen in normoxic conditions than in hypoxic conditions and oscillating conditions due to higher levels of PDH. In addition, since PDH is not directly regulated by HIF1-α but through PDK, its level may adapt more slowly than other proteins.

Since we have seen the impact of environmental oxygen on tumour growth, we now want to understand how genetic deregulations impact the results in the model. In that regard, tumour growth was initiated in normoxic conditions with constant normoxic Dirichlet boundary conditions during the simulation. Consequently, differences in tumour growth are only caused by the effect of genes. Three different genetic perturbations are investigated:

- Increased levels of HIF caused by a decreased activity of its two main regulators Prolyl Hydroxylases (PHD) and Factor-Inhibiting HIF (FIH) ($\gamma_{O\to h} = 8.0$) to depict the effect of oncogenetic deregulation leading to higher HIF levels.

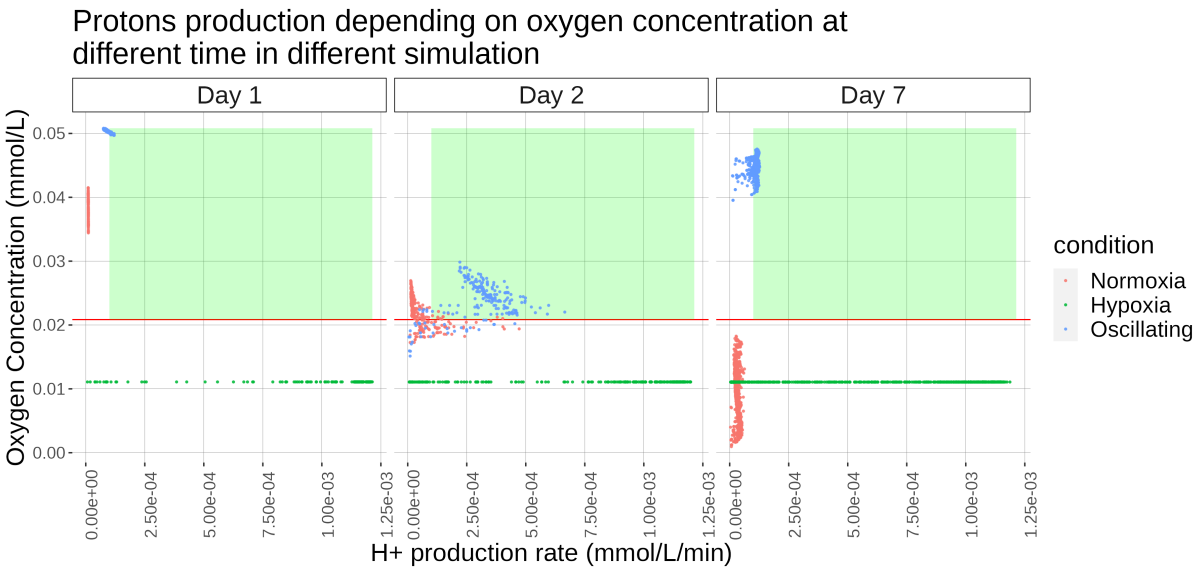

Figure 3.9: Acid production rate following oxygen extracellular concentrations at different times in different conditions. The red line indicates the hypoxia threshold (0.02085 mmol/L or 2% O_2). In oscillating conditions, oxygen concentration is slowly decreased from normoxia to hypoxia during 6h, then oxygen is increased to normoxia at the same rate. This process is repeated until the end of the simulation. Only living cells are represented on the graph. The green rectangle represents the region corresponding to a Warburg effect.

- A decreased sensitivity of LDH to HIF upregulation ($\gamma_{h\to l} = 2.0$) to represent the case where HIF mediated promotion of glycolytic pathways is not as efficient.
- A decreased sensitivity of PDH to PDK inhibition ($\gamma_{k\to q} = 0.7$) to investigate the case where HIF mediated inhibition of OXPHOS is not as strong.

Figure 3.11 shows that the diameter of the tumour is increased when reducing either LDH upregulation by HIF1-α or HIF degradation. The shape of the tumour after 7 simulated days changes very slightly. Compared to the normoxic simulation: reducing LDH upregulation maximises both the size of the proliferating part and the necrotic core, while reducing PDH inhibition by PDK decreases them. The smallest tumour size was found in hypoxic conditions with no deregulation. These results tend to show that glucose is a more critical nutrient for proliferation than oxygen. It seems that decreasing the reliance of cells on oxygen favours cell survival as cells with an increased level of PDH proliferate less than cells in the normoxic simulation. The gap widens when compared to cells with increased HIF levels. However, at the same time cells that can consume less glucose due to lower LDH levels proliferate more. It is probably due to a better availability of glucose as its depletion is slower.

Yet, how do these perturbations translate to the Warburg Effect and the H^+ secretion? As done in the previous section, we will look at the H^+ secretion of cells in the different simulation 3.12. In that figure, we see that the span of extracellular oxygen available for the different cells is wider than when only environmental oxygen is modified. Furthermore, after 2 days of simulated time cells in all simulations are under the hypoxic threshold. Still, after 7 days oxygen levels increase back to normoxia when HIF degradation or LDH upregulation are reduced. The most severe hypoxia is reached when PDH inhibition is reduced, probably because cells rely more on oxygen as a primary source of ATP and Consequently uptake more from the environment. In that case, a more severe hypoxia is also needed to trigger the usage of a glycolytic pathway. When LDH upregulation is reduced, the H^+ secretion is reduced as well thus limiting the appearance of a Warburg phenotype. In that same simulation, oxygen levels tend to be higher than when the tumour is grown in normoxia. Two possible explanations can be: the metabolism of these cells is more efficient and can adapt more easily to normoxic/hypoxic conditions; the tumour diameter being higher cells may be closer to the boundary of the domain and thus have more access to the oxygen diffused. Lastly, we see that when HIF1-α degradation is reduced the secretion of lactic acid in the environment is higher even when oxygen is widely available. Hence, we have the appearance of Warburg cells. It seems that the Warburg Effect is only triggered after a first period of hypoxia which may be attributable to the initial conditions. The initial protein levels used in all deregulation simulations

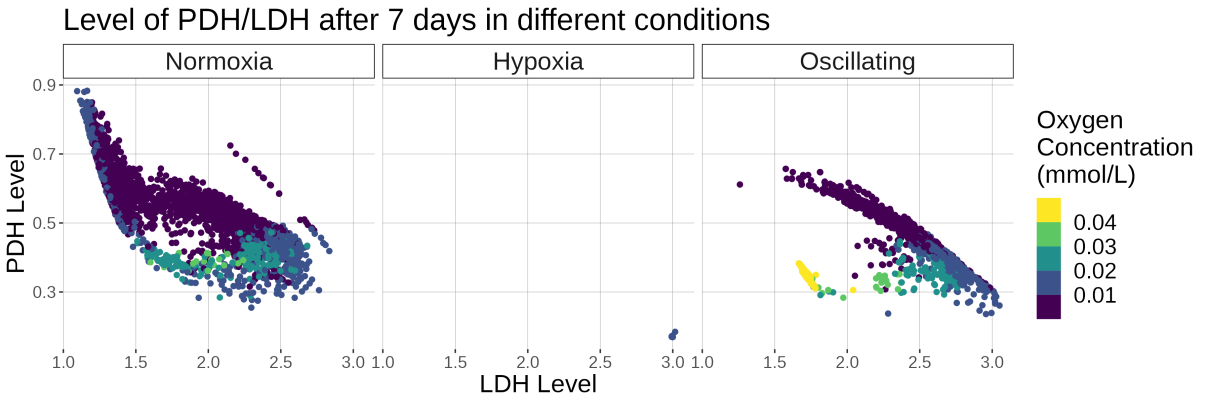

Figure 3.10: Plot of the level of PDH against the level of LDH coloured by the extracellular oxygen concentration. The graph shows the results after 7 days of growth for different conditions. In oscillating conditions, oxygen concentration is slowly decreased from normoxia to hypoxia during 6h, then oxygen is increased to normoxia at the same rate. This process is repeated until the end of the simulation.

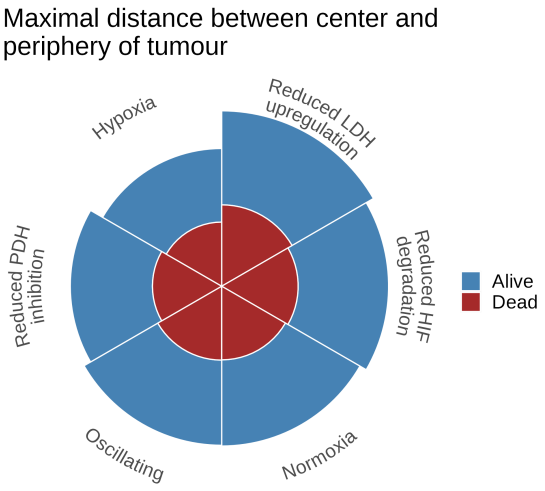

Figure 3.11: Pie chart of the maximal distance between the centre of the tumour and the periphery coloured by the state of the cell (living or dead). All simulations are included for comparison. Sizes are sorted from the largest proliferating distance to the smallest. Blue corresponds to the periphery of the tumour and brown to the dead necrotic core.

are equal to the one in the normoxic simulation. A first period of hypoxia will cause HIF to accumulate but because it is not efficiently degraded anymore, its level stays high and promotes a glycolytic state resulting in the adoption of a Warburg phenotype in normoxia.

Naturally, the different deregulation settings affect protein levels (figure 3.13). It is possible to see that deregulation of the $HIF \rightarrow LDH$ or $PDK \rightarrow LDH$ interactions only, affects the LDH/PDH of the glycolytic state. When LDH upregulation is reduced, the glycolytic state has lower values of LDH. When PDH inhibition is reduced, the glycolytic has higher values of PDH. It appears that the severe hypoxic conditions created when PDH inhibition is reduced prevent cells from adopting an OXPHOS state as a result of the severe hypoxia. Other deregulations do not affect the dynamic described earlier. It seems that more cells exhibit the intermediate metabolic state when HIF degradation is reduced even when oxygen availability should allow for the adoption of the OXPHOS one.

3.4 Conclusions: Appearance of the Warburg phenotype in the model

To conclude this chapter on the impact HIF has on metabolism, we have seen here a mathematical model combining genetic regulations with kinetic equations to understand how these processes at the cell scale reflect on the tissue scale. From these results, we can conclude that two factors can induce a Warburg Effect in the simulations: (1) creating stressful conditions for the cell forcing it to adopt a more acidic phenotype, and (2) reducing the degradation of HIF to impact the level of LDH and PDH. To recall, HIF can be upregulated by oncogenetic events including some that are frequent in the case of glioblastoma. These events include oncogene mutations or accumulation of onco-metabolites. We have seen that three metabolic states, mirrored by LDH and PDH levels, can be adopted by the cell: an OXPHOS state (high

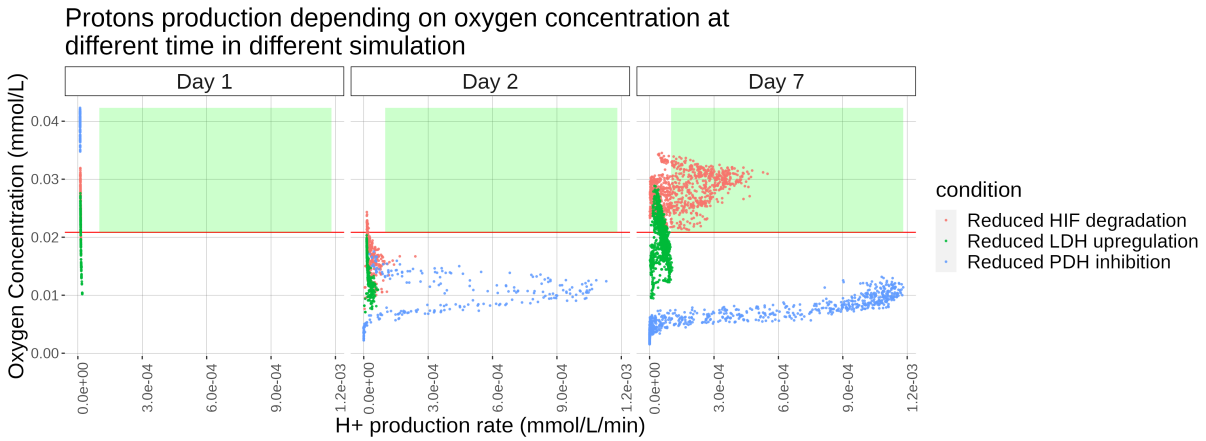

Figure 3.12: Acid production rate following oxygen extracellular concentrations at different times with different genetic perturbations. The red line indicates the hypoxia threshold (0.02085 mmol/L or 2% O_2). Only living cells are represented on the graph. Three genetics perturbations have been selected: reduced oxygen-induced degradation of HIF ($\gamma_{O\to h} = 8.0$), lower use of glycolysis in hypoxic conditions ($\gamma_{h\to l} = 2.0$) and lower effect of hypoxia on oxygen consumption ($\gamma_{k\to q} = 0.7$). Tumour growth was initiated in normoxia. The green rectangle represents the region corresponding to a Warburg effect.

PDH/low LDH), a glycolytic state (low PDH/high LDH) and an intermediate state (medium PDH/medium LDH). It seems that reduced regulation of HIF1-α by oxygen-dependent mechanisms tends to favour the intermediate state. Lastly, even though it did not trigger the Warburg phenotype, reducing LDH levels in the model increased cell proliferation but also reduced the secretion of H^+ . Stronger inhibition of LDH would probably lead to reduced generation of ATP leading to cell death. For example, the LDH-A represent an interesting target for a selective anti-cancer strategy as HIF mainly upregulates that isoform.

This chapter described the first implementation of the model with the intended goal. The next chapters of the Thesis describe how this first approach was extended with more genes and biological processes. A statistical analysis of glioblastoma data was performed to drive the review of the literature to find which processes could be valuable candidates for implementation in the model.

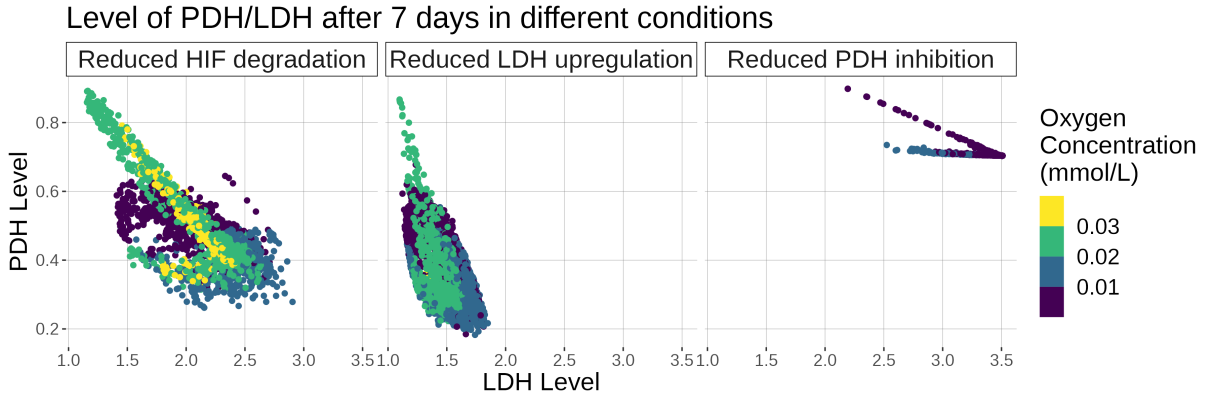

Figure 3.13: Plot of the level of PDH against the level of LDH coloured by the extracellular oxygen concentration. The graph shows the results after 7 days of growth with three different genetic perturbations: reduced oxygen-induced degradation of HIF ($\gamma_{O\to h} = 8.0$), lower use of glycolysis in hypoxic conditions ($\gamma_{h\to l} = 2.0$) and lower effect of hypoxia on oxygen consumption ($\gamma_{k\to q} = 0.7$) Tumour growth was initiated in normoxia.

4 Statistical analysis of glioblastoma transcriptomic data: new insights in biological processes

In the precedent chapter, I described a mathematical model to study the impact of Hypoxia Inducible Factor (HIF) on the metabolism in a multi-scale model. We have seen that the environmental oxygen variations and HIF stabilizing factors can impact the appearance of the Warburg Effect in the model. Yet we have seen that HIF impact on the cell is broad. To help facilitate the selection of processes to review in the literature and implement, a statistical analysis of transcriptomic data was conducted. It was decided to perform a statistical analysis of transcriptomic data. Some genes are said to be “coding” for a protein, meaning that they are implicated in the production of proteins involved in many biological functions. With the analysis of transcriptomic data and the information stored on online biological databases, it is possible to define what are the processes affected by a specific condition (here cancer). More specifically, we can assess which processes are the most frequently deregulated in glioblastoma and investigate in the literature their potential interaction with HIF to determine if they are relevant to implement in the model. In this chapter, I will describe the workflow of the statistical analysis performed: the data that were used, the results and what processes were selected with the reasons why.

4.1 From transcriptomics to deregulated pathways: the Workflow of analysis

4.1.1 The datasets

The production of protein in the cells is a two-step process that involves the generation of an intermediate molecule similar to DNA called mRNA (figure 4.1). A gene is said to be “expressed” when mRNA is produced and it is possible to quantify the expression of a gene by measuring the production of mRNA. Two methods are used to that end: microarray and RNA-Sequencing (RNA-Seq). While the first one uses probes attached to a surface that emits fluorescence when a gene is expressed, the second one sequences mRNA in the same way that DNA can [112, 113]. Both techniques can be used to assess the

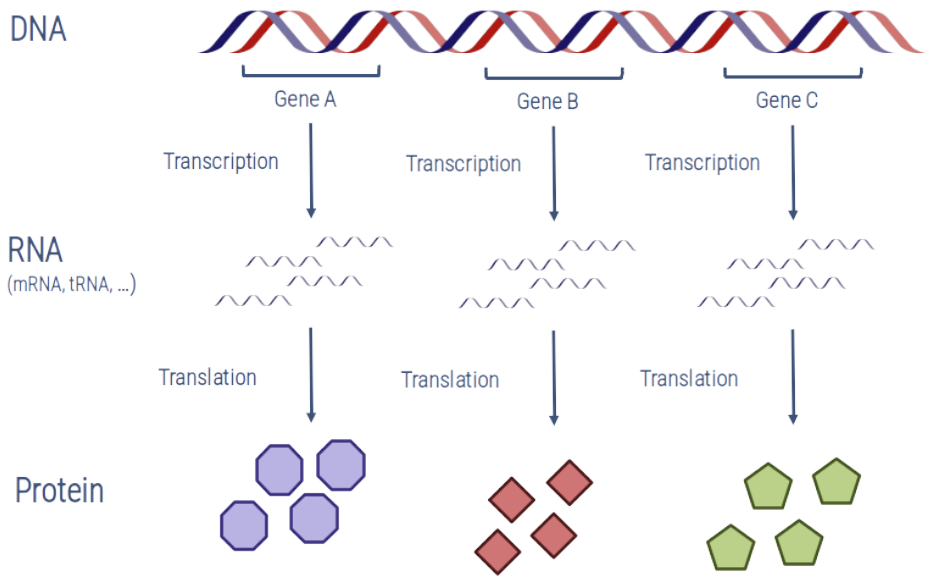

Figure 4.1: Diagram representing the production of protein. It is a two-step process: a gene is transcripted in RNA during the transcription, then the RNA is translated into a protein through the translation. Multiple types of RNA exist, yet only mRNA is translated into protein.

differences in gene expression between two conditions, yet their nature differs and as such the algorithm used to treat the data.

The data used for this work are RNA-Seq available on the The Cancer Genome Atlas (TCGA) database and RNA-Seq shared by the Institut du Cerveau et de la Moelle épinière (ICM). RNA-Seq experiments generate short sequences of an mRNA molecule called “read”. Reads are then mapped to their corresponding genes using reference genome sequences (figure 4.2) available on biological databases. The number of reads that are mapped to a gene constitutes the value quantifying the

Figure 4.2: Mapping of RNA-Seq reads against a reference genome.
Adapted from Haas et *al* [114].

expression of a gene in this experiment and gives a metric called the “raw-count”. This metric is influenced by factors that do not depend on the expression of a gene such as the gene length, total number of reads and sequencing bias. Two popular normalization methods include fragments per kilobase of exon per million mapped reads (FPKM) and transcripts per million (TPM).

fragments per kilobase of exon per million mapped reads (FPKM) **formulae :**

$$FPKM = \frac{q_g \times 10^9}{l_g \times \sum_{i=1}^{N} q_i}, \tag{4.1}$$

where q_g is the number of fragments counts, l_g is the gene length and $\sum_{i=1}^{N} q_i$ is the total number of mapped fragments.

transcripts per million (TPM) **formulae :**

$$TPM = \frac{q_g/l_g}{\sum_{i=1} Nq_i/l_i} \times 10^6, \tag{4.2}$$

where q_g is the number of reads mapped to a gene, l_g is the gene length and $\sum_{i=1} Nq_i/l_i$ is the sum of mapped reads to transcript normalized by the transcript length.
FPKM can be converted to TPM by: $\frac{FPKM}{\sum FPKM} \times 10^6$.

Subsequent steps to analyze the data in this work rely on a Differential Expression (DE) analysis to identify the altered biological processes in the data. FPKM and TPM normalize away the sequencing depth which is the number of sequenced reads for a given sample. These normalization methods are not suitable for performing comparison among samples and should not be used for DE analysis, therefore more suitable methods for this analysis were designed [115, 116].

The first dataset used in this work was established by the Institut du Cerveau et de la Moelle épinière (ICM) in Paris. It is composed of RNA-Seq of cells taken from the tumour of 20 patients diagnosed with glioblastoma that were cultured *in vitro*. One biological replicate is available per sample and approximately 20,000 genes were sequenced in the experiments. Compared to regular RNA-Seq that are obtained using a bulk tumour, the Patient Derived Cell Lines (PDCL) data are *single-cell* i. e.they are derived from one cell only. There are no matching controls and these had to be found online from a study on the characterization of different astrocytic models conducted by Lundin et *al* [117]. These controls contain four samples of different astrocyte cell lines with three biological replicates per sample. Some of these cell lines are commercial astrocytomas purchased while some are astrocyte cells generated by a protocol described by the authors. They first generated Long-Term Neural Epithelial Stem Cell

Accession Number	Sample name
GSM2927872	AF22_NES-Astro_d29_Br1
GSM2927873	AF22_NES-Astro_d29_Br2
GSM2927874	AF22_NES-Astro_d29_Br3
GSM2927881	iCellAstro_Br1
GSM2927882	iCellAstro_Br2
GSM2927883	iCellAstro_Br3
GSM2927884	CCF_Br1
GSM2927885	CCF_Br2
GSM2927886	CCF_Br3
GSM2927887	phaAstro_Br1
GSM2927888	phaAstro_Br2
GSM2927889	phaAstro_Br3

Table 4.1: List of controls used for comparison with Patient Derived Cell Lines (PDCL) data. The accession number is an ID that can be used to retrieve the data on the Gene Expression Omnibus (GEO) platform. The accession number of the experiment is GSE109001. The dataset contains 4 samples of different astrocyte cell lines with 3 replicates per sample. Data were last updated on January 29, 2019.

Culture (ltNES) cells from AF22 human induced Pluripotent Stem Cells (hiPSC) cells, then ltNES were differentiated into astrocytes [118].

The second dataset used was downloaded from the TCGA-GBM project on The Cancer Genome Atlas (TCGA) database. This project aimed to sequence the mutations and gene expressions for the glioblastoma (GBM) disease which is the first tumour sequenced by the TCGA initiative. The data contains information about primary and recurrent tumours. Sample acquisition and RNA-Seq output followed the TCGA guidelines and protocol described on their website (https://docs.gdc.cancer.gov/). Compared with PDCL, the dataset includes 169 tumour samples and 5 matching controls. There is generally less data on healthy brains although it is possible to find whole brain data from post-mortem brain samples. In this work, the data version 33.1 from May 31 2022 was used.

4.1.2 Differential Expression Analysis

One can use RNA-Seq data previously described to unravel the processes involved in different conditions such as disease. A popular method is to perform a Differential Expression (DE) analysis to determine genes that are over-expressed/upregulated or under-expressed/down-regulated. A gene is said to have its expression deregulated when the quantity of mRNA differs between two conditions tested (figure 4.3). Usually, the conditions are normal or disease. This step will produce a list of genes deregulated in GBM that can be used in the following analysis as an input to assess the deregulated pathways. For this step, two analyses are conducted: one at the population level, to assess the typical deregulation, and one at the individual level, to assess the heterogeneity among the samples.

Population-level analyses are now standard procedures and many softwares are available such as edgeR [119], Limma [120] and DESeq2 [121]. DESeq2 was selected for the study at the population level as the software is well-documented and its performance was published. It can perform an automatic detection of outliers. In addition, it includes a built-in normalization method suitable for comparing samples. This is an important advantage as we have seen earlier that the normalization method may influence the performance of the analysis. Because DESeq2 normalizes the gene expression values, it accepts only raw counts. DESeq2 fit a generalized linear model (GLM) for each gene using a negative binomial distribution (or gamma-Poisson distribution) to model gene counts. The variability between replicates is estimated using an empirical bayes shrinkage method. It uses the same method to estimate the change in expressions, called the fold-change, of genes with low expression as they are subjected to stronger variability. DESeq2 uses a Wald Test to test for differential expression: the estimated fold-change is divided by its standard error, giving a z-statistic, which is compared to a standard normal distribution. The Wald statistic computed by DESeq2 will be used as an input for pathway enrichment analysis. In the population analysis, all the tumour samples are compared against all the controls to assess the typical

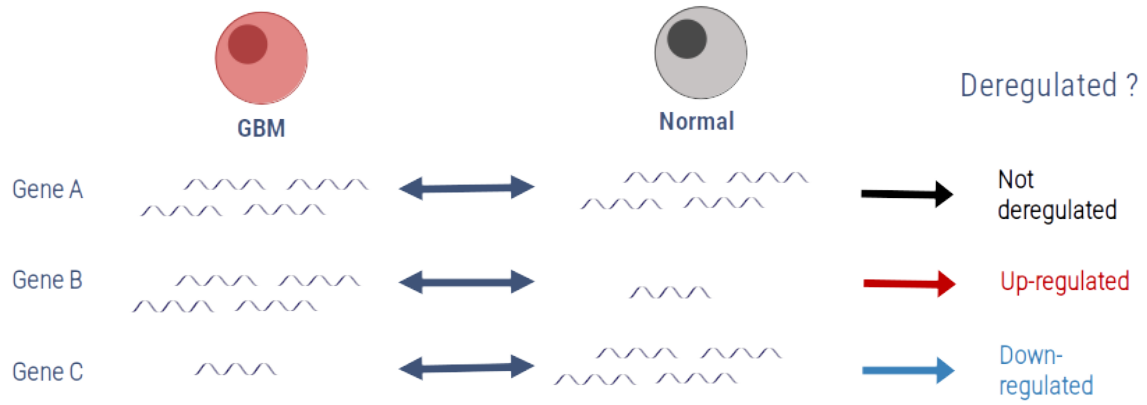

Figure 4.3: Representation of a Differential Expression analysis. The quantity of gene transcripted into an mRNA is compared between two conditions, here normal against glioblastoma (GBM). In this example: gene A is not affected, gene B is upregulated in cancer, the gene C is downregulated in cancer.

deregulation of GBM. Tumour samples are considered as “replicates of one GBM patient” to respect the hypothesis of DESeq2 which assumes several replicates per conditions to compute the dispersion as said before.

With the current focus on personalized medicine, more data analyses have tried to assess tumour heterogeneity. The development of single-cell RNA-Seq technology opened the path to a better understanding of cellular heterogeneity caused by different transcriptional programs. A study has shown that analyses of bulk tumours disagree with results using single-cell data. They show that the main variations were linked to the metabolic of mitochondrial or glycolytic pathways [122]. With that issue in mind, the analysis at the individual scale will determine which processes are frequently deregulated among the samples of our datasets. A method was recently developed to answer the growing interest and need in studying heterogeneity of gene expression. This method called Personalized Differential Analysis (PenDA) has been used for the analysis at the individual scale. PenDA is a rank-based method that was designed to perform DE analysis by comparing one sample of the tested condition with a group of controls. Population-scale tools like DESeq2 generally rely on modelling the data using a distribution law and then tests for differential expression. These methods were designed to extract typical gene deregulation patterns in a condition, consequently, they do not provide information at the individual scale. In addition, the DESeq2 method assumes that several biological replicates are present in the datasets to estimate the dispersion and normalize the data accordingly. Population-scale methods are usually sensitive to the normalization applied and the batch effect while PenDA is robust to these effects. Before assessing the differentially expressed gene, the PenDA method creates for each gene g the lists $L(g)$ and $H(g)$ of other genes with an expression *lower* or *higher* than the current gene g in at least 99% of the control samples. The list is then shrunk to the l genes with the closest expression to g. Then for each tumour sample T:

1. The expression $E(g,T)$ of each gene g in the tumour sample T is compared to the corresponding expression of genes in the list L and H to define four non-overlapping sets of genes L_d, L_u, H_d and H_u; where:

$$
\begin{aligned}
L_d &= \{g' \in L(g)\ E(g',T) < E(g,T)\} \\
L_u &= \{g' \in L(g)\ E(g',T) > E(g,T)\} \\
H_d &= \{g' \in H(g)\ E(g',T) < E(g,T)\} \\
H_u &= \{g' \in H(g)\ E(g',T) > E(g,T)\}
\end{aligned}
$$

   The ordering of g in T has changed compared to the control when $L_u \neq \emptyset$ or $H_d \neq \emptyset$.

2. The gene g is considered deregulated in the tumour T if, and only if:

$$
\left(\frac{|L_u|}{|L|} \geq h\right) \text{ or } \left(\frac{|H_d|}{|H|} \geq h\right),
$$

where $|X|$ is the cardinality of the ensemble X, h is the minimal proportion of genes in L and H whos relative ordering with g has changed. It this condition is satisfied, then g is downregulated if $|L_d| + |H_d| < |L|$, or upregulated if $|L_u| + |H_u| < |H|$. In the case where L and H are empty (for example in the tail of the ranked genes), the percentile method is used: the gene is considered downregulated if $E(g,T) < p_l/f$ or upregulated if $E(g,T) > p_u \times f$. p_l and p_u are respectively the x and $(100-x)$ percentile of the gene expression distribution in the normal samples ensemble, f is a factor allowing to expand the window of normal expression.

3. To minimize the bias caused by the deregulation of genes in the L or H, all genes detected as deregulated are then excluded from the L and H list. The precedent steps are then repeated until the list of deregulated genes converges or until a maximal number of iterations is reached.

Figure 4.4 describes graphically the list $L(g)$ and $H(g)$. Expression data submitted to PenDA were

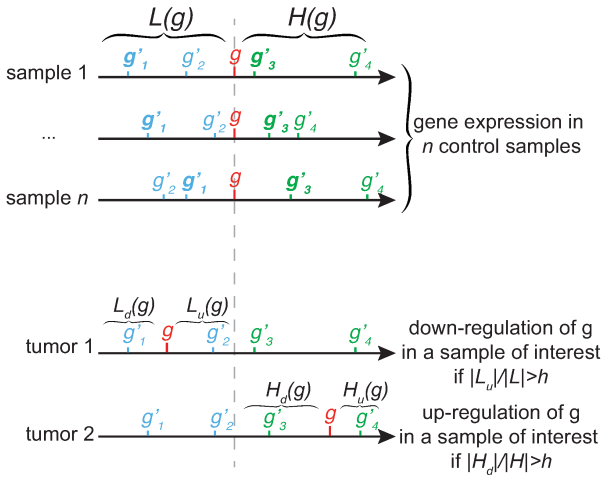

Figure 4.4: Diagram representing the $L(g)$ and $H(g)$ lists in Personalized Differential Analysis (PenDA). For each gene g in a control ensemble, genes with a lower expression or higher expression in at least 99 % of the samples are determined. Source: Richard et *al* [123].

normalized using the pseudo-count method from DESeq2 ($log_2(count+1)$). Lastly, PenDA only outputs whether a gene is down/up-regulated or not deregulated but does not compute any statistic or fold-change of expression. In the individual analysis, one tumour sample is compared to all the controls.

4.1.3 Pathway Enrichment analysis

Differential Expression analysis outputs the genes that are deregulated due to a specific condition, here glioblastoma. Yet, that list can contain thousands of genes depending on the datasets analyzed which proves to be difficult to interpret. Filtering genes based on the significance and the strength of the deregulation observed is a possible option to reduce the number of genes to investigate. However, a gene is often not linked to only one function in the cell. It is even more apparent when deregulated genes belong to signalling pathways. Furthermore, there is a lot of redundancy among the pathways, different genes may lead to similar functions. Pathway Enrichment analysis became a standard approach to analyze a list of deregulated genes and determines the deregulated pathways in a sample. Simply put, pathway enrichment turns a list of deregulated genes into a smaller list of pathways [124]. That method (1) reduces the complexity of the data to analyze and (2) facilitates the interpretation of deregulated mechanisms as the name of altered pathways is more tangible than a list of genes. Three generations of pathway enrichment have been developed to interpret high-throughput technology data (figure 4.5). All these tools typically need as input a list of genes of interest like the deregulated genes from a from a DE analysis. The first generation called Over-Representation Analysis (ORA) was developed as a response to the need for functional analysis to interpret microarray gene expression data. ORA methods statistically determine the fractions of genes in a pathway found among a set of gene of interest. For each pathway, the number of genes present in the pathway are counted. Then pathways are tested for over- or under-representation in the list of genes submitted. Common tests used in these methods are based on the hypergeometric, chi-square or binomial distribution. These methods present several limitations:

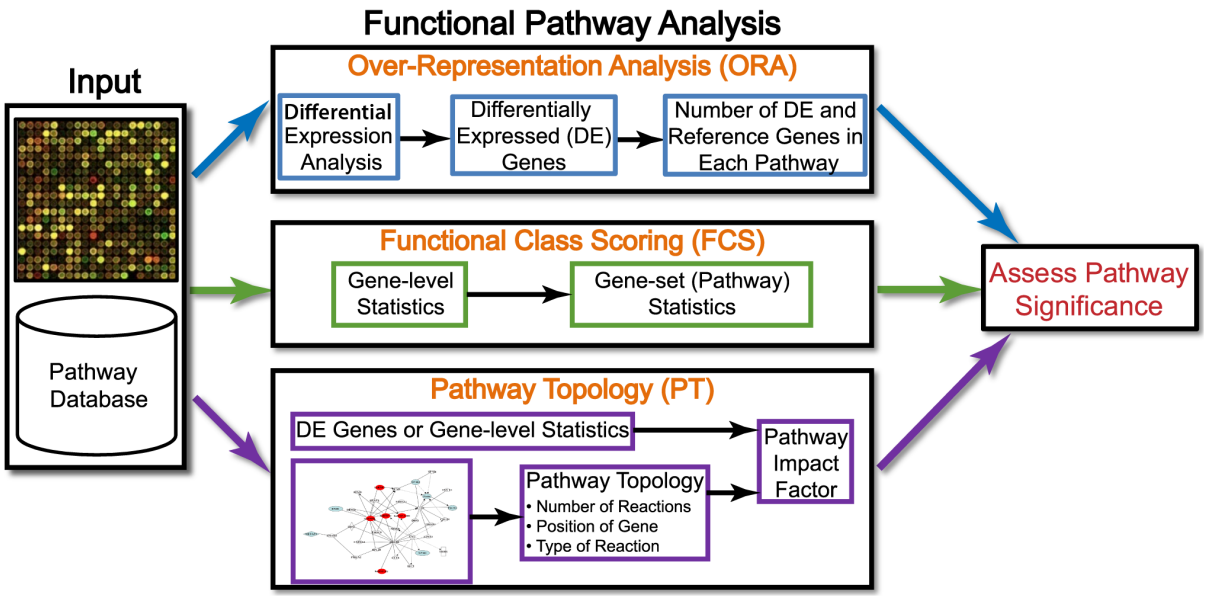

Figure 4.5: Overview of the different generations of pathways enrichment analysis.
Source: Khatri et *al* [125].

- They do not consider the measured change, they ignore any values that can be associated with a gene such as a fold-change of expression.
- Genes submitted are often filtered beforehand by their change of expression and the significance of the observed change, thus ORA method typically only uses the most significant genes.
- Each gene and pathway is treated as independent from each other which is not biologically accurate.

The second generation Functional Class Scoring (FCS) hypothesizes that not only individual large changes in gene expression can significantly alter a pathway, but that small coordinated changes in sets of genes can do as well. They use a gene-level statistic such as the signal-to-noise ratio, t-test and the z-statistic. A pathway level statistic is then computed for each pathway like the Kolmogorov-Smirnov statistic, sum, mean or median level of gene statistic. FCS method addresses some of the ORA limitations: they do not require any filtering keeping only the most significant genes and use a metric associated with each gene to detect altered pathways. They still have some limitations:

- Like for ORA, each pathway is assumed to be independent.
- Many FCS rank genes according to their change in expression but discard these changes in further analysis.

Lastly, Pathway Topology (PT) goes further than FCS by using information on the interactions of a gene network (activation/inhibition, cell compartment) to test for enrichment. PT methods are similar to FCS as they perform the same steps but the key difference is the additional information considered by PT. Despite more limitations are addressed, they still are limited by:

- Exhaustiveness of the information covered by the current database annotations.
- Their inability to account for the dynamical nature of biological systems.

For this step, two software were selected: G:Profiler [126], an ORA method, and gene set enrichment analysis (GSEA) [127], a FCS method. Both of these tools are used for the study at the population scale but only G:Profiler is used for the study at the individual scale. While DESeq2 is used for the DE analysis at the population level, PenDA is used for the individual study. Compared to DESeq2, PenDA does not rely on a statistical test to search for deregulation nor does it give the fold-change of expression. If G:Profiler only needs the name of the genes that are deregulated, GSEA requires that the list of genes is ranked by a metric. PenDA does not give metric hence only G:Profiler can be used for the individual analysis. The overall workflow is summarized in figure 4.6.

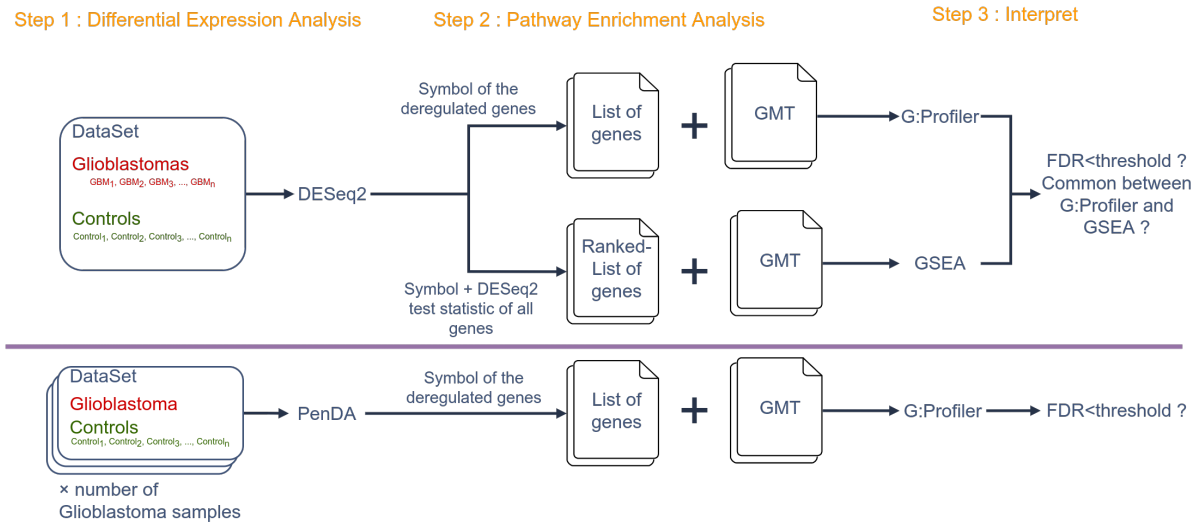

Figure 4.6: Diagram of the overall workflow to analyze the deregulated pathways in glioblastoma. The top panel describes the analysis at the population scale. Here, all the tumours are compared against all the controls using the DESeq2 tool. The result is then submitted to G:Profiler and GSEA. The bottom panel describes the analysis at the individual level. In that case, one tumour sample is compared to all the controls. DESeq2 assumption is not met as only one replicate is used for the tumour sample, therefore it cannot be used. PenDA does not compute a statistic nor the fold-change in expression. Because GSEA needs genes that are ranked using a metric it cannot be used for the individual analysis.

Both G:Profiler and GSEA require functional pathway information in the form of a gene matrix transposed (GMT) file. GMT files contain gene sets: a group of genes sharing a common biological feature (for a biological function or chromosomal location) which here is a pathway. In a GMT, each row corresponds to a gene set i. e.a pathway. The first column contains the ID of the pathway, the second column contains a description and the remaining columns contain the genes present in the pathway. An example of GMT can be found in table 4.2. Pathway information was downloaded from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [128] and Reactome [129], two biological databases with curated sets of data. A same pathway can have different definitions among databases. For example, glycolysis is defined as *canonical glycolysis* in Gene Ontology, *glycolysis* in Reactome and *Glycolysis/Gluconeogenesis* in KEGG. It was shown that the choice of database can impact the results [130], hence it is recommended to use more than one. It is also beneficial to exclude small and large pathways as: (1) small pathways are redundant with larger pathways while large pathways are often too general (metabolism, cell cycle) and (2) small pathways are never picked up by pathway enrichment tools while large pathways tend to always shows-up in the results. A pathway size of 10-15 to 200-500 genes is recommended yet sizes of 2,000 genes can be found in the literature [124]. In both the population and the individual studies, pathways with a size between 15 to 500 genes were kept. Furthermore, an increased number of pathways increases the correction applied to the p-value. The p-value is a statistical value associated with a test which represents the probability that an event is observed due to a random effect rather than truly due to the condition studied. In other terms, in a DE analysis, the p-value represents the probability that the fold-change is not different between the two conditions; in a pathway enrichment analysis, it is the probability that a pathway is not enriched. Consequently, the lower the p-value is, the more significant the test is. When doing multiple hypothesis testing, the p-value can sometimes be significant only by “chance” alone. Multiple-testing correction methods were designed to adjust the p-value and prevent this problem. The most common, and the one used during the analysis, is the False Discovery Rate (FDR) correction method. This correction method inflates the p-value depending on the number of test runs. As a consequence, the fewer pathways we are testing for enrichment, the lower the FDR correction is. Filtering unnecessary pathways can help reduce the correction applied to the p-value. Like the p-value, the FDR should be as low as possible. In this work, pathways were considered enriched when the FDR was below 0.05 (5 %). A threshold of 0.1 can be found in the literature but it seems too lenient.

G:Profiler is a tool accessible online via a web interface or libraries provided by the developers [126]. It uses the cumulative hypergeometric distribution to test for functional enrichment. Besides pathway information, G:Profiler only requires as an input a list of gene names. G:Profiler is updated regularly with

Pathway ID	Description	Genes
R-HSA-1234174	Cellular response to hypoxia	AJUBA ARNT CA9 CITED2 CREBBP CUL2 EGLN1 EGLN2 EGLN3 ELOB ELOC EP300 EPAS1 EPO HIF1A HIF1AN HIF3A ...
R-HSA-1650814	Collagen biosynthesis and modifying enzymes	ADAMTS14 ADAMTS2 ADAMTS3 BMP1 COL10A1 COL11A1 COL11A2 COL12A1 COL13A1 COL14A1 COL15A1 COL16A1 COL17A1 COL18A1 ...
R-HSA-8948216	Collagen chain trimerization	COL10A1 COL11A1 COL11A2 COL12A1 COL13A1 COL14A1 COL15A1 COL16A1 COL17A1 COL18A1 COL19A1 COL1A1 COL1A2 COL20A1 COL21A1 COL22A1 ...
R-HSA-1442490	Collagen degradation	ADAM10 ADAM17 ADAM9 COL12A1 COL13A1 COL14A1 COL15A1 COL16A1 COL17A1 COL18A1 COL19A1 COL23A1 COL25A1 COL26A1 COL9A1 COL9A2 COL9A3 ...
R-HSA-1474290	Collagen formation	ADAMTS14 ADAMTS2 ADAMTS3 BMP1 CD151 COL10A1 COL11A1 COL11A2 COL12A1 COL13A1 COL14A1 COL15A1 COL16A1 COL17A1 COL18A1 COL19A1 COL1A1 ...
R-HSA-2559586	DNA Damage/Telomere Stress Induced Senescence	ACD ASF1A ATM CABIN1 CCNA1 CCNA2 CCNE1 CCNE2 CDK2 CDKN1A CDKN1B EP400 H1-0 H1-1 H1-2 ...
R-HSA-5693606	DNA Double Strand Break Response	ABL1 APBB1 ATM BABAM1 BAP1 BARD1 BAZ1B BRCA1 BRCC3 BRE CHEK2 EYA1 EYA2 EYA3 EYA4 FAM175A H2AFX ...
R-HSA-5693532	DNA Double-Strand Break Repair	ABL1 APBB1 ATM ATR ATRIP BABAM1 BAP1 BARD1 BAZ1B BLM BRCA1 BRCA2 BRCC3 BRE BRIP1 CCNA1 CCNA2 CDK2 ...
R-HSA-70171	Glycolysis	AAAS ADPGK ALDOA ALDOB ALDOC BPGM ENO1 ENO2 ENO3 GAPDH GAPDHS GCK GCKR GNPDA1 GNPDA2 GPI HK1 HK2 HK3 NDC1 NUP107 NUP133 ...
R-HSA-71403	Citric acid cycle (TCA cycle)	ACO2 CS DLD DLST FAHD1 FH IDH2 IDH3A IDH3B IDH3G MDH2 ME2 ME3 NNT OGDH SDHA SDHB SDHC SDHD SUCLA2 ...

Table 4.2: Example of GMT downloaded from the Reactome database.

the most common databases such as KEGG, Reactome, WikiPathways, miRTarBase, Gene Ontology, etc. Except for a few databases due to licensing limitations, including KEGG, all built-in G:Profiler's gene sets can be downloaded from their website.

Gene set enrichment analysis (GSEA) is a rank-based method developed by Subramanian et *al* [127] that determines the pathway enriched in a ranked list of genes that are ordered using a metric: a positive or negative floating value. The fold-change in expression between a tumour and control is often used as the metric [124]. The list is sorted in a decreasing order so the genes with the strongest absolute value of the rank can be at either tail of the list. The GSEA approach determines whether the genes involved in a gene set are mostly distributed at the top or at the bottom of a gene list. The algorithm runs through the list to compute a score called Enrichment Score (ES), a weighted Kolmogorov-like statistic (figure 4.7). A gene metric is added to the score if it is present in the pathway, substracted otherwise. The maximum deviation from zero is the final ES. The score of the tested pathway is compared with the score of the null distribution to test for significance. The score of the null distribution is generated by permutating the genes in the dataset or the phenotype of the samples and then the ES is computed. GSEA includes built-in gene sets from the Molecular Signatures Database (MSigDB) which can be downloaded. A study from Zyla et *al* has shown that the metric used to rank the genes can impact the result of GSEA. They evaluated the performance of 16 different metrics and found that the absolute value of the Moderated

Welch Test statistic, the Minimum Significant Difference, the absolute value of the Signal to Noise ratio and the Baumgartner-Weiss-Schindler test statistic give the best outcome [131]. Most of these metrics tested are statistics computed during a test. In this study, the Wald metric computed by DESeq2 was used as it combine biological information and statistical significance. The Wald-Test statistic is (1) positive when a gene is upregulated and negative when it is downregulated; (2) the higher the absolute value of the statistic, the more significant the fold change (figure 4.8).

4.2 Exploration of the deregulated mechanisms at the population and individual scale

Before conducting a Differential Expression (DE) analysis, one may want to investigate whether the controls can be compared to the tumours. Indeed, the goal of such analysis is to assess the differences between a tumour and a healthy tissue which cannot be done if samples among the control group have an expression far from each other. In other words, there should be less differences between controls than between a control and a tumour sample. In addition, tumour samples and control samples should be clearly separable and not mixed. A Principal Component Analysis (PCA) was performed on the gene count normalized by the *pseudo-count* method (figure 4.9) to determine whether these two criteria are respected. As can be seen on the PCA, controls are well separated from the tumour samples in both datasets. However, controls are more sparse in the case of PDCL compared with TCGA.

Even if we want to maximize the distance between controls and tumours and reduce the distance between controls, the distribution of gene counts should be skewed in the same way for both tumour and control. A similar distribution of gene expression in different cancers from TCGA was observed [132]. The density of normalized gene counts is shown in figure 4.10 to compare the different samples. These data

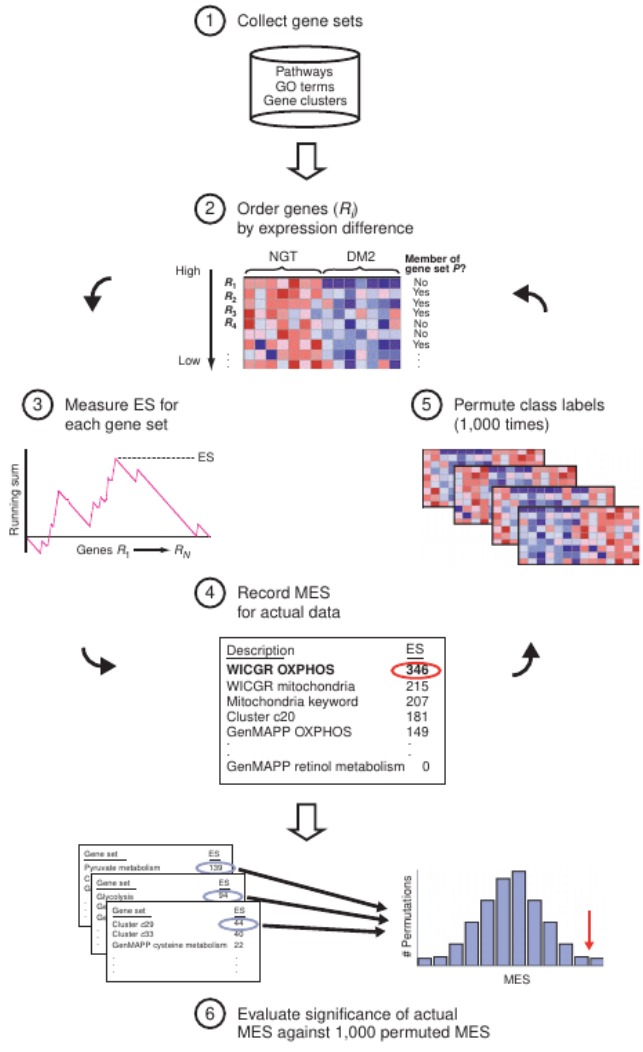

Figure 4.7: Algorithm of GSEA. (1) Gene sets are collected from biological databases. (2) Genes are ordered following a metric that usually represents expression strength or significance. (3) A Kolmogorov-Smirnov running sum statistic is computed for each pathways. The sum increases when the gene encountered is present in the gene set and decreases otherwise. The maximal deviation from 0 constitutes the Enrichment Score (ES). (4) The ES is computed for each pathways. (5) A null distribution is generated by swapping the phenotype of each sample or the genes and then the ES is calculated like in the previous step. (6) The significance of each pathway is assessed and the p-value is calculated. Source: Subramanian et *al* [127].

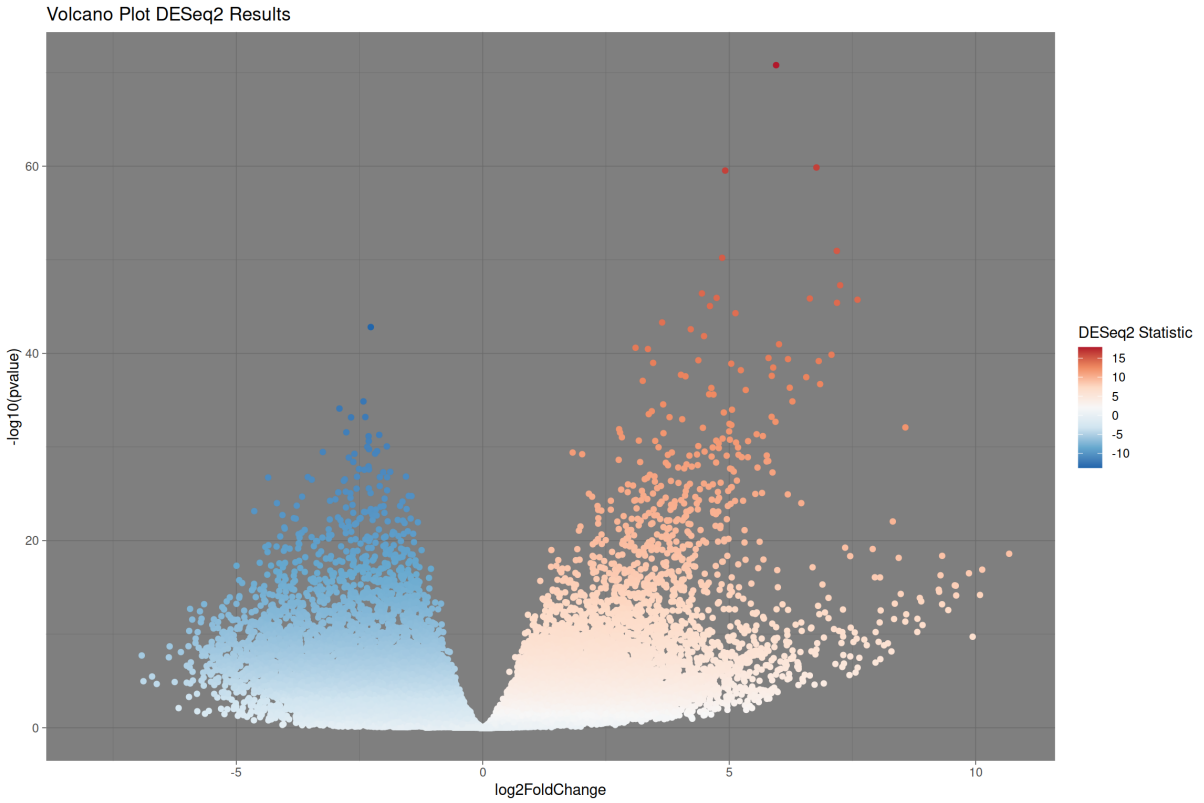

Figure 4.8: Volcano plot of the result of DESeq2 during the population study on the TCGA dataset. The x-axis is the change of expression, negative values indicate a downregulation and positive values indicate an upregulation. The y-axis represents the significance of the observed change, the higher the more significant. Each point represents the result of DESeq2 for one gene. Points were coloured by their computed Wald statistic: red shows a positive statistic and blue is a negative one. The stronger the shade, the higher the absolute value of the statistic is. It is possible to see that the sign of the statistic follows the one of the fold-change and that the most significant genes have a higher absolute statistic.

suggest that the distribution seems reasonably close for a comparison to be performed. Nevertheless, the observation with TCGA tends to indicate that better results can be yielded with TCGA rather than with PDCL. Although the gene distribution seems to show that comparison in the case of PDCL is still possible.

4.2.1 Analysis of the deregulated pathways at the population scale

In a first intent to assess roughly the quality of the results, it is possible to have a look at the p-value of a target pathway that is known to be deregulated in the condition tested. Zyla et *al* used the adjusted p-value of target pathways for different cancer types to assess the sensitivity of the different ranking metrics [131]. Here the *Glioma* entry (path:hsa0514) from the KEGG database will be used to appreciate the result given by the different pathway enrichment methods. The FDR is expected to be lower than 0.05. A lower FDR corresponds to a result of higher quality. With PDCL data, the *Glioma* entry has a FDR of 0.053 in G:Profiler and 0.83 with GSEA. With TCGA data, the *Glioma* entry has a FDR of 0.0009 in G:Profiler and 0.005 in GSEA. Based on this, the FDR for the *Glioma* entry tends to show a poor quality of results with the PDCL datasets. The control data used are from another study which introduces an additional variability that is not due to the condition tested. Furthermore, it includes astrocytoma cell lines, which are cancerous cells. On the opposite, TCGA data that includes matching controls (even if there are very few samples) display a lower FDR. This may partially explain why the FDR is above the threshold set and shows the importance of finding good controls for transcriptomic analyses but will have little impact on the work described here as the selected mechanism will be reviewed in the literature beforehand.

Two different databases (KEGG and Reactome) were used as different databases may have different entries for the same pathway, impacting the result. Figure 4.11 and 4.12 show that a good agreement between the pathway enrichment tools depends on the biological category and the data used. There is

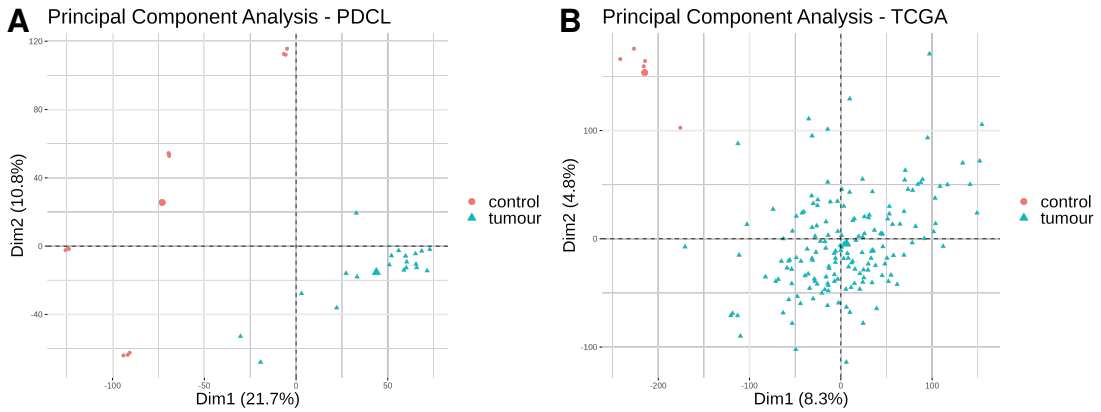

Figure 4.9: Principal Component Analysis (PCA) of the gene counts normalized by the *pseudo-count* method for (A) PDCL and (B) TCGA datasets.

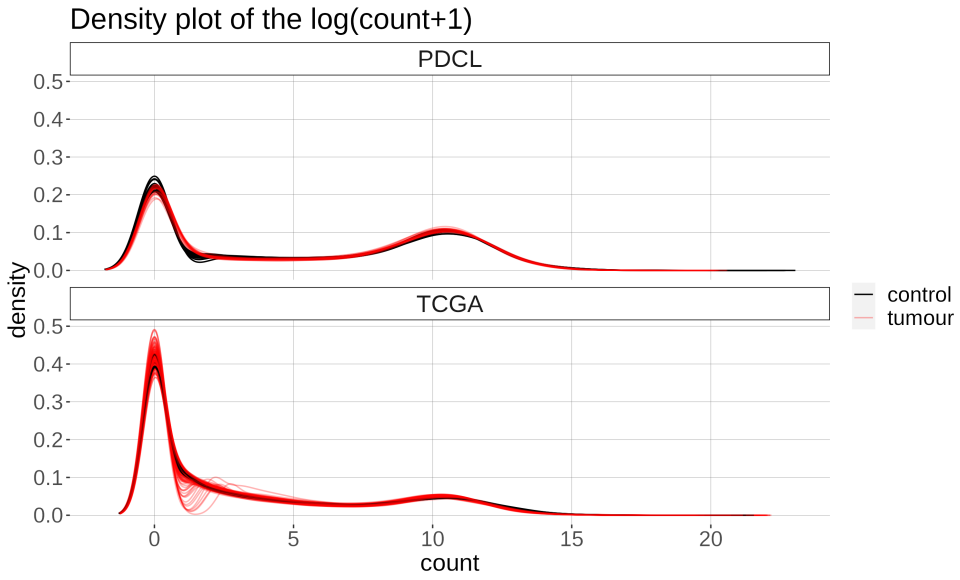

Figure 4.10: Density plot of the gene counts normalized by the *pseudo-count* method for PDCL and TCGA datasets. Each curve represents a different sample.

generally a better agreement between GSEA and G:Profiler in KEGG with TCGA data compared to PDCL with the same database. For example, the *Organismal Systems* category displays no agreement between G:Profiler and GSEA for PDCL data, while with TCGA data, the majority of pathways are enriched in both tools showing a good agreement. Similarly, the *Environmental Information Processing* category is only enriched with G:Profiler in PDCL while pathways enriched in both tools are present with TCGA data. Compared to KEGG, a few Reactome categories are not enriched in either pathway enrichment tool such as the *Circadian Clock* and the *Reproduction* categories. These categories could be removed from the GMT files to improve the analysis as a higher number of pathways to test increases the correction on the p-value. Interestingly, the disease category is only filled with pathways that are found enriched in both GSEA and G:Profiler when using PDCL data. Still with PDCL, *DNA Repair* and *Cell-Cycle* are the categories showing the most agreement. With TCGA data, the categories exhibiting the most agreement are the *Neuronal System*, the *Imune System*, the *Signal Transduction* and the *Cell-Cycle* (in that order). In most categories, the majority of pathways enriched are found by GSEA.

Figure 4.13 shows the FDR value of several pathways for both databases and both datasets. It would be expected that the *Oxidative phosphorylation* is altered as we have seen that cancer cells tend to display increased glycolytic activity. It seems though that no matter the datasets used, there is not a consensus reached between GSEA and G:Profiler as this pathway was enriched only with GSEA. Signalling pathways known to be frequently altered in GBM such as *PI3K-AKt signalling*, *RAS signalling*, *p53 signalling* or *VEGF signalling* are only found significant in TCGA which shows a difference between

the two datasets. Some of these pathways may not be detected in PDCL as some of the controls are astrocytomas as previously said. The *VEGF signalling* pathway may not be detected because the PDCL cells were cultivated in *in vitro* normoxic conditions before RNA-Sequencing. *Steroid biosynthesis* and the *Cholesterol biosynthesis* (two entries linked to Cholesterol metabolism) are only found enriched in PDCL. With a few examples, we see that the two datasets display differences in the deregulated pathways identified. Here, we only examine the *typical* deregulations in both datasets but we will see that in addition to the deregulated mechanisms, we can also determine the frequency of deregulation and in which samples. For example, the *p53 signalling* that is not enriched in PDCL is still found deregulated in some samples when conducting the analysis at the individual level.

4.2.2 Analysis of the deregulated pathways at the individual scale

The goal of the analysis at the individual scale is to assess the differences among the different samples and determine the frequency of deregulation for each pathway. The number of deregulated pathways varies between 0 and 40 per PDCL samples 4.14. The PDCL sample with the higher number of altered pathways is the *4339-p21* with mainly pathways linked to diseases. Aside from the disease category, majority of the deregulated pathways in *4339-p21* are associated to the *Cellular Processes* and *Environmental/Genetic Information Processing* from KEGG, or the *Cell-Cycle* from Reactome, showing again the importance of using different databases. Some PDCL samples are not enriched using either KEGG or Reactome. Like with PDCL data, the main category represented among the TCGA samples is the disease category from both KEGG and Reactome 4.15.
In addition, besides pathways linked to diseases, the *Cell-Cycle* from Reactome is one of the most

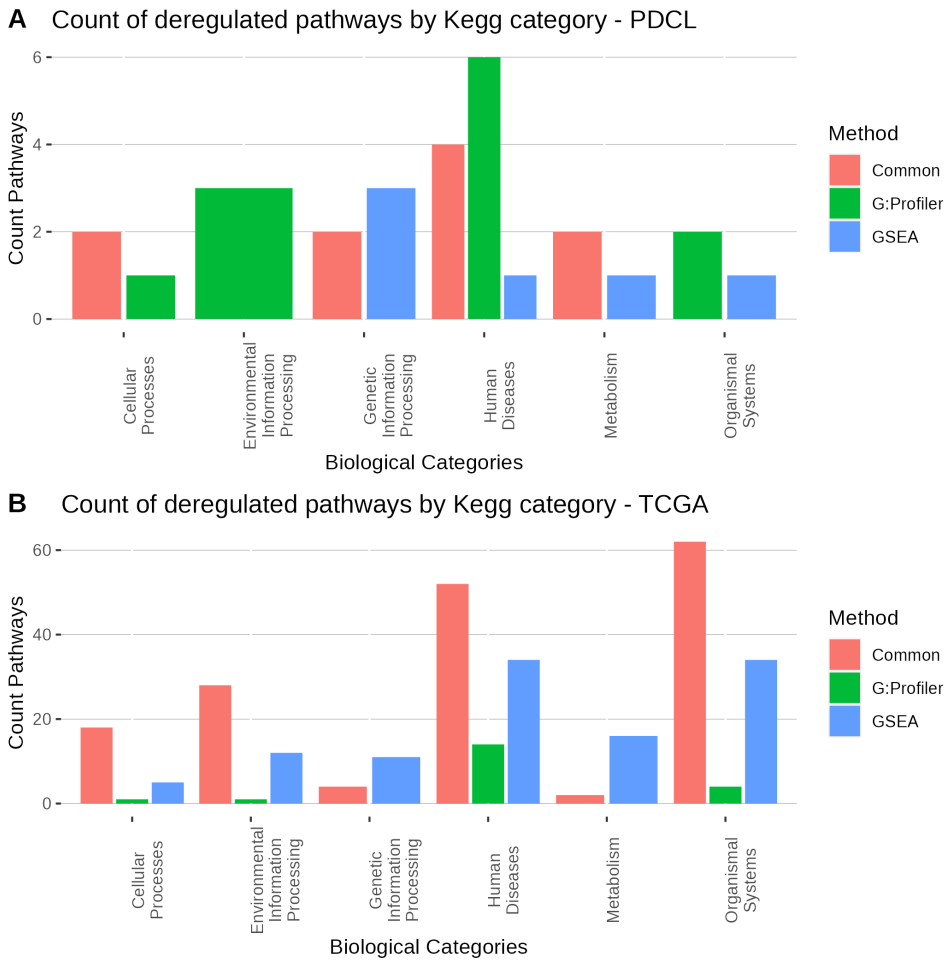

Figure 4.11: Barplot of the count of significantly deregulated pathways with the KEGG database using (a) PDCL and (b) TCGA data. Pathways are coloured whether they pass the FDR threshold only in G:Profiler (green), only in GSEA (blue) or in both tool (red).

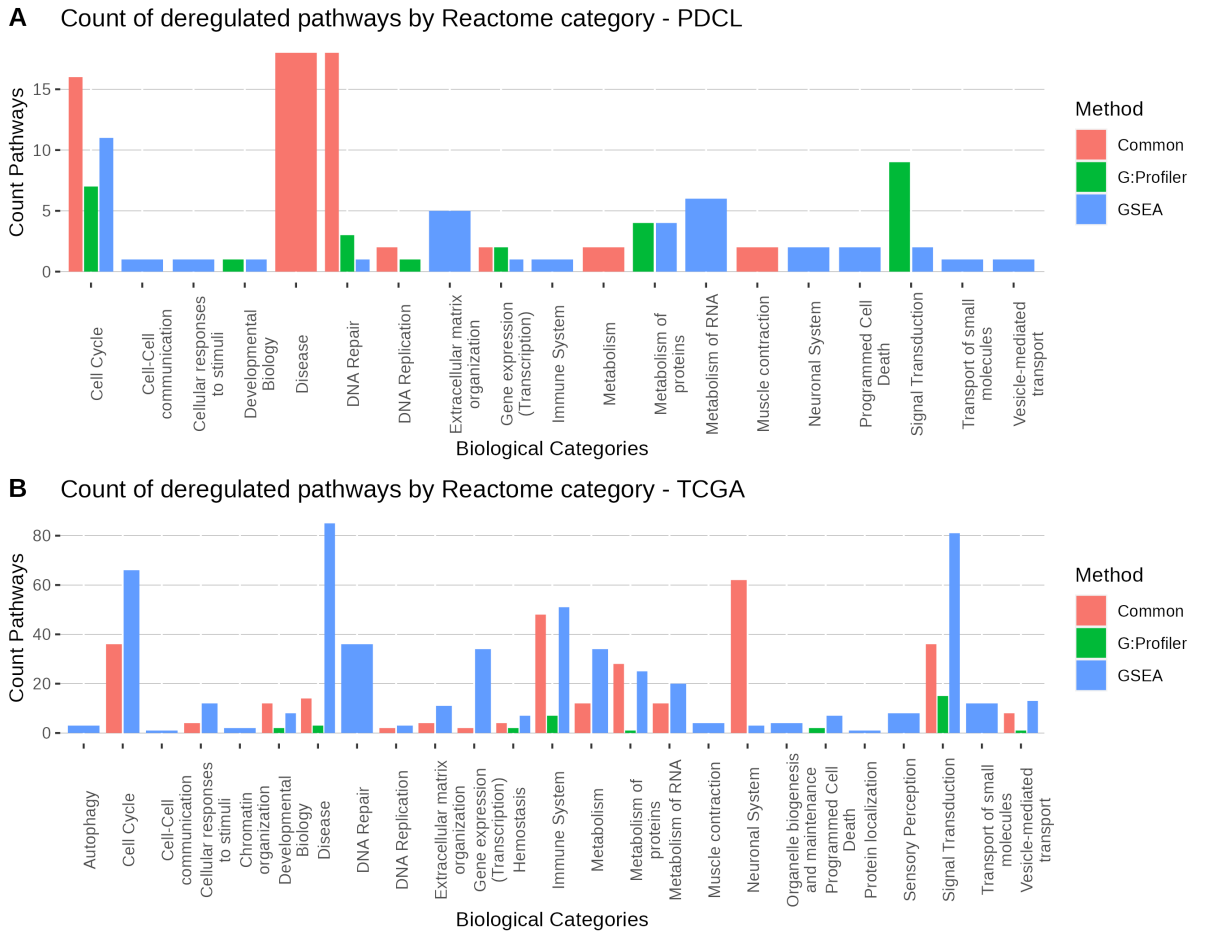

Figure 4.12: Barplot of the count of significantly deregulated pathways with Reactome pathways database using (a) PDCL and (b) TCGA data. Pathways are coloured whether they pass the FDR threshold only in G:Profiler (green), only in GSEA (blue) or in both tool (red).

represented categories. Compared to PDCL, all samples in TCGA have at least one pathway found enriched. In an attempt to understand why so many pathways linked to disease were present in the results, I investigated the contribution of a gene to a pathway (the number of times that a gene is associated with a pathway in the result). It demonstrated that few genes that were involved in disease pathway had a role in other biological processes like the *Cellular Processes* KEGG category or the *Signal transduction*/*Extracellular Matrix Organization* Reactome categories. Certain genes are even up- or down-regulated in 90 % of the samples in TCGA. The involvement of some genes in many functions could explain why disease pathways are over-represented in the results.

I also performed k-means clustering to understand how samples tend to group together based on their deregulated categories (figure 4.16). The number of altered pathways associated with one category in a sample was divided by the total number of pathways present in the GMT file to create a vector of values for each sample to use for the clustering. The k-means algorithm was run with three clusters. The number of clusters was chosen graphically with the correlation of each category to the principal component of the PCA on the percentage of deregulation of each category. It seems that clusters are better separated using KEGG categories than with Reactome where some overlap between clusters is visible. Filtering out unnecessary Reactome categories or merging them might improve clustering. Results with TCGA seem to group samples with dysregulated mechanisms involved in the *Metabolism*, the genetic processing like *DNA replication/repair* and processes involved in the interaction with the environment such as Extra-Cellular Matrix (ECM) *organisation* or *Environmental Information Processing*. A similar pattern can be observed with PDCL samples, although it seems that one cluster groups together samples with fewer deregulated pathways.

Now that we have seen the different categories affected in the samples of PDCL and TCGA, we want to go deeper and see what processes are often altered in our datasets to select which should be included

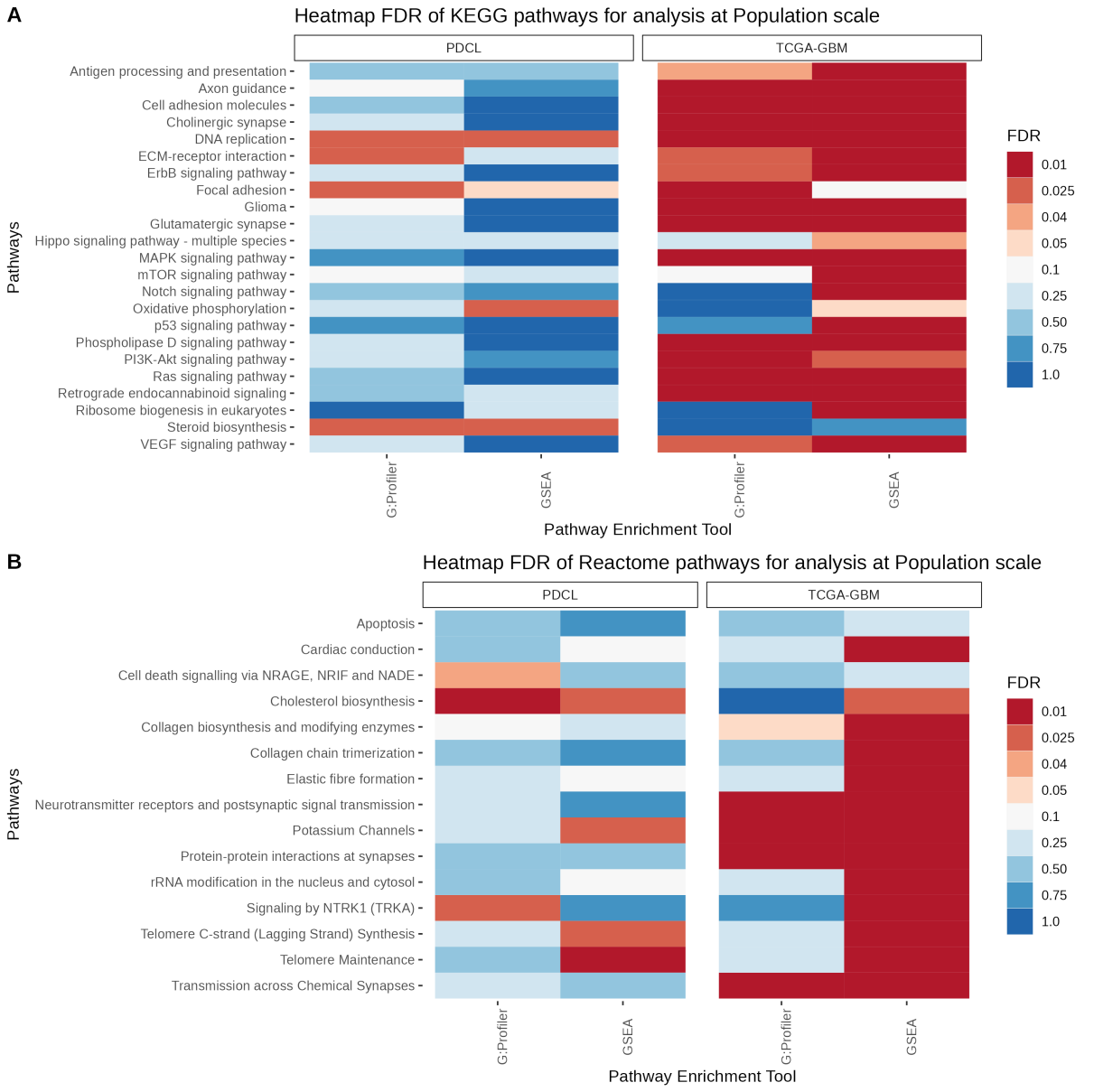

Figure 4.13: Heatmap of the FDR of a few deregulated pathways in PDCL and TCGA with (A) KEGG pathways and (B) Reactome pathways. FDR values below 0.05 are considered significant (red), FDR values above this threshold are not (blue). The stronger the shade of red, the more significant the result is. Inversely for blue.

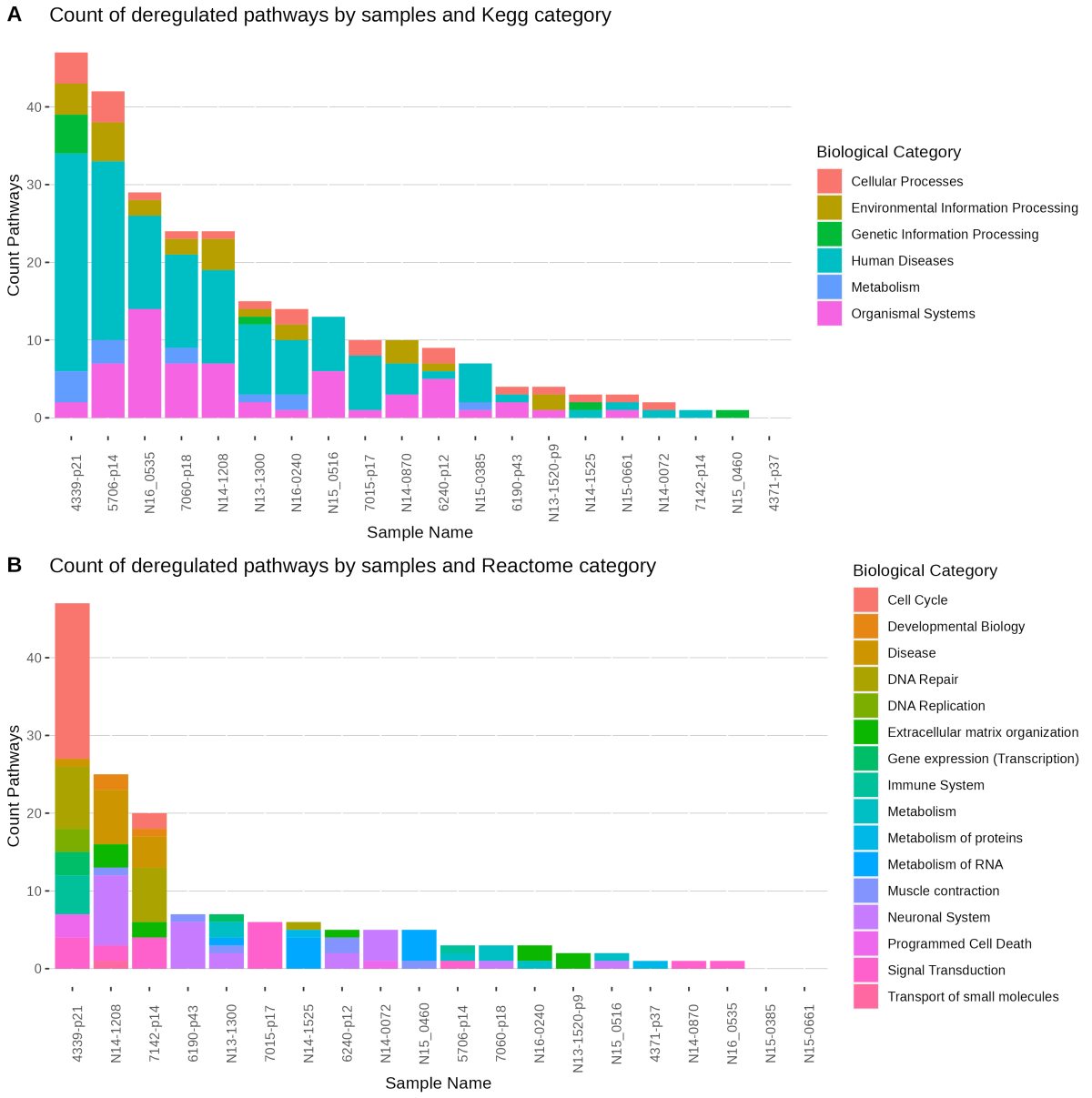

Figure 4.14: Number of deregulated pathways per category and sample in the PDCL dataset using (A) KEGG and (B) Reactome pathways. The *Human Diseases* and *Organismal Systems* categories are the most frequently deregulated KEGG categories. In Reactome, deregulated pathways are often associated with *Neuronal System*, *DNA Repair/Replication* and *Signal Transduction*. *4339-p21* is the sample with the most deregulated pathways in both KEGG and reactome. Pathways in this sample are associated with the *Cellular Processes* and *Environmental/Genetic Information Processing* KEGG, or the *Cell-Cycle* Reactome category.

in the model. For reading comfort, tables 4.3 and 4.4 only show a subset of the most often deregulated pathways I am presenting in the next paragraph. The complete list of the top 15 deregulated pathways for each dataset and pathway enrichment tools can be found in the appendix.
With the PDCL datasets (table 4.3), the most frequently deregulated pathways are the *Focal Adhesion* (13 samples) for KEGG and the *Transimission across Chemical Synapses* for Reactome.

With TCGA (table 4.4), the most frequently deregulated pathways are the *Cell-Cycle* for KEGG and the *Collagen Formation* for Reactome. We see that among the frequently deregulated pathways in PDCL and TCGA, several of them are signal pathways that have been documented to be deregulated in cancer. With PDCL, these include the *Hippo Signaling pathway* and the *Retrograde endocannabinoid signalling*, deregulated in 6 and 5 samples, respectively [133–135]. Surprisingly, these two pathways were not found enriched during the population analysis of PDCL. We saw that *p53 signalling*, a famous deregulated pathway in glioblastoma, was only found deregulated with GSEA in the TCGA dataset. While this pathway

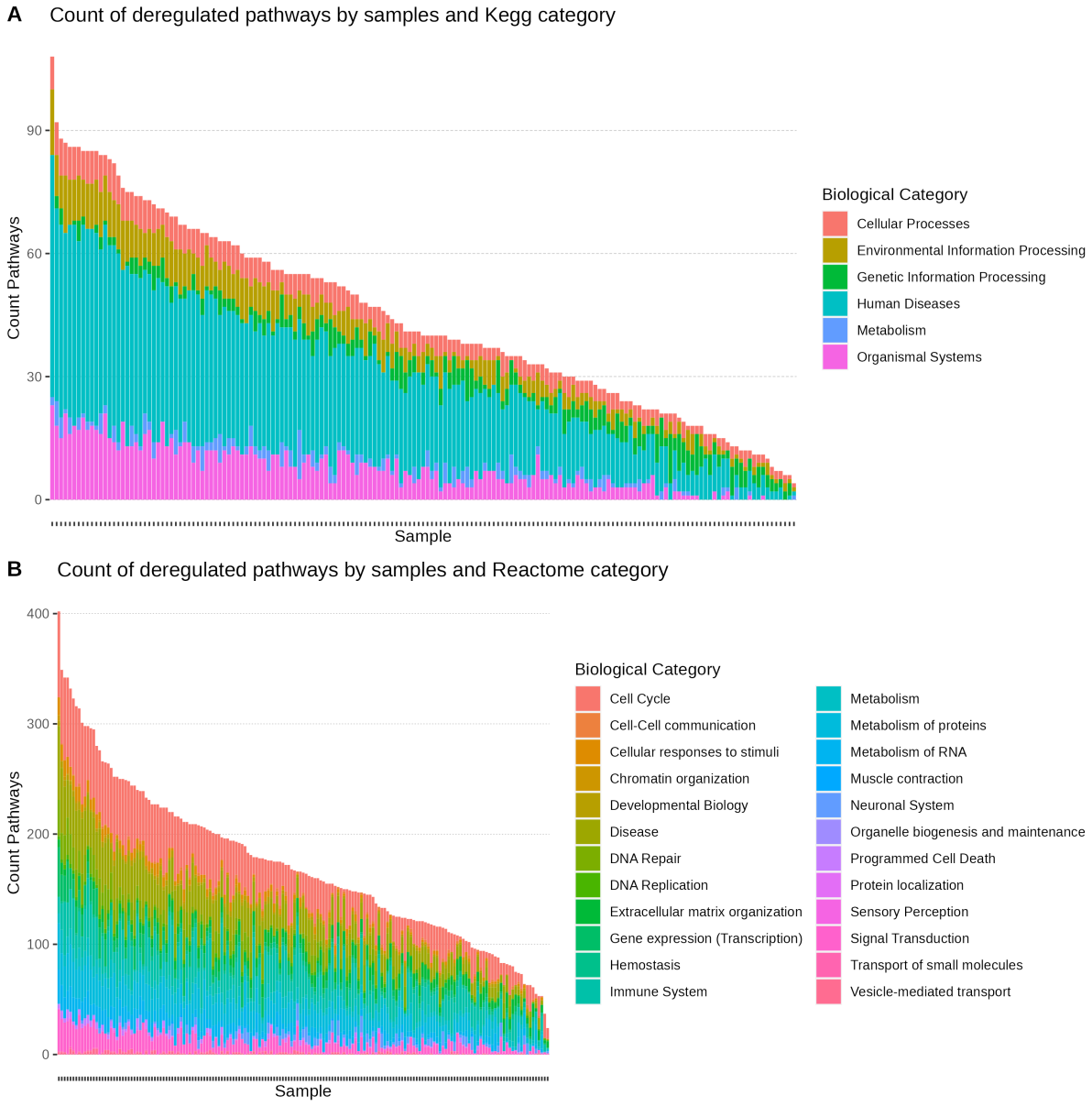

Figure 4.15: Number of deregulated pathways per category and per sample in the TCGA dataset using (A) KEGG and (B) Reactome pathways. Sample names are not displayed to avoid overploting. Like the result with PDCL samples, the *Disease* is among the most deregulated categories in both KEGG and Reactome. While with PDCL, Reactome pathways associated with the Cell-Cycle were predominantly deregulated in one sample, in TCGA the Cell-Cycle was greatly affected among all the samples.

is not present in the results with PDCL data at the individual level, results with TCGA data at the individual scale shows it is frequently deregulated (118 samples). Most of the pathways involved in the *Cell-Cycle* were deregulated in one PDCL sample, hence it is not surprising that they are not frequently deregulated in this dataset. At the opposite, the cell cycle was greatly affected in almost all the samples of TCGA. Taking a closer look at these, the results suggest that the parts of the cell cycle mainly affected are the *G1/S Phase transtion* and the *Cell-Cycle checkpoints*.

While the main focus of study in this thesis is the metabolism of cells, a few metabolic pathways came up in the top most frequently deregulated pathways in both PDCL and TCGA data. Among the few that do emerge, we see that the *Oxidative phosphorylation* from KEGG is deregulated in four PDCL samples but does not show up in the top frequently deregulated pathways with TCGA. That pathway was enriched by GSEA only for PDCL and TCGA in the population analysis. Still with PDCL data, the *Steroid biosynthesis* (from KEGG) and the *Cholesterol biosynthesis* (from Reactome) are deregulated in 4 and 3 samples respectively. This result is specific to PDCL data and was also enriched by GSEA and

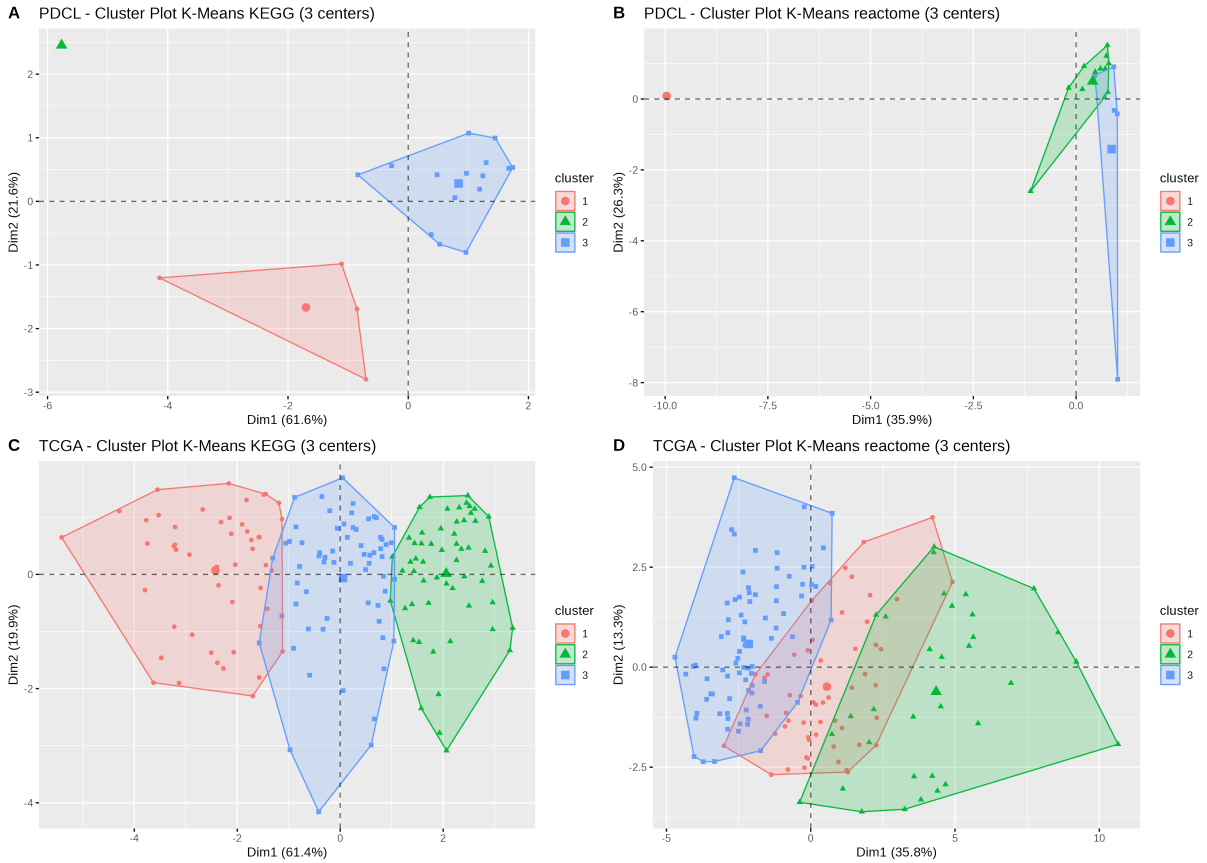

Figure 4.16: Clustering of (A-B) PDCL samples and (C-D) TCGA samples using the percentage of deregulation of each category. Results are shown with KEGG (A,C) and Reactome (B,D) categories. In A: cluster 1 contains samples that are associated with the *Organismal Systems/Environmental Information processing/Human Diseases*, cluster 2 contains samples associated *Metabolism/Cellular Processes/Genetic Information Processing*, and cluster 3 contains samples with few deregulated pathways. In B: cluster 1 contains samples associated with *Cell-Cycle/DNA replication and repair/Gene expression/Signal Transduction/Immune System*, cluster 2 is associated with *Metabolism*, and cluster 3 is associated with ECM*/Neuronal System/Disease/Transport of small molecules*. In C: cluster 1 contains samples associated with *Environmental Information Processing/Organismal Systems/Human Diseases/Cellular Processes*, cluster 2 is associated with *Metabolism*, and cluster 3 is associated with *Genetic Information Processing*. In D: cluster 1 contains samples associated with *Immune System/Programmed Cell Death/Metabolism/Signal Transduction/Cellular Response to Stimuli/Vesicle-mediated transport*, cluster 2 is associated with *Cell-Cycle/DNA replication/DNA repair/Chromatin organization/Gene Expression (Transcription)*, and the cluster 3 is associated with *Extra-Cellular Matrix Organisation/Cell-Cell communication/Hemostasis*.

G:Profiler in the population analysis. Controls in the PDCL dataset are composed of astrocyte cell lines. The brain is one of the most cholesterol-rich organs with 20 % of the total cholesterol body. Because cholesterol cannot pass through the Blood-Brain Barrier (BBB), it is synthesized *de novo* in the brain mainly by astrocyte cells [136]. We can hypothesize that cholesterol is only observed in the result of PDCL as they are single-cells compared with astrocytes while TCGA data are bulk-tumour. Bulk tumour data may hide the deregulation of this pathway as different cell lines are included in the result, showing more of a general trend rather the individual deregulation.

Another process that catches the eye in tables 4.3 and 4.4, is the one linked to interaction with the ECM. Indeed, we can see that pathways such as *Focal adhesion* and ECM*-receptor interaction* from KEGG, or the *Collagen biosynthesis and modifying enzymes* from Reactome are frequently deregulated in both PDCL and TCGA. Other pathways related to interaction with ECM such as *Collagen formation* from Reactome are frequently deregulated as well but in either dataset. It has been observed that the ECM collagen content in glioblastoma differs from the one in healthy tissue. As said in chapter 1, the matrix in GBM is stiffer. GBM are aggressive tumours with increased invasion that is mediated by interaction with the ECM.

Pathway ID	Description	Category	Number of samples
KEGG			
path:hsa04510	Focal adhesion	Cellular Processes	13
path:hsa04512	ECM-receptor interaction	Environmental Information Processing	5
path:hsa00100	Steroid biosynthesis	Metabolism	4
path:hsa00190	Oxidative phosphorylation	Metabolism	4
path:hsa04723	Retrograde endocannabinoid signaling	Organismal Systems	4
path:hsa04392	Hippo signaling pathway - multiple species	Environmental Information Processing	3
path:hsa04514	Cell adhesion molecules	Environmental Information Processing	3
Reactome			
R-HSA-112315	Transmission across Chemical Synapses	Neuronal System	5
R-HSA-1650814	Collagen biosynthesis and modifying enzymes	Extracellular matrix organization	4
R-HSA-8948216	Collagen chain trimerization	Extracellular matrix organization	4
R-HSA-191273	Cholesterol biosynthesis	Metabolism	3
R-HSA-109581	Apoptosis	Programmed Cell Death	2
R-HSA-1566948	Elastic fibre formation	Extracellular matrix organization	2
R-HSA-157579	Telomere Maintenance	Cell Cycle	2
R-HSA-174417	Telomere C-strand (Lagging Strand) Synthesis	Cell Cycle	2

Table 4.3: Table of the frequently deregulated pathways in PDCL.

4.3 Conclusions: New Insights and perspective for mathematical modelling

In the previous section, we explored the heterogeneity of different tumour samples using single-cell and bulk tumour RNA-Seq data. From the analysis at the population scale, we saw that the cholesterol and collagen metabolism pathways were deregulated in at least one dataset or pathway enrichment tool. Nonetheless, the analysis at the individual scale showed that these pathways were among the most frequently deregulated as well. A review of the literature on these two mechanisms has demonstrated their veracity. I decided to select these as candidates for mathematical modelling as both were documented to facilitate tumour growth or confer cells an invasive phenotype.

Numerous studies have documented the relevance of targeting cholesterol metabolism in glioblastoma therapy, hence this pathway seemed interesting to implement for various reasons. First, studies have shown that targeting components of cholesterol metabolism such as synthesis, uptake or trafficking can induce cell death, autophagy or cell cycle arrest in the G2/M phase [136–140]. Second, this pathway could be used in selective therapy as it was observed that glioma cells were sensitive to cholesterol synthesis inhibition while astrocytes were not [141]. Third, targeting this pathway could improve temozolomide (TMZ) therapy as it was shown that cholesterol could modulate TMZ efficiency through different mechanisms [142, 143]. Lastly, the frequent upregulation of EGFR in glioblastoma could impact cholesterol homeostasis resulting in GBM cell proliferation and invasion. If at first glance this process is relevant due to the reasons cited above, no evidence could link it to the HIF protein and the response to hypoxia.

In contrast, the interaction between the cell and the ECM was as much documented in glioblastoma for its effect on migration and also showed a direct link with the HIF protein. The *Collagen formation* Reactome pathway is deregulated in all the TCGA samples. Angiogenesis is reduced by inhibition of collagen cross-linking, resulting in reduced GBM progression without increasing cell invasion, hypoxia or

Pathway ID	Description	Category	Number of samples
KEGG			
path:hsa04110	Cell Cycle	Cellular Processes	153
path:hsa04512	ECM-receptor interaction	Environmental Information Processing	147
path:hsa04510	Focal adhesion	Cellular Processes	129
path:hsa03030	DNA replication	Genetic Information Processing	126
path:hsa04514	Cell adhesion molecules	Environmental Information Processing	124
path:hsa04115	p53 signaling pathway	Cellular Processes	118
Reactome			
R-HSA-1474290	Collagen formation	Extracellular matrix organization	159
R-HSA-453279	Mitotic G1 phase and G1/S transition	Cell Cycle	154
R-HSA-1650814	Collagen biosynthesis and modifying enzymes	Extracellular matrix organization	152
R-HSA-69206	G1/S Transition	Cell Cycle	151
R-HSA-2022090	Assembly of collagen fibrils and other multimeric structures	Extracellular matrix organization	149
R-HSA-69620	Cell Cycle Checkpoints	Cell Cycle	149

Table 4.4: Table of the frequently deregulated pathways in TCGA.

necrosis in mice tumours compared to VEGF targeting drugs. As said in Chapter 1, HIF can upregulate genes that are associated with collagen post-translational modification to increase matrix stiffness, collagen fibre stability and matrix-degradation to improve cell migration. Collagen-associated genes are frequently deregulated in either the PDCL and TCGA datasets (figure 4.17). Most of the genes are upregulated in a large number of TCGA samples. They are greatly deregulated in PDCL samples as well, although some deregulation does not seem to be coherent with the literature. For example, the genes COL1A1 and COL3A1 are frequently downregulated while we would expect the opposite. MMPs genes are the most deregulated genes among the ones presented in figure 4.17, being upregulated in almost all the TCGA samples. It may be due to the control used or the condition of culture. To talk strictly about HIF1-α , the subunit is upregulated in 5 PDCL and 150 TCGA samples, showing that HIF is greatly affected. Regarding metabolic genes, LDHA is deregulated (almost only upregulated) in all the PDCL samples and more than 100 TCGA samples. Taking the literature and the results together, collagen synthesis looks promising and will be selected for modelling in the current HIF-metabolic model to investigate how the metabolism interacts with cell migration.

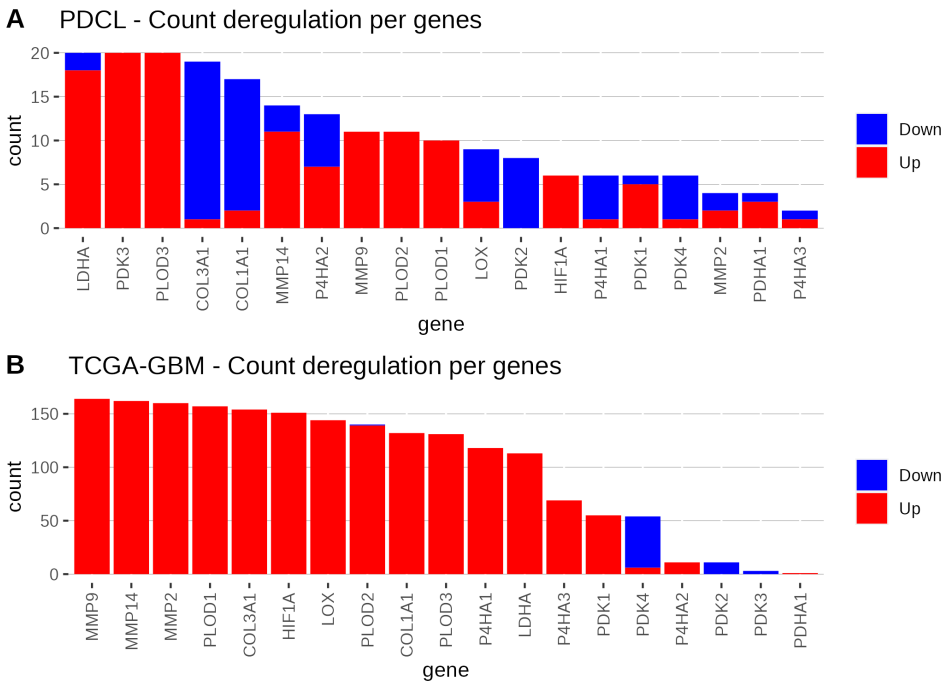

Figure 4.17: Count of the number of samples a gene is found upregulated (red) or downregulated (blue). The genes displayed are involved in collagen metabolism: MMPs are involved in collagen degradation, P4Hs and PLODs are involved in collagen post-translational modification and LOX catalyze collagen fibre cross-linking. Metabolic genes like LDH, PDK, PDH and HIF are also displayed.

5 Modelling the impact of HIF on cell Migration and Invasion

In chapter 3, I described a mathematical model of the impact of the Hypoxia Inducible Factor (HIF) on the cellular metabolism and the consequence for the Warburg Effect. We have seen that constant variations of environmental oxygen and reduced HIF degradation by oxygen-dependent mechanisms can induce a Warburg effect, i. e.increased lactate production in the presence of oxygen, in the model. In chapter 4, I described a statistical analysis of transcriptomic data to assess the deregulated mechanism in the case of glioblastoma (GBM). We found that collagen metabolism is frequently deregulated in the two datasets we used with a few key genes involved in collagen post-translational modification being regulated by HIF.

In this chapter, I will present how the mechanisms identified in chapter 4 are integrated to the model developed in chapter 3. The structure of the Extra-Cellular Matrix (ECM), the nature of collagen fibres, their formation and the cellular migration were covered in chapter 1. Hence, this chapter will focus on the explanation of the biological processes related to the impact of HIF on the collagen content (concentration and state). I will recapitulate: what genes are regulated by HIF, how they interact, what are the consequences on the extracellular collagen and the cell migration speed. Furthermore, I will detail how these processes were turned into rules in the existing model and how they were implemented. Lastly, we will analyze the simulation results and see how these processes affect the growth of the tumour.

5.1 State of Extra-Cellular Matrix modelling

Metastasis is one of the hallmarks of cancer. It is widely seen as an obstacle to therapy and leads to tumour spread and recurrence. Cellular migration is a complex process involving many changes such as loss of cell-cell interaction and adhesion, acquisition of a motile phenotype and increased cell-matrix interactions. It is not surprising that interactions between the cell and the ECM have been the topic of many mathematical models in the hope of better understanding cellular migration and its implication in cancer development. Thus, different aspects of cellular migration have been studied such as cell-cell adhesion [144, 145].

Regarding collagen, its role in the ECM structure makes it a valuable target for mathematical description. Collagen has been described as a substrate that does not diffuse nor decay to understand its impact on the cell speed and the migration distance [54]. Cell-fibre interactions were implemented in an individual-based model to understand how fibres dictate the migrating behaviour of the cell (for example, in which area does cell preferentially migrates) [77]. In that case, fibres are modelled as an individual structure and interact with cells in a force-based approach: contact between a cell and a fibre generates traction forces that allow the cell to move and rearrange the matrix fibres. This allows the study of the impact of the number of fibres (density) and their alignment on cell migration at a smaller scale. A module was created for PhysiCell to represent fibres as rod-shaped agents in the release 1.13 (`https://github.com/MathCancer/PhysiCell/releases/tag/1.13.1`). Models in the literature further studied cell-collagen interactions by including matrix degradation mechanisms to show how they favours invasion [67, 92, 145]. Matrix degradation was even depicted in a model that does not describe cell invasion but instead the supply of MT1-MMP [146].

It became apparent that the mathematical community has shown some interest in defining the interaction between the cell and the matrix to clarify cell invasion. Collagen can be included in various way: as a substrate, with fibres represented as individual and through protease-mediated matrix degradation. Yet, as explained in chapter 2, the major part of HIF modelling focuses on a specific topic (interaction between HIF and the cell cycle, transcriptional activity of HIF). Consequently, the role of HIF on the collagen content and the consequences for cellular invasion has not been documented. The influence of collagen fibre cross-linking and fibre degradation on cellular migration have been highlighted in the literature. Nonetheless, it seems that no model implemented this mechanism. Another shady area concerns the upregulation by HIF of genes that promote collagen stability, collagen cross-linking (in other words collagen resistance to degradation) and collagen degradation. It would be expected that hypoxia promotes either an increased stability/density of collagen or its degradation but not both at the same time.

Taken together, the promotion of both collagen cross-linking and collagen degradation by HIF leads to the question: Why does HIF promote collagen stability and degradation? What is the interplay between these two mechanisms? How does it impact the cell migration speed? The model of metabolism described in chapter 3 will then be adapted to consider these issues of cell invasion.

5.2 The genetic regulations involved in collagen content: interactions with HIF

We have seen in chapter 1 that cancer cell invasion is a process involving modification of the structure of the ECM. These modifications are made by genes related to the processing of collagen (post-translational modification, cross-linking and degradation). An increased number of invasive cells is observed in hypoxia which is now widely seen as a factor favouring cell migration. It is expected that hypoxia-activated genes include genes linked to collagen modification. Three families of genes involved in collagen modification and upregulated by hypoxia are: prolyl 4-hydroxylase (P4H), matrix metalloproteinase (MMP) and lysyl oxidase (LOX). One candidate per family has been selected for modelling: P4HA1, MT1-MMP (also called MMP14) and LOX [53, 57, 66, 147].

Both P4HA1 and P4HA2 isoforms are regulated by HIF1-α [8, 17, 59, 66, 148, 149] and no noticeable difference was documented, thus only P4HA1 is modelled. The MMP family is large with several members being regulated by HIF. Among them, the frequently documented proteases are: MMP2 [148, 150], MMP9 [53] and MT1-MMP [151–154]. While MMP2 and MMP9 were found to be oncogenes [70, 74, 155], only the MT1-MMP will be included in the model, a choice motivated by several reasons. First, The activity of several MMPs is compensatory and redundant. It appears that modelling different types of MMPs would then increase the complexity of the model by introducing redundancy without an actual gain. Second, MMP2 activation necessitates both TIMP2 and MT1-MMP [73, 146, 156], therefore it is reasonable to suppose that MT1-MMP is the limiting factor to MMP2 activation. Third, it has been shown in epithelial cells and fibroblast that only MT1-MMP could promote cellular invasiveness whereas soluble MMPs (MMP1, MMP2, MMP8 and MMP13) failed to do so [73]. Finally, there is some ambiguity regarding the collagenolytic activity of MMP2. It is a gelatinase that may be important for the clearance of degraded and denatured collagen that has been partially cleaved by MT1-MMP [69, 71]. Although HIF upregulates three members of the LOX family, it seems that the cross-linking activity is most often attributed to the LOX enzymes. Consequently, LOX is the last enzyme added.
Their integration within the model will follow the same mathematical formalism explained earlier.

Reminder: mathematical model of protein regulations from chapter 3

$$\frac{dh}{dt} = \alpha_h\ \beta_h - d_h\ S(O, s_{O\to h}, \gamma_{O\to h})\ h,$$
$$\frac{dl}{dt} = \alpha_l\ S(h, s_{h\to l}, \gamma_{h\to l}) - d_l\ l,$$
$$\frac{dk}{dt} = \alpha_k\ S(h, s_{h\to k}, \gamma_{h\to k}) - d_k\ k,$$
$$\frac{dq}{dt} = \alpha_q\ S(k, s_{k\to q}, \gamma_{k\to q}) - d_q\ q,$$

with h, l, k and q the levels of HIF, LDH, PDK and PDH. $S(Y, s, \gamma)$ is the Shifted-Hill function, a non-linear term to account for genetic regulations:

$$S(Y, s, \gamma) = \frac{s^n}{s^n + Y^n} + \gamma\frac{Y^n}{s^n + Y^n}.$$

The parameters were already described in section 3.1.1 and values given in table 5.1.

Besides HIF-mediated upregulations, different mechanims are involved in the regulation of MT1-MMP quantity. Like every transmembrane MMP, MT1-MMP is inhibited by TIMP-2, TIMP-3 and TIMP-4 but not by TIMP-1 [73]. At high concentrations, the pericellular collagen can induce the expression of MT1-MMP by interacting with receptors on the cell [71]. MT1-MMP can cleave itself to release its active site, this process called ectodomain shedding constitutes a major mechanism of downregulation for membrane-bound proteins [73, 156]. As we are interested in the effect HIF can have upon the matrix degradation, these processes were not implemented to only reflect the impact of HIF. Equations for the new variables are of the form:

$$\frac{dp}{dt} = \alpha_p \; S(h, s_{h\to p}, \gamma_{h\to p}) - d_p \; p, \tag{5.1}$$
$$\frac{dm}{dt} = \alpha_m \; S(h, s_{h\to m}, \gamma_{h\to m}) - d_m \; m, \tag{5.2}$$
$$\frac{dr}{dt} = \alpha_r \; S(h, s_{h\to r}, \gamma_{h\to r}) - d_r \; r, \tag{5.3}$$

where p, m and r describe the levels of P4HA1, MT1-MMP and LOX.

The Hill coefficient n, production rate α and degradation rate d used were the same as in chapter 3. Yet P4HA1, MT1-MMP and LOX were not originally implemented in the model from Li et *al* in 2020 [109]. Their parameters for the Shifted-Hill function have been fitted using the transcriptomic data of glioblastoma from the The Cancer Genome Atlas (TCGA) database screened during the statistical analysis in chapter 4. Parameters of the genes previously included in the model were also re-defined to conserve the same dynamic as in the data. Parameters were fitted following the procedure:

1. Transcriptomic data were normalized using the pseudo-count method from the DESeq2 package ($log(count + 1)$).
2. Normalized values were interpolated with the formulae : $\hat{x} = \frac{x - x_{min}}{x_{max} - x_{min}}$ to obtain values between 0 and 1.
3. The Shifted-Hill function is the summation of two Hill terms: an activating term and an inhibiting term. When $\gamma > 1$, only the active term is effective, when $\gamma < 1$ only the inhibiting term is effective. Therefore, depending on whether $\gamma > 1$ or $\gamma < 1$ we fit the parameter s on one or the other term:

$$h(Y) = \frac{Y^n}{s^n + Y^n} \quad \text{if the gene } Y \text{ upregulate the gene } X,$$
$$h(Y) = \frac{s^n}{s^n + Y^n} \quad \text{if the gene } Y \text{ downregulate the gene } X,$$

   where Y is the regulating gene, n is the Hill coefficient, s is the concentration at which the effect of the regulating gene Y on the regulated gene X is half-maximal. The Hill coefficient is fixed to 4, the value used in the precedent chapter.
4. The parameter s estimated is then rescaled to the dimensions of the data before interpolation (pseudo-counts).
5. The parameter γ is estimated by measuring the fold-change of expression:

$$\text{fold-change} = \frac{x_{max}}{x_{min}} \quad \text{if the gene } X \text{ is upregulated,}$$
$$\text{fold-change} = \frac{x_{min}}{x_{max}} \quad \text{if the gene } X \text{ is downregulated.}$$

The parameter $s_{O\to h}$ for the Shifted-Hill equation describing the increase of HIF degradation by oxygen was not modified. The parameter $\gamma_{O\to h}$ describing the strength of HIF degradation by oxygen-dependent mechanisms was estimated like the γ parameters of the other protein regulation described in the procedure above. Initial conditions for gene levels associated with HIF, LDH, PDK, PDH, P4HA1, MT1-MMP and LOX were set to 1. The list of parameters is summarised in table 5.1.

Parameter	Value	Dimension	Parameter	Value	Dimension
α_h	0.005	min^{-1}	d_h	0.005	min^{-1}
α_l	0.005	min^{-1}	d_l	0.005	min^{-1}
α_k	0.005	min^{-1}	d_k	0.005	min^{-1}
α_q	0.005	min^{-1}	d_q	0.005	min^{-1}
α_p	0.005	min^{-1}	d_p	0.005	min^{-1}
α_m	0.005	min^{-1}	d_m	0.005	min^{-1}
α_r	0.005	min^{-1}	d_r	0.005	min^{-1}
$s_{O\to h}$	0.02085	mmol/L	$s_{h\to l}$	13.04	-
$s_{h\to k}$	13.34	-	$s_{k\to q}$	9.94	-
$s_{h\to p}$	13.68	-	$s_{k\to m}$	13.25	-
$s_{h\to r}$	13.44	-			
$\gamma_{O\to h}$	25.01	-	$\gamma_{h\to l}$	16.06	-
$\gamma_{h\to k}$	17.87	-	$\gamma_{k\to q}$	0.78	-
$\gamma_{h\to p}$	16.87	-	$\gamma_{k\to m}$	20.43	-
$\gamma_{h\to r}$	25.46	-			
β_h	25.01	-	n	4	-

Table 5.1: Parameters for genetic regulations in the model of the impact of HIF on the cellular invasion. The symbol "-" stands for dimensionless. Table cells coloured indicate either a new parameter or a parameter that has changed.

5.3 Integration of the collagen in the Extra-Cellular Matrix

Reminder: mathematical model of metabolism from chapter 3

Oxygen (O) consumption:

$$f_O = \Psi_O V_O \frac{O}{K_O + O},$$
$$\Psi_O = \frac{\Phi_O - \phi_O}{1 + \exp\left(-\lambda_q (q - q_0)\right)} + \phi_O.$$

Glucose (G) consumption:

$$f_G = \left(\frac{\Psi_G A_0}{2} - \frac{29 f_O}{10}\right) \frac{G}{K_G + G},$$
$$\Psi_G = \frac{\Phi_G - \phi_G}{1 + \exp\left(-\lambda_l (l - l_0)\right)} + \phi_G.$$

ATP (A) **production**:

$$f_A = 2 f_G + \frac{29}{5} f_O.$$

H⁺ (H) secretion:

$$f_H = 2 f_G \, k_H.$$

Extra-cellular quantities:

$$\frac{\partial O}{\partial t} = D_O \nabla^2 O - \sum_N^{i=1} f_O^i,$$
$$\frac{\partial G}{\partial t} = D_G \nabla^2 G - \sum_N^{i=1} f_G^i,$$
$$\frac{\partial H}{\partial t} = D_H \nabla^2 H + \sum_N^{i=1} f_H^i.$$

Generation of ATP **at each timestep**:

$$\frac{dA}{dt} = f_A.$$

PhysiCell itself does not include the ECM but rather substances that diffuse in the environment and can interact with cells (uptake, secretion, chemotaxis). A paper from Gonçalves et *al* in 2021 [54]

used PhysiCell to study the impact of collagen density on cell migration, they replicated results from cellular migration experiments in collagen gels. They include the density of collagen as a simple substrate that does not diffuse nor decay. It has been decided to follow the same approach as a more complex description of collagen density is not relevant to the study.

The aim is not to characterize deeply the interaction between the cells and the collagen fibers but rather to observe the changes in collagen content depending on the HIF quantity. Modelling collagen as a mere substrate allows the model to remain simple without compromising the flexibility to permit two states of collagen: normal or cross-linked. Cross-linked collagen increases matrix stiffness but most importantly it is resistant to degradation by MMP [69, 71]. This can lead to changes of migration speed for the cell, therefore cross-linking should be implemented in the model.

Procollagen α chains are coded by the COL genes (COL1A1, COL1A3, etc) and undergo post-translational modification by P4Hs genes to promote the formation of the collagen fibres. Here, it is hypothesized that collagen that has not been processed by P4HA1 is unstable and does not participate in the collagen content. Thus, collagen secretion by the cell will be directly driven by the level of P4HA1 in a linear relation. Similarly, collagen degradation and collagen cross-linking are influenced by the level of MT1-MMP and LOX with a linear relation. Collagen is either added by the effect of the gene P4HA1 or removed due to the degradation by MT1-MMP or the cross-linking by LOX. Cross-linked collagen is resistant to protease degradation, thus it can only be added to the environment by the effect of LOX. This relationship is graphically represented in figure 5.1.

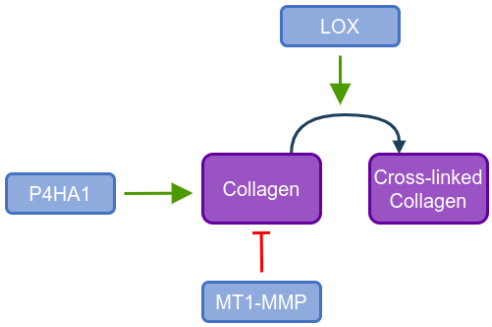

Figure 5.1: Secretion, degradation and cross-linking of collagen mediated by P4H1A, MT1-MMP and LOX.

The addition of collagen will follow a simple linear dynamic that depends on a base secretion rate and the level of P4HA1 gene. However, compared to the secretion of collagen that solely depends on genetic parameters and does not need a substrate (in the model), degradation and cross-linking processes require collagen as a substrate. The rate of degradation/cross-linking then should (1) increase as the level of the related enzyme increases, and (2) decrease to become null when collagen density falls. Consequently, like for oxygen and glucose consumption, degradation/cross-linking of collagen will be modelled using Michaelis-Menten. The Michaelis-Menten dynamic was used previously in the literature to describe MT1-MMP mediated degradation [146]. The resulting function to describe collagen fate is:

$$f_C = \beta_p p - \beta_m m \frac{C}{K_m + C} - \beta_r r \frac{C}{K_r + C}, \tag{5.4}$$

$$f_{C_r} = \beta_r r \frac{C}{K_r + C}. \tag{5.5}$$

Here f_C describes the density of collagen C, and f_{C_r} cross-linked collagen C_r. The variables p, m and r represent the level of P4HA1, MT1-MMP and LOX. The parameter β_p is the base rate of collagen secretion, β_m is the rate of degradation by m and β_r is the rate of cross-linking by r. K_m is the level of m at which the degradation rate of collagen is half-maximal, and K_r is the level of r at which the cross-linking rate is half-maximal. Compared to functions 3.14 and 3.16, there is no term to tune the effect of the p, m and r variables. These two terms aimed to ensure that the effect of LDH and PDH regulations would stay in the boundary defined by the model: oxygen consumption cannot increase more than the maximum V_o; the increased reliance of cells on the glycolytic phenotype would respect the interval defined by Robertson-Tessi [95]. Here, this rule is not needed as the Michaelis-Menten term prevents degradation of collagen when the density is null and there is no need to impose a maximum boundary on the rates.

While degradation of the ECM is essential for cellular migration, ECM acts as an important scaffolding for migration, hence its degradation must be localized [73]. MT1-MMP is a membrane-bound protein usually located on the invapodia or lamellipodia: the migration front of the cell [73, 146]. Consequently, the distance of MT1-MMP mediated degradation should be limited by the cell in the model. The cell will only degrade the collagen of the voxel it is currently inside. For simplicity of implementation, LOX-mediated cross-linking will follow the same rule. PhysiCell does not have built-in interaction between substances, which means that it does not describe any interactions of the type "the **Substance A** influences the concentration of a **Susbtance B**". Hence, interactions between extracellular LOX and collagen would need to be implemented manually. The unit to quantify the density of collagen and cross-linked collagen is mg/mL, the unit the most often reported when conducting gel experiments. The extracellular density of collagen is described by the partial differential equations (PDEs):

$$\frac{\partial C}{\partial t} = \sum_{N}^{i=1} f_C^i, \tag{5.6}$$

$$\frac{\partial C_r}{\partial t} = \sum_{N}^{i=1} f_{C_r}^i. \tag{5.7}$$

Simulations were run with the following initial conditions:

- Oxygen (equation 3.20): $O(x, y, t_0) = 0.056$ mmol/L (normoxia or 5 % O_2) [16].
- Glucose (equation 3.21): $G(x, y, t_0) = 5.0$ mmol/L (serum glucose) [91].
- H^+ (equation 3.22): $H^+(x, y, t_0) = 3.98 \times 10^{-5}$ mmol/L (pH 7.4).
- Collagen (equation 5.6): $C(x, y, t_0) = 2.5$ mg/mL (optimal collagen density for growth [157]).
- Cross-linked Collagen (equation 5.6): $C_r(x, y, t_0) = 0.0$ mg/mL.

And the following Dirichlet-Boundary conditions:

- Oxygen (equation 3.20): $O(x_0, y, t) = O(x_L, y, t) = O(x, y_0, t) = O(x, y_L, t) =$ 0.056 mmol/L.
- Glucose (equation 3.21): $G(x_0, y, t) = G(x_L, y, t) = G(x, y_0, t) = G(x, y_L, t) = 5.0$ mmol/L.
- H^+ (equation 3.22): $H^+(x_0, y, t) = H^+(x_L, y, t) = H^+(x, y_0, t) = H^+(x, y_L, t) = 3.98 \times 10^{-5}$ mmol/L (pH 7.4).
- Collagen and Cross-linked collagen: No boundary conditions as they do not diffuse.

The list of parameters for the model can be found in table 5.2.

5.4 Integration of the cell migration

It is now common knowledge that ECM collagen content is an important factor that influences both the ability of the cell to migrate and its speed [52]. Therefore, collagen density-dependent variations of the cell migration speed will be implemented in the model. As migration differs in 2D and 3D and depends not only on collagen quantity but on its properties as well, literature may be hard to apprehend at first. For example, 2D experiments indicate that matrix stiffness hinders cell invasion while 3D experiments suggest it facilitates it [55]. A study has shown that inhibition of LOX collagen cross-linking by βAPN inhibits metastasis in mouse models [53] while cells cultured for three days in non-crosslinked collagen exhibited a higher migration distance than cells in crosslinked hydrogel [62].

Laforgue et *al* [75] studied the effect of collagen concentration on the cell migration speed of three bladder cancer cell lines. They observed a constant decrease in the cell migration speed as collagen increased for concentrations between 0.95 and 4.5 mg/mL. However, their experiment was limited to 100 minutes to prevent the activation of MMPs. Collagen degradation is known to affect cancer cell migration, thus their experiment may not accurately represent cell migration on a longer timescale [75]. I first implemented the effect of collagen on the migration speed as Gonçalves et *al* [54]: increased collagen density leads to slower cell migration. Yet, based on the literature I decided to drop this implementation

Parameter	Value	Unit
Oxygen (3.14,3.15)		
V_O	0.01875	$mmol/L/min$
K_O	0.0075	$mmol/L$
Φ_O	1	-
ϕ_O	0	-
λ_q	15	-
q_0	0.575	-
Glucose (3.16,3.17)		
K_G	0.04	$mmol/L$
A_0	0.10875	$mmol/L/min$
Φ_G	50	-
ϕ_G	1	-
λ_l	4	-
l_0	8.03	-
H^+ (3.19)		
K_H	$2.5 \cdot 10^{-4}$	-
collagen (5.4 - 5.5)		
β_p	4.04×10^{-6}	$mg/mL/min$
β_m	2.16×10^{-5}	$mg/mL/min$
K_m	8.7×10^{-11}	mg/mL
β_r	2.91×10^{-10}	$mg/mL/min$
K_r	6.75×10^{-9}	mg/mL
Diffusion (3.20 - 3.22)		
D_O	$109,200$	$\mu m^2/min$
D_G	$30,000$	$\mu m^2/min$
D_{H+}	$27,0000$	$\mu m^2/min$

Table 5.2: Parameters for metabolism. The symbol "-" stands for dimensionless. Table cells coloured indicate either a new parameter or a parameter that has changed.

for another that seemed better suited to describe how fast the cell migrate. I will not describe the results obtained with this first implementation as only a few simulations were run with it.

Nonetheless, other observations documented in the literature have shown that the relation between collagen density and cell migrating capability is in fact a biphasic response. Schor et *al* [157] determined the effect of collagen on the migration speed of human fibroblasts and hamster melanoma cells using 3D gels composed of type I collagen extracted from rat tendons. Unlike Laforgue, they concluded that the effect of collagen was biphasic with a maximal cell migration speed reached at 2.0 mg/mL of collagen [157]. Lang et *al* [55] similarly observed a biphasic effect of collagen concentration on cellular migration although, in their experiment, maximal cell invasiveness was reached at 1.2 mg/mL. In addition, they investigated the effect of matrix stiffness and pore size using glutaraldehyde, a collagen cross-linking agent that does not change pore size. They showed that when collagen was cross-linked with glutaraldehyde, lower collagen concentrations and higher pore size enhanced cell invasion [55]. Therefore, the effect of collagen on cell migration speed seems to be a biphasic response dependent on collagen concentrations, stiffness induced by cross-linking and pore sizes. The results of their experiments are shown in figure 5.2. It seems sensible to hypothesize that the cell has to meet a "target collagen content" to achieve optimal migration and invasion. As said in section 1.4.3, the cell tightly regulates the collagen content as fibres provide a scaffold for the cell to migrate but also impose resisting forces. Consequently, the cell may create space for migration by degrading fibres but avoid total degradation through cross-linking to protect them. The increased stiffness might increase the traction forces generated when pulling the fibres. The

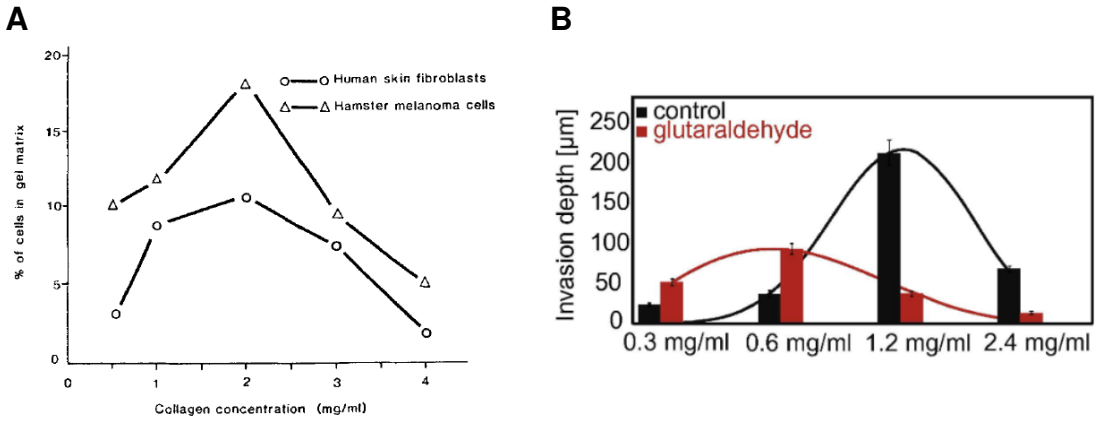

Figure 5.2: Results of the experiment from (A) Schor et *al* [157] showing the effect of collagen on human fibroblasts and hamster melanoma and (B) Lang et *al* [55] showing the biphasic effect of collagen concentration and cross-linking on MDA-MB 231 cells.

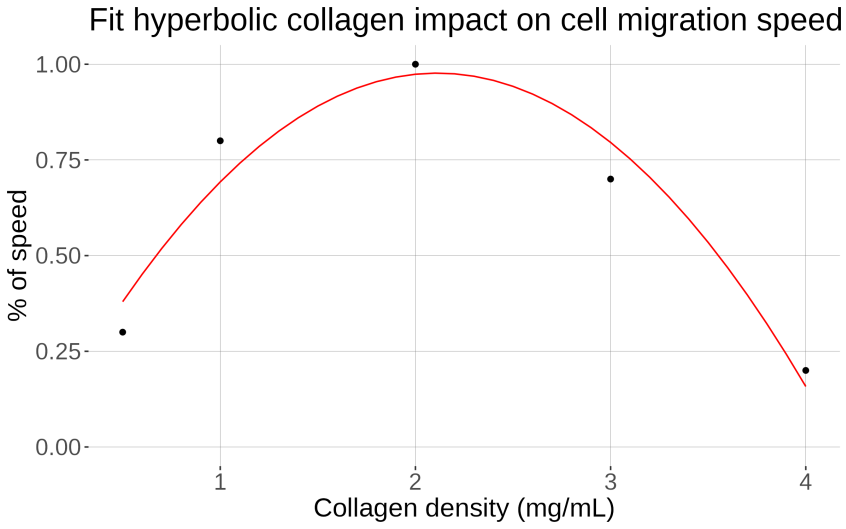

Figure 5.3: Result of the fit of a second-degree polynomial on the data from Schor et *al* [157].

final choice was to implement the migration speed of the cell as a parabolic function dependent on the collagen density.

Moreover, a study from Kaufman et *al* [61] revealed that cells are less invasive in a softer matrix (0.5 mg/mL of collagen). Soft collagen gels may fail to promote invasion as (1) they are too fragile for the cell to generate the traction forces required to move, and (2) the cell cannot polarize and migrate [55, 157]. Whereas when collagen concentration is higher, cell migration is hindered by the resisting forces created by collagen fibres [54, 55, 61, 157]. It is expected that the pore size, therefore the available space for the cell to migrate, is reduced when collagen concentration increases [55, 61].

Owing to this evidence, cell migration was allowed only when the density of collagen is comprised between 0.5 and 4.0 mg/mL. The collagen interval where the cell is allowed to migrate corresponds to the observations from Schor et *al* . With their data, we fit a second-degree polynomial with values in $[0, 1]$ to modify the speed of the cell from 0 up to a maximum. Their data were normalized between zero and 1 prior to fitting. From the curve obtained, shown in figure 5.3, we can see that the cell speed will first increase as the density of collagen increases up to a maximum (2.0 mg/mL). Passed that maximum, the cell speed decreases as the density of collagen increases.

The speed of the cell is represented by the following piecewise function that has been based on the experimental data described in equation 5.8:

$$\nu = \begin{cases} 0 & \text{if } C_t \notin [0.5, 4.0] \text{ mg/mL}, \\ \nu_0(-0.23x^2 + 0.971x - 0.048) & \text{if } C_t \in [0.5, 4.0] \text{ mg/mL}, \end{cases} \tag{5.8}$$

where ν_0 is the base speed of the cell and C_t the total density of collagen ($C_t = C + C_r$). Here ν_0 is equal to 0.8 µm/min [61].

5.5 Exploration of the environmental and genetic impact on tumour growth and cellular migration

A simulation was run with the parameters and initial conditions defined in sections 5.2 and 5.3 to establish a reference. The resulting tumour, shown in figure 5.4, resembles the tumour obtained with the model presented in chapter 3: a necrotic core with dead and quiescent cells surrounded by a ring of proliferative cells. Here the layer of quiescent cells, mainly located on the bottom of the tumour, seems a bit more visible than in the previous model. On the opposite, a group of necrotic swelling cells is present on the upper part of the tumour.

Some cells tend to detach from the tumour's main body and move slightly away, although it is not significant enough to call those cells "invasive". This observation is similar to the results from Rubenstein et *al* [92] with their Potts model of cell migration (figure 5.5).

Because collagen is an important aspect of the model, it appears essential to investigate how the tumour impacts extracellular collagen density. Collagen density within the tumour ranges from 0 to 2.0 mg/mL. The highest density of collagen inside the tumour is found at its centre where tumour growth started. At the beginning of the simulation, cells are in normoxia and may not alter much the collagen content. Because cells at the centre of the tumour are the first to die due to harsh conditions, they may not had the time to alter significantly their environment. Collagen density within the tumour displays a pattern composed of circles, like a travelling wave. This may be caused by successive cell division: when cells divide, they push each other away, distributing them in circles as the tumour grows rather than uniformly. Some areas with low collagen density are present within the tumour. If at first a possible impact of the genes was hypothesized, it turned out that this is caused by a higher local cellular density. This is clearly

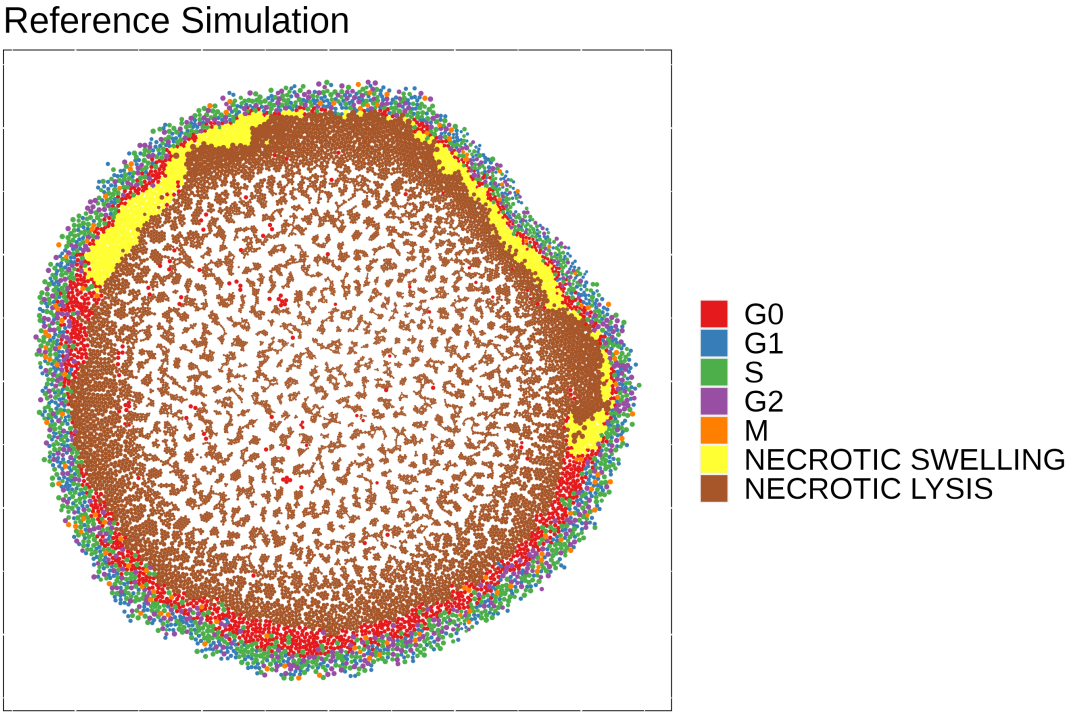

Figure 5.4: Tumour after 14 days grew in normoxia with the default initial conditions and parameters.

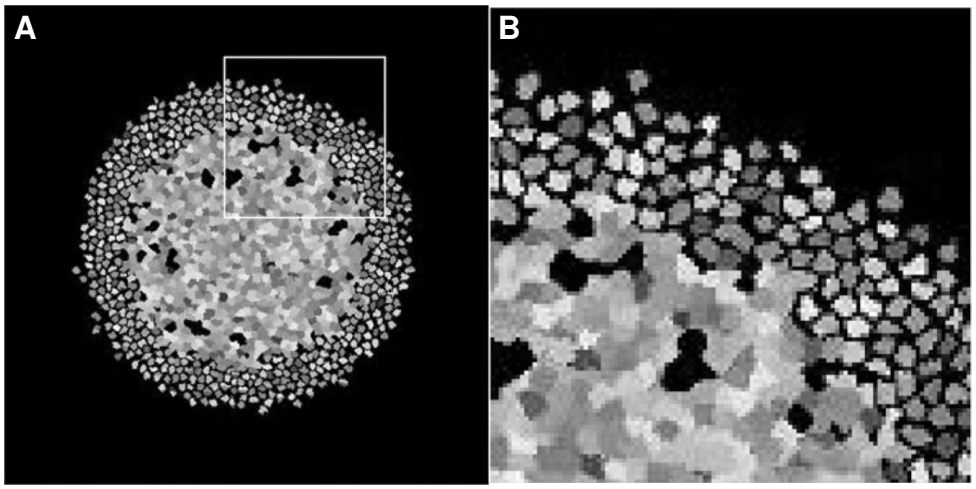

Figure 5.5: (A) Tumour simulated with a Potts model and (B) enlarged view of a part of the tumour. Source from Rubenstein et *al* [92].

visible when comparing the collagen density and the cellular density as shown in figure 5.6, where a matching pattern between the cellular density and the collagen density shows up.

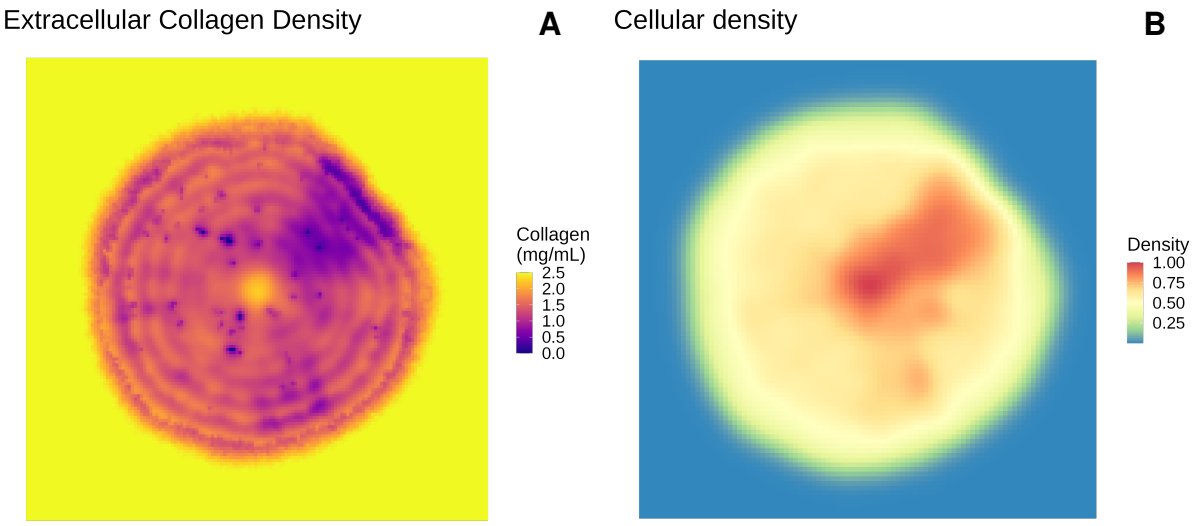

Figure 5.6: Plot of (A) the collagen density inside the tumour, and (B) the normalised cellular density.

To understand the impact of environmental collagen and oxygen on tumour growth, different environmental settings were tested:

- Bi-Gel: The microenvironment is split into two collagen densities like a collagen gel with a soft part (2.5 mg/mL) and a stiff one (5.0 mg/mL) to assess the impact two densities may have on the tumour.
- Lower/Upper Oxygen: Oxygen only diffuses at the top and the bottom of the domain to determine how cells with lower access to oxygen respond.
- Bi-Gel + Lower/Upper Oxygen: Combination of the Bi-Gel condition and the Lower/Upper Oxygen settings to observe the impact of both conditions.
- Complex matrix: Collagen density is randomly generated following a uniform distribution between 0.5 and 4.0 mg/mL to simulate a heterogeneous environment comparable to *in vivo* conditions.
- No Cell-Cell adhesion: In that case, cell-cell adhesion forces are null as migrating cells are known to have reduced cell-cell adhesion and increased cell-matrix interactions.

These different collagen conditions are summarized graphically in figure 5.7.

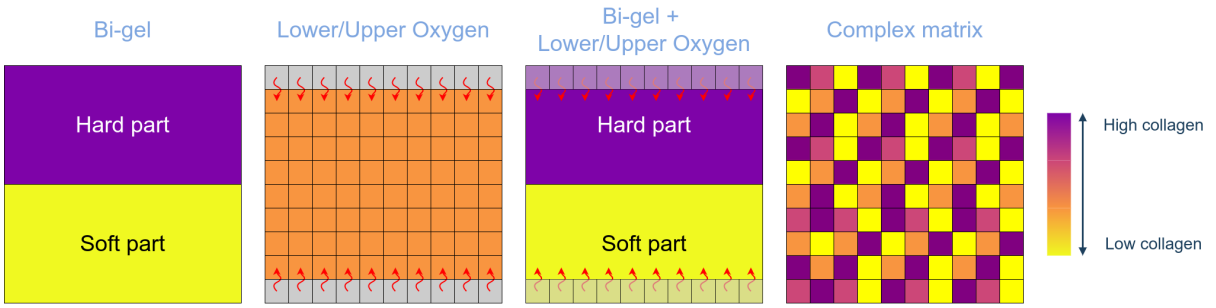

Figure 5.7: Different environmental settings represented graphically: Bi-Gel with a stiff part and soft part; Lower/Upper oxygen with diffusion at the top and the bottom of the domain; Combination of Bi-Gel and Lower/Upper oxygen; Complex matrix with randomly initiated values of collagen.

The tumour after 14 days of growth under different environmental settings is shown in figure 5.8. As can be seen, the tumour's shape is influenced by the initial collagen distribution. Oxygen, however, seems to have no impact on the tumour shape, while we will see later that it greatly influences the proliferation of the tumour. Interestingly, cells tend to pack together when dying in the necrotic core of the Reference tumour. Yet, when there is no cell-cell adhesion, cell spatial distribution shows the same wave pattern as extracellular collagen in figure 5.6. This result is similar to the result generated by the model from Anderson et *al* [145] (figure 5.9).

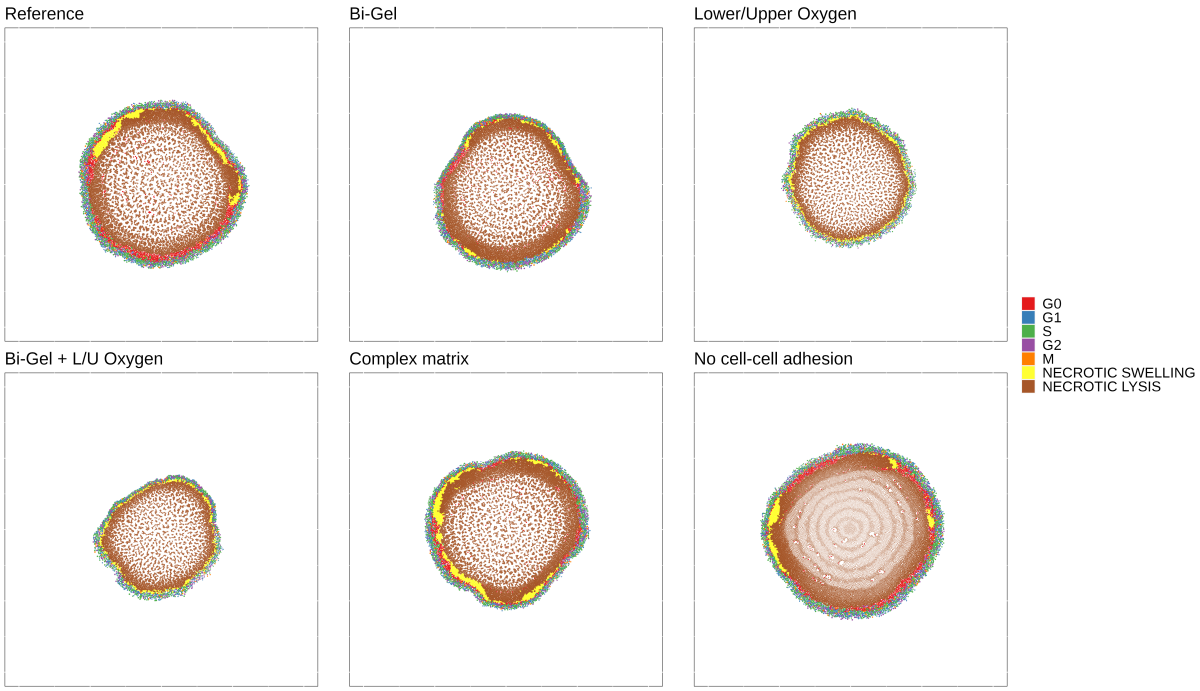

Figure 5.8: Tumour growth in different conditions.

While in the reference the tumour has a "roundish" shape, the tumour shape in the Bi-Gel resembles a "pear" and seems to display "branching" in a Complex Matrix. In Bi-Gel conditions, the upper part of the matrix is stiffer than the lower part. This difference in collagen density causes the cells to migrate faster at the bottom as collagen conditions are more favourable. A look at the cell migration speed inside the tumour grown on a Bi-Gel, shown in figure 5.10, confirms this hypothesis. Like in the case of the Bi-Gel, different migration speeds might be responsible for the shape of the tumour growing on a Complex Matrix. Hollow areas on the periphery of the tumour correlate with lower migration speed.

The distribution of migration speed in the different settings is compared in figure 5.11 to further investigate the impact of different collagen conditions. In the Reference simulation, migration speed ranges from almost 0 to 0.8 µm/min with most of the values distributed around 0.7 µm/min. No cell had a null speed in the Lower-Upper Oxygen, the Complex Matrix or the simulation with No Cell-Cell Adhesion.

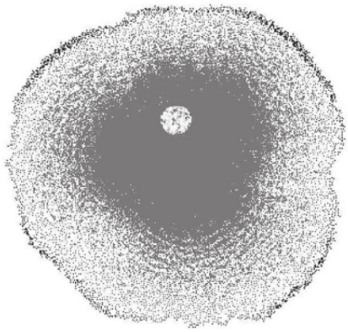

Figure 5.9: Simulation result of tumour growth by Anderson [145].

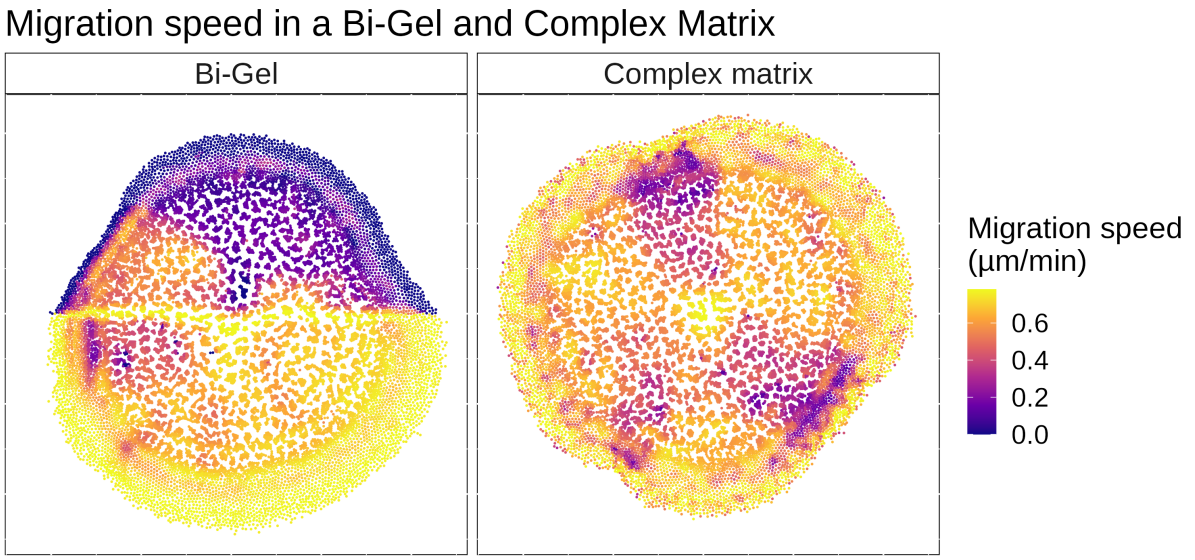

Figure 5.10: Cells migration speed in tumours grown on a Bi-Gel or a Complex Matrix.

The simulations where the tumour was grown on a Bi-Gel are the only cases where we can observe two peaks around which speed values seem to aggregate. This can be explained because cells are grown in two different collagen densities depending on whether they are at the top or the bottom of the tumour, enforcing different speeds on the cell. The same applies when a tumour is grown on a Complex Matrix, yet the distribution of migration speed is more diverse owing to the random initialization of the matrix. A peak around 0.65 is still present. The distribution of migration speed values in the simulation with no cellular adhesion tends to look more like the one in the Reference simulation. But compared to the Reference and like with the Complex Matrix, no cell has a speed of 0. The case with oxygen diffusing only at the top and bottom of the domain strongly selects for migration speed values comprised between 0.6 and 0.8. Probably cells cannot have lower migration speed because they die beforehand due to a lack of nutrients as almost all oxygen is depleted at the end of the simulation. This effect is also visible when combining a Bi-Gel with poorer oxygen conditions. Two peaks corresponding to the two different collagen densities in the environment are still present but, as can be seen, the distribution of cells with speed between 0.4 and 0.6 is narrower when combining a bi-gel and lower oxygen than with the bi-gel alone.

If we already see an effect of collagen and oxygen on the migrative capability of the cell, they do affect proliferation as well. The impact of the different conditions on proliferation is determined by the number of cells (living and dead) at the end of the simulation, and the radius of the tumour (figure 5.12). The radius of the tumour is calculated by measuring the distance between the centre and the cell the furthest from the center for both the proliferative and the necrotic layers. Tumour growth is maximized in the model when cells have no adhesion and minimized when combining a bi-gel with lower oxygen diffusion. The simulation with no cellular adhesion has the highest number of cells and tumour radius. The tumour radius is identical in the Reference and the Bi-Gel conditions. An explanation could be that the lower part of the tumour in the Bi-Gel proliferates in the same conditions as the Reference tumour. Because

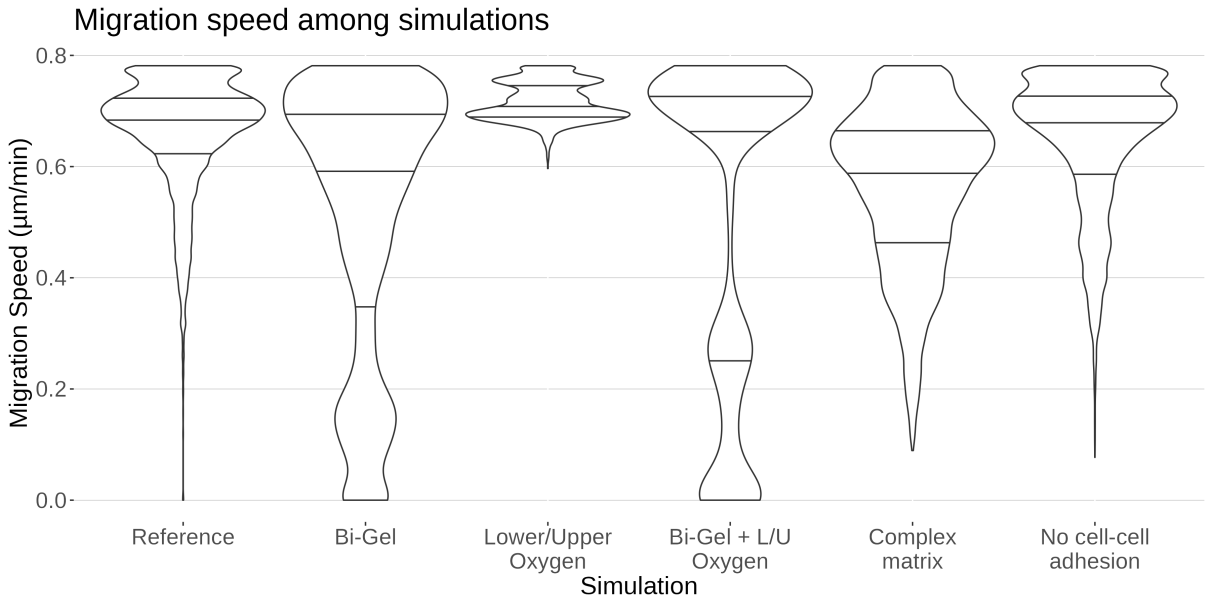

Figure 5.11: Violin plot of the migration speed in the different collagen conditions.

there is no distinction between the radius at the bottom and the top of the tumour, only the proliferative distance in the softer part might be considered. The number of cells at the end of the simulation is lower in the Bi-Gel than in the Reference confirming that proliferation is affected. Surprisingly, while the tumour radius is lower with the complex matrix, the number of cells after 14 simulated days is higher than with the Bi-Gel. Taken together, these results show that collagen conditions alone can impact the proliferation of the tumour. Tumour radius and cell number are reduced when oxygen conditions are poorer, showing that oxygen alone can reduce cell proliferation as well. The effect is even more marked when poorer oxygen condition is combined with bi-gel conditions, suggesting a dual effect of both collagen and oxygen.

We now have confirmed that oxygen and collagen impact tumour proliferation but do they impact protein levels as well? Here, distributions of protein levels are similar between all conditions except for the cases with poorer oxygen conditions (figure 5.13). In these cases, protein levels group around one value showing no diversity. This suggests that hypoxia strongly select for a specific phenotype, a result that was already observed in chapter 3 and present in figure 5.11.

We have seen that tumour growth is impacted in the model by different collagen conditions and oxygen conditions. Oxygen and collagen seem to have a dual effect on proliferation and migration speed. It seems that only collagen does not strongly impact the genetic regulations implemented in the model.

Now that the impact of the environment on the tumour has been determined, we want to understand if different genetic regulations could lead to different outcomes. An issue that comes to mind is: Is there a gene more critical to tumour growth than the others? To answer that question, the qualitative effect of genetic regulations must be addressed. It was decided to give the same effect of hypoxia on the regulation of P4HA1, MT1-MMP and LOX; and to give the same impact of these genes on the modification of environmental collagen. A study on the rate of secretion of collagen by cells during wound healing showed that after several days, despite the collagen secretion rate rised by 4, the collagen content did not changed further. The authors observed that the rate of degradation of collagen matched the rate of secretion [158]. Setting the genetic parameters to be equal would facilitate the interpretation of the qualitative role for each gene, and replicate to some extent the phenomenon described above. To that end, the following parameters were changed to have the same values:

- Strength of regulation by HIF: $\gamma_{h\to p} = \gamma_{h\to m} = \gamma_{h\to r} = 4$.
- Level of HIF at which the regulated gene level is half-maximum: $s_{h\to p} = s_{h\to m} = s_{h\to r} = 17.5$.
- Collagen secretion degradation and cross-linking rates: $\beta_p = \beta_m = \beta_r = 2.16 \times 10^{-5}$mmol/L/min.
- Density of collagen at which the activity of MT1-MMP and LOX is half-maximal: $K_m = K_r = 8.7 \times 10^{-11}$mmol/L.

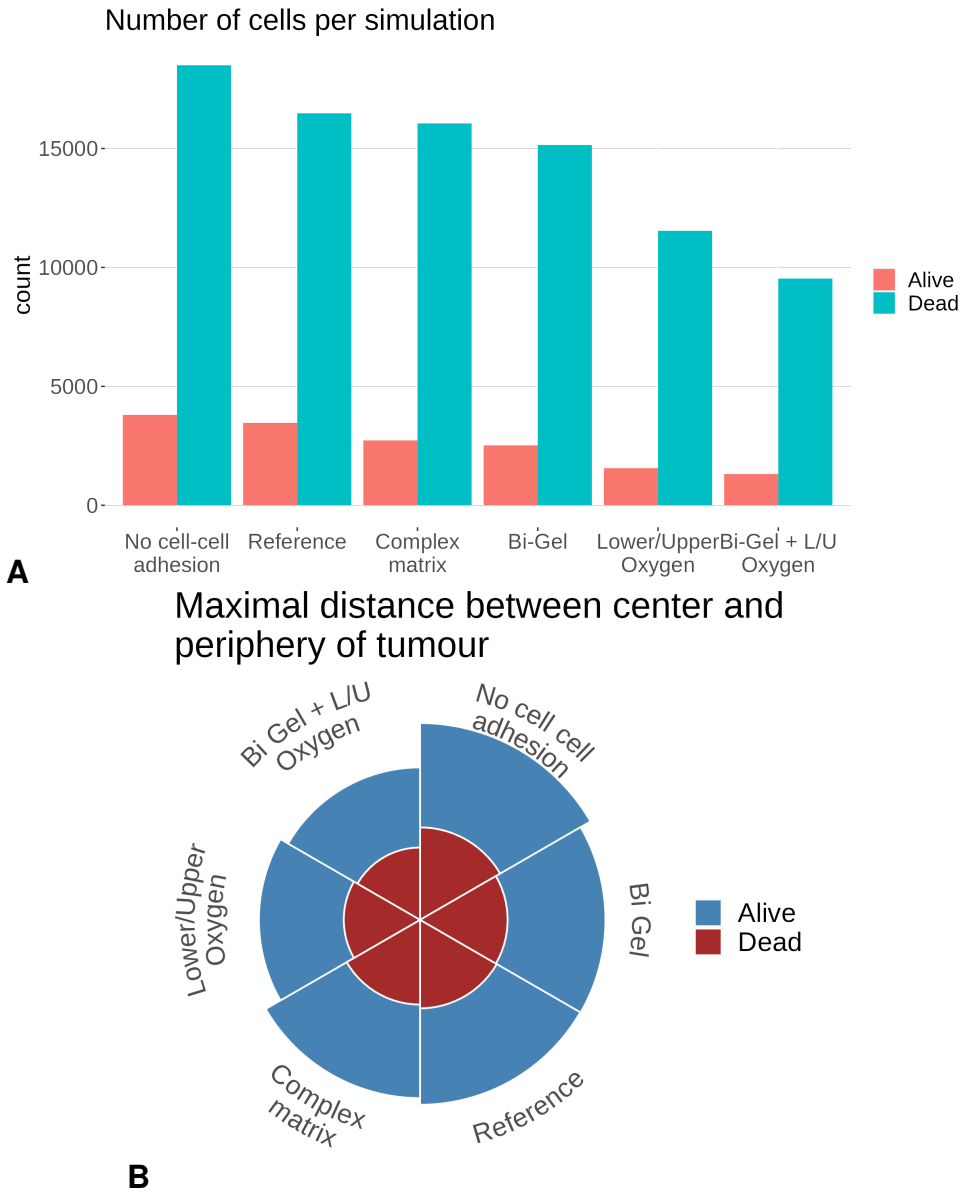

Figure 5.12: Proliferation of the tumour in the different collagen conditions with (A) a barplot of the number of cells and (B) a pie chart of the maximal proliferation distance. In both cases, the colours represent living and dead cells.

A simulation was run with this set of parameters to establish a new reference with “Equal Genetics” in regular environmental conditions (homogenous collagen matrix and normoxia). Then, the $\gamma_{h\to p}, \gamma_{h\to m}, \gamma_{h\to r}$ parameters have been successively set to:

1. 8 one at a time, to reflect different sensitivity to hypoxia by the genes P4HA1, MT1-MMP and LOX.
2. 1 one at a time, to reflect an absence (inhibition) of sensitivity to hypoxia by the genes P4HA1, MT1-MMP and LOX.

It seems that in that setting, the different genetic regulations had little impact on the proliferation of the tumour with not only a close number of cells at the end (both living and dead) but a similar tumour radius as well (figure 5.14). The shape of the tumours was similar among the simulations. These observations may suggest that all genes are equally important or that stronger deregulations are required to disturb tumour growth in the model.

Interestingly, migration speed does not vary in the tumour when genetic regulations are equal or when only the sensitivity of LOX to hypoxia is affected (figure 5.15). When the sensitivity of MT1-MMP to hypoxia is increased, cells migrate faster compared to the reference with equal genetic parameters. On the opposite, increasing the sensitivity of P4HA1 to hypoxia slows down cell migration compared to the

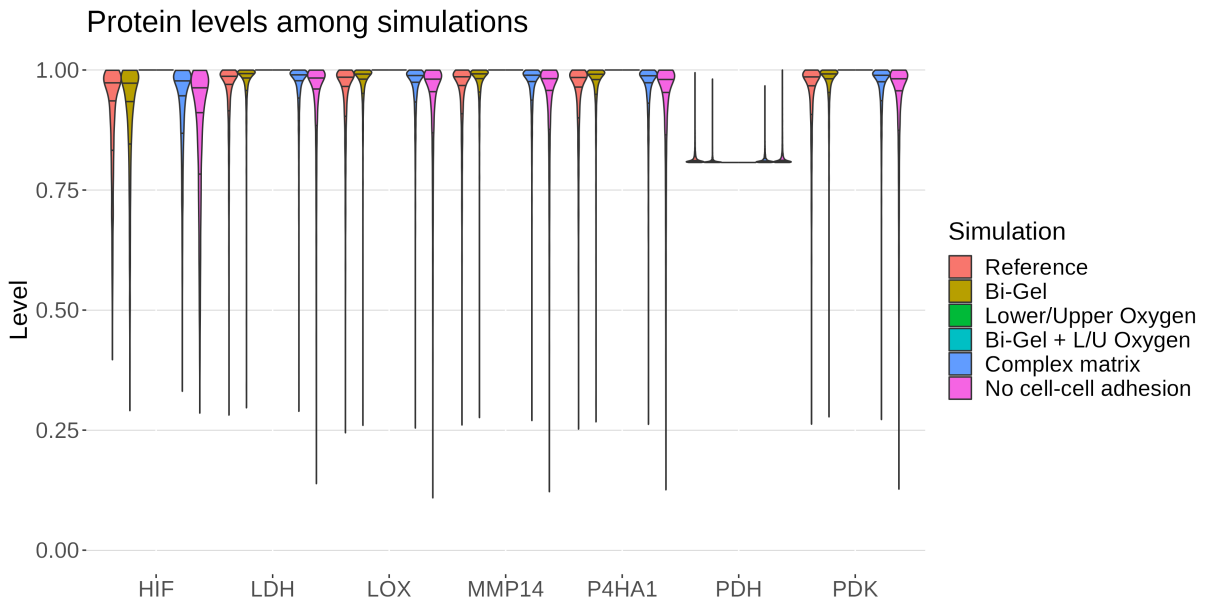

Figure 5.13: Violin plot of the protein levels in different collagen conditions. Levels are normalized to 1 to avoid different scales as they have different parameters.

reference with equal genetic parameters. Moreover, the distribution of cell migration speed is wider when P4HA1 sensitivity to hypoxia is increased compared to other simulations. Inhibiting MT1-MMP and P4HA1 reversed the effect on cell migration: P4HA1 inhibition increased cell migration speed compared to the equal genetic reference, MT1-MMP inhibition decreased the cell migration speed instead. It suggests that degradation of collagen favours migration while increased secretion of collagen tends to hinder it. LOX deregulation did not alter the distribution of cell speed compared to the new reference in the current environmental setting. A significant effect may be observed when varying the initial density of cross-linked collagen.

Distributions of extracellular collagen and cross-linked collagen are shown in figure 5.16 to determine how different sensitivities to hypoxia impact extracellular collagen. In the case where all genes have the same parameters, extracellular collagen has densities mainly spread between 1.75 - 2.25 mg/mL, and extracellular cross-linked collagen ranges mainly between 0.4 - 0.8 mg/mL. Increasing the sensitivity of MT1-MMP or LOX by hypoxia widens the distribution of extracellular collagen and centres them around lower values. Inhibiting the sensitivity of P4HA1 to hypoxia showed the same result. Increasing the sensitivity of P4HA1 to hypoxia seems to normalize extracellular collagen around 2.2 mg/mL. The extracellular cross-linked collagen distribution is narrower around 0.6 mg/mL when either P4HA1 or MT1-MMP sensitivity to hypoxia is affected. When MT1-MMP level is increased, collagen may be degraded before it can be cross-linked leading to reduced extracellular collagen. Surprisingly, increasing the upregulation of P4HA1 by hypoxia has the same effect. While it would be expected that higher extracellular collagen would result in higher cross-linked collagen owing to better availability. A possible explanation for that result is that collagen is degraded faster by MT1-MMP than it is cross-linked by LOX even with the same parameters. Naturally, the distribution of extracellular cross-linked collagen is wider when LOX upregulation by hypoxia is stronger, or narrower with reduced extracellular cross-linked collagen when LOX is not sensitive to hypoxia. Hence, increased LOX levels lead to more heterogeneous levels of extracellular cross-linked collagen. Inhibiting the sensitivity of MT1-MMP and LOX allowed for higher extracellular collagen in the same way.

Since the model has changed from 3, it is interesting to see how these changes affected the metabolism. Does collagen impact the appearance of the Warburg effect? Do the genes involved in collagen biosynthesis affect the metabolism? Is it possible to induce the Warburg Effect by fast variations of oxygen or by reducing the degradation of HIF as in section 3.3?

Figure 5.17 shows the production of H^+ depending on the extracellular oxygen concentrations at different times in the following simulations: Reference Simulation, Lower/Upper oxygen diffusion, Complex Matrix, Reference with Equal Genetic, Oscillating oxygen and Reduced HIF degradation. The latter two cases are simulations run in the same conditions as in chapter 3 to reflect how the new parameters have impacted the model. In the Oscillating simulation, the tumour growth starts with normoxic Dirichlet

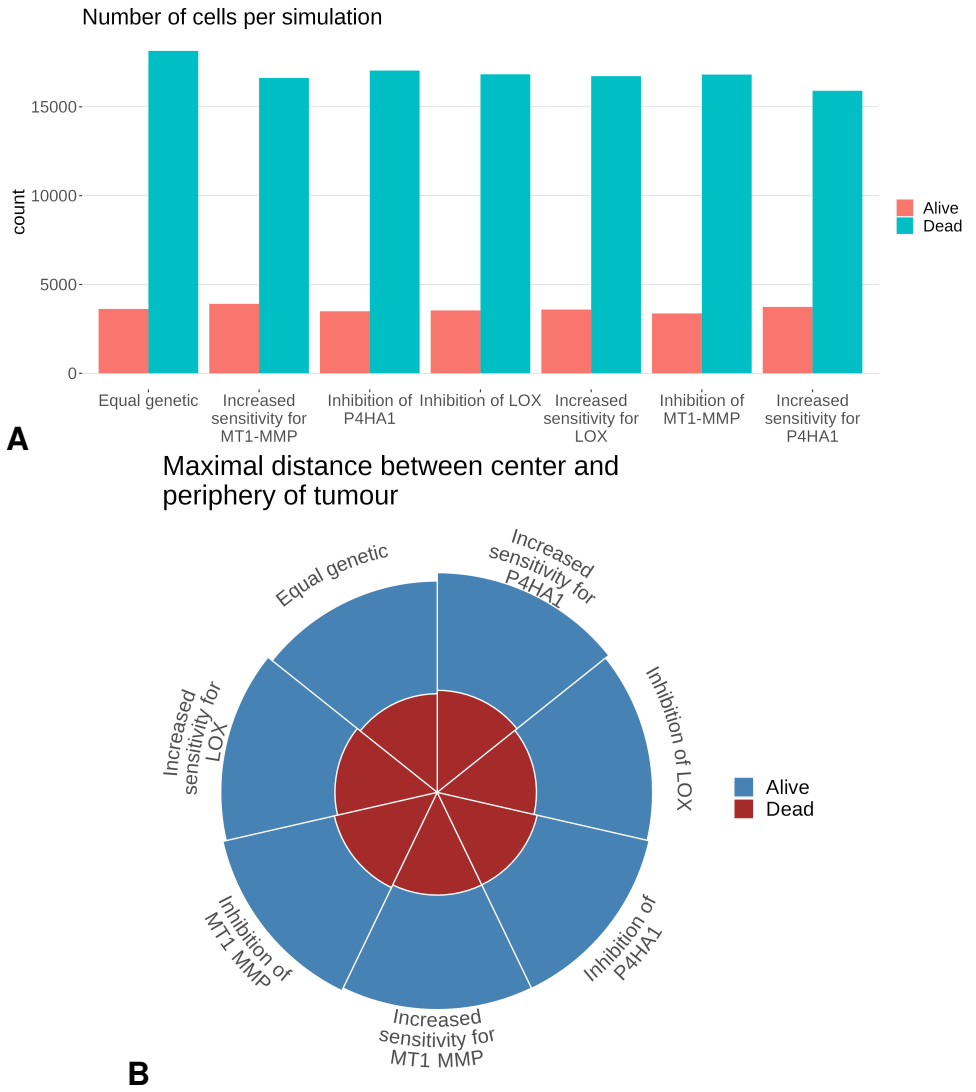

Figure 5.14: Proliferation of the tumour depending on different gene sensitivities to hypoxia with (A) a barplot of the number of cells and (B) a pie chart of the maximal proliferation distance. In both cases, the colours represent living and dead cells. The genetic settings tested are: increased sensitivity of P4HA1 to hypoxia ($\gamma_{h\to p} = 8$), increased sensitivity of MT1-MMP to hypoxia ($\gamma_{h\to m} = 8$), and increased sensitivity of LOX to hypoxia ($\gamma_{h\to r} = 8$), no sensitivity of P4HA1 to hypoxia ($\gamma_{h\to p} = 1$), no sensitivity of MT1-MMP to hypoxia ($\gamma_{h\to m} = 1$), and no sensitivity of LOX to hypoxia ($\gamma_{h\to r} = 1$).

Boundary conditions. The oxygen diffusing at the border of the domain is slowly decreased over 6 hours down to 1 % O_2 (0.01112 mmol/L) and then is increased back to 5 % O_2 (0.056 mmol/L). This process is repeated until the end of the simulation to create stressful conditions for the cell. The simulation with Reduced HIF degradation by oxygen simulates the effect of oncogenic mutations stabilizing HIF in normoxia. While cells are still in normoxia after one day, except for the simulation with poorer oxygen conditions, all cells are below the hypoxic threshold after the second day and remain in hypoxia. Compared to previously, there are no cells that exhibit a Warburg Effect even in the same conditions that triggered one with the precedent version of the model. Cells grown in poorer oxygen conditions tend to have access to almost no oxygen. When cells are below the threshold of hypoxia, they start to increase their production of H^+ as oxygen levels decrease. The rate of H^+ secretion decreases when glucose starts lacking in the environment. This pattern is visible in all simulations, except in the Lower/Upper oxygen one. Because the same pattern is observed across all simulations it suggests that collagen has little impact on the metabolism in the model. Nonetheless, the pattern observed is similar to the results obtained in section 3.3 when inhibition of pyruvate dehydrogenase (PDH) by HIF is reduced. It should be reminded that the parameter controlling the sensitivity of PDH to HIF has changed from $\gamma_{h\to q} = 0.14$ to $\gamma_{h\to q} = 0.78$ when genetic parameters were fitted to transcriptomic data. When inhibition of PDH by hypoxia is reduced, the

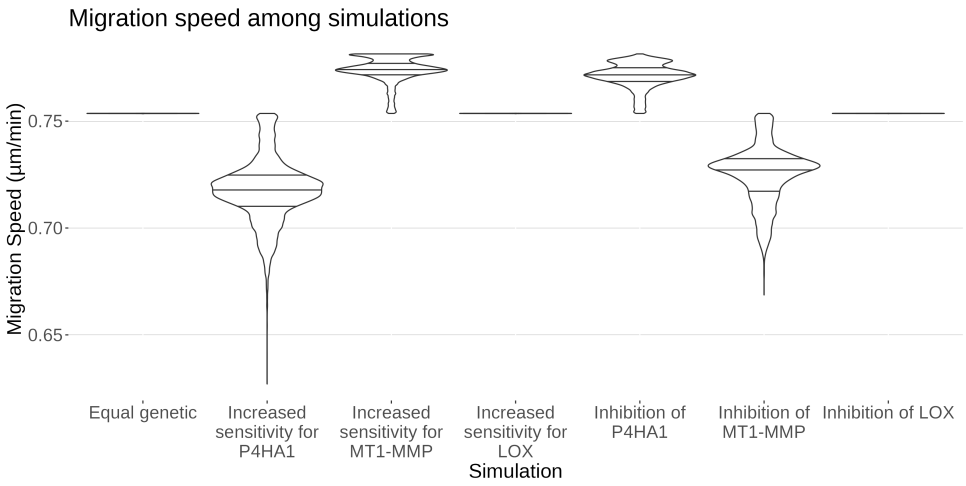

Figure 5.15: Violin plot of the migration speed depending on different gene sensitivities to hypoxia. The genetic settings tested are: increased sensitivity of P4HA1 to hypoxia ($\gamma_{h\to p} = 8$), increased sensitivity of MT1-MMP to hypoxia ($\gamma_{h\to m} = 8$), and increased sensitivity of LOX to hypoxia ($\gamma_{h\to r} = 8$), no sensitivity of P4HA1 to hypoxia ($\gamma_{h\to p} = 1$), no sensitivity of MT1-MMP to hypoxia ($\gamma_{h\to m} = 1$), and no sensitivity of LOX to hypoxia ($\gamma_{h\to r} = 1$).

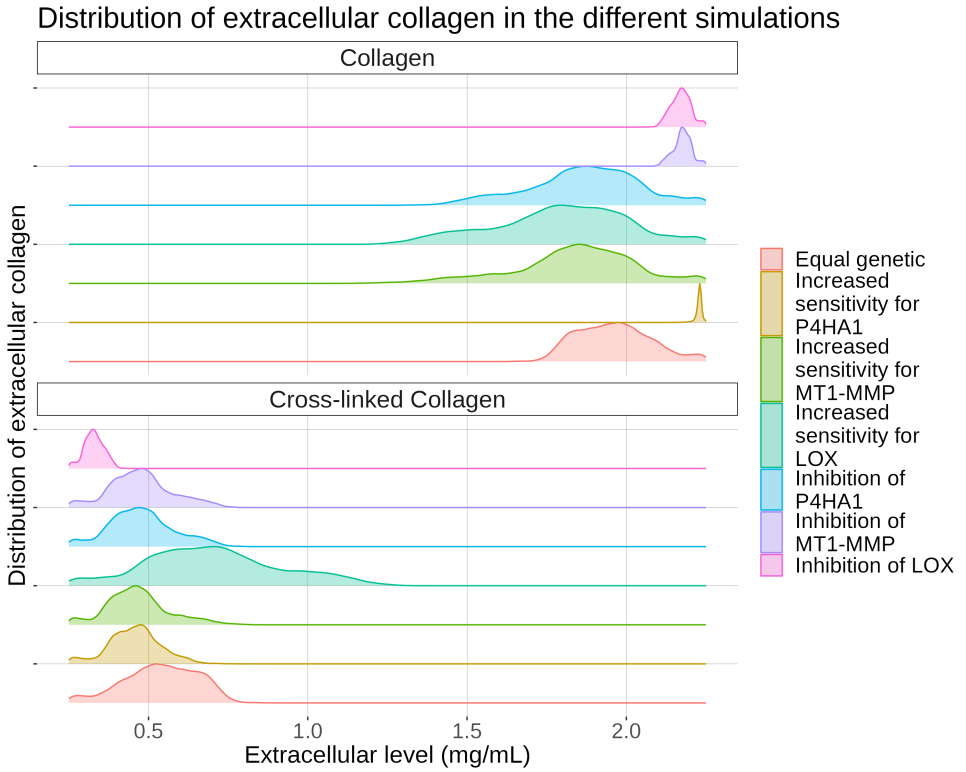

Figure 5.16: Distribution of extracellular collagen and cross-linked collagen depending on different sensitivities to hypoxia. The genetic settings tested are: increased sensitivity of P4HA1 to hypoxia ($\gamma_{h\to p} = 8$), increased sensitivity of MT1-MMP to hypoxia ($\gamma_{h\to m} = 8$), and increased sensitivity of LOX to hypoxia ($\gamma_{h\to r} = 8$), no sensitivity of P4HA1 to hypoxia ($\gamma_{h\to p} = 1$), no sensitivity of MT1-MMP to hypoxia ($\gamma_{h\to m} = 1$), and no sensitivity of LOX to hypoxia ($\gamma_{h\to r} = 1$).

cell may rely more on oxygen than cells with stronger inhibition leading to lower oxygen levels. In that case, oxygen may not go back to normoxic conditions and the Warburg phenotype cannot be observed. This result shows that if environmental conditions can induce a Warburg phenotype, genetic regulation may influence the ability of the cell to adopt it. It would explain why the Warburg phenotype does not appear in the results.

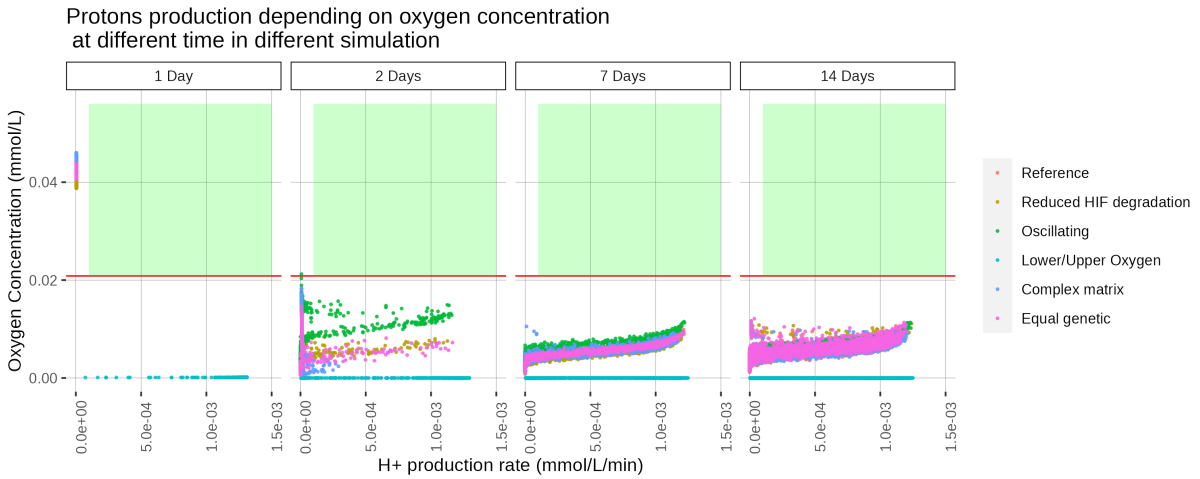

Figure 5.17: Rate of production of H^+ at 1, 2, 7 and 14 Days in different simulations. The red line indicates the hypoxia threshold (0.02085 mmol/L or 2% O_2). In oscillating conditions, oxygen concentration is slowly decreased from normoxia (5 % O_2) to severe hypoxia (1 % O_2) during 6h, then oxygen is increased to normoxia at the same rate. This process is repeated until the end of the simulation. Only living cells are represented on the graph. The green rectangle represents the region corresponding to a Warburg effect.

Levels of lactate dehydrogenase (LDH) and pyruvate dehydrogenase (PDH), shown in the figure 5.18, seem to support this theory. Like in section 3.3, three metabolic states are observed. They correspond to the oxidative state (high PDH/low LDH), the glycolytic state (low PDH/high LDH), and an intermediate state in between the two. This intermediate state is adopted by the cell during its transition from the oxidative to the glycolytic state. This dynamic was already observed previously and given that different collagen conditions do not influence it supports the idea that genetic regulations prevail over collagen onto the metabolism. Interestingly, cells growing in Oscillating conditions seem to be trapped in the glycolytic state. This is a different result from what has been observed before.

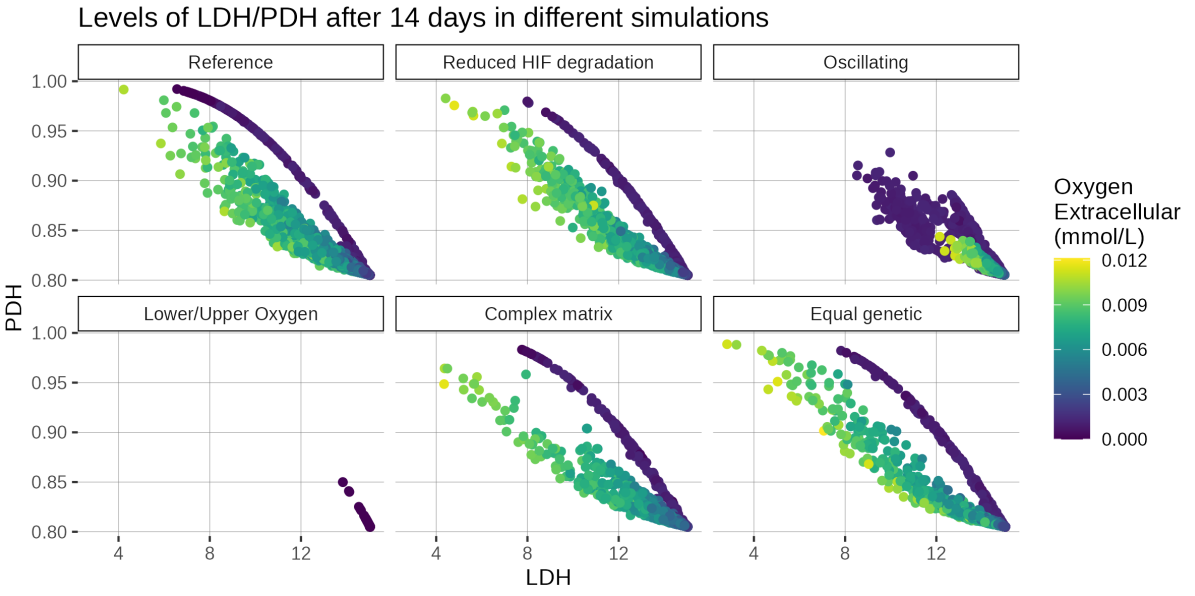

Figure 5.18: LDH and PDH levels after 14 days of growth in different simulations. In oscillating conditions, oxygen concentration is slowly decreased from normoxia (5 % O_2) to severe hypoxia (1 % O_2) during 6h, then oxygen is increased to normoxia at the same rate. This process is repeated until the end of the simulation.

5.6 Conclusions: Influence of collagen and genetic parameters on migration and metabolism

To conclude, collagen regulation has been added to the model following the results obtained with the statistical analysis of transcriptomic data. Collagen was modelled as a substrate that does not diffuse nor decay, its density in the environment is then only influenced by the cells. Extracellular collagen and cross-linked collagen inside the tumour were mainly influenced by the local cell density which in turn drove the cellular migration speed. As a consequence, it affected the tumour shape at the end of the simulation.

Results obtained showed that the extracellular environment and the cell-cell adhesion forces are the major factors driving tumour proliferation in the model. Tumour growth was maximized when cell-cell adhesion stopped, which favours the migration of the cell and is coherent with the previous findings in the literature. Extracellular density and oxygen had an impact on tumour proliferation as less favourable conditions (lower oxygen diffusion, stiffer collagen gel) tended to reduce proliferation. In addition, the combination of stiffer collagen and lower oxygen reduced tumour size even further suggesting that both factors have an additive effect. In particular, the initial collagen matrix influenced the tumour shape. Since the model simulated 14 days of growth, tumour shape differences after longer proliferation may be more visible. The extracellular collagen had little impact on the protein levels, except for oxygen which strongly enforced specific levels of protein as in the model in chapter 3.

Genetic parameters and impact of P4HA1, MT1-MMP and LOX on the collagen content were normalized to be equal to study qualitatively their impact. It was found that different sensitivity to hypoxia did not significantly alter the proliferation of the tumour with the condition tested. This seems to suggest that they are equally relevant to tumour proliferation. Even so, cell migration speed was impacted by different sensitivity of P4HA1 and MT1-MMP to hypoxia while LOX did not have an effect. This may indicate that the secretion and degradation of collagen are the most important factors, at least in the model, for a faster cell speed. All genetic parameters tested influenced the distribution of extracellular collagen and cross-linked collagen. Higher levels of MT1-MMP seemed to induce lower levels of collagen and lower cross-linked collagen density. On the opposite, higher levels of P4HA1 induced higher densities of extracellular collagen but had the same effect on extracellular cross-linked collagen as MT1-MMP. Inhibition of sensitivity to hypoxia for P4HA1 decreased extracellular collagen and MT1-MMP increased it, they both decreased extracellular cross-linked collagen. Higher LOX quantity simultaneously decreased regular collagen and increased cross-linked collagen while its inhibition had the opposite effect. Consequently, P4HA1 and MT1-MMP both mainly impact the extracellular collagen while LOX mainly regulates cross-linked collagen. Deregulation of P4HA1 and MT1-MMP sensitivity to hypoxia had the same effect on extracellular cross-linked collagen.

Compared to the previous version of the model, here the Warburg Effect has not been adopted by cells in any simulations. The results showed that the production of H^+ in hypoxia followed the same trend as the results obtained previously when PDH inhibition by hypoxia was reduced. Different collagen conditions could not induce a Warburg effect. Oscillating oxygen and reduced HIF degradation could induce a Warburg phenotype precedently. Here this induction has abrogated. In this chapter, the parameters of the model have been fitted to transcriptomic data leading to a reduced PDH inhibition by HIF after fitting. This may be the leading cause of the Warburg Effect disappearance and suggests that genetic regulation is an important factor driving cellular metabolism. No effect of the environment has been observed, thus its effect may be limited by genetic regulations. This was supported by LDH and PDH levels that showed the same dynamics between the two versions of the model with three metabolic states: an oxidative state, a glycolytic state and an intermediate state. The intermediate state is adopted by the cell while transit from one to the other was still observed in the result as well.

Conclusions and perspectives

In this Thesis, we explored some aspects of the impact of the Hypoxia Inducible Factor (HIF) on tumour growth. I started by investigating the effect of hypoxia on cellular metabolism using a mathematical model describing a network of genes regulated by HIF which includes the two important metabolic genes lactate dehydrogenase (LDH) and pyruvate dehydrogenase (PDH), to reflect the effect of HIF on the appearance of the Warburg Effect. An anisotropic diffusion of oxygen has been implemented to depict a more realistic oxygen gradient through the tumour. With that model, we could test different environmental conditions and genetic regulation to see which could induce a Warburg Effect and saw that both rapid variations of extracellular oxygen and increased levels of HIF are enough to trigger the appearance of a Warburg phenotype. In addition, I performed a statistical analysis of transcriptomic data from The Cancer Genome Atlas (TCGA) database and a Patient Derived Cell Lines (PDCL) dataset, to unravel the deregulated mechanisms in glioblastoma. Several processes were found to be frequently deregulated in glioblastoma (GBM), some even only deregulated in PDCL data, yet collagen biosynthesis was selected as the candidate for modelling. Following the result of the statistical analysis, the model has then been extended to reflect the effect of HIF on cellular migration through its action on Extra-Cellular Matrix (ECM) remodelling. The genes P4HA1, MT1-MMP and LOX have been implemented to impact the secretion, degradation and cross-linking of collagen. The results indicated that collagen density has a substantial effect on the proliferation of the tumour and its shape. If it was already established that oxygen can impact the growth of the tumour and was shown again in the result, it seems that the effect of oxygen combined with the effect of collagen. Increased P4HA1 and MT1-MMP quantity, respectively decreased and increased cellular migration speed while LOX seemed to have little effect. Interestingly, a Warburg phenotype was not observed in the simulations with the new parameters. This is likely to be caused by the new parameters regulating the sensitivity of PDH to HIF. Similarities between the production of H^+ when the sensitivity of PDH to HIF was reduced in the model in chapter 3 with the new results support this hypothesis. Furthermore, the LDH and PDH protein levels dynamic were unchanged. This indicates that if microenvironmental conditions favour the appearance of a Warburg phenotype, genetic regulations may prevail.

The work presented in this Thesis is not devoid of limitations, some being shared between the two versions of the mathematical model. Concerning the statistical analysis performed, a clear limitation is the control cases in both datasets. Indeed, we saw that in the PDCL dataset, controls have been taken from a different study, and for TCGA data, there are only five matching cases which may be not sufficient for good results. The analysis would definitely benefit from a comparison with more numerous and more adapted samples. Unfortunately, transcriptomic data of healthy samples can be hard to find, especially on the brain as this organ is very sensitive. Initiatives to make these data available do exist but the hunt for good data has yet to become easy.

Limitations of the mathematical models are mainly shared between the two versions. Simulations of tumour growth were run in 2D to avoid the computational cost generated by 3D. Yet oxygen diffusion is closer to *in vivo* conditions in spheroids, showing mechanisms that are not correctly depicted by regular cell cultures. 2D simulations may not describe these mechanisms and oxygen diffusion may be more accurately represented with 3D simulations. More work on the implementation of the anisotropic diffusion would be necessary to optimize the computing load. Here, the modelling approach focused on replicating *in vitro* conditions and angiogenesis has not been implemented. It is known that the dysregulated vasculature can create local hypoxia as the new vessels formed are often less organized and consequently less efficient. Thus, feedback between HIF and angiogenic processes could greatly influence the results. Regarding representing heterogeneity, the model did not implement any mechanisms to show spontaneous mutations via slight changes of genetic parameters during division. Oncogenic mutations stabilizing HIF, not present at the beginning of the simulation, could appear during tumour growth. It would show how phenotype varies over time.

Specifically on the model of migration centred on HIF, several processes could impact the level of HIF independently of oxygen. Indeed, it has been shown in the literature that P4HA1 or MT1-MMP could promote HIF stabilization, creating a positive feedback where HIF increase the levels of these proteins

which in turn increase the levels of HIF. Moreover, regulation of MT1-MMP is complex (auto-regulation with domain shedding, upregulation by collagen density) and has been simplified to only reflect the effect of HIF in the model. HIF feedback loop and MT1-MMP complete regulation were out of the scope of this model but their benefit in enhancing the results should not be ignored. Lastly, migration speed was impacted identically by collagen and cross-linked collagen. However, it could be possible that cross-linked collagen is more robust due to increased stability, making it more able to bear the mechanical forces exerted by the cell. A more realistic gradient of migration velocity could then be described if cross-linking was confirmed to be more adapted for cell migration in the literature.

The first model developed in this Thesis allowed to question the impact of HIF on the metabolism and more specifically on the Warburg Effect, a topic that is often addressed in the literature. If the Warburg Effect has been thoroughly investigated through mathematical modelling and biological experiments, it seems that HIF has received more attention from biological experimenters than from mathematical modellers. We saw that HIF modelling has been done yet it was mainly to understand its regulation and mechanisms, rather than to understand a general role or effect during tumour growth. In this sense, this work is innovative as we explicitly describe HIF quantity and link it to changes in the metabolism. Furthermore, we saw how we could lead the extension of a model through a data analysis-guided investigation of the literature. Data analysis using statistical methods has proven to be efficient in understanding the features of a dataset. In the case of cancer, such analyses are valuable to describe and represent the heterogeneity among the cells. It seems that statistical analysis and mathematical modelling are conducted in separate studies too often. We saw in this work that both methods are valuable and can complement each other to orient the research.

This Thesis opens new perspectives of study to better define the underlying mechanisms of hypoxia on tumour proliferation. The genetic regulation implemented follows a simple mathematical framework that has been described in the literature and successfully applied to study state transition in biological systems. We saw that the model is simple to extend, more complex systems could be modelled to depict with better accuracy the heterogeneity of metabolic/invasive states induced by HIF. It was the spirit of the Thesis to keep the model as simple as possible. Yet the changes induced by HIF are numerous and this work aimed to propose a simple mathematical implementation of HIF impact on biological functions, yet this work can be seen as providing a basis for more mathematical modelling studies of HIF. The complete role and consequences of hypoxia in tumour growth are not yet fully understood. consequently, this work provides new ideas and hypotheses to explore in the hope of better harnessing the interest hypoxia targeting therapies can hold.

A Appendix

A.1 Published Article: Modelling the role of HIF in the regulation of metabolic key genes LDH and PDH: Emergence of Warburg Phenotype.

Modelling the role of HIF in the regulation of metabolic key genes LDH and PDH: Emergence of Warburg Phenotype.

Kévin SPINICCI[1,2], Pierre JACQUET[1], Gibin POWATHIL[2], and Angélique STEPHANOU[1]

[1]Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France

[2]Department of Mathematics - Swansea University

May 2022

Abstract

Oxygenation of tumours and the effect of hypoxia in cancer cell metabolism is a widely studied subject. Hypoxia Inducible Factor (HIF), the main actor in the cell response to hypoxia, represents a potential target in cancer therapy. HIF is involved in many biological processes such as cell proliferation, survival, apoptosis, angiogenesis, iron metabolism and glucose metabolism. This protein regulates the expressions of Lactate Dehydrogenase (LDH) and Pyruvate Dehydrogenase (PDH), both essential for the conversion of pyruvate to be used in aerobic and anaerobic pathways. HIF upregulates LDH, increasing the conversion of pyruvate into lactate which leads to higher secretion of lactic acid by the cell and reduced pH in the microenvironment. HIF indirectly downregulates PDH, decreasing the conversion of pyruvate into Acetyl Coenzyme A which leads to reduced usage of the Tricarboxylic Acid (TCA) cycle in aerobic pathways. Upregulation of HIF may promote the use of anaerobic pathways for energy production even in normal extracellular oxygen conditions. Higher use of glycolysis even in normal oxygen conditions is called the Warburg effect. In this paper, we focus on HIF variations during tumour growth and study, through a mathematical model, its impact on the two metabolic key genes PDH and LDH, to investigate its role in the emergence of the Warburg effect. Mathematical equations describing the enzymes regulation pathways were solved for each cell of the tumour represented in an agent-based model to best capture the spatio-temporal oxygen variations during tumour development caused by cell consumption and reduced diffusion inside the tumour. Simulation results show that reduced HIF degradation in normoxia can induce higher lactic acid production. The emergence of the Warburg effect appears after the first period of hypoxia before oxygen conditions return to a normal level. The results also show that targeting the upregulation of LDH and the downregulation of PDH could be relevant in therapy.

Introduction

Cells rely on two main processes to produce ATP: Oxidative Phosphorylation (OXPHOS) by using oxygen, and glycolysis by using glucose. Glycolysis is a pathway generating both ATP and pyruvate using glucose as input [1, 2, 3]. Pyruvate produced by glycolysis can then be used to fuel the Tricarboxylic Acid (TCA) cycle and produce the compounds involved in OXPHOS, the aerobic pathway. If oxygen is not present, pyruvate is turned into lactate, this process is called fermentation [4]. Lactate formed during fermentation is secreted into the microenvironment which causes a decrease in extracellular pH.

In 1927, Otto Warburg observed that the tumour consumed more glucose and produced more lactic acid than normal tissues [5]. At first Warburg's observation did not consider the presence of oxygen, yet since increased lactic acid production was also observed when oxygen is available, it has slowly been associated with aerobic glycolysis [6]. Nowaday, high rate of glycolysis, even if oxygen is available, is known as the Warburg Effect [7, 8]. In this paper, we will retain this definition. Tumours can develop anywhere, yet harsh conditions favour tumour appearance [9]. Most tumours have median oxygen levels falling below 2%, the threshold at which the hypoxic response is half-maximal [10]. For this reason, a lot of interest has been put in the effect of oxygenation on tumour metabolism and specifically on the Hypoxia Inducible Factor (HIF) protein. This protein, being the main actor in the cell response to hypoxia, is interesting to explore as a potential target for cancer therapy since hypoxic cells are more radioresistant [10, 7].

HIF Structure and Mechanism of action

The HIF protein was discovered by Semenza and co-workers during a study on the erythropoietin (EPO) gene, a gene encoding for the erythropoietin hormone involved in red blood cells production, in 1991 [11]. They found DNA sequences in the gene important for its transcriptional activation in hypoxic conditions, now called Hypoxia Response Elements (HRE). The HIF protein is a heterodimer composed of two subunits HIF-1α and HIF-1β, it acts as a transcription factor by binding to HRE in hypoxic conditions. The subunit HIF-1α is oxygen-sensitive and degraded in presence of oxygen, compared to the constitutively expressed HIF-1β subunit. Three isoforms of the α subunit have been identified: HIF-1α, HIF2-α and HIF3-α. HIF-1α and HIF2-α are the most studied of the three homologs, HIF-1α is expressed ubiquitously in the body while HIF2-α expression is tissue-specific [11]. It has been demonstrated that overexpression or suppression of HIF-1α or HIF2-α influence each other *in vitro* and one homolog can be more expressed than the other. Kidney lesions with early VHL inactivation show more activation of HIF-1α than HIF-2α but this balance can change [12]. Transcriptional activity of HIF-1α requires the binding of the co-factor CBP/p300 to the C-TAD domain of HIF-1α, then HIF will bind to HRE and activate the transcription of its target genes [11, 13, 14].

HIF regulation

Oxygen-dependent regulation of HIF-1α is mainly done by Prolyl Hydroxylase (PHD) and FIH-1 enzymes. They act at the posttranslational level by inducing its degradation or disrupting its interaction

with co-factors. Prolyl Hydroxylase (PHD) proteins catalyze the hydroxylation of proline residues, targeting HIF-1α for proteasomal degradation by the Von Hippel-Lindau (VHL) tumour suppressor protein. Hydroxylation of asparagine residues by Factor Inhibiting HIF-1 (FIH-1) inhibits the interaction between HIF-1α and the important co-factor CBP/p300, preventing regulation of HIF-1α target genes. Since PHD and FIH-1 need oxygen to hydroxylate HIF-1α residues, they act as oxygen sensors in the cell response to hypoxia. Hypoxia promotes HIF-1α protein stability and transcriptional activity. Reactive Oxygen Species (ROS) and oncometabolites such as succinate, fumarate, lactate upregulate HIF-1α [14].

Oxygen-independent mechanisms regulating HIF-1α transcription and translation include PI3K/Akt/mTOR and RAS/RAF/MEK/ERK pathways. Multiple growth factors, oncogenes, mutations (such as in the tumour suppressor genes PTEN and p53) or ROS may increase HIF-1α levels through PI3K and RAS signalling cascade [14, 11, 13]. A study by The Cancer Genome Atlas (TCGA) identified the most altered genes in glioblastoma, it reveals that RTK/RAS/PI3K are among the frequently altered pathways in this disease [15]. It suggests that HIF is a strong candidate for cancer therapy, not only because of its role in the cellular response to hypoxia but also for its frequent deregulation in cancer as well. HIF regulation is summarized in figure 1.

Impact on cellular biological functions

The cell response to hypoxia initiated by HIF affects many biological processes such as cell proliferation, survival, apoptosis, angiogenesis, iron metabolism and glucose metabolism [13]. Pathway enrichment analysis of 98 HIF target genes revealed 20 pathways including those implicated in cancer, glycolysis/gluconeogenesis and metabolism of carbohydrates [16].

HIF can prevent G1/S transition through the regulation of cyclin-dependent kinase inhibitors (p21, p27) and cyclin proteins (cyclin G2, cyclin E) [17]. Cyclin E downregulation is mediated through the inhibition of cyclin D by HIF causing a slowing down or arrest of the cell cycle in the G1 phase and promoting the entry into quiescence, which can be a mechanism to escape chemotherapy [18].

The Tricarboxylic Acid (TCA) cycle (also called Citric Acid or Krebs Cycle) is a circular process fueled by AcetylCoA generating NADH and $FADH_2$ for its use in the Oxidative Phosphorylation (OXPHOS) pathway. Although OXPHOS is the main pathway generating ATP, TCA produce energy in the form of GTP (equivalent of ATP). These processes represent the aerobic pathways used by the cell when oxygen is present for ATP production. Pyruvate produced by the last steps of the glycolysis is turned into Acetyl Coenzyme A by Pyruvate Dehydrogenase (PDH) to fuel the TCA cycle, promoting an oxidative metabolism [4, 19]. However, Pyruvate Dehydrogenase Kinase (PDK) an inhibitor of PDH is upregulated by HIF [20].

When oxygen is not present, the Lactate Dehydrogenase (LDH) enzyme catalyze the reaction in which pyruvate formed by the glycolysis is turned into lactate to generate NAD^+. This last step allows glycolysis to continue in anaerobic condition since NAD^+ is required for pyruvate production. In presence of oxygen, NAD^+ availability is ensured by OXPHOS [4, 19].

Different isoforms of both LDH and PDH enzymes exist. Those isoforms present several differences like kinetics parameters, the tissue or the cellular compartment where the isoforms is expressed. For

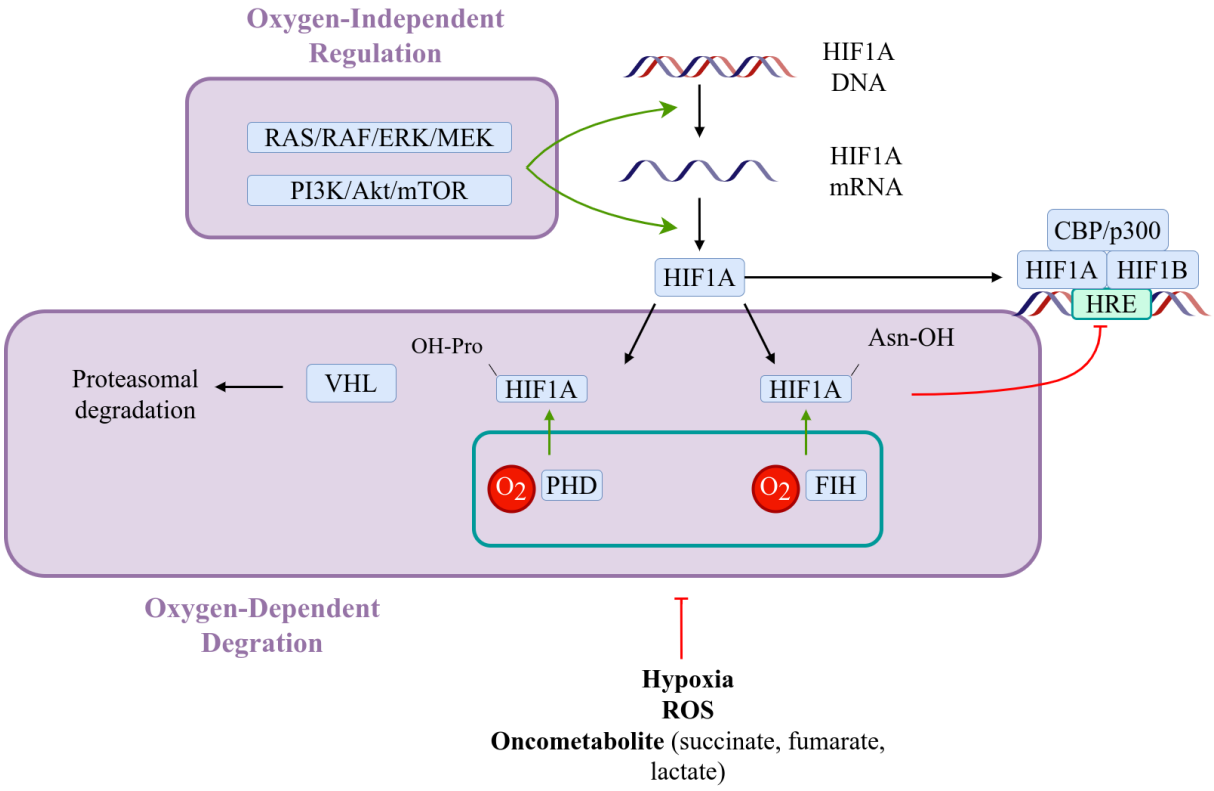

Figure 1: Regulation of Hypoxia Inducible Factor by oxygen-dependent and oxygen-independent mechanisms. PI3K/Akt/mTOR and RAS/RAF/ERK/MEK signalling pathways increase HIF transcription and translation in an oxygen-independent way. The oxygen-dependent regulation relies mainly on the two enzymes: PHD and FIH-1. PHD catalyzes the oxygen-dependent hydroxylation of proline residues on the HIF protein, which is then targeted for proteasomal degradation by the VHL. FIH-1 catalyzes the oxygen-dependent hydroxylation of asparagine residues, which inhibits the interaction between the HIF protein and the CBP/p300 co-factor. Hydroxylation of HIF residues by PHD and FIH-1 is inhibited by hypoxia, ROS and oncometabolites such as succinate, fumarate and lactate.

example, LDH-A is expressed in the skeletal muscle LDH-B in heart [21]. LDH-A is also the isoform comonly upregulated in cancer [22]. The differences between isoforms add a level of complexity. However, this is out of the scope of this study.

The Warburg effect is caused by an increase in glucose utilization by the cells, the glycolysis being one of the pathways affected by hypoxia. HIF increases the expression of glucose transporters GLUT1 and GLUT3 which contain HRE in their promoters, resulting in higher glucose uptake [23]. Furthermore, HIF induces the overexpression of specific glycolytic isoforms for each enzyme involved in all the steps of the glycolysis [21]. Thus, HIF upregulates the expression of LDH, resulting in higher lactate secretion which acidifies the microenvironment. Not only hypoxia will increase the use of glycolysis by the cell, but it will also reduce the use of TCA cycle.

In this paper, we want to study how genetic (or epigenetic) regulations, between HIF and its two targets LDH and PDH, may affect the emergence of the Warburg effect. The Warburg effect results in an increased production of lactic acid by the tumour by metabolizing glucose, even in normoxia [24, 7, 8, 25, 26].

Material and Method

Genetic Regulations

Here, we assume that HIF plays a major role in mediating the cell response to hypoxia. We have selected LDH and PDH to model the effect of hypoxia on metabolism since (1) they are key enzymes for the conversion of pyruvate into lactate/AcetylCoA respectively after the glycolysis, and (2) they are both regulated by HIF directly or indirectly. PDH is downregulated by HIF through its inhibitor PDK, therefore PDK will be included in the model (see figure 2). Genetic regulations are based on the model described by Li *et al* [20]. All genetic regulations are described by the following equations:

$$\frac{du}{dt} = A_u - D_u \times H_{O_2 \to u} \times u \tag{1}$$

$$\frac{dv}{dt} = A_v \times H_{u \to v} - D_v \times v \tag{2}$$

$$\frac{dw}{dt} = A_w \times H_{u \to w} - D_w \times w \tag{3}$$

$$\frac{dz}{dt} = A_z \times H_{w \to z} - D_z \times z \tag{4}$$

where u, v, w and z are HIF, LDH, PDK and PDH levels, respectively; A is a parameter for gene production and D for gene degradation. LDH and PDK upregulations by HIF and PDH downregulations by PDK are described with a non-linear function named the shifted-Hill function. In the same way, the increased HIF protein degradation in normoxia is described using the same function. The shifted-Hill function has the form:

$$H_{Y \to Z} = \frac{S^n}{S^n + Y^n} + \gamma_{Y \to Z} \frac{Y^n}{S^n + Y^n}. \tag{5}$$

Here, n is the Hill coefficient. S is the gene level with a half-threshold of production. The positive parameter γ represents an activation if > 1 or an inhibition if < 1. $H_{Y \to Z}$ represents the effect of the

Parameter	Value	Dimension	Parameter	Value	Dimension
A_u	0.05	1/min	D_u	0.005	1/min
A_v	0.005	1/min	D_v	0.005	1/min
A_w	0.005	1/min	D_w	0.005	1/min
A_z	0.005	1/min	D_z	0.005	1/min
$S_{O_2 \to u}$	0.02085	mmol/L	$S_{u \to v}$	4.48	-
$S_{u \to w}$	5.0	-	$S_{w \to z}$	2.2	-
$\gamma_{O_2 \to u}$	10.0	-	$\gamma_{u \to v}$	3.61	-
$\gamma_{u \to w}$	6.97	-	$\gamma_{w \to z}$	0.14	-

Table 1: Parameters used in genetics regulations. The symbol "-" stands for dimensionless.

regulating gene Y on the regulated gene Z, it can be an upregulation if γ is > 1, or a downregulation if γ is < 1. All genes levels are dimensionless, parameters used in the equation above are summarised in table 1.

Cell Metabolism

Nutrient consumption rates change over time depending on microenvironment conditions. In normoxia, glycolysis transforms glucose to pyruvate, then pyruvate is converted to AcetylCoA by PDH enzymes to feed the TCA cycle. The TCA cycle works in cooperation with OXPHOS to produce ATP using oxygen, which constitutes the aerobic pathway [19]. Since the conversion of pyruvate to AcetylCoA is catalyzed by the PDH enzyme, its availability bounds the use of TCA and should be reflected in the consumption of oxygen. In hypoxia, glucose consumption is increased to produce the ATP needed using aerobic pathways. Pyruvate formed by glycolysis is then turned into lactate by LDH enzymes, increasing acidity in the microenvironment [19]. Like PDH, increased LDH levels should reflect an increased usage of anaerobic pathways with higher consumption of glucose. As PDH and LDH play an important role in the fate of pyruvate, their respective levels should impact cells metabolism in our model.

In equations 6 and 7, we define p_O and p_G, two terms to describe the impact of LDH and PDH on glucose and oxygen consumptions using a sigmoid function based on the logistic funtion. p_O and p_G will adjust the consumptions rates of oxygen and glucose defined in equations 8 and 9, by increasing or decreasing the maximal rates according to the level of PDH and LDH.

$$p_O = \frac{\phi_O - \psi_O}{1 + \exp\left(-l_z(z - z_0)\right)} + \psi_O \tag{6}$$

$$p_G = \frac{\phi_G - \psi_G}{1 + \exp\left(-l_v(v - v_0)\right)} + \psi_G \tag{7}$$

Here, ϕ_O and ϕ_G are the maximal values for p_O and p_G. ψ_O and ψ_G are the minimal values for p_O and p_G. z and v are the current level of PDH and LDH. z_0 and v_0 represent the midpoint of p_O and p_G. l_z and l_v represent the steepness of the curve for p_O and p_G.

Cells consumption and production are described following the functions from the model defined by [27] (a brief description of the complete model is available in supplementary material).

Oxygen consumption is determined using a Michaelis-Menten function [27]:

$$f_O = p_O V_O \frac{O_e}{O_e + K_O} \tag{8}$$

PDH allows the pyruvate to enter the TCA cycle as Acetyl Coenzyme A, it is a limiting step in the aerobic pathway. This is included in the model by adjusting the maximum oxygen consumption rate V_O using the term p_O to represent PDH level effect on metabolism. O_e is the extracellular oxygen concentration. K_O is the extracellular oxygen concentration at which the cell oxygen consumption rate is half-maximum.

Following Robertson-Tessi *et al* [27], we assume that ATP demand drives glucose consumption. In low oxygen conditions, the cell will consume more glucose to produce ATP in the last step of the glycolysis, then pyruvate is turned into lactate by the LDH enzyme. An increase of LDH indicates an upregulation of anaerobic pathways which means here, an increase in glucose consumption. We use the term p_G to increase glucose consumption in the equation 9 when levels of LDH increase [27].

$$f_G = \left(\frac{p_G A_0}{2} - \frac{29 f_O}{10} \right) \frac{G_e}{G_e + K_G} \tag{9}$$

A_0 is the target ATP production. G_e is the extracellular glucose concentration. K_G is the extracellular glucose concentration at which the glucose consumption rate is half-maximal.

In this paper, we are studying how HIF can impact the interplay between aerobic (TCA + OXPHOS) and anaerobic (glycolysis +lactate secretion) pathways to generate ATP, due to its PDH and LDH enzymes important for conversion of pyruvate. Therefore, we do not directly model aerobic and anaerobic pathways but rather we compute the theoretical level of ATP generated by both processes (equation 12). We take the same stoichiometric coefficients as in [27]: glycolysis uses 1 mole of glucose produces 2 moles of ATP, aerobic pathway uses 1 mole of glucose and 5 moles of oxygen to produce 29 moles of ATP. We can compute the ATP produced from the nutrients consumed using the yield from glycolysis and aerobic pathway [27]:

$$f_A = 2 f_G + \frac{29 f_O}{5} \tag{10}$$

Glycolysis produces 2 moles of pyruvate with 1 mole of glucose. If oxygen is absent, pyruvate is turned into lactate, giving a total of 2 moles of lactate [4]. Lactic acid production is given by the glucose consumed:

$$f_{H^+} = k_H 2 f_G \tag{11}$$

k_H is a fixed parameter for proton buffering (dimensionless).

Quantity of ATP produced by the cell is modelled by an ODE, and extracellular quantities of the three molecules oxygen (O), glucose (G) and protons (H^+) are described by PDEs in the following equations:

$$\frac{dA}{dt} = f_A \tag{12}$$

$$\frac{\partial O}{\partial t} = D_O \nabla^2 O - \sum_{k=1}^{N_i} f_O^k \tag{13}$$

$$\frac{\partial G}{\partial t} = D_G \nabla^2 G - \sum_{k=1}^{N_i} f_G^k \tag{14}$$

$$\frac{\partial H^+}{\partial t} = D_{H^+} \nabla^2 H^+ + \sum_{k=1}^{N_i} f_{H^+}^k \tag{15}$$

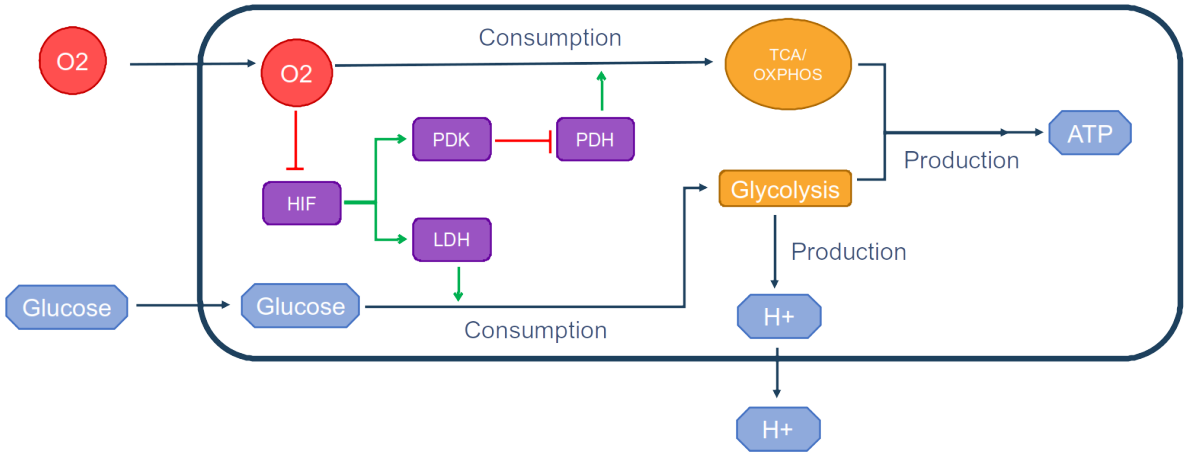

Figure 2: Cell metabolism and genetic regulations implemented in the model. Green arrows represent upregulation, red arrows represent inhibition.

D_O, D_G and D_{H^+} are the diffusion coefficient for each molecules. N_i is the number of cells in the voxel i. Initial values for oxygen in equation 13 are $O(x, y, 0) = 0.056$ mmol/L in normoxia and 0.012 mmol/L in hypoxia. Initial value for glucose in equation 14 is $G(x, y, 0) = 5.0$ mmol/L. Initial value for protons in equation 15 is $H^+(x, y, 0) = 3.98.10^{-5}$ mmol/L (pH 7.4). Let x_0 and y_0 the lower boundary of the domain in x and y, x_L and y_L the upper boundary in x and y. Boundary values for oxygen in equation 13 are $O(x_0, y, t) = O(x_L, y, t) = O(x, y_0, t) = O(x, y_L, t) = 0.056$ mmol/L in normoxia and 0.012 mmol/L in hypoxia. Boundary values for glucose in equation 14 are $G(x_0, y, t) = G(x_L, y, t) = G(x, y_0, t) = G(x, y_L, t) = 5.0$ mmol/L. Boundary values for H^+ in equation 15 are $H^+(x_0, y, t) = H^+(x_L, y, t) = H^+(x, y_0, t) = H^+(x, y_L, t) = 3.98.10^{-5}$ mmol/L (pH 7.4). Parameters used in those functions are summarised in Table 2. The schematic in figure 2 shows the cellular metabolism and the genetics regulation implemented in the model.

Numerical Implementation

The tumour microenvironment plays a vital role in the growth and progression of tumour cells. As the tumour grows, intracellular and intercellular interactions influence the changes in its microenvironment, which can further result in cells dynamic. Here, we aim to develop a modelling framework to simulate the growth of a large population of cells cultured *in vitro*, each cell having its metabolism influenced by the microenvironment conditions to represent accurately the resources dynamics in the tumour. Therefore, the numerical implementation of the model must have sufficient performance to simulate the behaviour of thousands of cells. In this regard, we selected Physicell, an open-source C++ framework designed to run simulations containing a large population of cells. This framework has good performance with a low memory footprint, allows the user to implement his custom code and define custom cell types, run a multi-agent-based simulation in 2D or 3D [28].

Most aspect of the model are handled by the physicell software [28], this includes : cell division and progression through the cell cycle, cell adhesion and repulsion, substrate diffusion, cell exchanges with

Parameter	Value	Unit
V_O	0.01875	mmol/L/min
K_O	0.0075	mmol/L
K_G	0.04	mmol/L
k_H	2.5e-4	-
A0	0.10875	mmol/L/min
ϕ_G	50	-
ψ_G	1	-
l_G	4	-
v_0	2.35	-
ϕ_O	1	-
ψ_O	0	-
l_O	15	-
z_0	0.575	-
D_O	109,200	μm^2/min
D_G	30,000	μm^2/min
D_{H^+}	27,0000	μm^2/min

Table 2: Parameters for metabolism. The symbol "-" stands for dimensionless.

the environment (secretion and consumption). Cells are modelled with the shape of a sphere that cannot deform, adhesion and repulsion are implemented using a potential function. The cell division process is implemented as a cycle, where the user can define each steps and and progression between them. There is no condition on the neighbourhood, a cell will divide even if it is surrounded by other cells as long as there is sufficient nutrient. As a consequence, certain regions of the tumour will exhibit a higher cell density. We implemented in Physicell an heterogeneous diffusion with respect to local cellular density. In the model, phases duration are 5h in G1, 8h in S, 4h in G2 and 1h in M, for a total of 18h to complete a cell cycle [29].

Here, the impact of extracellular oxygen concentration is studied considering different boundary conditions: physiological normoxia at 0.056 mmol/L (5% O_2), pathological hypoxia at 0.01112 mmol/L (1% O_2) and a last where boundary conditions are modified during the simulation from physiological normoxia to pathological hypoxia. The hypoxia threshold is set at 0.02085 (2% O_2), the level at which HIF has a half-maximal response [10].

The governing ODEs (equation 1 - 4 and 12) and PDEs (equation 13 - 15) are run at each timestep to compute cell nutrient consumption, energy and acidity production for that period. After each time step, the cell state is updated according to the quantity of ATP generated and the extracellular pH. Therefore, cells can proliferate and divide only if they were able to generate enough ATP and if extracellular pH is higher than the acid resistance of the cell (6.1 [27]). If the quantity of ATP generated is less than a threshold $ATP_{quiescence}$, the cell enters quiescence and is then prevented to complete the G1 phase. If the

quantity of ATP generated is less than a threshold ATP_{death} or if the pH is less than a threshold pH_{death}, the cell dies and enters into the death cycle where it is progressively removed from the microenvironment by lysis.

To simulate the cell entry into quiescence, we created a phase G0 with a reversible link to the G1 phase of the cell cycle. If the condition for proliferation are not met, we set the transition rate from G1 to G0 at a maximum value and the rate from G0 to G1 at 0. The cell is forced to enter the G0 phase and is prevented to transit to the G1 phase to continue its division. Once the level of ATP rises again, we revert the transition rates values to allow the cell to leave the G0 phase and divide anew. The cell can only transit to the G0 phase from the G1 phase, thus it will complete its cycle once it leaved the G1 phase and will divide even if ATP levels fall while the division process is ongoing.

Results

Qualitative exploration of the model at the cell scale

A well-known phenomenon is the Warburg effect, increased production of lactic acid by the tumour [5] even in normoxia [24, 7, 8]. A qualitative study of the genetic deregulations at the cell scale would reveal how it impacts lactic-acid production to investigate the appearance of the Warburg effect. The primary aim of this study is to investigate the role of genetic regulations in cell metabolic changes.

In our mathematical model, the regulating effect of a gene on another is mainly driven by the γ parameter in the shifted-Hill function. Setting this parameter equal to 1 simulate a loss of the regulating function. An over-sensitivity of a gene by its regulator is modelled by setting the γ parameter to 40, the maximum defined in the model from [20]. Results of a few regulations are shown in figure 3.

When no genetic deregulations are applied to the model (figure 3.A), protons production range from 0.0001 mmol/L/min to 0.001 mmol/L/min with normal γ parameters. Around 0.01 mmol/L oxygen (1%), the cell progressively increases its H^+secretion rate from 0.0001 mmol/L/min to the maximum 0.001 mmol/L/min.

In our model, when HIF is not subjected to oxygen degradation (figure 3.B), the rate of H^+production is only influenced by glucose concentration. In this case, cell's lactic-acid secretion rate can reach 0.001 mmol/L/min even in normal oxygen pressure, as a result of the Warburg effect. Increased degradation of HIF in oxygen (figure 3.C) reduces the oxygen threshold at which the cell has a lactic-acid secretion rate of 0.001 mmol/L/min. Lower levels of oxygen are needed to reach the maximal secretion rate compared to the normal degradation rate of HIF. With no deregulation (figure 3.A), the lactic-acid secretion rate starts to increase at around 0.019 mmol/L of oxygen and reach a maximum at around 0.08 mmol/L. With increased HIF degradation by oxygen (figure 3.C), this span is reduced and lactic-acid secretion increases at around 0.012 mmol/L of oxygen. Similar to our result, a model from [25] shows that a lower degradation rate of HIF increases the chance that cells use glycolysis instead of OXPHOS, which will increase lactic acid secretions.

Inhibiting LDH sensitivity to HIF (figure 3.D) causes the maximum lactic-acid secretion rate to fall to 0.0008 mmol/L/min. Increasing LDH sensitivity to HIF does not permit the cell to have a higher

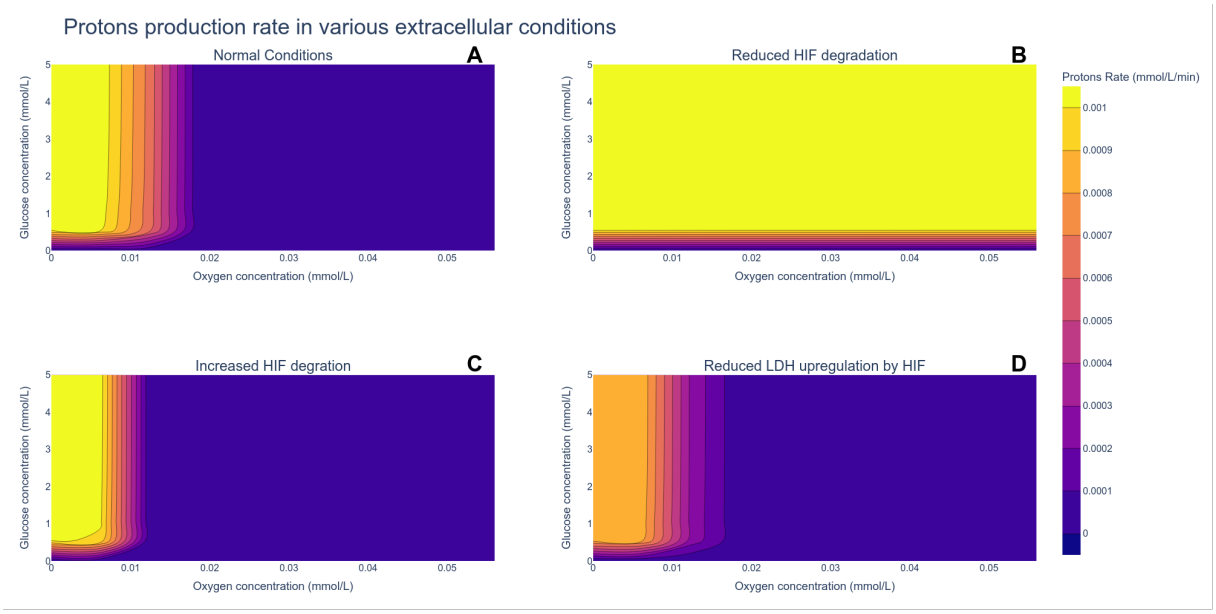

Figure 3: Influence of genetic upregulation or inhibition on the production rate of protons at different glucose and oxygen concentrations. (A) Result with no genetic deregulation. ($\gamma_{O_2 \to u} = 10.0$, $\gamma_{u \to v} = 3.61$, $\gamma_{u \to w} = 6.97$, $\gamma_{w \to z} = 0.14$) (B) Result with inhibition of the oxygen-dependant degradation of HIF. ($\gamma_{O_2 \to u} = 1.0$, $\gamma_{u \to v} = 3.61$, $\gamma_{u \to w} = 6.97$, $\gamma_{w \to z} = 0.14$) (C) Result with over-degradation of HIF by oxygen. ($\gamma_{O_2 \to u} = 40.0$, $\gamma_{u \to v} = 3.61$, $\gamma_{u \to w} = 6.97$, $\gamma_{w \to z} = 0.14$) (D) Result with loss of upregulation of LDH by HIF. ($\gamma_{O_2 \to u} = 10.0$, $\gamma_{u \to v} = 3.0$, $\gamma_{u \to w} = 6.97$, $\gamma_{w \to z} = 0.14$)

H^+production rate in normoxia, while a decrease prevents a high H^+production rate in hypoxia (results not shown).

Interfering with PDK sensitivity to HIF or PDH sensitivity to PDK seems to have no effect on acid production in the model but on oxygen consumption by the cell (results not shown).

Exploration of environment and genetic properties on the emergence of the Warburg phenotype

Influence of environmental oxygen conditions

The Warburg effect is currently defined as high production of acidity due to the use of glycolysis even in normoxia [25, 26, 9]. We ran several simulations with different environmental oxygen conditions to assess whether microenvironmental conditions only, can induce a Warburg effect in the model.

Figure 4 shows how oxygen conditions affect tumour growth. In oscillating conditions, the oxygen concentration varies between physiological normoxia and pathological hypoxia and reverse every 6 hours until the end of the simulation. Kinetics of HIF show a peak after 6h and a decrease to an equilibrium state after 24h-48h. We choose to simulate 6h-period of hypoxia/normoxia to avoid the cell reaching an equilibrium and to simulate stressful conditions with a high response to a low level of oxygen. Constant hypoxia slows down tumour growth and reduces tumour diameter compared to normoxia. In all 3 different conditions, the centre of the tumour is composed of dead cells surrounded by living cells at the

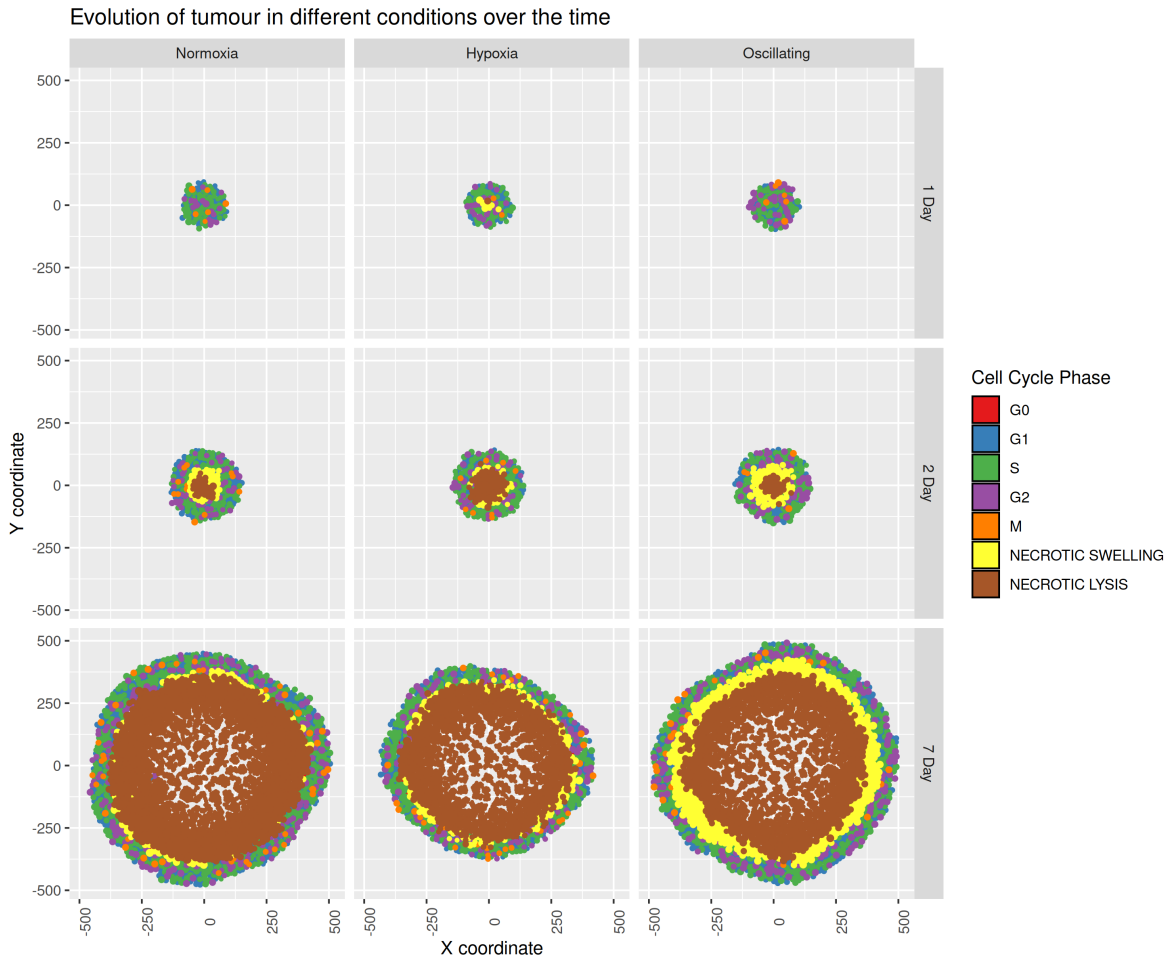

Figure 4: Evolution of tumour growth at different times in different conditions. In oscillating conditions, the oxygen concentration is slowly decreased from normoxia to hypoxia during 6 hours, then cells are slowly put back in normoxia at the same rate. This process is repeated until the end of the simulation.

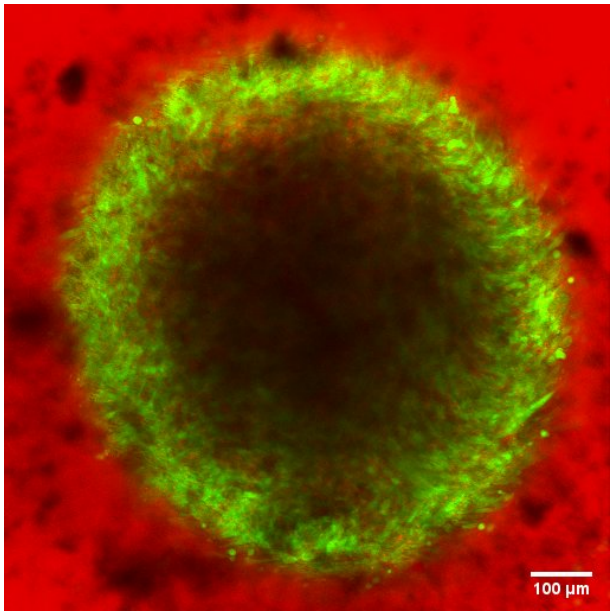

Figure 5: Picture of a spheroid grown for 30 days in an experiment run in the laboratory. Cells were marked using the fluorescent proteins Green FLuorescent Protein (GFP) and Sulforhodamine B (SRB). Living cells are colored in green, dying cells appear in red. The centre of the tumour is composed of hypoxic and dead cells, both do not emit fluorescence.

periphery. Only in normoxia and varying oxygen conditions, some cells in the centre of the tumour do continue to divide (only visible after 7 days of growth). This may be due to the changes in the tumour microenvironment with the increased cell death at the centre. As the cells die, more nutrients will be available to quiescent cells to enable them to renter proliferating phase. Moreover, spatial changes due to the shrinkage of dead cells can influence the availability of nutrients at the centre. This might show a mechanism by which the tumour can grow back after a period of harsh conditions, for example quiescence can be a mechanism to avoid drugs effect for the tumour cell [30]. Necrotic core has been observed in biological experiments run in the lab (figure 5).

It seems that varying the concentration of oxygen from normoxia to hypoxia, and reversing this process, every 6 hours does not affect the diameter of the tumour at the end of the simulation. However, a ring of necrotic cells in the swelling phase appears thicker than in other conditions.

Results in figure 6 show acid production according to the extracellular oxygen concentration. Red line y-axis intercept is equal to 0.02085 mmol/L (2% O_2), which corresponds to the threshold of hypoxia in physiological conditions. It is the level at which HIF has a half-maximal response as well [10]. Cells above this level are considered to be in normoxia while the rest of the cells are in hypoxia. Levels of extracellular oxygen fall below the hypoxia threshold after 2 days of growth in normoxic conditions (a necrotic core in the centre of the tumour has already formed). Due to poor oxygen concentration, cells with higher glycolytic activity appear and reach a H^+production rate of almost 5×10^{-4}mmol/L/min. The maximum glycolytic activity of cells falls at 7 days of growth because of reduced glucose availability. When tumour growth is started in hypoxic conditions, high glycolytic activity is present after only one day of growth. In those conditions, the way the cell produces its energy is influenced only by glucose

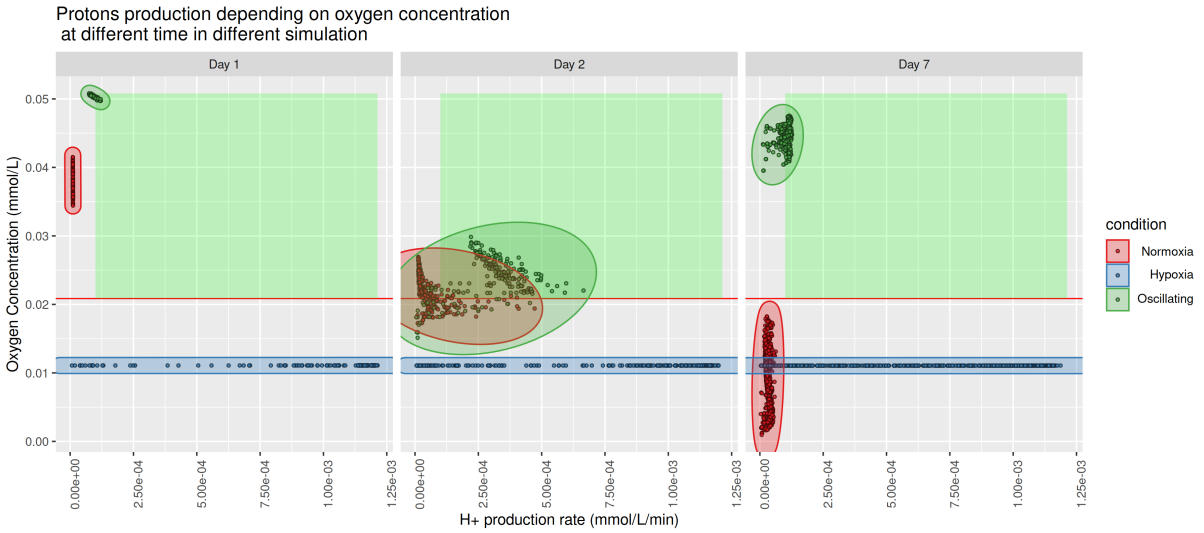

Figure 6: Acid production rate following oxygen extracellular concentrations at different times in different conditions. The red line indicates the hypoxia threshold. In oscillating conditions, oxygen concentration is slowly decreased from normoxia to hypoxia during 6h, then oxygen is increased to normoxia at the same rate. This process is repeated until the end of the simulation. Only living cells are represented on the graph. The green rectangle represents the region corresponding to a Warburg effect.

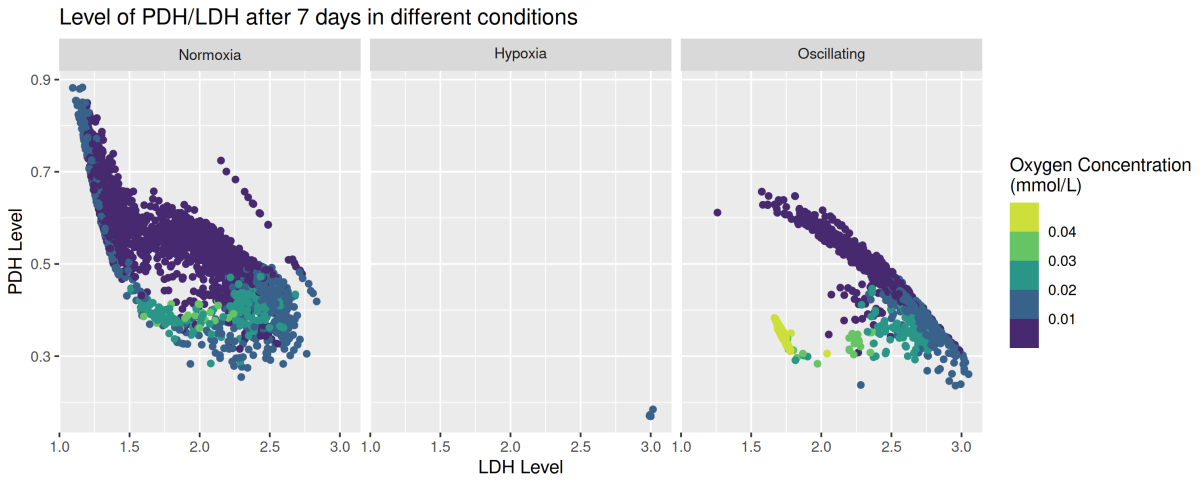

Figure 7: Plot of level of PDH against the level of LDH coloured by the extracellular oxygen concentration. The graph shows the results after 7 days of growth for different conditions. In oscillating conditions, oxygen concentration is slowly decreased from normoxia to hypoxia during 6h, then oxygen is increased to normoxia at the same rate. This process is repeated until the end of the simulation. Only living cell are represented on the graph.

concentrations (similar to the result shown in figure 3). Therefore, hypoxic conditions directly select cells with high glycolytic activity.

The fact that, in the model, hypoxia may select cells with high glycolytic activity is supported by the levels of LDH/PDH genes presented in figure 7. In normoxia, cells have a level of LDH and PDH of 1 for both, it can be associated to an oxidative state. In hypoxia, LDH level reaches 3.0 and PDH level falls to 0.25, it can be associated to a glycolytic state. At the beginning of the simulation in normoxic conditions, cells have 1:1 LDH/PDH levels. As the simulation goes, oxygen becomes less available. Thus LDH level increases while PDH level decreases. The result in normoxic conditions shows that cells migrate from an oxidative to a glycolytic state as oxygen concentration decreases. Cells around 2:0.5 LDH/PDH levels have a hybrid state where they rely on both nutrients to produce ATP. Again hypoxia selects for cells with high levels of LDH and low levels of PDH, suppressing the possibility for the cell to adopt a hybrid state.

Interestingly, extracellular oxygen concentration after 7 days is higher when oxygen varies between normoxia and hypoxia every 6 hours than in constant normoxia. Since cells are put in hypoxia several times a day, they rely more on glycolysis and consume less oxygen. Cells with higher glycolytic activity (2.5×10^{-4}mmol/L/min) even above the threshold of hypoxia appear at 2 days. It suggests that the Warburg Effect can be caused by environmental conditions with rapid variations. Combined with figure 7, genetic levels seems to indicate that cells cannot enter a complete oxidative state and are trapped either in a hybrid or a glycolytic state.

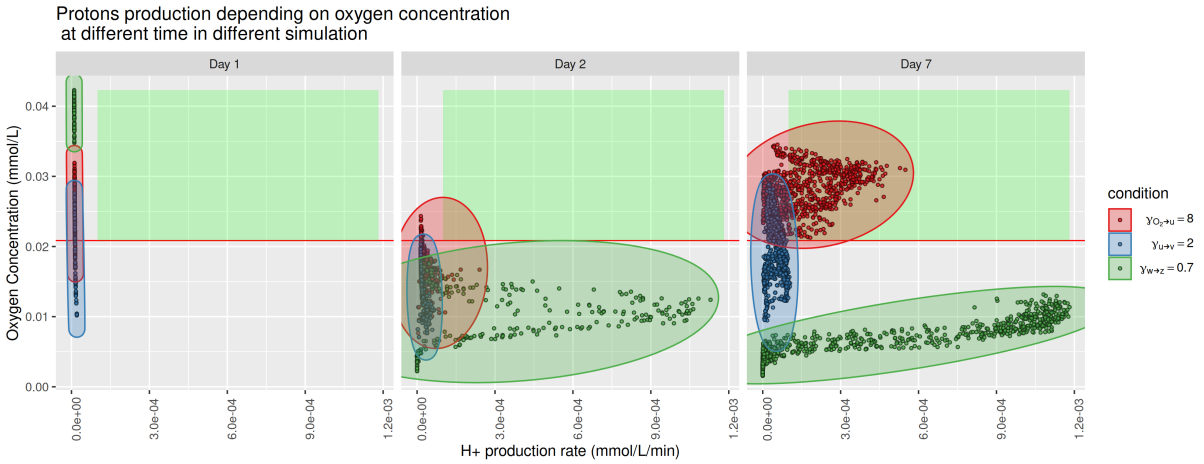

Figure 8: Acid production rate following oxygen extracellular concentrations at different times with different genetic perturbations. The red line indicates the hypoxia threshold. Only living cell are represented on the graph. Three genetics pertubations have been selected: reduced oxygen induced degradation of HIF ($\gamma O_2 \to HIF = 8.0$), lower use of glycolysis in hypoxic conditions ($\gamma HIF \to LDH = 2.0$) and lower effect of hypoxia on oxygen consumption ($\gamma PDK \to PDH = 0.7$). Tumour growth was initiated in normoxia. The green rectangle represents the region corresponding to a Warburg effect.

Influence of the intrinsic genetic properties of the cell

Here, tumour growth is initiated in normoxia (results not shown). Extracellular oxygen concentrations only vary due to cells consumption and reduced diffusion in the tumour. Only genetics regulations have been modified between each simulation to assess the impact of different genetic deregulations (mutations or epigenetic alterations) on tumour growth and cell metabolism. Results are similar to normoxic conditions with no genetic deregulations (presented in figure 4). When reducing inhibition of PDH by PDK, tumour radius at 7 days of growth is lower than in normoxia and higher than in hypoxia with no mutation.

Figure 8 shows that cells start to become hypoxic after day 1, reaching a majority by day 2. After 7 days with a reduced HIF degradation rate by oxygen, extracellular oxygen goes back to normoxic levels yet cells have a higher acid production rate that corresponds to a Warburg effect. In this case, we suppose that cells slowly drain oxygen levels in the environment to a point where hypoxia is reached. Due to poor oxygen conditions, cells adapt their metabolism to enter a glycolytic state that they keep even if the oxygen supply goes back above 2 $\%O_2$. Together with the result in figure 9, this might be caused by a delay in the response from returning to normal conditions since HIF regulation by O_2 is affected. While some cells have levels of LDH greater than 2 and PDH lower than 0.50 (hybrid to glycolytic state), some

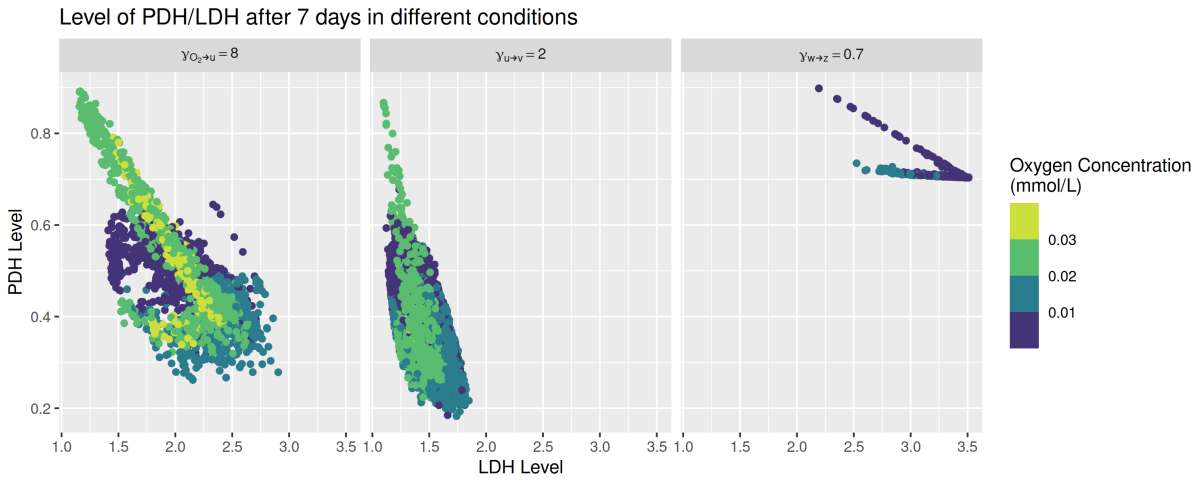

Figure 9: Plot of level of PDH against the level of LDH coloured by the extracellular oxygen concentration. The graph shows the results after 7 days of growth with three different genetic perturbations: reduced oxygen induced degradation of HIF ($\gamma O_2 \rightarrow HIF = 8.0$), lower use of glycolysis in hypoxic conditions ($\gamma HIF \rightarrow LDH = 2.0$) and lower effect of hypoxia on oxygen consumption ($\gamma PDK \rightarrow PDH = 0.7$). Tumour growth was initiated in normoxia.

have a ratio of LDH/PDH almost equal to 1:1. This suggests that the Warburg Effect is not irreversible with a reduced HIF degradation rate by oxygen alone.

As expected, reducing the increase in LDH levels due to HIF response does not induce a high acidification rate in normoxia but affects the maximum acid production rate and level of LDH. Instead of inducing a glycolytic phenotype, it seems to repress it.

Reducing the inhibiting power of PDK on PDH allows the cell to keep a higher PDH level, a key enzyme for oxygen consumption and oxidative state in the model. Cells exhibit an acid production rate similar to those in hypoxic conditions after 2 and 7 days, compared to other genetics deregulation. While in normoxia with no genetic deregulation cells seem to fluctuate around the threshold of hypoxia, here they are all below this level. Since PDH is not effectively regulated by HIF, the cell tends to stay in an oxidative state and rely less on glycolysis. We can suppose that cells consume oxygen even when the level fall, creating further harder conditions. Results also show that adaptation to hypoxia is delayed and the cell only adopts a glycolytic state at oxygen conditions near-pathological hypoxia. PDH levels do not fall far below 0.75 even after 7 days of growth compared to others conditions, indicating that cells can only adopt an oxidative or hybrid state.

Discussion

In this paper, we formulated a mathematical model to study the impact of HIF on LDH and PDH, key enzymes of glycolysis and TCA cycle and thus investigating its role in cellular metabolism. Since its discovery, HIF has been actively studied by the scientific community. There are several modelling approaches to study the effects of HIF [27, 20, 18, 25] and here, we investigate its role using a multi-agent

model, considering a heterogeneous environment that changes over time. Furthermore, the model is used to investigate the impact of genes on metabolism and the effect of different environmental conditions and different genetic deregulations (such as mutations or epigenetic alterations) can have on the Warburg Effect, an overproduction of acidity due to and increased glycolysis even in normoxia. Over-production of lactate can also be caused by reduced use of pyruvate in the mitochondria, remaining pyruvate is then turned into lactate.

Using the level of LDH and PDH genes as markers, we can define three different metabolic states like [20, 25]: oxidative, glycolytic and hybrid. The oxidative state corresponds to a high level of PDH and a low level of LDH, and inversely in a glycolytic state. The hybrid state then corresponds to medium levels of both enzymes, 2:0.5 for LDH and PDH respectively. As expected, normoxia strongly selects for the first state while hypoxia selects for the second one. The hybrid state is observed as the oxygen levels change over time due to tumour growth. Thus it appears that the cell adopts this state when adapting to changing oxygen conditions or when oxygen levels vary between normoxia and hypxia several times during tumour growth (oscillating conditions in the model).

We observed some differences between our model and the model in a recent paper from Li *et al* [20]: (1) they identified a normal state with a level of LDH at 1 and a level of PDH at 0.1, (2) their oxidative and glycolytic states have different levels of genes than those present in our model. This difference in the result can be explained by the fact that we only include a small fraction of their gene regulation network in our model, to only account for the effect caused by HIF.

We have simulated tumour growth when oxygen supply does not vary over time, hence differences in extracellular oxygen level can only be caused by cell consumption or reduced diffusion owing to higher cell density. We found that when there are rapid changes in oxygen supply to the tumour, cells with higher glycolytic rates above the threshold of hypoxia appear. It shows that varying microenvironmental conditions are sufficient to induce a Warburg phenotype for the cell. The results are inline with the findings by Damaghi *et al* [9]. However, the model does not include sudden genetic mutation which can be caused by harsh conditions. Therefore, in our case cell would not be trapped into a Warburg phenotype and this state can be reversed to a normal state if the cell is given enough time in favourable conditions. Lactate secretion, which decreases the extracellular pH, depends on glucose consumption. A study from Casciari *et al* [31] has shown that a lower extracellular pH decreases dramatically glucose consumption, the Warburg effect could also be inhibited by low pH (6.95). We may suppose that after difficult conditions genes may be over-expressed or inhibited which will force the cell to adopt a Warburg phenotype.

The importance of HIF degradation in normoxia is further highlighted by the model results. We were also able to induce a Warburg effect by reducing the degradation rate of HIF by oxygen-dependent enzymes. Our results show that this effect only appears after a first period of hypoxia. It suggests that HIF accumulation forces the cell to adopt a glycolytic state and prevent it from returning to an oxidative state in normoxia. HIF inhibition therapy would prevent the appearance of Warburg cell type in cancer. PI3K and mTOR, two genes that increase HIF level independently of the level of oxygen [11, 14, 13], are studied as potential targets in anti-cancer therapy due to their altered expression in cancer and their role

in signalling pathways affecting many biological functions [32, 33], possibly causing HIF overexpression. AMPK enzyme is known to interact with HIF [25] and inhibits its expression, some evidence link this gene to anti-tumour activity [34]. Those interactions could be added in further modelling work to study their impact on the Warburg effect as they may be important players interacting with HIF.

It has been shown that extracellular pH can (1) influence the cell metabolism (reduce glucose consumption, increase the cells doubling time) [31], (2) affect the ability of tumour cells to form metastasis, invade other tissue or migrate [35] and (3) could be a mechanism of invasion [36]. Currently, therapy targeting extracellular pH in the tumour are under development. Moreover, pH also affects the efficiency of different drugs such as temozolomide [37]. Reducing the increase in LDH level by the cell response to hypoxia lowered the rate of acid production in our simulation. Inhibitor of LDH could be used in combination with pH targeting therapy to improve treatment outcomes.

Reducing the down-regulation of PDH by HIF in the model forces the cell to rely as much as possible on oxygen to produce its energy. Herein, changes in metabolism toward glycolytic activity requires lower levels of oxygen. A study has shown that inhibition of HIF resulted in reduced lactate production, increase in oxygen consumption and radiotherapy sensitivity [7]. Whether increasing oxygen consumption by PDH upregulation would result in better outcomes in therapy in the model remains to be studied.

Conclusion

The main interest of the model is its ability to qualitatively describe HIF expression in tumour development over time with oxygen diffusion that depends on both cell consumption and cell density in the tumour to obtain a more realistic diffusion. Results of the model show that varying oxygen levels and reduced HIF degradation can cause increased glycolytic activity in normal oxygen levels. In the model, the emergence of the Warburg effect is preceded by a first period of hypoxia before returning to normoxic concentrations. This suggests that adaptation to environmental conditions is the primary phenomenon to understanding the Warburg effect. Interfering with the genetic activity of HIF or its effect on LDH and PDH may be used in therapy to induce specific behaviour in the cell.

Acknowledgements

We thank Alaa Tafech for providing the picture of the spheroid in Figure 5. This project has received financial support from CNRS through the MITI interdisciplinary programs.
Kévin SPINICCI gratefully acknowledges the support of Swansea University Strategic partnership Research Scholarship and the support of IDEX Université Grenoble Alpes.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] R. A. Bender. "Glycolysis". In: *Brenner's Encyclopedia of Genetics: Second Edition* 2 (2013), pp. 346–349. DOI: 10.1016/B978-0-12-374984-0.00659-8.

[2] Thomas C. King. "Cell Injury, Cellular Responses to Injury, and Cell Death". In: *Elsevier's Integrated Pathology* (2007), pp. 1–20. DOI: 10.1016/b978-0-323-04328-1.50007-3.

[3] F K Zimmermann. "Glycolysis in Saccharomyces cerevisiae". In: *Encyclopedia of Genetics* (2001), pp. 885–888. DOI: 10.1006/rwgn.2001.0570.

[4] Blanco, Antonio and Gustavo Blanco. "Medical Biochemistry . Academic Press, 2018." In: (2017), pp. 283–323. DOI: 10.1016/B978-0-12-803550-4/00014-8.

[5] Otto Warburg, Franz Wind, and Erwin Negelein. "The metabolism of tumors in the body". In: *Journal of General Physiology* 8.6 (1927), pp. 519–530. ISSN: 15407748. DOI: 10.1085/jgp.8.6.519. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2140820/.

[6] Pierre Jacquet and Angélique Stéphanou. "Metabolic Reprogramming, Questioning, and Implications for Cancer". In: *Biology 2021, Vol. 10, Page 129* 10.2 (Feb. 2021), p. 129. ISSN: 20797737. DOI: 10.3390/BIOLOGY10020129. URL: https://www.mdpi.com/2079-7737/10/2/129/htm%20https://www.mdpi.com/2079-7737/10/2/129.

[7] Eric Leung et al. "Metabolic targeting of HIF-dependent glycolysis reduces lactate, increases oxygen consumption and enhances response to high-dose single-fraction radiotherapy in hypoxic solid tumors". In: *BMC Cancer* 17.1 (Dec. 2017), p. 418. ISSN: 1471-2407. DOI: 10.1186/s12885-017-3402-6. URL: https://bmccancer.biomedcentral.com/articles/10.1186/s12885-017-3402-6.

[8] Ian F. Robey et al. "Hypoxia-inducible factor-1α and the glycolytic phenotype in tumors". In: *Neoplasia* 7.4 (2005), pp. 324–330. ISSN: 15228002. DOI: 10.1593/neo.04430. URL: /pmc/articles/PMC1501147/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1501147/.

[9] Mehdi Damaghi et al. "The Harsh Microenvironment in Early Breast Cancer Selects for a Warburg Phenotype". In: *bioRxiv* (Apr. 2020), p. 2020.04.07.029975. DOI: 10.1101/2020.04.07.029975. URL: https://doi.org/10.1101/2020.04.07.029975.

[10] S. R. McKeown. *Defining normoxia, physoxia and hypoxia in tumours - Implications for treatment response*. Mar. 2014. DOI: 10.1259/bjr.20130676. URL: /pmc/articles/PMC4064601/%20/pmc/articles/PMC4064601/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4064601/.

[11] Georgina N. Masoud and Wei Li. "HIF-1α pathway: Role, regulation and intervention for cancer therapy". In: *Acta Pharmaceutica Sinica B* 5.5 (2015). ISSN: 22113843. DOI: 10.1016/j.apsb.2015.05.007. URL: https://www.sciencedirect.com/science/article/pii/S2211383515000817?pes=vor.

[12] J. Xu et al. "Epigenetic regulation of HIF-1α in renal cancer cells involves HIF-1α/2α binding to a reverse hypoxia-response element". In: *Oncogene* 31.8 (2012). ISSN: 09509232. DOI: 10.1038/onc.2011.305. URL: https://pubmed.ncbi.nlm.nih.gov/21841824/%20https://www.nature.com/articles/onc2011305.

[13] Ji-Won Lee et al. "Hypoxia-inducible factor (HIF -1)alpha : its protein stability and biological function s". In: *Experimental and Molecular Medicine* 36.1 (2004), pp. 1–12. DOI: 10.1038/emm.2004.1. URL: https://www.nature.com/articles/emm20041.

[14] Yoshihiro Hayashi et al. "Hypoxia/pseudohypoxia-mediated activation of hypoxia-inducible factor-1α in cancer". In: *Cancer Science* 110.5 (May 2019), pp. 1510–1517. ISSN: 1347-9032. DOI: 10.1111/cas.13990. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cas.13990.

[15] Roger McLendon et al. "Comprehensive genomic characterization defines human glioblastoma genes and core pathways". In: *Nature* 455.7216 (2008), pp. 1061–1068. ISSN: 00280836. DOI: 10.1038/nature07385. URL: https://www.nature.com/articles/nature07385.

[16] Lucija Slemc and Tanja Kunej. "Transcription factor HIF1A: downstream targets, associated pathways, polymorphic hypoxia response element (HRE) sites, and initiative for standardization of reporting in scientific literature". In: *Tumor Biology* 37.11 (2016). ISSN: 14230380. DOI: 10.1007/s13277-016-5331-4. URL: https://link.springer.com/article/10.1007/s13277-016-5331-4.

[17] Nobuhito Goda, Sara J. Dozier, and Randall S. Johnson. *HIF-1 in cell cycle regulation, apoptosis, and tumor progression*. July 2003. DOI: 10.1089/152308603768295212. URL: https://www.liebertpub.com/doi/abs/10.1089/152308603768295212.

[18] B. Bedessem and A. Stéphanou. "A mathematical model of HiF-1α-mediated response to hypoxia on the G1/S transition". In: *Mathematical Biosciences* 248.1 (2014), pp. 31–39. ISSN: 00255564. DOI: 10.1016/j.mbs.2013.11.007. URL: http://dx.doi.org/10.1016/j.mbs.2013.11.007.

[19] Jerry J. Zimmerman, Amélie von Saint André-von Arnim, and Jerry McLaughlin. *Cellular Respiration*. Fourth Edi. Elsevier, 2011, pp. 1058–1072. ISBN: 9780323073073. DOI: 10.1016/B978-0-323-07307-3.10074-6. URL: http://dx.doi.org/10.1016/B978-0-323-07307-3.10074-6.

[20] Wenbo Li and Jin Wang. "Uncovering the Underlying Mechanisms of Cancer Metabolism through the Landscapes and Probability Flux Quantifications". In: *iScience* 23.4 (Apr. 2020), p. 101002. ISSN: 25890042. DOI: 10.1016/j.isci.2020.101002. URL: https://doi.org/10.1016/j.isci.2020.101002.

[21] Alvaro Marin-Hernandez et al. "HIF-1alpha Modulates Energy Metabolism in Cancer Cells by Inducing Over-Expression of Specific Glycolytic Isoforms". In: *Mini-Reviews in Medicinal Chemistry* 9.9 (Aug. 2009), pp. 1084–1101. ISSN: 13895575. DOI: 10.2174/138955709788922610. URL: http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1389-5575&volume=9&issue=9&spage=1084.

[22] C. Granchi et al. “Inhibitors of Lactate Dehydrogenase Isoforms and their Therapeutic Potentials”. In: *Current Medicinal Chemistry* 17.7 (2010), pp. 672–697. ISSN: 09298673. DOI: 10.2174/092986710790416263. URL: https://pubmed.ncbi.nlm.nih.gov/20088761/.

[23] Shabnam Heydarzadeh et al. *Regulators of glucose uptake in thyroid cancer cell lines*. June 2020. DOI: 10.1186/s12964-020-00586-x. URL: https://doi.org/10.1186/s12964-020-00586-x.

[24] Kieran Smallbone et al. “Metabolic changes during carcinogenesis: potential impact on invasiveness”. In: *Journal of theoretical biology* 244.4 (Feb. 2007), pp. 703–713. ISSN: 0022-5193. DOI: 10.1016/J.JTBI.2006.09.010. URL: https://pubmed.ncbi.nlm.nih.gov/17055536/.

[25] Dongya Jia et al. “Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways”. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.9 (Feb. 2019), pp. 3909–3918. ISSN: 10916490. DOI: 10.1073/pnas.1816391116. URL: https://pubmed.ncbi.nlm.nih.gov/30733294/.

[26] Rupert Courtnay et al. “Cancer metabolism and the Warburg effect: the role of HIF-1 and PI3K”. In: *Molecular biology reports* 42.4 (2015), pp. 841–851. ISSN: 15734978. DOI: 10.1007/s11033-015-3858-x.

[27] Mark Robertson-Tessi et al. “Impact of Metabolic Heterogeneity on Tumor Growth, Invasion, and Treatment Outcomes”. In: *Cancer Research* 75.8 (Apr. 2015), pp. 1567–1579. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-14-1428. URL: http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-14-1428.

[28] Ahmadreza Ghaffarizadeh et al. “PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems”. In: *PLOS Computational Biology* 14.2 (Feb. 2018), e1005991. ISSN: 1553-7358. DOI: 10.1371/JOURNAL.PCBI.1005991. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005991.

[29] Geoffrey M Cooper. “The Eukaryotic Cell Cycle”. In: (2000). URL: https://www.ncbi.nlm.nih.gov/books/NBK9876/.

[30] Tomás Alarcón and Henrik Jeldtoft Jensen. “Quiescence: A mechanism for escaping the effects of drug on cell populations”. In: *Journal of the Royal Society Interface* 8.54 (2011), pp. 99–106. ISSN: 17425662. DOI: 10.1098/rsif.2010.0130. arXiv: 1002.4579.

[31] Joseph J. Casciari, Stratis V. Sotirchos, and Robert M. Sutherland. “Variations in tumor cell growth rates and metabolism with oxygen concentration, glucose concentration, and extracellular pH”. In: *Journal of Cellular Physiology* 151.2 (1992), pp. 386–394. ISSN: 10974652. DOI: 10.1002/jcp.1041510220.

[32] Jing Yang et al. *Targeting PI3K in cancer: Mechanisms and advances in clinical trials*. 2019. DOI: 10.1186/s12943-019-0954-x.

[33] Tian Tian, Xiaoyi Li, and Jinhua Zhang. *mTOR signaling in cancer and mtor inhibitors in solid tumor targeting therapy*. 2019. DOI: 10.3390/ijms20030755.

[34] Weidong Li et al. *Targeting AMPK for cancer prevention and treatment*. 2015. DOI: 10.18632/oncotarget.3629.

[35] S. D. Webb, J. A. Sherratt, and R. G. Fish. "Mathematical Modelling of Tumor Acidity: Regulation of Intracellular pH". In: *Journal of Theoretical Biology* 196.2 (Jan. 1999), pp. 237–250. ISSN: 0022-5193. DOI: 10.1006/JTBI.1998.0836. URL: https://doi.org/10.1006/jtbi.1998.0836.

[36] Kieran Smallbone, Robert A. Gatenby, and Philip K. Maini. "Mathematical modelling of tumour acidity". In: *Journal of Theoretical Biology* 255.1 (2008), pp. 106–112. ISSN: 00225193. DOI: 10.1016/j.jtbi.2008.08.002.

[37] Angélique Stéphanou and Annabelle Ballesta. "pH as a potential therapeutic target to improve temozolomide antitumor efficacy : A mechanistic modeling study". In: *Pharmacology Research and Perspectives* 7.1 (Feb. 2019). ISSN: 20521707. DOI: 10.1002/PRP2.454.

A.2 Writing Article: Exploration of altered mechanism of the tumour cell using Differential Expression and Pathway Enrichment analysis: Application to glioblastoma

Exploration of altered mechanism of the tumour cell using Differential Expression and Pathway Enrichment analysis : Application to glioblastoma

Kévin Spinicci[1,2], Jorge Bretones-Santamarina[3], Clémentine Descamps[1], Maite Verreault[4], Magali Richard[1], Annabelle Ballesta[3], Angélique Stéphanou[1], and Gibin Powathil[2]

[1]Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France

[2]Department of Mathematics - Swansea University

[3]Institut Curie, 75248 Paris, France

[4]Institut du Cerveau, CNRS, UMR 7225, 75013 Paris, France

2022

Abstract

RNA-Seq technology has become a standard to study gene expressions and discover how diseases alter biological mechanisms. Despite its wide utilization and range of applications, study and interpretation of RNA-Seq data still remain a challenge for most scientists. Due to the variety of protocols described in the literature, it is especially hard for newcomers. A common approach to interpreting RNA-Seq data is to use Pathway Enrichment analysis, a statistical method that finds enriched pathways for a list of genes of interest. In other words, this method turns a large list of genes into a smaller list of pathways. In this paper, we describe a workflow to analyze RNA-Seq data. We applied this workflow RNA-Seq data from 20 patients diagnosed with glioblastoma. We also used RNA-Seq data from the TCGA-GBM project, an initiative to sequence glioblastoma mutation and gene expression, from the The Cancer Genome Atlas (TCGA) database. The workflow starts with a Differential Expression (DE) analysis to determine which genes are deregulated in the case of glioblastoma. With this step we deduce a list of genes of interest. We selected the two tools DESeq2 and Personalized Differential Analysis (PenDA) for their performance and ease of use. We submit the list of genes to G:Profiler and Gene Set Enrichment Analysis (GSEA), two pathway enrichment methods, to assess which pathways emerge from this list. Because this method determines a typical profile of deregulations for glioblastoma, we performed an analysis where we compare each sample of the PDCL dataset one by one to all controls. This second analysis better accounts for heterogeneity among the samples. We brielfy describe the result obtained with G:Profiler and GSEA and compare

the differences between TCGA and Patient Derived Cell Lines (PDCL). We found that most of the mechanisms that are found altered in both TCGA and PDCL datasets are involved in the cell-cycle, DNA replication and repair. In the PDCL dataset, the most commonly deregulated pathway is the *focal adhesion* of the cell to the matrix. *Collagen formation* and *cholesterol synthesis* are pathways frequently deregulated as well. A review of the literature shows that collagen is involved in glioblastoma growth, invasion and migration while cholesterol has been linked to glioblastoma growth. In addition, cholesterol synthesis has been suggested as a potential drug target for therapy ignoring healthy cells.

Introduction

RNA-Sequencing (RNA-Seq) is now a standard technology used by the Life Sciences community to obtain information about gene expressions in samples. With a wide range of applications, variations of RNA-Seq protocols and analyses described in the literature, new users can find RNA-Seq studies hard to conduct [1]. Omics experiments generated data are growing at a fast rate, allowing the scientific community to find mechanisms related to a particular condition. Nevertheless, these experiments often produce a long list of genes to analyze which still represents a challenge for researchers. Pathway enrichment analysis, a method that turns such a large list of genes into a smaller list of pathways, became a standard approach to interpret omics data [2].

This method is very interesting as (1) it reduces the complexity of the data to analyze and (2) interpreting altered pathways is more tangible than a list of genes. Three generations of pathway enrichment have been developed to interpret high-throughput technologies data. The first generation, called Over-Representation Analysis (ORA), statistically test for pathways that are found over or under-represented in a list of genes. These methods generally use a hypergeometric, chi-square or binomial distribution for their test. The second generation, called Functional Class Scoring (FCS), hypothesizes that not only individual large changes in gene expressions can significantly affect a pathway, but small coordinated changes in sets of genes can do as well. FCS methods use a gene-level statistic to compute each pathway's test statistic. The last generation, called Pathway Topology (PT), goes further than the previous generation by using information about the interaction between genes (activation/inhibition, cell compartment) to test for enrichment. Although more and more limitations have been addressed through the different generations of pathway enrichment, they are still limited by (1) the exhaustiveness of the information covered by current databases annotations and (2) the inability of these methods to account for the dynamic nature of a biological system [3].

In this paper we will apply this methodology to address glioblastoma, an aggressive brain tumour classified grade IV by the World Health Organization (WHO) known to be the most common as well as the deadliest brain cancer [4, 5]. The current standard of care consists of surgery followed by radiotherapy combined with cycles of temozolomide chemotherapy [6]. Intensive research failed to improve the patient outcome and the median survival for treated patients is only 14 months [7]. Even after tumour resection patients are not out of trouble as almost all tumours recur with a decreased sensibility to therapy [8].

In this article, we will present a method to study the deregulated pathways from RNA-Seq data

using statistical methods. We apply the method to the TCGA-GBM project data available for use from the The Cancer Genome Atlas (TCGA) platform, and to a dataset of Patient Derived Cell Lines (PDCL) from the Institut du Cerveau et de la Moelle épinière (ICM). While the TCGA-GBM contains bulk glioblastoma tumour expression, the PDCL data consists of single cell RNA-Seq from glioblastoma tumours. The method we use mainly consists of a Differential Expression (DE) analysis followed by a Pathway Enrichment analysis. In a first time, we compare the dysregulation between both datasets at a population level using the DESeq2 statiscal tool to assess the typical deregulation of glioblastoma. In a second approach, we assess the per samples deregulations to better assess the frequence of pathway deregulations. For that we use a tool called Personalized Differential Analysis (PenDA) designed toward this goal. To our knowledge, this is the only tool available to carry out Differential Expression analysis at the sample level. The authors show that in most cases, PenDA outperforms other statiscal tool like DESeq2 [9]. Understanding of glioblastoma heterogeneity is an important axis of research as heterogeneity is a cause of therapeutic failure when treating this kind of tumour [10]. In the disussion we will address the *Cholesterol Metabolism*, a pathway deregulated in PDCL but not in TCGA, and the *Collagen Synthesis*, a pathway frequently deregulated in both our datasets with a role in cell migration. We briefly review the litterature on this two pathways and how they represent potential target for therapy. The goal of this contribution is to present a method to analyze expression data illustrated by a case study.

Materials and Methods

Data Sets

Patient-Derived Cell Line (PDCL)

This database has been established by the Institut du Cerveau et de la Moelle épinière (ICM, Paris). Cells come from tumours of 20 patients diagnosed with glioblastoma, they were then cultured *in vitro* before RNA-Seq analysis was conducted. Results include the raw count value of approximately 20,000 genes for each sample. Only one replicate per sample is available.

Because the PDCL dataset does not include matching control samples, we compared them to astrocytes RNA-Seq data from a study on the characterization of different astrocytic models by Lundin *et al* [11]. The dataset contains 4 samples of different astrocyte cell lines with 3 replicates per sample, all information about the samples including the accession number are listed in table 1. Long-Term Neural Epithelial Stem Cell Culture (ltNES) were generated from AF22 human induced pluripotent stem cells (hiPSC) obtained from a middle-aged female [12]. The medium was changed every other day and cells were passaged once they reached 80% confluence; 7–9 passages during the differentiation protocol of 28 days. ltNES cells were differentiated into astrocytes using a previously described protocol [11] as follows: ltNES cells were plated at 60,000 $cells/cm^2$ on 2 $\mu g/cm^2$ poly-L-ornithine and 0.2 $\mu g/cm^2$ laminin (PLO-Laminin) (Sigma, St. Louis, MO) double-coated culture vessels in FHIA-differentiation medium; DMEM/F12, N2 supplement (1:100; Invitrogen, Carlsbad, CA), B27 (1:100; Invitrogen), FGF2 (8 ng/mL; PeproTech, Rocky Hill, NJ), heregulin 1β (10 ng/mL; Sigma), IGF1 (200 ng/mL; Sigma),

Accession Number	Sample name
GSM2927872	AF22_NES-Astro_d29_Br1
GSM2927873	AF22_NES-Astro_d29_Br2
GSM2927874	AF22_NES-Astro_d29_Br3
GSM2927881	iCellAstro_Br1
GSM2927882	iCellAstro_Br2
GSM2927883	iCellAstro_Br3
GSM2927884	CCF_Br1
GSM2927885	CCF_Br2
GSM2927886	CCF_Br3
GSM2927887	phaAstro_Br1
GSM2927888	phaAstro_Br2
GSM2927889	phaAstro_Br3

Table 1: List of samples with their accession number used as the control condition. Each sample can be either downloaded separately or as one raw count file in text format using the accession number GSE109001 on Gene Expression Omnibus (GEO). The accession number is an ID to retrieve data on the GEO database. The dataset contains 4 samples of different astrocyte cell lines with 3 replicates per sample. At the time of the publication, data were last updated on January 29, 2019.

activin A (10 ng/mL; PeproTech). RNAsequencing was performed on cells harvested on day 29 in 3 replicates. The iCell, CCF-STTG1 and and the human brain astrocytes were purchased from Cellular Dynamics International, ATCC and Neuromics. All commercial cell lines were cultivated according the manufacturer's recommendation.

The Cancer Genome Atlas (TCGA)

The dysregulations found in this dataset were compared with glioblastoma data from the TCGA-GBM project available on the TCGA database (version 33.1, May 31 2022). The TCGA-GBM project aim to sequence the mutations and gene expressions for the glioblastoma disease. Glioblastoma tumour was the first to be sequenced by the TCGA initiative. The dataset contains gene expressions for primary/recurrent tumours and solid normal tissues. All samples have been processed following the bioinformatic pipeline defined by TCGA and described on their website. Few normal transcriptomes profiling are available for the brain, hence this dataset is over-represented by tumourous samples with 169 tumour samples and 5 control samples.

Differential Expression Analysis

This analysis consist of comparing the expression of genes in a tumour against a control to look for genes whose expressions are deregulated (over or under expressed). This generate a list of genes as a result that is then used in the second step to perform pathways enrichment analysis. Different portions of RNA,

called reads, are sequenced during RNA-Seq experiment. To quantify the expression of a gene, those reads are then mapped to the genes of a reference genome. The number of mapped reads is computed for each gene giving a raw value called raw counts. Factors such as transcript (mRNA that went through post-transcriptional modification) length, the total number of reads and sequencing biases can impact the raw counts value, therefore this value is normalized to obtain a value suitable for comparison [1]. Different tools are available to perform DE analysis such as edgeR [13], Limma [14], DESeq2 [15] and PenDA [9]. In this study, we selected DESeq2 [15] and PenDA to carry out the DE analysis.

DE analysis was performed on the two datasets PDCL and TCGA where we compared all the tumour samples against all the normal samples (glioblastoma versus normal). With this approach, we identify a typical profile or the typical gene deregulations in the case of glioblastoma. Glioblastomas are known to be heterogeneous tumours, hindering therapeutic success [10, 7, 4]. Since the approach described does not account for heterogeneity between patients, we compared each sample one by one with all the controls for both datasets. This second approach defines the deregulation for each sample, giving the ability to determine the biological functions that are frequently deregulated or cell line specific. However, DESeq2 hypothesis assumes several replicates are available in the dataset. This condition is not met since in this case we compare only one sample with no replicate against all the controls, leading to reduced accuracy. Thus we will use PenDA, a statistical method designed for this purpose. Because this tool does not compute a metric to perform its test (rank comparison method) the results from PenDA are then fed to G:Profiler only.

DESeq2

DESeq2 uses a negative binomial distribution (also called gamma-Poisson distribution) to model read counts and perform a Wald Test to test for differential expression [15]. As DESeq2 includes a built-in normalization method, only raw counts can be used as input to the tool. It also performs an automated detection of outliers, detection of genes with low counts, estimation of model parameters and gene expression dispersion. The log_2 fold change value is a numeric value used to quantify the strength and direction of deregulation (up or down) between a condition and the control. We selected DESeq2 for the features cited above, its ease of use and its overall performance [15]. The Wald-Test statistic has two useful features : (1) the value is positive when a gene is upregulated and negative when it is downregulated; (2) the higher the absolute value of the statistic is, the more significant the fold change is, therefore it combines biological and statistical informations at the same time.

Personalized Differential Analysis (PenDA)

PenDA is a rank-based method to perform DE using a sample from the condition to test for differential expression and a reference dataset [9]. In comparison, population-scale analysis tools like DESeq2 usually model the data using a distribution law and perform a statistical test to detect differential expression. Those methods were designed to investigate typical gene deregulations patterns, consequently, they do not provide information at the individual scale. They are also sensitive to batch effects and the choice of normalization applied. With all these limitations in mind, PenDA was designed to investigate gene

expression data at the individual scale with robustness to batch and normalization effects. For our PenDA analysis, we normalize the counts using the pseudo-count method ($log_2(count + 1)$) from the DESeq2 package before filtering and ranking with PenDA.

Pathway Enrichment Analysis

Here, we use the list of deregulated genes identified by the Differential Expression analysis as input to investigate the deregulated pathways in the PDCL and TCGA datasets. They were both analyzed at the population and the patient scale. Here we selected the two tools G:Profiler and Gene Set Enrichment Analysis (GSEA) to run the pathway enrichment analysis. Both tools give as an output a list of pathways with a p-value associated to each pathways. The p-value represents the probability that a pathway is found enriched due to a random effect rather than truly due to the condition studied. Therefore the lower the p-value is, the more significant the result. When statistical testing is repeated, the p-value can be found significant by chance alone. Multiple-testing correction methods adjust the significance of the p-value to correct for this effect. The most commonly used method, and the one used in this paper, is Benjaminini-Hochberg False Discovery Rate (FDR) multiple-hypothesis correction method. Like for the p-value, we want the FDR to be the lowest possible, generally below 0.05 yet 0.1 is also found in the literature [2]. To reduce the number of pathways to investigate in the literature and to ensure the reliability of the results, we selected the results with a FDR below 0.05 in G:Profiler and GSEA.

G:Profiler

G:Profiler is a tool for finding biological processes enriched in a list of genes (for example a list of deregulated genes or a list of mutated genes) accessible either online using their web interface or using their API and libraries [16]. Only the names of the genes are needed as input. It searches for pathways or biological categories which are over-represented in a list of genes using the cumulative hypergeometric test. G:Profiler includes multiple hypothesis correction methods with a built-in method called g:SCS, developed by the G:Profiler team, or the more known Bonferroni and FDR correction method. We choose the FDR correction method as GSEA use the same method to be consistent between tools.

Gene Set Enrichment Analysis (GSEA)

GSEA is a rank-based method developed by Subramanian *et al* [17] that determines the pathways enriched in a list of genes ordered by a metric. This ranking metric is a float value that can be positive or negative, for example, the fold change in expression between a tumour and a control. The list is sorted in decreasing order, thus genes the most strongly or significantly upregulated will be at the top of the list and at bottom for the downregulated. The GSEA approach determines whether the genes involved in a gene set are mostly distributed at the top or the bottom of a gene list. The algorithm of GSEA runs through the list of genes to compute a score called enrichment score (ES), similar to a weighted Kolmogorov-Smirnov-like statistic. A gene metric is added to the score if it is present in the pathways, substracted otherwise. The score is then determined by the maximum deviation from zero. The ES of a null distribution, generated by permutating the genes in the dataset or the phenotype of the samples, is computed using

the same algorithm. The score of the tested pathways is compared with the score of the null distribution to compute the significance of a result. GSEA use the FDR method to adjust for multiple hypothesis testing. Despite both of them need a list of genes, G:Profiler needs only the genes symbol as input while for GSEA, each gene must be assigned a metric value. A study from Zyla *et al* has shown that the GSEA metric used can impact the results [18]. They evaluated the performance of 16 different metrics and found that the absolute value of the Moderated Welch Test statistic, the Minimum Significant Difference, the absolute value of the Signal to Noise ratio and the Baumgartner-Weiss-Schindler test statistic give the best outcome. Most of the metrics tested are statistics computed during a test. Wald-Test statistic combines both biological and statistical information, suggesting it is a valuable metric for running GSEA analysis. Thus, we ranked the genes of the dataset with the statistic of the Wald-Test computed by DESeq2 before we run GSEA.

Gene Matrix Transposed (GMT)

Biological information must be provided to pathway enrichment tools as gene sets, a group of genes sharing a common biological feature (for example a biological function or chromosomal location). Both G:Profiler and GSEA need a gene supplied in a tabular format called Gene Matrix Transposed (GMT) where each row is a gene set, in this study gene sets are pathways. The first column of the file contains the pathway ID, the second column contains a description and the remaining column are all the genes involved in the pathway. G:Profiler is updated regularly with the most common databases such as KEGG, Reactome, WikiPathways, miRTarBase, Gene Ontology, etc. Except for a few databases due to licensing limitations, including KEGG, all built-in G:Profiler's gene sets can be downloaded from their website. GSEA includes built-in gene sets from the Molecular Signatures Database (MSigDB) which can be downloaded. G:Profiler and GSEA allow the user to submit custom GMT files as well. We downloaded gene sets from KEGG [19] and Reactome [20], two biological databases with curated sets of data. It is recommended to use different databases as the same pathways can have different definitions among databases. For example, the glycolysis is defined as *canonical glycolysis* in Gene Ontology, *glycolysis* in Reactome and *Glycolysis/Gluconeogenesis* in KEGG. It is beneficial to exclude small and large pathways as : (1) small pathways are redundant with larger pathways while large pathways are often too general (metabolism, cell-cycle); (2) small pathways are never picked up by pathway enrichment tools while large pathways tend to always shows-up in the results; (3) higher number of pathways in the data makes multiple-testing correction more strict [2]. Pathway sizes of 10-15 to 200-500 genes are recommended yet sizes of 2,000 genes can be found in the literature [2]. In this paper, we kept pathways of 15 to 500 genes.

Implementation of the Workflow and Interpretation of the Results

The protocol used in this study can be split into three steps :

1. Differential Expression (DE) analysis is used to analyze RNA-Seq data, *i.e.* to produce a list of the genes that are deregulated in glioblastoma. We compared, gene expression between samples taken from a patient's tumour with healthy astrocyte cells.

2. Pathway Enrichment allows to find the deregulated pathways in the list of genes defined in the DE analysis. Pathway enrichment analysis searches for pathways that are enriched in a list of genes, in other words, pathways that are deregulated. For this step, we used two methods documented in the literature : G:Profiler and Gene Set Enrichment Analysis (GSEA). We downloaded pathway information from Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome, two pathway databases that defined biological processes of the cell.

3. We compare the biological processes found deregulated in our datasets to identify the impaired mechanisms. In this regard, we mapped each pathway to their highest ancestors in the hierarchy of pathways to define biological categories for each database. For example, the entry *G1/S phase transition* in Reactome is a child pathway of the entry *Cell-Cycle*. Thus, we class this pathway in the Cell-Cycle category. To avoid false positives and reduce the number of pathways found, we only considered the pathways that are below the significance threshold in both G:Profiler and GSEA during the analysis of the results.

All the data analyses were carried out with the R programming language. G:Profiler pathway enrichment were performed with the official R librarie (version 0.2.1), and GSEA pathway enrichment with the *fgsea* R package available on Bioconductor (version 1.12.0) [21]. Plots and reports, available as supplementary materials, were generated with the ggplot2 (version 3.3.6) and rmarkdown (version 2.14) R library. The workflow presented is summarized in figure 1.

Results

Validation of the controls

When conducting Differential Expression analysis, the user has to ensure that controls are separated from the tumours. Homogeneity among the controls, similarity between the control and tumour counts distribution are also factors that can improve the results. Despite that the controls should be separated from the tumours, it has been observed for different types of cancer in TCGA that the gene expression follows a similar distribution between the two conditions [22].

Figure 2 shows the results of a Principal Component Analysis (PCA) performed on the counts normalized using the pseudo-log method from the DESeq2 package ($log_2(count + 1)$). It can be seen that the controls are separated from the tumours in both the PDCL and the TCGA datasets by the first two components of the PCA. The first component explain more variances in the data in the PDCL datasets with 21% compared to 8% with TCGA. In the TCGA datasets, the tumours samples tend to spread more accross the second component than the controls which seem to pack together while in the PDCL dataset, controls are more scattered.

Figure 3 shows the distribution of normalized gene counts in controls and tumours among both datasets. Distribution of counts is similar among controls and tumours in both datasets with a similar intensity on the peak between controls and tumours with respect to the dataset.

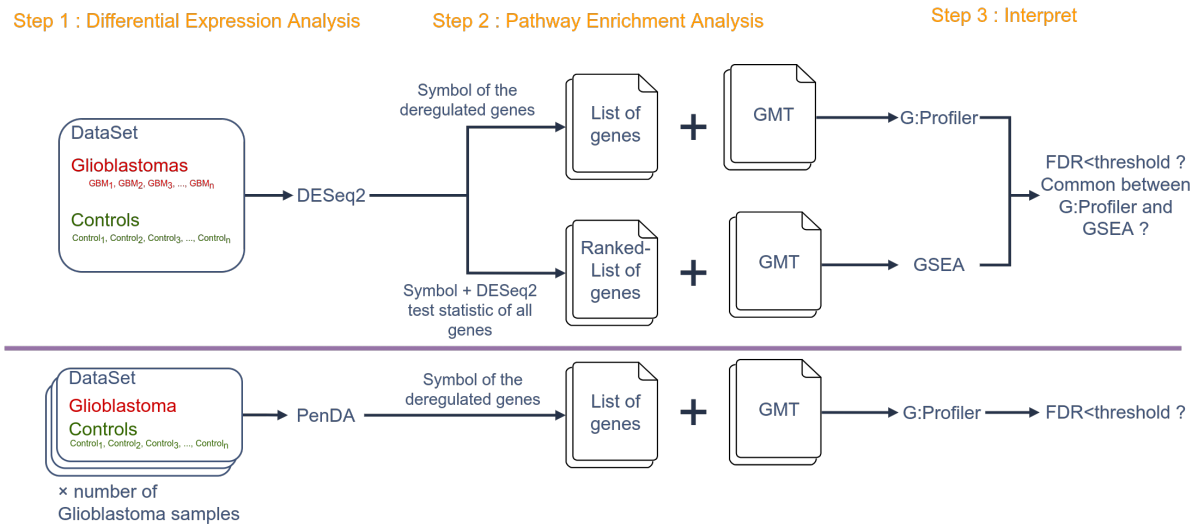

Figure 1: Diagram of the overall approach to search for the altered biological functions in glioblastoma. The top panel describes the workflow to find the typical deregulation in glioblastoma : a global analysis at the population level. In this case, we compare all the controls against all the glioblastoma samples in the dataset. We use DESeq2 to perform the Differential Expression analysis with each tumour and control samples considered as a "replicate of a typical patient or normal tissue". Then, we use G:Profiler and GSEA on the result of DESeq2 to perform pathway enrichment. The bottom panel describes the workflow adopted to assess the specific deregulations of each samples in both datasets : a personalized analysis. Here we use PenDA, a tool designed to perform DE analysis at the individual scale. In this case, we only apply G:Profiler on the PenDA results.

Figure 2: Plot of the PCA on the counts of control and glioblastoma samples of the PDCL and TCGA datasets. Counts were normalized using the pseudo-log method of the DESeq2 package ($log_2(count + 1)$).

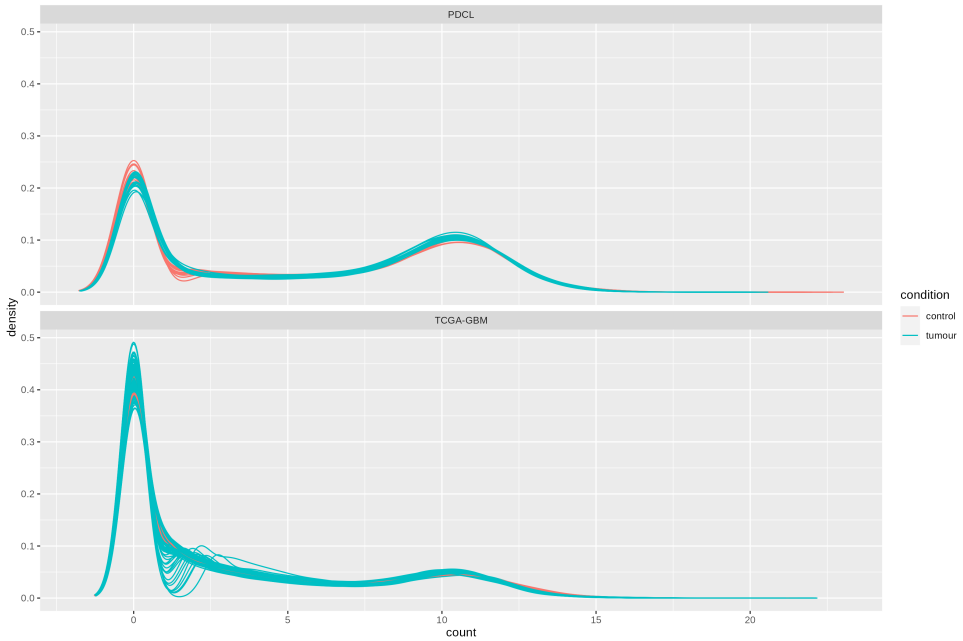

Figure 3: Plot of the density of the counts in control and glioblastoma samples of the PDCL and TCGA datasets. Counts were normalized using the pseudo-log method of the DESeq2 package ($log_2(count + 1)$). Each curve represents a different sample.

Together, these data seems to indicate that the controls choosen are suitable to perform the Differential Expression analysis then pathway enrichment.

Analysis of the deregulated pathways at the population scale

Target pathways show quality of enrichment among the results

The KEGG database documents the processes affected in the case of many diseases including glioma. During pathway enrichment analysis, one can take one or several target pathways to assess the quality of the results. For example, Zyla *et al* used the adjusted p-value of the *Glioma* entry (path:hsa05214) as a target pathway to assess the sensitivity of different ranking metrics when studying microarray data from brain tissue [18]. Here, as we are comparing RNA-Seq data from glioblastoma samples to normal controls, we should expect the adjusted p-value (FDR in this study) to be below our definied threshold. The FDR value for *Glioma* is 0.053 in G:Profiler and 0.83 in GSEA for the PDCL dataset. With TCGA data, the FDR value for *Glioma* is 0.0009 in G:Profiler and 0.005 in GSEA. It suggests that TCGA data have a better quality than the PDCL ones. Controls for the PDCL dataset are taken from another study [11], introducing experimental and biological variability independent from disease deregulation. This can be seen on figure 2, as the controls are more spread in PDCL than in TCGA. Controls in the TCGA dataset come from matching samples taken from the original tissue following the exact same protocol as the tumour samples to reduce such variability. Futhermore, some replicates in the controls of the PDCL datasets comes from astrocytomas cell lines. Therefore they are not completely suitable for Differential Expression analysis. It shows that a good set of controls can influence the quality of the results given.

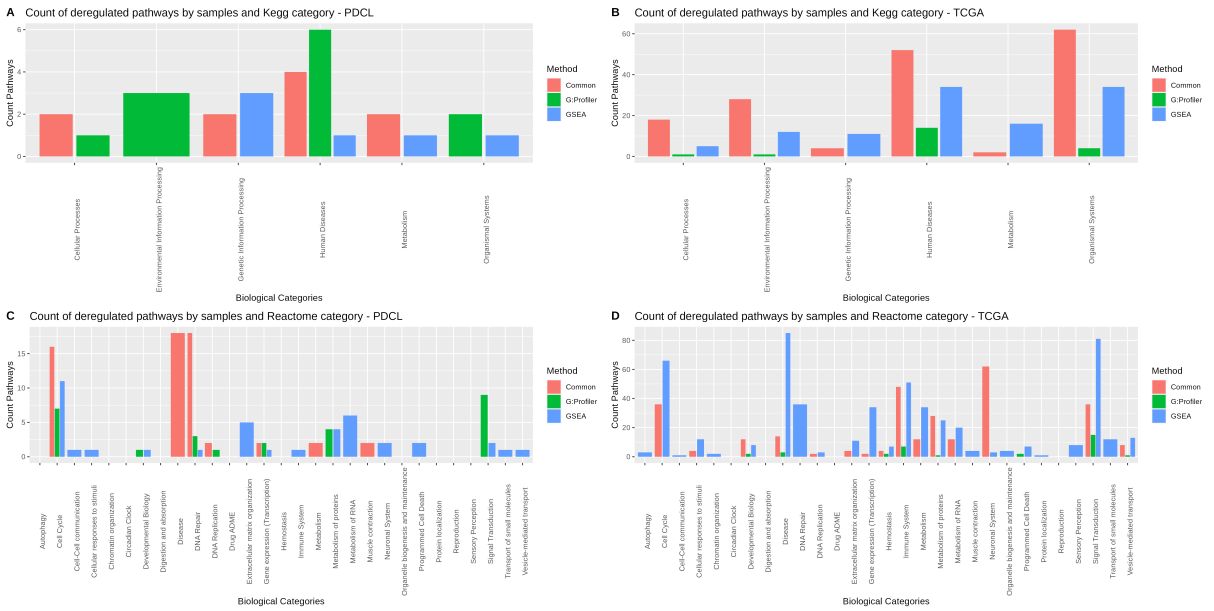

Figure 4: Barplot of the count of significantly deregulated pathways for (A) KEGG categories in PDCL; (B) KEGG categories in TCGA; (C) Reactome categories in PDCL; (D) Reactome categories in TCGA. Pathways are colored whether they are specific to G:Profiler or GSEA, they are significantly enriched in one of the tools, or common, they are significantly enriched in both tools.

Differences in deregulated categories among the databases and the tools

As can be seen in figure 4, G:Profiler and GSEA yield different results which are also influenced by the database choosen. We recommand investigating more than one database when performing pathway enrichment as the choice of the databse may impact the results [23]. Most of the time, more pathways are significantly enriched in GSEA compared to G:Profiler. In some cases, there is no correlation betwen G:Profiler and GSEA results. For example, the *Organismal Systems* KEGG category is only enriched with pathways specific to G:Profiler or to GSEA with PDCL data. Still with the same dataset and database, the *Environmental information Processing* category is only found enriched with G:profiler. Inversely, the *Disease* Reactome category is enriched with only pathways found by both G:Profiler and GSEA in PDCL data. In addition, Reactome categories like *Circadian Clock* or *Reproduction* are not found enriched in any of the datasets nor tools. In this study, pathways in the GMT are filtered only on their size. Yet, it may be benificial to remove pathways which are not involved in the biological condition being investigated. The FDR multiple hypothesis correction method is influenced by the number of null-hypothesis tested, here the number of genes tested for differential expression or pathway tested for enrichment. Therefore, the more pathways are being tested, the stronger the correction applied on the p-value is.

Investigation of selected pathways

FDR values for different pathways from KEGG and Reactome given by G:Profiler and GSEA in both datasets is presented in figure 5. As before, we see that a same pathways can be found enriched in one

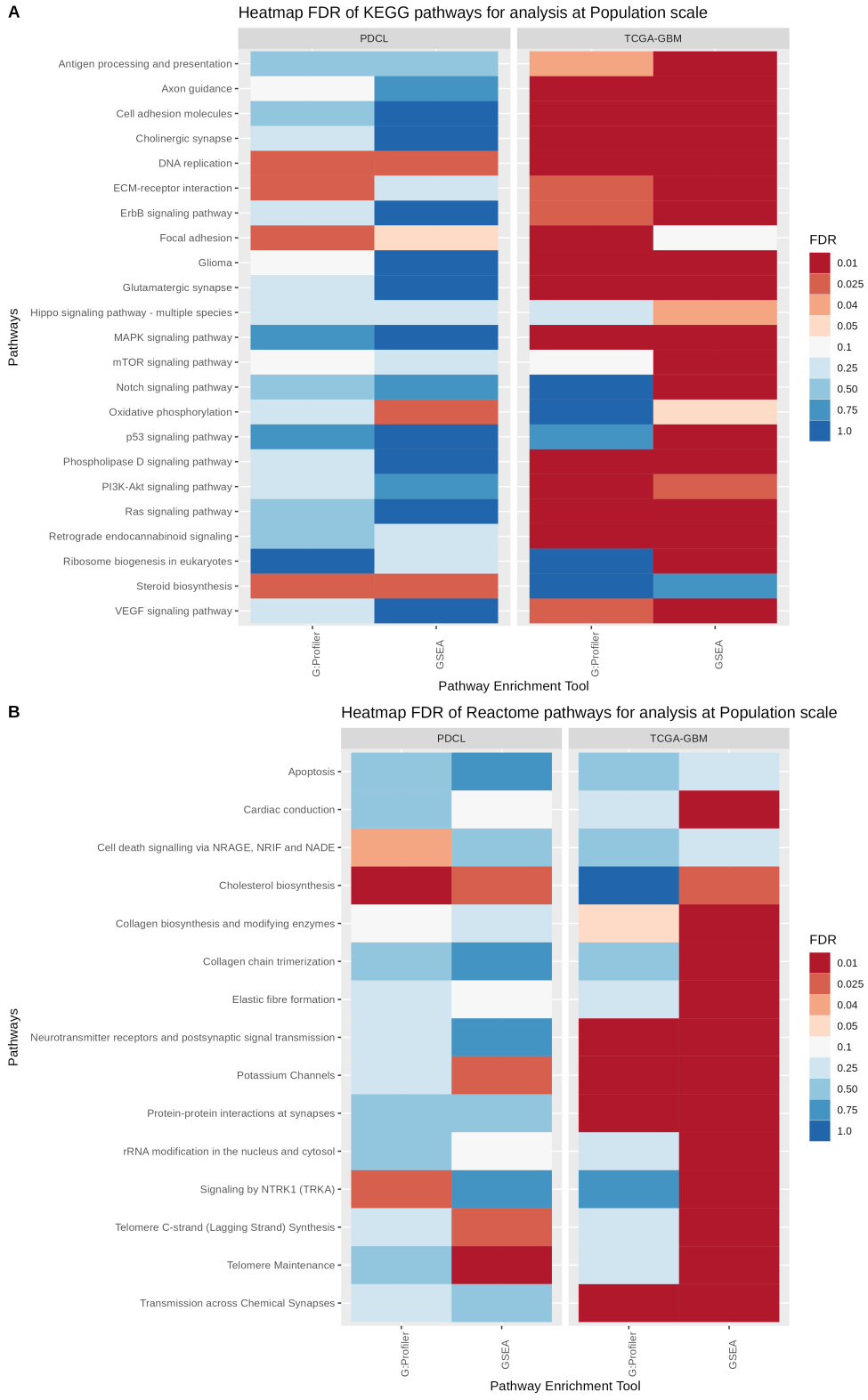

Figure 5: Heatmap of deregulated pathways in the PDCL and the TCGA datasets for (A) the KEGG database and (B) the Reactome database. FDR values below 5% are considered significant (red) and their are not considered significant if greater than 5% (blue).

enrichment tool but not in another. For example, the FDR for *Oxidative phosphorylation* is significant in both dataset with GSEA but not with G:Profiler. In TCGA, the *p53 signaling* pathway is found significant only with GSEA and it is not enriched with PDCL data. Yet it is found deregulated in one sample during the personnalized analysis of PDCL samples (described below) suggesting it is rather due to the dataset. Using pathway enrichment we can also highlights the differences between both datasets. As an example, many well-known signaling pathways such as *PI3K-Akt signaling*, *RAS signaling* and *VEGF signaling* are found enriched only in TCGA. At the opposites, pathways involved in cholesterol metabolism such as *Steroid biosynthesis* and *Cholesterol biosynthesis* are found enriched only in PDCL. We will show below, that we can assess which samples are associated with an altered cholesterol synthesis.

Analysis of the deregulated pathways at the individual level

Cell adhesion to the ECM is the most frequently deregulated pathway in PDCL samples

Table 2 and 3 show the most frequently deregulated pathways in PDCL. The *Focal adhesion* (path:hsa04510) is the most frequently deregulated KEGG pathway with 13 samples affected while *Transmission across Chemical Synapses* (R-HSA-112315) is the most frequently deregulated Reactome pathway with 5 samples affected. Results contain pathways linked to processes influencing the cell motility and its fixation on the matrix such as: *Focal adhesion* (path:hsa04510), *ECM-receptor interaction* (path:hsa04512), *Cell adhesion molecules* (path:hsa04514), *Collagen synthesis* (R-HSA-165814, R-HSA-8948216) and *elastic fibre formation* (R-HSA-1566948). Glioblastoma cells have the ability to migrate and invade surrounding tissue, sometimes migrating to other organs as well [24], supporting the validity of these results.

In a similar way, *Steroid biosynthesis* and *Cholesterol biosynthesis* are found deregulated in 4 and 3 different samples, respectively. *Cholesterol metabolism* has been documented in the litterature to be affected in the case of glioblastoma.

The *Oxidative phosphorylation* (path:hsa00190), well-known for its role in the generation of ATP, is found deregulated in 4 samples. This pathway has been widely studied due to its important role in the generation of ATP and its potential implication in the Warburg Effect, an increased use of glycolysis to produce ATP even though oxygen is available [25]. It was first hypothesized that mitochondria were defective in tumour cells, yet it was later shown that mitochondrial function in cancer cells was similar to normal cells [26]. The *Hippo signaling pathway* (path:hsa04392, path:hsa04390) and *Retrograde endocannnabinoid signaling* (path:hsa04723), two signaling pathway whose role in cancer have been documented in the litterature, are found deregulated in 6 and 5 different samples, respectively [27, 28, 29].

As can be seen in figure 6, the sample *4339-p21* is the most deregulated with around 40 pathways significantly enriched in both KEGG and Reactome. Most of the pathway found deregulated with the KEGG database, across all the PDCL samples, are associated to *Human Diseases*. *Human Diseases* aside, the majority of deregulated pathways in the sample *4339-p21* are associated to the *Cellular Processes* and *Environmental/Genetic Information Processing* KEGG, or the *Cell-Cycle* Reactome category.

Database	Pathway ID	Description	Category	Number of PDCL
KEGG	path:hsa04510	Focal adhesion	Cellular Processes	13
KEGG	path:hsa04360	Axon guidance	Organismal Systems	10
KEGG	path:hsa04512	ECM-receptor interaction	Environmental Information Processing	5
KEGG	path:hsa04724	Glutamatergic synapse	Organismal Systems	5
KEGG	path:hsa04725	Cholinergic synapse	Organismal Systems	5
KEGG	path:hsa00100	Steroid biosynthesis	Metabolism	4
KEGG	path:hsa00190	Oxidative phosphorylation	Metabolism	4
KEGG	path:hsa04612	Antigen processing and presentation	Organismal Systems	4
KEGG	path:hsa04723	Retrograde endocannabinoid signaling	Organismal Systems	4
KEGG	path:hsa03008	Ribosome biogenesis in eukaryotes	Genetic Information Processing	3
KEGG	path:hsa04012	ErbB signaling pathway	Environmental Information Processing	3
KEGG	path:hsa04072	Phospholipase D signaling pathway	Environmental Information Processing	3
KEGG	path:hsa04392	Hippo signaling pathway - multiple species	Environmental Information Processing	3
KEGG	path:hsa04514	Cell adhesion molecules	Environmental Information Processing	3
KEGG	path:hsa04714	Thermogenesis	Organismal Systems	3

Table 2: Table of the frequently KEGG pathways deregulated in the PDCL. *Focal adhesion* is the most frequently deregulated pathway with 13 samples affected.

Database	Pathway ID	Description	Category	Number of PDCL
Reactome	R-HSA-112315	Transmission across Chemical Synapses	Neuronal System	5
Reactome	R-HSA-1650814	Collagen biosynthesis and modifying enzymes	Extracellular matrix organization	4
Reactome	R-HSA-5576891	Cardiac conduction	Muscle contraction	4
Reactome	R-HSA-8948216	Collagen chain trimerization	Extracellular matrix organization	4
Reactome	R-HSA-112314	Neurotransmitter receptors and postsynaptic signal transmission	Neuronal System	3
Reactome	R-HSA-191273	Cholesterol biosynthesis	Metabolism	3
Reactome	R-HSA-6790901	rRNA modification in the nucleus and cytosol	Metabolism of RNA	3
Reactome	R-HSA-6794362	Protein-protein interactions at synapses	Neuronal System	3
Reactome	R-HSA-109581	Apoptosis	Programmed Cell Death	2
Reactome	R-HSA-1296071	Potassium Channels	Neuronal System	2
Reactome	R-HSA-1566948	Elastic fibre formation	Extracellular matrix organization	2
Reactome	R-HSA-157579	Telomere Maintenance	Cell Cycle	2
Reactome	R-HSA-174417	Telomere C-strand (Lagging Strand) Synthesis	Cell Cycle	2
Reactome	R-HSA-187037	Signaling by NTRK1 (TRKA)	Signal Transduction	2
Reactome	R-HSA-204998	Cell death signalling via NRAGE, NRIF and NADE	Signal Transduction	2

Table 3: Table of the frequently Reactome pathways deregulated in the PDCL. *Transmission across Chemical Synapses* is the most frequently deregulated pathway with 5 samples affected. Collagen pathways are dysregulated in 4 different samples.

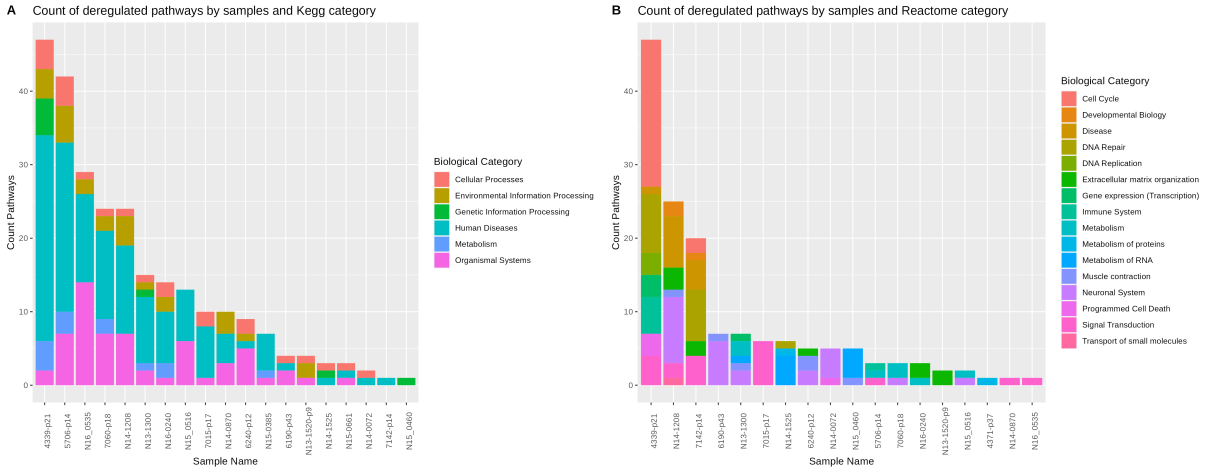

Figure 6: Number of deregulated pathways per category and per sample in the PDCL dataset. The *Human Diseases* and *Organismal Systems* categories are the most frequently deregulated KEGG categories. In Reactome, deregulated pathways are often associated with *Neuronal System*, *DNA Repair/Replication* and *Signal Transduction*. *4339-p21* is the sample with the most deregulated pathways in both KEGG and reactome. Pathways in this sample are associated to the *Cellular Processes* and *Environmental/Genetic Information Processing* KEGG, or the *Cell-Cycle* Reactome category.

Pathways linked to the Cell-Cycle and the ECM interactions are the most frequenlty deregulated in TCGA-GBM samples

Tables 4 and 5 show the most frequently deregulated pathways in the TCGA dataset. While *Focal adhesion* was the most frequently altered KEGG pathway among the PDCL samples, it is fourth with TCGA. Furthermore, no pathways linked to the metabolism appear in the table for KEGG pathways. As expected, the *p53 signalling* pathway is frequently altered with 118 samples out of 169 samples affected. Similar to the result obtained with PDCL samples, 3 Reactome pathways linked to the collagen protein synthesis are among the pathways often dysregulated. Still with the Reactome database, 3 pathways involved in the *Cell-Cycle* are deregulated in at least 149 samples and seem to indicate that the *G1/S Phase Transition* is the main part of the cell cycle affected by glioblastoma.

Like for the PDCL dataset, we investigated what were the most affected biological categories for each samples and show the results in figure 7. Same as in the PDCL, the *Disease* category is higly enriched in boh KEGG and Reactome. However, more different Reactome categories are present in the results with TCGA. In addition, while the *Cell-Cycle* Reactome category was mostly observed affected in the PDCL sample named *4339-p21* (which was also the most affected sample), this category is grealty altered in all sample for TCGA. TCGA results are also marked by an increased alteration of protein's metabolism across all the samples.

Sample clustering by their deregulated categories

For each sample, we divided the number of pathways deregulated in a category by the total number of pathways associated to this category in the GMT file. We used the vector of values generated to

Database	Pathway ID	Description	Category	Number of Samples
KEGG	path:hsa04110	Cell Cycle	Cellular Processes	153
KEGG	path:hsa04512	ECM-receptor interaction	Environmental Information Processing	147
KEGG	path:hsa04145	Phagosome	Cellular Processes	130
KEGG	path:hsa04510	Focal adhesion	Cellular Processes	129
KEGG	path:hsa03030	DNA replication	Genetic Information Processing	126
KEGG	path:hsa04640	Hematopoietic cell lineage	Organismal Systems	125
KEGG	path:hsa04672	Intestinal immune network for IgA production	Organismal Systems	125
KEGG	path:hsa04514	Cell adhesion molecules	Environmental Information Processing	124
KEGG	path:hsa04612	Antigen processing and presentation	Organismal Systems	124
KEGG	path:hsa03010	Ribosome	Genetic Information Processing	123
KEGG	path:hsa04115	p53 signaling pathway	Cellular Processes	118
KEGG	path:hsa04380	Osteoclast differentiation	Organismal Systems	94
KEGG	path:hsa04218	Cellular senescence	Cellular Processes	92
KEGG	path:hsa04658	Th1 and Th2 cell differentiation	Organismal Systems	87
KEGG	path:hsa04061	Viral protein interaction with cytokine and cytokine receptor	Environmental Information Processing	86

Table 4: Table of the frequently KEGG pathways deregulated in the TCGA. Unsurprisingly, the *Cell Cycle* pathway is the most deregulated one with 153 out of 169 samples affected. Similar to the PDCL dataset, the *Focal adhesion* is one of the most frequenlty deregulated pathway with 129 samples. Ohter pathways involved in the interaction between the cell and its environment are also frequenlty deregulated. p53 signalling, a pathway known to be frequently deregulated in the case of glioblastoma, is also found in the results.

Database	Pathway ID	Description	Category	Number of Samples
Reactome	R-HSA-1474290	Collagen formation	Extracellular matrix organization	159
Reactome	R-HSA-1799339	SRP-dependent cotranslational protein targeting to membrane	Metabolism of proteins	156
Reactome	R-HSA-453279	Mitotic G1 phase and G1/S transition	Cell Cycle	154
Reactome	R-HSA-1650814	Collagen biosynthesis and modifying enzymes	Extracellular matrix organization	152
Reactome	R-HSA-69206	G1/S Transition	Cell Cycle	151
Reactome	R-HSA-2022090	Assembly of collagen fibrils and other multimeric structures	Extracellular matrix organization	149
Reactome	R-HSA-6798695	Neutrophil degranulation	Immune System	149
Reactome	R-HSA-69620	Cell Cycle Checkpoints	Cell Cycle	149
Reactome	R-HSA-72695	Formation of the ternary complex, and subsequently, the 43S complex	Metabolism of proteins	149
Reactome	R-HSA-72662	Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	Metabolism of proteins	147
Reactome	R-HSA-72649	Translation initiation complex formation	Metabolism of proteins	146
Reactome	R-HSA-72689	Formation of a pool of free 40S subunits	Metabolism of proteins	146
Reactome	R-HSA-156827	L13a-mediated translational silencing of Ceruloplasmin expression	Metabolism of proteins	145
Reactome	R-HSA-2408522	Selenoamino acid metabolism	Metabolism	145
Reactome	R-HSA-72702	Ribosomal scanning and start codon recognition	Metabolism of proteins	145

Table 5: Table of the frequently Reactome pathways deregulated in the TCGA. Similar to the results obtained with the PDCL dataset, pathways linked to the collagen protein are found frequently deregulated (in more than 140 samples). Results in the table seems to indicate that *G1/S phase transition* is the main pahse of the cell cycle dysregulated in glioblastoma.

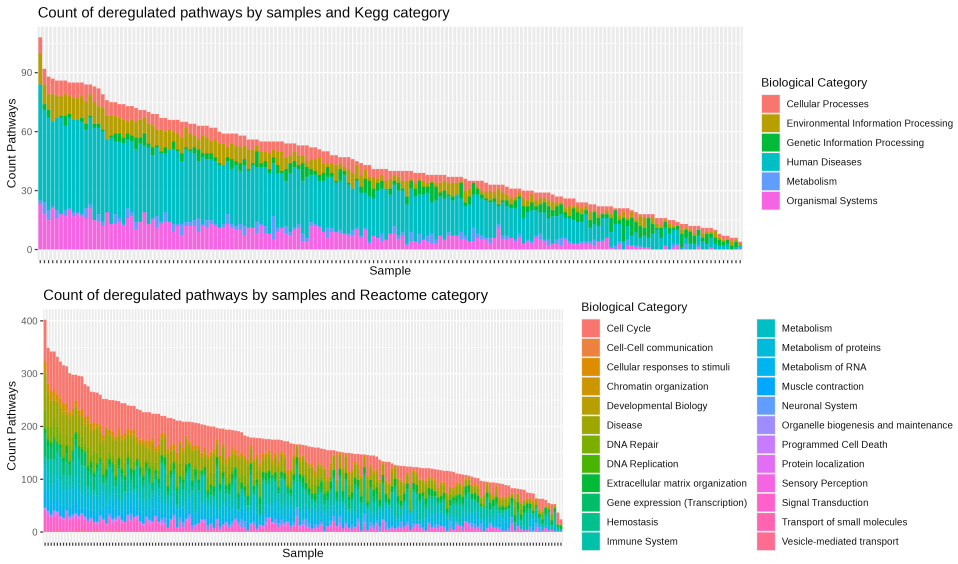

Figure 7: Number of deregulated pathways per category and per sample in the TCGA dataset. Sample names are not displayed to avoid overploting. Like the result with PDCL samples, the *Disease* is among the most deregulated categories in both KEGG and Reactome. While with PDCL, Reactome pathways associated to the Cell-Cycle were predominantly deregualted in one sample, in TCGA the Cell-Cycle was greatly affected among all the samples.

perform clustering of the samples using the K-Means algorithm with 3 clusters. The number of clusters was choosen graphically by studying the correlation of each categories to the principal components and with the coordinate of the samples in the new space. Clustering results are shown in figure 8. Results are displayed in two dimensions using the first 2 component of a PCA on the the same data used for clustering. There is a better delimitation of the cluster using KEGG categories than with Reactome in two dimensions for both datasets. There is some overlap between the clusters defined using Reactome categories in both datasets. Reactome has more categories defined and used for clustering than KEGG, thus filtering out some categories that are not relevant for a study may facilitate interpretation of the results.

Clustering of TCGA samples seems to separate samples with dysregulated functions involved in the *Metabolism*, genetic processing like *DNA replication/repair* and processes linked to the extracellular environment like *ECM organization* or *Environmental information processing* accross both pathway databases. A similar pattern is observed with PDCL sample clustering with Reactome, although it seems that *Metabolism* is less associated with deregulation than other categories (such as *Signal Processing*). Interestingly, when using KEGG categories to cluster PDCL samples, one cluster seems to group samples that have less deregulated pathways and the PCA tend to separate samples mainly on this criterion.

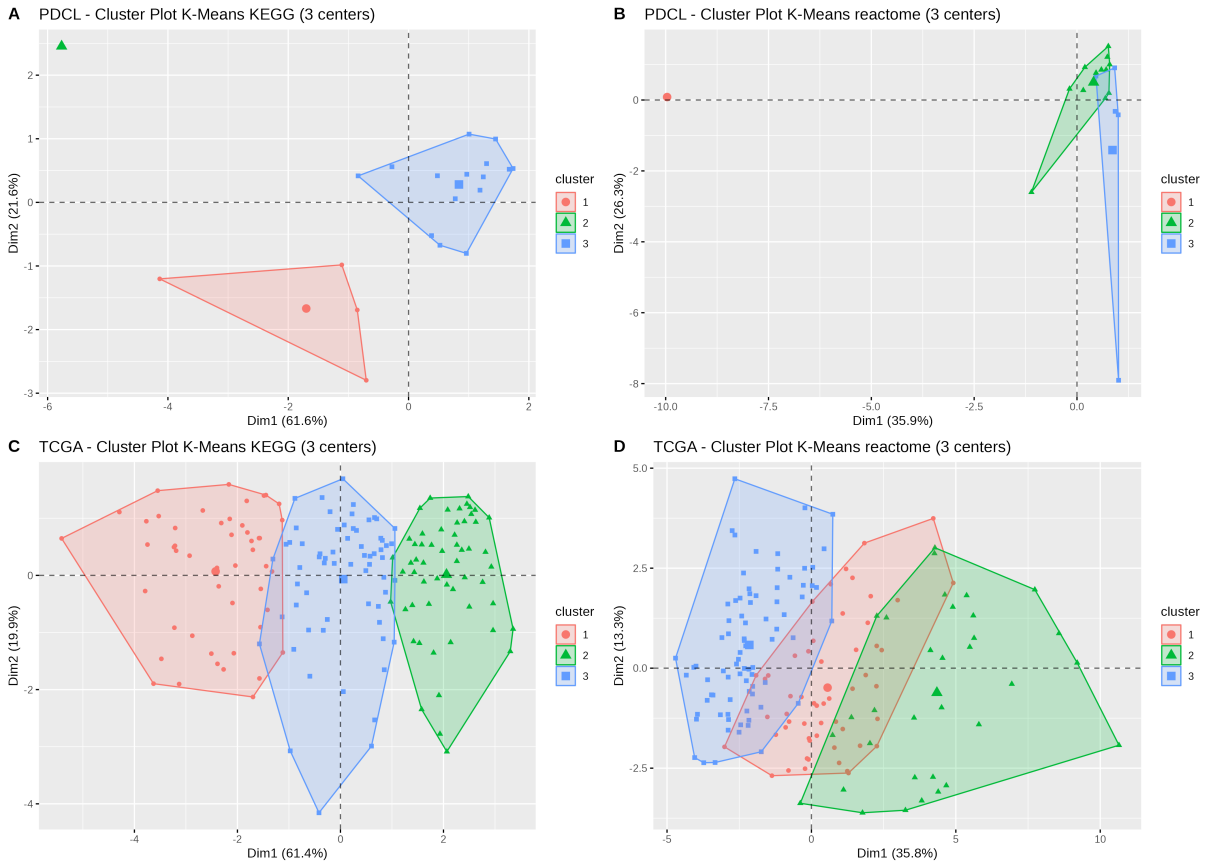

Figure 8: Clustering of (A-B) PDCL and (C-D) TCGA samples using the percentage of deregulation per biological categories. Results are shown with KEGG and Reactome categories. To generate a vector of values for each samples, we divided the number of pathways found significantly enriched for a category by the number of total pathways in this category present in the GMT file. Clustering was done using the K-Means method with 3 clusters. In A (PDCL samples with KEGG categories), cluster 1 seems to contains samples pathways are associated the *Organismal Systems/Environmental Information processing/Human Diseases*, the cluster 2 contain samples with deregulated pathways associated with *Metabolism/Cellular Processes/Genetic Information Processing*, and the cluster 3 contains samples with few deregulated pathways. In B (PDCL with Reactome categories), the cluster 1 contains samples with dysregulated pathways associated with *Cell-Cycle/DNA replication and repair/Gene expression/Signal Transduction/Immune System*, the cluster 2 with mainly *Metabolism*, and the cluster 3 with mainly *ECM/Neuronal System/Disease/Transport of small molecules*. In C (TCGA with KEGG), the cluster 1 contain samples with deregulated pathways mainly associated with *Environmental Information Processing/Organismal Systems/Human Diseases/Cellular Processes*, the cluster 2 with *Metabolism*, and the cluster 3 with *Genetic Information Processing*. In D (TCGA with Reactome), the cluster 1 contain samples with deregulated pathways mainly associated with *Immune System/Programmed Cell Death/Metabolism/Signal Transduction/Cellular Response to Stimuli/Vesicle-mediated transport*, the cluster 2 with *Cell-Cycle/DNA replication/DNA repair/Chromatin organization/Gene Expression (Transcription)*, and the cluster 3 with *Extra-Cellular Matrix Organisation/Cell-Cell communication/Hemostasis*.

Discussion

In this study we presented a methodology using two statiscal methods, Differential Expression analysis and Pathway Enrichment, to analyze RNA-Seq data and retrieve the altered biological functions. We will briefly review the literature on two results having a role in the Extra-Cellular Matrix (ECM) and Metabolism to show how this method helped identify documented processes favouring glioblastoma growth.

Focal adhesion, the adhesion of the cell to the ECM, is the most deregulated pathway in the PDCL datasets. Linked to the ECM, pathways involved in the collagen biosynthesis, a constituent of the ECM, are frequently deregulated. Evidence in the literature supports this result, showing the role of collagen and cell interaction in glioblastoma growth, invasiveness and migration. Furthermore, the constituent of the matrix in the tumour microenvironment differs from the composition of the matrix in the normal brain with higher collagen concentration [30]. Studies in the literature describe different mechanisms by which collagen can influence tumour growth. Collagen concentration and structure (crosslinking) have been documented to influence glioblastoma growth and invasion [31, 32, 33]. Glioblastoma adopts a rounded shape in collagen-IV while they adopt a spindle shape in collagen-I/III, showing an effect of the type of collagen on tumour's morphology [33]. Genes associated with fibrillar collagen and coding for collagen processing enzymes are upregulated in glioblastoma compared to grade III gliomas. As an example, a protein that binds to collagen to internalize it, called Endo180, is found upregulated in grade IV gliomas and correlates with collagen I deposition. This protein is lowly expressed in the normal brain but is strongly associated with the mesenchymal subclass of gliomas which is associated with grade IV gliomas, thus it has potential as a biomarker [34]. Although its expression differs between cell lines, collagen XVI has been found upregulated in glioblastoma compared to normal cells. Inhibition of endogenous expression of collagen XVI results in reduced cell adhesion to surface [35]. Collagen has been reported to affect angiogenesis, the formation of new blood vessels, important for tumour growth. Inhibition of collagen crosslinking and expression reduce angiogenesis and progression of glioblastoma without increasing cell invasion, hypoxia or necrosis in mice tumour compared to VEGF targeting anti-angiogenic drugs [30]. Thus, targeting collagen structure and expression appears to be a potential therapy to decrease angiogenesis without the adverse effects of regular angiogenesis therapy. These evidence combined with the results of our study suggest that targeting ECM of the glioblastoma microenvironment is a viable strategy in therapy.

A specificity of the PDCL dataset is the altered cholesterol metabolism. Brain is the most cholesterol-rich organ with approximately 20% of the whole body cholesterol. Its cholesterol metabolism differs from other organs as the cholesterol cannot pass through the Blood-Brain Barrier (BBB). Hence, cholesterol must be synthesized de novo in the brain [36, 37, 38]. As for interactions between the cell and the Extra-Cellular Matrix, few evidence in the literature describe varied mechanisms by which high levels of cholesterol may promote tumour growth. Firstly, densely packed normal astrocytes inhibit cholesterol synthesis while glioma cells maintain high cholesterol levels. Glioma cells were sensitive to cholesterol synthesis inhibition while astrocytes were not, suggesting that this pathway can be a viable target in

therapy [39]. Studies describe the antitumoral effect of many drugs targeting cholesterol metabolism in glioblastomas. For example, Liver X Receptor (LXR) agonist LXR-623 is able to induce cell death in glioblastoma mouse model and spare normal cells by downregulating Low-Desnity Lipoprotein Receptor (LDLR) mediated cholesterol intake [36, 38]. Phytol and retinol exhibit cytotoxic activity in glioblastoma cell lines (U87MG, A172 and T98G) in a dose-dependent way by down-regulating genes involved in cholesterol and/or fatty acid synthesis [40]. Targeting cholesterol trafficking with itraconazole suppresses the growth of glioblastoma by inducing autophagy [41] while alkylphospholipids also induced cell cycle arrest in G2/M phase in U87MG cells [42]. Cholesterol has been observed to modulate the efficiency of Temozolomide (TMZ), a drug widely used in treatment of glioblastoma. Its interaction with the BBB membrane increases its rigidity, slowing down the distribution of the TMZ active metabolite (MTIC) which cannot reach cytotoxic dose in tumoural tissue [43]. Higher levels of intracellular cholesterol enhance TMZ apoptosis activation through the DR5 plasma membrane death receptor [37]. The RTK/RAS/PI3K signalling pathway is frequently altered in glioblastoma with its component EGFR upregulated [44]. EGFR frequent upregulation causes the overexpression of YTHDF2, a downstream effector of the EGFR/SRC/ERK signalling. As a result, downregulation of the LXRα inhibits LXRα-dependent cholesterol homeostasis which promotes glioblastoma proliferation and invasion.

Our workflow starts with a Differential Expression analysis where we merely compare the gene expression of samples belonging to one condition with a control condition (usually a normal sample). Thus the quality of the controls used greatly influence the results. In this paper, cholesterol synthesis is found downregulated in the PDCL dataset which seems contradictory to the statement found in the literature (an increased cholesterol metabolism conferring an advantage to glioblastoma proliferation and invasion). Glioblastoma cells tend to increase their cholesterol intake rather than increase the synthesis of cholesterol, made by astrocytes cells [36, 38]. Controls used in this dataset were downloaded from another study and are composed of astrocyte cell lines which might explain this observation. We observed a high amount of pathways linked to diseases enriched in the results. This can be due to the biological variability introduced by using controls from a different study in the case of PDCL, by the lack of controls or an effect of using bulk tumour in the case of TCGA. We investigated the contribution of a gene to a pathway and determined how many times a gene is associated with a pathway in the results to determine whether some genes might be reponsible for this observation. We found that some group of genes are involved in several different categories including the Disease categorie which may partially explain the amount of pathways linked to dieases. More can be seen in the supplementary materials. Although we kept those pathways in our study, they should therefore filtered out of the GMT before any analysis.

We have not found in the literature any publication that has evaluated the performance of the DESeq2 statistic as the gene metric for GSEA. Generally, the fold change of a gene is used as a metric [2], but a study on the performance of 16 different metrics seems to show that tests statistic are better suited to rank genes for GSEA [18]. We first used the fold change to rank genes in our dataset, which lead to poor results with few pathways common between GSEA and G:Profiler (results not shown). In comparison, the DESeq2 statistic yields much better performance and more comparable results between the 2 enrichment tools. The FDR value for the *Glioma* entry in KEGG with the TCGA dataset indicate that this statistic

might be used as a potential metric to rank genes before enrichment with GSEA. More investigations to evaluate the performance of this metric are necessary.

In this paper we used a tool called Personalized Differential Analysis (PenDA) designed to perform single sample Differential Expression analysis. To our knowledge, this is the only tool available to carry such studies and outperforms more conventionnal tools like DESeq2 according to the authors [9]. This approaches can be usefull to better determine the intra-tumour heterogeneity. It can be applied to single-cell data or to spatial transcriptomic to assess cellular and spatial heterogeneity in a tumour. If it is unfortunately not possible to assess the difference in expression as PenDA does not compute any Fold Change, such tools will provide valuable information on the differences between cancer types. For example, the developers of PenDA identified biomarkers for several types of cancer available on the TCGA platform.

Conclusion

In this article, we present a method relying on a statistical test to analyze biological data produced by RNA-Seq experiments. We apply this method to glioblastoma, an aggressive and invasive kind of brain tumour with poor patient outcomes. We use a Patient Derived Cell Lines (PDCL) dataset consisting of RNA-Seq data from 20 patients diagnosed with glioblastoma. We compare this dataset with a dataset composed of RNA-Seq data from the TCGA-GBM project accessible on the TCGA database. Our workflow starts by running a Differential Expression analysis on the RNA-Seq data to find genes whose expressions are affected by the glioblastoma disease. This analysis produces a list of genes whose regulation is altered. To carry on this step, we selected the tools DESeq2 and PenDA as they both show good performances. The second step finds the biological pathways enriched in this list of genes. We submit the list of genes to G:Profiler and GSEA, two pathway enrichment tools. While G:Profiler only need the name of the deregulated genes, GSEA needs that each gene in the list is associated with a metric for its statistical test. We use the statistic computed by DESeq2 for its test as it gives the direction of deregulation (up or down) and the significance of the result (the further from zero, the more significant the deregulation is). A same mechanism can be defined slightly differently among the databases, hence we downloaded pathway informations from Reactome and KEGG to ensure the exhaustiveness of the results. We find that most of the mechanisms altered in both TCGA and PDCL are involved in the Cell-Cycle, DNA replication and Repair. In both cases, the *focal adhesion* and *Collagen formation* pathways are deregulated. Despite few pathways involved in the metabolism are deregulated, we found the *Cholesterol synthesis* affected by glioblastoma only in the PDCL dataset. When further studying the differences between the samples in the PDCL dataset, we show that *focal adhesion* is the most commonly deregulated pathway. Pathways involved in collagen synthesis and cholesterol metabolism are also among the most frequently deregulated. A review of the literature validated these results and show different mechanisms by which collagen and cholesterol can promote tumour growth, thus suggesting their potential in glioblastoma therapy.

Acknowledgements

Conflicts of Interest

The authors declare no conflict of interest.

Data availability

PDCL data cannot be shared in the publication due to licensing limitations.

References

[1] Ana Conesa et al. *A survey of best practices for RNA-seq data analysis*. Jan. 2016. DOI: 10.1186/s13059-016-0881-8. URL: /pmc/articles/PMC4728800/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4728800/.

[2] Jüri Reimand et al. "Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap HHS Public Access". In: *Nat Protoc* 14.2 (2019), pp. 482–517. DOI: 10.1038/s41596-018-0103-9. URL: https://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC6607905&blobtype=pdf%20https://www.nature.com/articles/s41596-018-0103-9.

[3] Purvesh Khatri, Marina Sirota, and Atul J. Butte. *Ten years of pathway analysis: Current approaches and outstanding challenges*. Feb. 2012. DOI: 10.1371/journal.pcbi.1002375. URL: /pmc/articles/PMC3285573/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3285573/.

[4] Addison Quinones et al. "The Multifaceted Metabolism of Glioblastoma". In: *Advances in Experimental Medicine and Biology* 1063 (2018), pp. 59–72. ISSN: 22148019. DOI: 10.1007/978-3-319-77736-8_4. URL: https://link.springer.com/chapter/10.1007/978-3-319-77736-8_4.

[5] H. W. Cheng et al. "Identification of thioridazine, an antipsychotic drug, as an antiglioblastoma and anticancer stem cell agent using public gene expression data". In: *Cell Death & Disease 2015 6:5* 6.5 (May 2015), e1753–e1753. ISSN: 2041-4889. DOI: 10.1038/cddis.2015.77. URL: https://www.nature.com/articles/cddis201577.

[6] Emilie Le Rhun et al. *Molecular targeted therapy of glioblastoma*. 2019. DOI: 10.1016/j.ctrv.2019.101896.

[7] P. D. Delgado-López and E. M. Corrales-García. *Survival in glioblastoma: a review on the impact of treatment modalities*. Nov. 2016. DOI: 10.1007/s12094-016-1497-x. URL: https://pubmed.ncbi.nlm.nih.gov/26960561/.

[8] B. Campos et al. "A comprehensive profile of recurrent glioblastoma". In: *Oncogene* 35.45 (2016), pp. 5819–5825. ISSN: 14765594. DOI: 10.1038/onc.2016.85.

[9] Magali Richard et al. “PenDA, a rank-based method for personalized differential analysis: Application to lung cancer”. In: *PLoS Computational Biology* 16 (5 May 2020). ISSN: 15537358. DOI: 10.1371/journal.pcbi.1007869.

[10] C. Neftel et al. “An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma”. In: *Cell* 178.4 (Aug. 2019), 835–849.e21. ISSN: 10974172. DOI: 10.1016/j.cell.2019.06.024. URL: https://www.sciencedirect.com/science/article/pii/S0092867419306877?pes=vor&entityID=https%3A%2F%2Fshibidp.cit.cornell.edu%2Fidp%2Fshibboleth%20https://reader.elsevier.com/reader/sd/pii/S0092867419306877?token=BB6AA03C2AF2757168747ECFEE29078B6084C7

[11] Anders Lundin et al. “Human iPS-Derived Astroglia from a Stable Neural Precursor State Show Improved Functionality Compared with Conventional Astrocytic Models”. In: *Stem Cell Reports* 10 (3 Mar. 2018), pp. 1030–1045. ISSN: 2213-6711. DOI: 10.1016/J.STEMCR.2018.01.021.

[12] Anna Falk et al. “Capture of Neuroepithelial-Like Stem Cells from Pluripotent Stem Cells Provides a Versatile System for In Vitro Production of Human Neurons”. In: *PLOS ONE* 7 (1 Jan. 2012), e29597. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0029597. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029597%20http://www.stammzellen.nrw.de/;L-072.0055%20http://www.ghst.de/%20http://www.vr.se%20http://www.hjarnfonden.se%20http://www.embo.org%20http://www.nihr.ac.uk%20http://www.medschl.cam.ac.uk.

[13] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. “edgeR: a Bioconductor package for differential expression analysis of digital gene expression data”. In: *Bioinformatics* 26 (1 Jan. 2010), pp. 139–140. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTP616. URL: https://academic.oup.com/bioinformatics/article/26/1/139/182458.

[14] Matthew E. Ritchie et al. “limma powers differential expression analyses for RNA-sequencing and microarray studies”. In: *Nucleic Acids Research* 43.7 (Apr. 2015), e47–e47. ISSN: 0305-1048. DOI: 10.1093/NAR/GKV007. URL: https://academic.oup.com/nar/article/43/7/e47/2414268.

[15] Michael I. Love, Wolfgang Huber, and Simon Anders. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology* 15.12 (Dec. 2014), pp. 1–21. ISSN: 1474760X. DOI: 10.1186/s13059-014-0550-8. URL: https://link.springer.com/article/10.1186/s13059-014-0550-8.

[16] Uku Raudvere et al. “g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)”. In: *Nucleic Acids Research* 47.W1 (July 2019), W191–W198. ISSN: 0305-1048. DOI: 10.1093/NAR/GKZ369. URL: https://academic.oup.com/nar/article/47/W1/W191/5486750.

[17] Aravind Subramanian et al. “Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.43 (Oct. 2005), pp. 15545–15550. ISSN: 00278424. DOI: 10.1073/pnas.0506580102. URL: /pmc/articles/PMC1239896/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239896/.

[18] Joanna Zyla et al. “Ranking metrics in gene set enrichment analysis: do they matter?” In: *BMC Bioinformatics 2017 18:1* 18.1 (May 2017), pp. 1–12. ISSN: 1471-2105. DOI: 10.1186/S12859-017-1674-0. URL: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1674-0.

[19] Minoru Kanehisa et al. “New approach for understanding genome variations in KEGG”. In: *Nucleic Acids Research* 47 (D1 Jan. 2019), pp. D590–D595. ISSN: 0305-1048. DOI: 10.1093/NAR/GKY962. URL: https://academic.oup.com/nar/article/47/D1/D590/5128935.

[20] Marc Gillespie et al. “The reactome pathway knowledgebase 2022”. In: *Nucleic Acids Research* 50 (D1 Jan. 2022), pp. D687–D692. ISSN: 0305-1048. DOI: 10.1093/NAR/GKAB1028. URL: https://academic.oup.com/nar/article/50/D1/D687/6426058.

[21] Gennady Korotkevich et al. “Fast gene set enrichment analysis”. In: *bioRxiv* (Feb. 2021), p. 060012. DOI: 10.1101/060012. URL: https://www.biorxiv.org/content/10.1101/060012v3%20https://www.biorxiv.org/content/10.1101/060012v3.abstract.

[22] Clémentine Decamps. “Développements méthodologiques pour la caractérisation de l’hétérogénéité tumorale”. In: (Nov. 2021). URL: https://theses.hal.science/tel-03601942%20https://theses.hal.science/tel-03601942/document.

[23] Sarah Mubeen et al. “The Impact of Pathway Database Choice on Statistical Enrichment Analysis and Predictive Modeling”. In: *Frontiers in Genetics* 10 (Nov. 2019), p. 1203. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.01203. URL: https://www.frontiersin.org/article/10.3389/fgene.2019.01203/full.

[24] Tamara T. Lah, Metka Novak, and Barbara Breznik. “Brain malignancies: Glioblastoma and brain metastases”. In: *Seminars in Cancer Biology* 60 (October 2019 2020), pp. 262–273. ISSN: 10963650. DOI: 10.1016/j.semcancer.2019.10.010.

[25] Kévin Spinicci et al. “Modeling the role of HIF in the regulation of metabolic key genes LDH and PDH: Emergence of Warburg phenotype”. In: *Computational and Systems Oncology* 2 (3 Sept. 2022), e1040. ISSN: 2689-9655. DOI: 10.1002/CSO2.1040. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/cso2.1040%20https://onlinelibrary.wiley.com/doi/abs/10.1002/cso2.1040%20https://onlinelibrary.wiley.com/doi/10.1002/cso2.1040.

[26] Rob A. Cairns, Isaac S. Harris, and Tak W. Mak. *Regulation of cancer cell metabolism*. Feb. 2011. DOI: 10.1038/nrc2981. URL: https://www.nature.com/articles/nrc2981.

[27] B. Wei et al. “Co-mutated pathways analysis highlights the coordination mechanism in glioblastoma multiforme.” In: *Neoplasma* 61 (4 Jan. 2014), pp. 424–432. ISSN: 0028-2685. DOI: 10.4149/NEO_2014_052. URL: https://europepmc.org/article/med/25027741.

[28] Zhijun Liu et al. “Induction of store-operated calcium entry (SOCE) suppresses glioblastoma growth by inhibiting the Hippo pathway transcriptional coactivators YAP/TAZ”. In: *Oncogene 2018 38:1* 38 (1 Aug. 2018), pp. 120–139. ISSN: 1476-5594. DOI: 10.1038/s41388-018-0425-7. URL: https://www.nature.com/articles/s41388-018-0425-7.

[29] Upal Basu-Roy et al. “Sox2 antagonizes the Hippo pathway to maintain stemness in cancer cells”. In: *Nature Communications 2015 6:1* 6 (1 Apr. 2015), pp. 1–14. ISSN: 2041-1723. DOI: 10.1038/ncomms7411. URL: https://www.nature.com/articles/ncomms7411.

[30] Tadanori Mammoto et al. “Role of collagen matrix in tumor angiogenesis and glioblastoma multiforme progression”. In: *American Journal of Pathology* 183.4 (2013), pp. 1293–1305. ISSN: 00029440. DOI: 10.1016/j.ajpath.2013.06.026. URL: http://dx.doi.org/10.1016/j.ajpath.2013.06.026.

[31] Pranita Kaphle, Yongchao Li, and Li Yao. “The mechanical and pharmacological regulation of glioblastoma cell migration in 3D matrices”. In: *Journal of Cellular Physiology* 234.4 (2019), pp. 3948–3960. ISSN: 10974652. DOI: 10.1002/jcp.27209.

[32] Laura J. Kaufman et al. “Glioma expansion in collagen I matrices: Analyzing collagen concentration-dependent growth and motility patterns”. In: *Biophysical Journal* 89.1 (2005), pp. 635–650. ISSN: 00063495. DOI: 10.1529/biophysj.105.061994. URL: http://dx.doi.org/10.1529/biophysj.105.061994.

[33] Shreyas S. Rao et al. “Glioblastoma behaviors in three-dimensional collagen-hyaluronan composite hydrogels”. In: *ACS Applied Materials and Interfaces* 5.19 (2013), pp. 9276–9284. ISSN: 19448252. DOI: 10.1021/am402097j.

[34] Ivo J. Huijbers et al. “A role for fibrillar collagen deposition and the collagen internalization receptor endo180 in glioma invasion”. In: *PLoS ONE* 5.3 (2010), pp. 1–12. ISSN: 19326203. DOI: 10.1371/journal.pone.0009808.

[35] Volker Senner et al. “Collagen XVI expression is upregulated in glioblastomas and promotes tumor cell adhesion”. In: *FEBS Letters* 582.23-24 (2008), pp. 3293–3300. ISSN: 00145793. DOI: 10.1016/j.febslet.2008.09.017.

[36] Genaro R. Villa et al. “An LXR-Cholesterol Axis Creates a Metabolic Co-Dependency for Brain Cancers”. In: *Cancer Cell* 30.5 (2016), pp. 683–693. ISSN: 18783686. DOI: 10.1016/j.ccell.2016.09.008. URL: http://dx.doi.org/10.1016/j.ccell.2016.09.008.

[37] Yutaro Yamamoto et al. “Involvement of intracellular cholesterol in temozolomide-induced glioblastoma cell death”. In: *Neurologia Medico-Chirurgica* 58.7 (2018), pp. 296–302. ISSN: 13498029. DOI: 10.2176/nmc.ra.2018-0040.

[38] Leila Pirmoradi et al. “Targeting cholesterol metabolism in glioblastoma: A new therapeutic approach in cancer therapy”. In: *Journal of Investigative Medicine* 67.4 (2019), pp. 715–719. ISSN: 17088267. DOI: 10.1136/jim-2018-000962.

[39] Diane M. Kambach et al. “Disabled cell density sensing leads to dysregulated cholesterol synthesis in glioblastoma”. In: *Oncotarget* 8.9 (2017), pp. 14860–14875. ISSN: 19492553. DOI: 10.18632/oncotarget.14740.

[40] Gustavo Facchini et al. “Toxic effects of phytol and retinol on human glioblastoma cells are associated with modulation of cholesterol and fatty acid biosynthetic pathways”. In: *Journal of Neuro-Oncology* 136.3 (2018), pp. 435–443. ISSN: 15737373. DOI: 10.1007/s11060-017-2672-9. URL: http://dx.doi.org/10.1007/s11060-017-2672-9.

[41] Rui Liu et al. “Itraconazole suppresses the growth of glioblastoma through induction of autophagy: Involvement of abnormal cholesterol trafficking”. In: *Autophagy* 10.7 (2014), pp. 1241–1255. ISSN: 15548635. DOI: 10.4161/auto.28912.

[42] Pablo Ríos-Marco et al. “Alkylphospholipids deregulate cholesterol metabolism and induce cell-cycle arrest and autophagy in U-87 MG glioblastoma cells”. In: *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1831.8 (2013), pp. 1322–1334. ISSN: 18792618. DOI: 10.1016/j.bbalip.2013.05.004. URL: http://dx.doi.org/10.1016/j.bbalip.2013.05.004.

[43] Maria João Ramalho et al. “Biophysical interaction of temozolomide and its active metabolite with biomembrane models: The relevance of drug-membrane interaction for Glioblastoma Multiforme therapy”. In: *European Journal of Pharmaceutics and Biopharmaceutics* 136.August 2018 (2019), pp. 156–163. ISSN: 18733441. DOI: 10.1016/j.ejpb.2019.01.015. URL: https://doi.org/10.1016/j.ejpb.2019.01.015.

[44] Roger McLendon et al. “Comprehensive genomic characterization defines human glioblastoma genes and core pathways”. In: *Nature* 455.7216 (2008), pp. 1061–1068. ISSN: 00280836. DOI: 10.1038/nature07385. URL: https://www.nature.com/articles/nature07385.

Supplementary Material

Gene functions are overlapping in several biological categories. To better understand the reason why many pathways associated with diseases emerge in the results despite not all of them are associated with cancer. To that end we tried to assess wheter genes that were associated with biological functions that are often deregulated in cancer were also present in pathways associated to dieases. Figures 9 and 10 shows that genes are raely associated to one pathway and one biological category only in both Reactome and KEGG. In KEGG, genes involved in *Genetic Information Processing* or in *Cellular Processes* are also involved in pathways linked to diseases. In addition, some of these genes are also upregulated in 90% of the samples. In Reactome, we see that genes that are involved in the *Extracellular Matrix Organization*

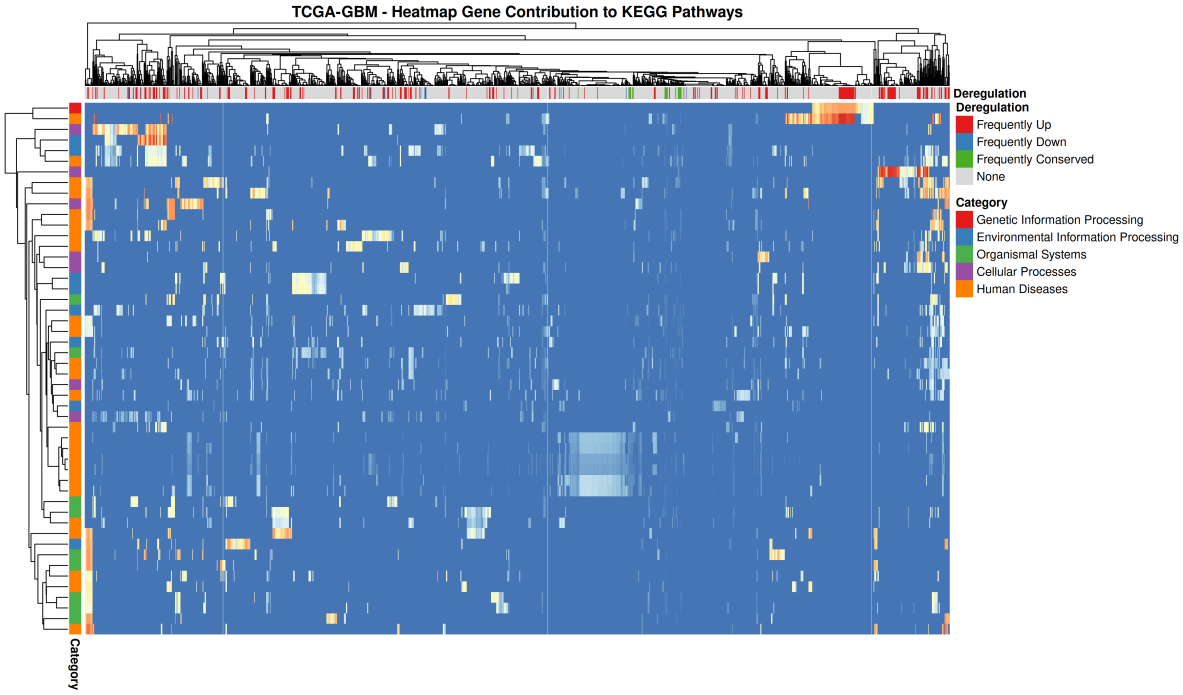

Figure 9: Heatmap of the gene contributions for each pathways with TCGA data and KEGG categories. Genes contribution to a pathway is computed by calculating how many times a gene appear in the enrichment result for a pathway. Red colour indicates a strong contribution and blue colour indicates a week contribution. Only genes with a total contributions higher than the third quartile and pathways where the total gene contribution is higher than the third quartile are kept. Pathways are colored by their categories. Genes are colored by their deregulation frequency : red indicates genes that are up-regulated in 90% of the samples, blue indicates genes that are down-regulated in 90% of the samples, green indicates genes that are not deregulated in 90% of the samples and grey genes that correspond to none of the above.

and in the *Metabolism of proteins* are also involed in pathways linked to diseases. Like for KEGG, some are upregulated in 90% of the samples as well.

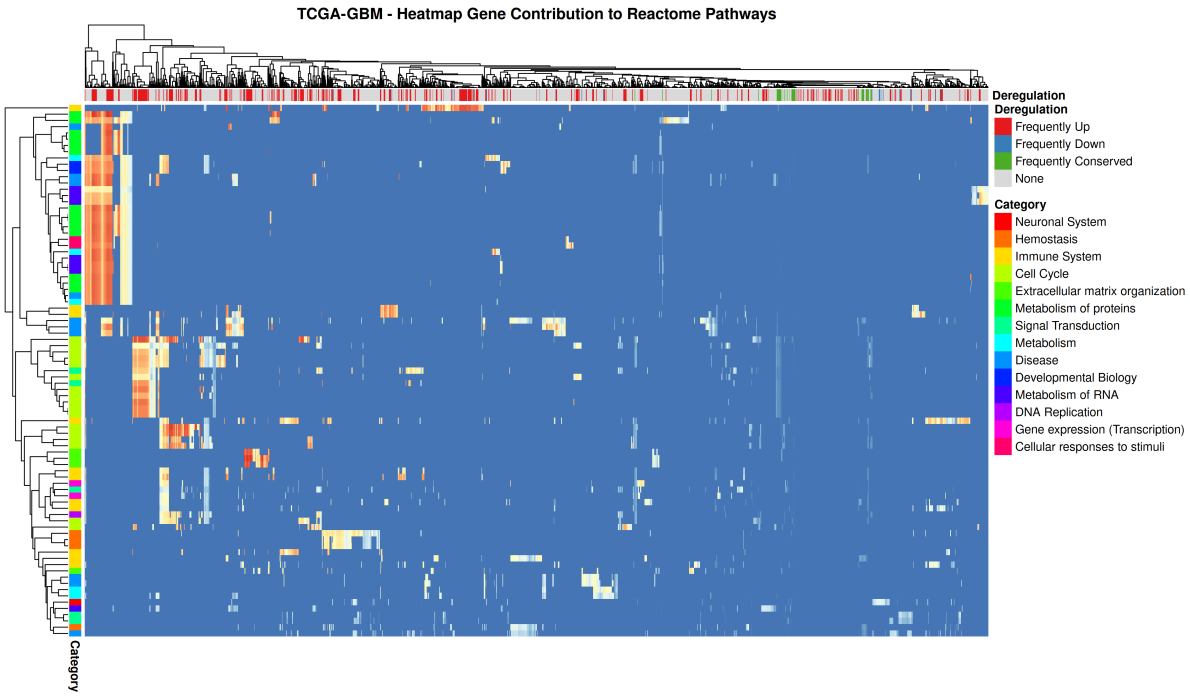

Figure 10: Heatmap of the gene contributions for each pathways with TCGA data and Reactome categories. Genes contribution to a pathway is computed by calculating how many times a gene appear in the enrichment result for a pathway. Red colour indicates a strong contribution and blue colour indicates a week contribution. Only genes with a total contributions higher than the 90% percentile and pathways where the total gene contribution is higher than the third quartile are kept. Pathways are colored by their categories. Genes are colored by their deregulation frequency : red indicates genes that are up-regulated in 90% of the samples, blue indicates genes that are down-regulated in 90% of the samples, green indicates genes that are not deregulated in 90% of the samples and grey genes that correspond to none of the above.

A.3 Complete tables of frequently deregulated pathways in TCGA and PDCL data

Pathway ID	Description	Category	Number of PDCL
path:hsa04510	Focal adhesion	Cellular Processes	13
path:hsa04360	Axon guidance	Organismal Systems	10
path:hsa04512	ECM-receptor interaction	Environmental Information Processing	5
path:hsa04724	Glutamatergic synapse	Organismal Systems	5
path:hsa04725	Cholinergic synapse	Organismal Systems	5
path:hsa00100	Steroid biosynthesis	Metabolism	4
path:hsa00190	Oxidative phosphorylation	Metabolism	4
path:hsa04612	Antigen processing and presentation	Organismal Systems	4
path:hsa04723	Retrograde endocannabinoid signaling	Organismal Systems	4
path:hsa03008	Ribosome biogenesis in eukaryotes	Genetic Information Processing	3
path:hsa04012	ErbB signaling pathway	Environmental Information Processing	3
path:hsa04072	Phospholipase D signaling pathway	Environmental Information Processing	3
path:hsa04392	Hippo signaling pathway - multiple species	Environmental Information Processing	3
path:hsa04514	Cell adhesion molecules	Environmental Information Processing	3
path:hsa04714	Thermogenesis	Organismal Systems	3

Table of the frequently Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways deregulated in the PDCL. *Focal adhesion* is the most frequently deregulated pathway with 13 samples affected.

Pathway ID	Description	Category	Number of PDCL
R-HSA-112315	Transmission across Chemical Synapses	Neuronal System	5
R-HSA-1650814	Collagen biosynthesis and modifying enzymes	Extracellular matrix organization	4
R-HSA-5576891	Cardiac conduction	Muscle contraction	4
R-HSA-8948216	Collagen chain trimerization	Extracellular matrix organization	4
R-HSA-112314	Neurotransmitter receptors and postsynaptic signal transmission	Neuronal System	3
R-HSA-191273	Cholesterol biosynthesis	Metabolism	3
R-HSA-6790901	rRNA modification in the nucleus and cytosol	Metabolism of RNA	3
R-HSA-6794362	Protein-protein interactions at synapses	Neuronal System	3
R-HSA-109581	Apoptosis	Programmed Cell Death	2
R-HSA-1296071	Potassium Channels	Neuronal System	2
R-HSA-1566948	Elastic fibre formation	Extracellular matrix organization	2
R-HSA-157579	Telomere Maintenance	Cell Cycle	2
R-HSA-174417	Telomere C-strand (Lagging Strand) Synthesis	Cell Cycle	2
R-HSA-187037	Signaling by NTRK1 (TRKA)	Signal Transduction	2
R-HSA-204998	Cell death signalling via NRAGE, NRIF and NADE	Signal Transduction	2

Table of the frequently Reactome pathways deregulated in the PDCL. *Transmission across Chemical Synapses* is the most frequently deregulated pathway with 5 samples affected. Collagen pathways are dysregulated in 4 different samples.

Pathway ID	Description	Category	Number of PDCL
path:hsa04110	Cell Cycle	Cellular Processes	153
path:hsa04512	ECM-receptor interaction	Environmental Information Processing	147
path:hsa04145	Phagosome	Cellular Processes	130
path:hsa04510	Focal adhesion	Cellular Processes	129
path:hsa03030	DNA replication	Genetic Information Processing	126
path:hsa04640	Hematopoietic cell lineage	Organismal Systems	125
path:hsa04672	Intestinal immune network for IgA production	Organismal Systems	125
path:hsa04514	Cell adhesion molecules	Environmental Information Processing	124
path:hsa04612	Antigen processing and presentation	Organismal Systems	124
path:hsa03010	Ribosome	Genetic Information Processing	123
path:hsa04115	p53 signaling pathway	Cellular Processes	118
path:hsa04380	Osteoclast differentiation	Organismal Systems	94
path:hsa04218	Cellular senescence	Cellular Processes	92
path:hsa04658	Th1 and Th2 cell differentiation	Organismal Systems	87
path:hsa04061	Viral protein interaction with cytokine and cytokine receptor	Environmental Information Processing	86

Table of the frequently KEGG pathways deregulated in the TCGA. Unsurprisingly, the *Cell Cycle* pathway is the most deregulated one with 153 out of 169 samples affected. Similar to the PDCL dataset, the *Focal adhesion* is one of the most frequently deregulated pathways with 129 samples. Other pathways involved in the interaction between the cell and its environment are also frequently deregulated. p53 signalling, a pathway known to be frequently deregulated in the case of glioblastoma, is also found in the results.

Pathway ID	Description	Category	Number of PDCL
R-HSA-1474290	Collagen formation	Extracellular matrix organization	159
R-HSA-1799339	SRP-dependent cotranslational protein targeting to membrane	Metabolism of proteins	156
R-HSA-453279	Mitotic G1 phase and G1/S transition	Cell Cycle	154
R-HSA-1650814	Collagen biosynthesis and modifying enzymes	Extracellular matrix organization	152
R-HSA-69206	G1/S Transition	Cell Cycle	151
R-HSA-2022090	Assembly of collagen fibrils and other multimeric structures	Extracellular matrix organization	149
R-HSA-6798695	Neutrophil degranulation	Immune System	149
R-HSA-69620	Cell Cycle Checkpoints	Cell Cycle	149
R-HSA-72695	Formation of the ternary complex, and subsequently, the 43S complex	Metabolism of proteins	149
R-HSA-72662	Activation of the mRNA upon binding of the cap-binding complex and eIFs, and subsequent binding to 43S	Metabolism of proteins	147
R-HSA-72649	Translation initiation complex formation	Metabolism of proteins	146
R-HSA-72689	Formation of a pool of free 40S subunits	Metabolism of proteins	146
R-HSA-156827	L13a-mediated translational silencing of Ceruloplasmin expression	Metabolism of proteins	145
R-HSA-2408522	Selenoamino acid metabolism	Metabolism	145
R-HSA-72702	Ribosomal scanning and start codon recognition	Metabolism of proteins	145

Table of the frequently Reactome pathways deregulated in the TCGA. Similar to the results obtained with the PDCL dataset, pathways linked to the collagen protein are found frequently deregulated (in more than 140 samples). Results in the table seem to indicate that *G1/S phase transition* is the main phase of the cell cycle dysregulated in glioblastoma.

Bibliography

[1] *Cancers : les chiffres clés - Qu'est-ce qu'un cancer ?* URL: https://www.e-cancer.fr/Comprendre-prevenir-depister/Qu-est-ce-qu-un-cancer/Chiffres-cles (visited on 10/10/2023).

[2] *Cancer Statistics for the UK*. Cancer Research UK. May 13, 2015. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk (visited on 10/10/2023).

[3] *Le ou les cancers ? - Qu'est-ce qu'un cancer ?* URL: https://www.e-cancer.fr/Comprendre-prevenir-depister/Qu-est-ce-qu-un-cancer/Le-ou-les-cancers (visited on 10/10/2023).

[4] Douglas Hanahan and Robert A. Weinberg. "Hallmarks of Cancer: The Next Generation". In: *Cell* 144.5 (Mar. 4, 2011). Publisher: Elsevier, pp. 646–674. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2011.02.013. URL: http://www.cell.com/article/S0092867411001279/fulltext (visited on 09/27/2021).

[5] Addison Quinones, Anne Le, A Quinones, and A Le. "The Multifaceted Metabolism of Glioblastoma". In: *Advances in Experimental Medicine and Biology* 1063 (2018). Publisher: Springer, Cham, pp. 59–72. ISSN: 22148019. DOI: 10.1007/978-3-319-77736-8_4. URL: https://link.springer.com/chapter/10.1007/978-3-319-77736-8_4 (visited on 04/13/2022).

[6] P. D. Delgado-López and E. M. Corrales-García. "Survival in glioblastoma: a review on the impact of treatment modalities". In: *Clinical and Translational Oncology* 18.11 (Nov. 1, 2016). Publisher: Springer-Verlag Italia s.r.l., pp. 1062–1071. ISSN: 16993055. DOI: 10.1007/s12094-016-1497-x. URL: https://pubmed.ncbi.nlm.nih.gov/26960561/ (visited on 01/05/2021).

[7] Tamara T. Lah, Metka Novak, and Barbara Breznik. "Brain malignancies: Glioblastoma and brain metastases". In: *Seminars in Cancer Biology* 60 (October 2019 2020), pp. 262–273. ISSN: 10963650. DOI: 10.1016/j.semcancer.2019.10.010.

[8] Ana Rita Monteiro, Richard Hill, Geoffrey J. Pilkington, and Patrícia A. Madureira. "The Role of Hypoxia in Glioblastoma Invasion". In: *Cells 2017, Vol. 6, Page 45* 6.4 (Nov. 22, 2017). Publisher: Multidisciplinary Digital Publishing Institute, p. 45. ISSN: 2073-4409. DOI: 10.3390/CELLS6040045. URL: https://www.mdpi.com/2073-4409/6/4/45/htm (visited on 01/11/2023).

[9] Roger McLendon et al. "Comprehensive genomic characterization defines human glioblastoma genes and core pathways". In: *Nature* 455.7216 (2008), pp. 1061–1068. ISSN: 00280836. DOI: 10.1038/nature07385. URL: https://www.nature.com/articles/nature07385.

[10] The Cancer Genome Atlas Research Network. "Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas". In: *New England Journal of Medicine* 372.26 (June 25, 2015). Publisher: Massachusetts Medical Society, pp. 2481–2498. ISSN: 0028-4793. DOI: 10.1056/NEJMOA1402121/SUPPL_FILE/NEJMOA1402121_DISCLOSURES.PDF. URL: https://www.nejm.org/doi/10.1056/NEJMoa1402121 (visited on 11/04/2022).

[11] C. Neftel et al. "An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma". In: *Cell* 178.4 (Aug. 8, 2019). Publisher: Cell Press, 835–849.e21. ISSN: 10974172. DOI: 10.1016/j.cell.2019.06.024. URL: https://www.sciencedirect.com/science/article/pii/S0092867419306877?pes=vor&entityID=https%3A%2F%2Fshibidp.cit.cornell.edu%2Fidp%2Fshibboleth (visited on 10/01/2020).

[12] B. Campos, L. R. Olsen, T. Urup, and H. S. Poulsen. "A comprehensive profile of recurrent glioblastoma". In: *Oncogene* 35.45 (2016). Publisher: Nature Publishing Group, pp. 5819–5825. ISSN: 14765594. DOI: 10.1038/onc.2016.85.

[13] Arabel Vollmann-Zwerenz, Verena Leidgens, Giancarlo Feliciello, Christoph A. Klein, and Peter Hau. "Tumor Cell Invasion in Glioblastoma". In: *International Journal of Molecular Sciences 2020, Vol. 21, Page 1932* 21.6 (Mar. 12, 2020). Publisher: Multidisciplinary Digital Publishing Institute, p. 1932. ISSN: 1422-0067. DOI: 10.3390/IJMS21061932. URL: https://www.mdpi.com/1422-0067/21/6/1932/htm (visited on 01/09/2023).

[14] Emilie Le Rhun, Matthias Preusser, Patrick Roth, David A. Reardon, Martin van den Bent, Patrick Wen, Guido Reifenberger, and Michael Weller. “Molecular targeted therapy of glioblastoma”. In: *Cancer Treatment Reviews* 80 (2019). ISSN: 15321967. DOI: 10.1016/j.ctrv.2019.101896.

[15] H. W. Cheng et al. “Identification of thioridazine, an antipsychotic drug, as an antiglioblastoma and anticancer stem cell agent using public gene expression data”. In: *Cell Death & Disease 2015 6:5* 6.5 (May 7, 2015). Publisher: Nature Publishing Group, e1753–e1753. ISSN: 2041-4889. DOI: 10.1038/cddis.2015.77. URL: https://www.nature.com/articles/cddis201577 (visited on 09/06/2022).

[16] S. R. McKeown. “Defining normoxia, physoxia and hypoxia in tumours - Implications for treatment response”. In: *British Journal of Radiology* 87.1035 (Mar. 1, 2014). Publisher: British Institute of Radiology. ISSN: 00071285. DOI: 10.1259/bjr.20130676. URL: /pmc/articles/PMC4064601/ (visited on 02/10/2021).

[17] Ming Zhu Jin and Wei Lin Jin. “The updated landscape of tumor microenvironment and drug repurposing”. In: *Signal Transduction and Targeted Therapy 2020 5:1* 5.1 (Aug. 25, 2020). Publisher: Nature Publishing Group, pp. 1–16. ISSN: 2059-3635. DOI: 10.1038/s41392-020-00280-x. URL: https://www.nature.com/articles/s41392-020-00280-x (visited on 01/18/2023).

[18] Eric Leung et al. “Metabolic targeting of HIF-dependent glycolysis reduces lactate, increases oxygen consumption and enhances response to high-dose single-fraction radiotherapy in hypoxic solid tumors”. In: *BMC Cancer* 17.1 (Dec. 15, 2017). Publisher: BioMed Central Ltd., p. 418. ISSN: 1471-2407. DOI: 10.1186/s12885-017-3402-6. URL: https://bmccancer.biomedcentral.com/articles/10.1186/s12885-017-3402-6 (visited on 01/05/2021).

[19] Young Heon Kim, Ki Chun Yoo, Yan Hong Cui, Nizam Uddin, Eun Jung Lim, Min Jung Kim, Seon Young Nam, In Gyu Kim, Yongjoon Suh, and Su Jae Lee. “Radiation promotes malignant progression of glioma cells through HIF-1alpha stabilization”. In: *Cancer Letters* 354.1 (Nov. 1, 2014). Publisher: Elsevier, pp. 132–141. ISSN: 0304-3835. DOI: 10.1016/J.CANLET.2014.07.048. (Visited on 01/10/2023).

[20] Hui Xu et al. “Activation of hypoxia signaling induces phenotypic transformation of glioma cells: implications for bevacizumab antiangiogenic therapy”. In: *Oncotarget* 6.14 (Mar. 14, 2015). Publisher: Impact Journals, pp. 11882–11893. ISSN: 1949-2553. DOI: 10.18632/ONCOTARGET.3592. URL: https://www.oncotarget.com/article/3592/text/ (visited on 01/12/2023).

[21] Rupert Courtnay, Darleen C. Ngo, Neha Malik, Katherine Ververis, Stephanie M. Tortorella, and Tom C. Karagiannis. “Cancer metabolism and the Warburg effect: the role of HIF-1 and PI3K”. In: *Molecular biology reports* 42.4 (2015), pp. 841–851. ISSN: 15734978. DOI: 10.1007/s11033-015-3858-x.

[22] Georgina N. Masoud and Wei Li. “HIF-1α pathway: Role, regulation and intervention for cancer therapy”. In: *Acta Pharmaceutica Sinica B* 5.5 (2015). ISSN: 22113843. DOI: 10.1016/j.apsb.2015.05.007. URL: https://www.sciencedirect.com/science/article/pii/S2211383515000817?pes=vor (visited on 10/14/2020).

[23] J. Xu et al. “Epigenetic regulation of HIF-1α in renal cancer cells involves HIF-1α/2α binding to a reverse hypoxia-response element”. In: *Oncogene* 31.8 (2012). ISSN: 09509232. DOI: 10.1038/onc.2011.305. URL: https://pubmed.ncbi.nlm.nih.gov/21841824/ (visited on 12/17/2020).

[24] Laura D’Ignazio, Michael Batie, and Sonia Rocha. “Hypoxia and inflammation in cancer, focus on HIF and NF-κB”. In: *Biomedicines* 5.2 (June 1, 2017). Publisher: MDPI AG. ISSN: 22279059. DOI: 10.3390/biomedicines5020021. URL: /pmc/articles/PMC5489807/?report=abstract (visited on 10/21/2020).

[25] Lucija Slemc and Tanja Kunej. “Transcription factor HIF1A: downstream targets, associated pathways, polymorphic hypoxia response element (HRE) sites, and initiative for standardization of reporting in scientific literature”. In: *Tumor Biology* 37.11 (2016). ISSN: 14230380. DOI: 10.1007/s13277-016-5331-4. URL: https://link.springer.com/article/10.1007/s13277-016-5331-4 (visited on 12/11/2020).

[26] Ji-Won Lee, Seong-Hui Bae, Joo-Won Jeong, Se-Hee Se-Hee Kim1 Kim, and Kyu-Won Kim. “Hypoxia-inducible factor (HIF -1)alpha : its protein stability and biological function s”. In: *Experimental and Molecular Medicine* 36.1 (2004), pp. 1–12. DOI: 10.1038/emm.2004.1. URL: https://www.nature.com/articles/emm20041.

[27] Baptiste Bedessem. *Contributions à l'étude de la réponse moléculaire à l'hypoxie : Modélisation mathématique et expérimentations sur cellules FUCCI*. Université Grenoble Alpes, Oct. 23, 2015. URL: https://tel.archives-ouvertes.fr/tel-01318127 (visited on 10/29/2020).

[28] Alvaro Marin-Hernandez, Juan Gallardo-Perez, Stephen Ralph, Sara Rodriguez-Enriquez, and Rafael Moreno-Sanchez. "HIF-1alpha Modulates Energy Metabolism in Cancer Cells by Inducing Over-Expression of Specific Glycolytic Isoforms". In: *Mini-Reviews in Medicinal Chemistry* 9.9 (Aug. 1, 2009). Publisher: Bentham Science Publishers Ltd., pp. 1084–1101. ISSN: 13895575. DOI: 10.2174/138955709788922610. URL: http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1389-5575&volume=9&issue=9&spage=1084 (visited on 01/25/2021).

[29] Yoshihiro Hayashi, Asumi Yokota, Hironori Harada, and Gang Huang. "Hypoxia/pseudohypoxia-mediated activation of hypoxia-inducible factor-1α in cancer". In: *Cancer Science* 110.5 (May 23, 2019). Publisher: Blackwell Publishing Ltd, pp. 1510–1517. ISSN: 1347-9032. DOI: 10.1111/cas.13990. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cas.13990 (visited on 01/07/2021).

[30] Balaji Krishnamachary et al. *Regulation of Colon Carcinoma Cell Invasion by Hypoxia-Inducible Factor 1 1*. Publication Title: CANCER RESEARCH Volume: 63. 2003, pp. 1138–1143. URL: https://cancerres.aacrjournals.org/content/63/5/1138 (visited on 04/22/2021).

[31] Kévin Spinicci, Pierre Jacquet, Gibin Powathil, and Angélique Stéphanou. "Modeling the role of HIF in the regulation of metabolic key genes LDH and PDH: Emergence of Warburg phenotype". In: *Computational and Systems Oncology* 2.3 (Sept. 1, 2022). Publisher: John Wiley & Sons, Ltd, e1040. ISSN: 2689-9655. DOI: 10.1002/CSO2.1040. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/cso2.1040 (visited on 08/25/2022).

[32] Esha Madan et al. "HIF-transcribed p53 chaperones HIF-1α". In: *Nucleic Acids Research* 47.19 (Nov. 4, 2019). Publisher: Oxford Academic ISBN: 1021210234, pp. 10212–10234. ISSN: 0305-1048. DOI: 10.1093/NAR/GKZ766. URL: https://academic.oup.com/nar/article/47/19/10212/5563951 (visited on 01/04/2023).

[33] Nobuhito Goda, Sara J. Dozier, and Randall S. Johnson. "HIF-1 in cell cycle regulation, apoptosis, and tumor progression". In: *Antioxidants and Redox Signaling* 5.4 (July 5, 2003). Publisher: Mary Ann Liebert Inc., pp. 467–473. ISSN: 15230864. DOI: 10.1089/152308603768295212. URL: https://www.liebertpub.com/doi/abs/10.1089/152308603768295212 (visited on 05/19/2021).

[34] Wen Wen et al. "Suppression of cyclin D1 by hypoxia-inducible factor-1 via direct mechanism inhibits the proliferation and 5-fluorouracil-induced apoptosis of A549 cells". In: *Cancer Research* 70.5 (2010), pp. 2010–2019. ISSN: 00085472. DOI: 10.1158/0008-5472.CAN-08-4910.

[35] B. Bedessem and A. Stéphanou. "A mathematical model of HiF-1α-mediated response to hypoxia on the G1/S transition". In: *Mathematical Biosciences* 248.1 (2014). Publisher: Elsevier Inc., pp. 31–39. ISSN: 00255564. DOI: 10.1016/j.mbs.2013.11.007. URL: http://dx.doi.org/10.1016/j.mbs.2013.11.007.

[36] Tomás Alarcón and Henrik Jeldtoft Jensen. "Quiescence: A mechanism for escaping the effects of drug on cell populations". In: *Journal of the Royal Society Interface* 8.54 (Jan. 6, 2011). Publisher: Royal Society, pp. 99–106. ISSN: 17425662. DOI: 10.1098/rsif.2010.0130. arXiv: 1002.4579. URL: https://royalsocietypublishing.org/doi/10.1098/rsif.2010.0130 (visited on 04/29/2021).

[37] Sen Li, Jinlan Gao, Xinbin Zhuang, Chenyang Zhao, Xiaoyu Hou, Xuesha Xing, Chen Chen, Qi Liu, Shuang Liu, and Yang Luo. "Cyclin G2 inhibits the warburg effect and tumour progression by suppressing LDHA phosphorylation in glioma". In: *International Journal of Biological Sciences* 15.3 (2019). Publisher: Ivyspring International Publisher, pp. 544–555. ISSN: 14492288. DOI: 10.7150/ijbs.30297. URL: /pmc/articles/PMC6367585/ (visited on 04/22/2021).

[38] R. A. Bender. "Glycolysis". In: *Brenner's Encyclopedia of Genetics: Second Edition* 2 (2013). ISBN: 9780080961569, pp. 346–349. DOI: 10.1016/B978-0-12-374984-0.00659-8.

[39] F K Zimmermann. "Glycolysis in Saccharomyces cerevisiae". In: *Encyclopedia of Genetics* (2001), pp. 885–888. DOI: 10.1006/rwgn.2001.0570.

[40] Antonio Blanco and Gustavo Blanco. "Chapter 14 - Carbohydrate Metabolism". In: *Medical Biochemistry*. Ed. by Antonio Blanco and Gustavo Blanco. Academic Press, Jan. 1, 2017, pp. 283–323. ISBN: 978-0-12-803550-4. DOI: 10.1016/B978-0-12-803550-4.00014-8. URL: https://www.sciencedirect.com/science/article/pii/B9780128035504000148 (visited on 08/11/2023).

[41] Pierre Jacquet. “Métabolisme des tumeurs : apport de la modélisation dans la compréhension de l’adaptabilité spatiotemporelle”. In: (Nov. 2022). URL: https://theses.hal.science/tel-03937196%20https://theses.hal.science/tel-03937196/document.

[42] Dongya Jia, Mingyang Lu, Kwang Hwa Jung, Jun Hyoung Park, Linglin Yu, José N. Onuchic, Benny Abraham Kaipparettu, and Herbert Levine. “Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways”. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.9 (Feb. 26, 2019). Publisher: National Academy of Sciences, pp. 3909–3918. ISSN: 10916490. DOI: 10.1073/pnas.1816391116. URL: https://pubmed.ncbi.nlm.nih.gov/30733294/ (visited on 12/18/2020).

[43] Otto Warburg, Franz Wind, and Erwin Negelein. “The metabolism of tumors in the body”. In: *Journal of General Physiology* 8.6 (1927), pp. 519–530. ISSN: 15407748. DOI: 10.1085/jgp.8.6.519. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2140820/ (visited on 01/11/2022).

[44] Otto Warburg. “On the origin of cancer cells”. In: *Science* 123.3191 (Feb. 24, 1956). Publisher: American Association for the Advancement of Science, pp. 309–314. ISSN: 00368075. DOI: 10.1126/SCIENCE.123.3191.309/ASSET/A8D38B53-799F-4009-AAD3-E77CEF33D301/ASSETS/SCIENCE.123.3191.309.FP.PNG. URL: https://www.science.org/doi/abs/10.1126/science.123.3191.309 (visited on 05/25/2022).

[45] Pierre Jacquet and Angélique Stéphanou. “Metabolic Reprogramming, Questioning, and Implications for Cancer”. In: *Biology 2021, Vol. 10, Page 129* 10.2 (Feb. 7, 2021). Publisher: Multidisciplinary Digital Publishing Institute, p. 129. ISSN: 20797737. DOI: 10.3390/BIOLOGY10020129. URL: https://www.mdpi.com/2079-7737/10/2/129/htm (visited on 01/26/2022).

[46] Ian F. Robey, Anthony D. Lien, Sarah J. Welsh, Brenda K. Baggett, and Robert J. Gillies. “Hypoxia-inducible factor-1α and the glycolytic phenotype in tumors”. In: *Neoplasia* 7.4 (2005). Publisher: Nature Publishing Group, pp. 324–330. ISSN: 15228002. DOI: 10.1593/neo.04430. URL: /pmc/articles/PMC1501147/?report=abstract (visited on 01/05/2021).

[47] Ebbe Boedtkjer and Stine F. Pedersen. “The Acidic Tumor Microenvironment as a Driver of Cancer”. In: *https://doi.org/10.1146/annurev-physiol-021119-034627* 82 (Feb. 10, 2020). Publisher: Annual Reviews, pp. 103–126. ISSN: 15451585. DOI: 10.1146/ANNUREV-PHYSIOL-021119-034627. URL: https://www.annualreviews.org/doi/abs/10.1146/annurev-physiol-021119-034627 (visited on 01/12/2023).

[48] Xiaofeng Li, Xiaozhou Yu, Dong Dai, Xiuyu Song, Wengui Xu, Xiaofeng Li, Xiaozhou Yu, Dong Dai, Xiuyu Song, and Wengui Xu. “The altered glucose metabolism in tumor and a tumor acidic microenvironment associated with extracellular matrix metalloproteinase inducer and monocarboxylate transporters”. In: *Oncotarget* 7.17 (Mar. 17, 2016). Publisher: Impact Journals, pp. 23141–23155. ISSN: 1949-2553. DOI: 10.18632/ONCOTARGET.8153. URL: https://www.oncotarget.com/article/8153/text/ (visited on 01/18/2023).

[49] Martina Conte and Christina Surulescu. “Mathematical modeling of glioma invasion: acid- and vasculature mediated go-or-grow dichotomy and the influence of tissue anisotropy”. In: *Applied Mathematics and Computation* 407 (Oct. 15, 2021). Publisher: Elsevier, p. 126305. ISSN: 0096-3003. DOI: 10.1016/J.AMC.2021.126305. arXiv: 2007.12204. (Visited on 01/05/2023).

[50] Veronica Estrella et al. “Acidity generated by the tumor microenvironment drives local invasion”. In: *Cancer Research* 73.5 (Mar. 1, 2013). Publisher: American Association for Cancer Research, pp. 1524–1535. ISSN: 00085472. DOI: 10.1158/0008-5472.CAN-12-2796/650773/AM/ACIDITY-GENERATED-BY-THE-TUMOR-MICROENVIRONMENT. URL: https://aacrjournals.org/cancerres/article/73/5/1524/586608/Acidity-Generated-by-the-Tumor-Microenvironment (visited on 01/18/2023).

[51] Charles C. Wykoff et al. “Hypoxia-inducible expression of tumor-associated carbonic anhydrases”. In: *Cancer Research* 60.24 (2000), pp. 7075–7083. ISSN: 00085472.

[52] Karin Pfisterer, Lisa E. Shaw, Dörte Symmank, and Wolfgang Weninger. “The Extracellular Matrix in Skin Inflammation and Infection”. In: *Frontiers in Cell and Developmental Biology* 9 (2021). ISSN: 2296-634X. URL: https://www.frontiersin.org/articles/10.3389/fcell.2021.682414 (visited on 08/25/2023).

[53] Daniele M. Gilkes, Gregg L. Semenza, and Denis Wirtz. “Hypoxia and the extracellular matrix: drivers of tumour metastasis”. In: *Nature Reviews Cancer 2014 14:6* 14.6 (May 15, 2014). Publisher: Nature Publishing Group, pp. 430–439. ISSN: 1474-1768. DOI: 10.1038/nrc3726. URL: https://www.nature.com/articles/nrc3726 (visited on 01/05/2023).

[54] Inês G. Gonçalves and Jose Manuel Garcia-Aznar. “Extracellular matrix density regulates the formation of tumour spheroids through cell migration”. In: *PLOS Computational Biology* 17.2 (Feb. 26, 2021), e1008764. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008764.

[55] Nadine R. Lang, Kai Skodzek, Sebastian Hurst, Astrid Mainka, Julian Steinwachs, Julia Schneider, Katerina E. Aifantis, and Ben Fabry. “Biphasic response of cell invasion to matrix stiffness in three-dimensional biopolymer networks”. In: *Acta Biomaterialia* 13 (Feb. 1, 2015). Publisher: Elsevier, pp. 61–67. ISSN: 1742-7061. DOI: 10.1016/J.ACTBIO.2014.11.003. (Visited on 02/28/2023).

[56] Daniele M. Gilkes, Saumendra Bajpai, Pallavi Chaturvedi, Denis Wirtz, and Gregg L. Semenza. “Hypoxia-inducible f0actor 1 (HIF-1) promotes extracellular matrix remodeling under hypoxic conditions by inducing P4HA1, P4HA2, and PLOD2 expression in fibroblasts”. In: *Journal of Biological Chemistry* 288.15 (Apr. 12, 2013). Publisher: Elsevier, pp. 10819–10829. ISSN: 00219258. DOI: 10.1074/jbc.M112.442939. URL: http://www.jbc.org/article/S0021925820673045/fulltext (visited on 01/06/2023).

[57] Stéphane Germain, Catherine Monnot, Laurent Muller, and Anne Eichmann. “Hypoxia-driven angiogenesis: Role of tip cells and extracellular matrix scaffolding”. In: *Current Opinion in Hematology* 17.3 (May 2010), pp. 245–251. ISSN: 10656251. DOI: 10.1097/MOH.0B013E32833865B9. URL: https://journals.lww.com/co-hematology/Fulltext/2010/05000/Hypoxia_driven_angiogenesis__role_of_tip_cells_and.15.aspx (visited on 01/06/2023).

[58] Tadanori Mammoto, Amanda Jiang, Elisabeth Jiang, Dipak Panigrahy, Mark W. Kieran, and Akiko Mammoto. “Role of collagen matrix in tumor angiogenesis and glioblastoma multiforme progression”. In: *American Journal of Pathology* 183.4 (2013). Publisher: American Society for Investigative Pathology, pp. 1293–1305. ISSN: 00029440. DOI: 10.1016/j.ajpath.2013.06.026. URL: http://dx.doi.org/10.1016/j.ajpath.2013.06.026.

[59] Gaofeng Xiong, Rachel L. Stewart, Jie Chen, Tianyan Gao, Timothy L. Scott, Luis M. Samayoa, Kathleen O’Connor, Andrew N. Lane, and Ren Xu. “Collagen prolyl 4-hydroxylase 1 is essential for HIF-1α stabilization and TNBC chemoresistance”. In: *Nature Communications 2018 9:1* 9.1 (Oct. 26, 2018). Publisher: Nature Publishing Group, pp. 1–16. ISSN: 2041-1723. DOI: 10.1038/s41467-018-06893-9. URL: https://www.nature.com/articles/s41467-018-06893-9 (visited on 01/09/2023).

[60] Achilleas D. Theocharis, Spyros S. Skandalis, Chrysostomi Gialeli, and Nikos K. Karamanos. “Extracellular matrix structure”. In: *Advanced Drug Delivery Reviews* 97 (Feb. 2016), pp. 4–27. ISSN: 0169-409X. DOI: 10.1016/J.ADDR.2015.11.001.

[61] Laura J. Kaufman, C. P. Brangwynne, K. E. Kasza, E. Filippidi, V. D. Gordon, T. S. Deisboeck, and D. A. Weitz. “Glioma expansion in collagen I matrices: Analyzing collagen concentration-dependent growth and motility patterns”. In: *Biophysical Journal* 89.1 (2005). Publisher: Elsevier, pp. 635–650. ISSN: 00063495. DOI: 10.1529/biophysj.105.061994. URL: http://dx.doi.org/10.1529/biophysj.105.061994.

[62] Pranita Kaphle, Yongchao Li, and Li Yao. “The mechanical and pharmacological regulation of glioblastoma cell migration in 3D matrices”. In: *Journal of Cellular Physiology* 234.4 (2019), pp. 3948–3960. ISSN: 10974652. DOI: 10.1002/jcp.27209.

[63] Arely León-López, Alejandro Morales-Peñaloza, Víctor Manuel Martínez-Juárez, Apolonio Vargas-Torres, Dimitrios I. Zeugolis, and Gabriel Aguirre-Álvarez. “Hydrolyzed Collagen—Sources and Applications”. In: *Molecules* 24.22 (Jan. 2019). Number: 22 Publisher: Multidisciplinary Digital Publishing Institute, p. 4031. ISSN: 1420-3049. DOI: 10.3390/molecules24224031. URL: https://www.mdpi.com/1420-3049/24/22/4031 (visited on 08/15/2023).

[64] Hsiang Hsi Hong, Nicole Pischon, Ronaldo B. Santana, Amitha H. Palamakumbura, Hermik Babakhanlou Chase, Donald Gantz, Ying Guo, Mehmet Ilhan Uzel, Daniel Ma, and Philip C. Trackman. “A role for lysyl oxidase regulation in the control of normal collagen deposition in differentiating osteoblast cultures”. In: *Journal of Cellular Physiology* 200.1 (July 1, 2004). Publisher: John Wiley & Sons, Ltd, pp. 53–62. ISSN: 1097-4652. DOI: 10.1002/JCP.10476. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/jcp.10476 (visited on 03/27/2023).

[65] Sang Taek Jung, Moon Suk Kim, Ji Yeon Seo, Hyung Chul Kim, and Youngho Kim. “Purification of enzymatically active human lysyl oxidase and lysyl oxidase-like protein from Escherichia coli inclusion bodies”. In: *Protein Expression and Purification* 31 (2 Oct. 2003), pp. 240–246. ISSN: 1046-5928. DOI: 10.1016/S1046-5928(03)00217-1.

[66] Carmen Chak Lui Wong et al. “Hypoxia-inducible factor 1 is a master regulator of breast cancer metastatic niche formation”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.39 (Sept. 27, 2011). Publisher: National Academy of Sciences, pp. 16369–16374. ISSN: 00278424. DOI: 10.1073/PNAS.1113483108/SUPPL_FILE/PNAS.201113483SI.PDF. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1113483108 (visited on 01/09/2023).

[67] Niall E. Deakin and Mark A.J. Chaplain. “Mathematical modeling of cancer invasion: The role of membrane-bound matrix metalloproteinases”. In: *Frontiers in Oncology* 3 APR (Apr. 3, 2013). Publisher: Frontiers, p. 70. ISSN: 2234943X. DOI: 10.3389/FONC.2013.00070/ABSTRACT. (Visited on 01/20/2023).

[68] Wenjun Guo and Filippo G. Giancotti. “Integrin signalling during tumour progression”. In: *Nature Reviews Molecular Cell Biology 2004 5:10* 5.10 (Oct. 2004). Publisher: Nature Publishing Group, pp. 816–826. ISSN: 1471-0080. DOI: 10.1038/nrm1490. URL: https://www.nature.com/articles/nrm1490 (visited on 08/30/2022).

[69] Sabrina Amar, Lyndsay Smith, and Gregg B. Fields. “Matrix metalloproteinase collagenolysis in health and disease”. In: *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1864.11 (Nov. 1, 2017). Publisher: Elsevier, pp. 1940–1951. ISSN: 0167-4889. DOI: 10.1016/J.BBAMCR.2017.04.015. (Visited on 06/06/2023).

[70] Giuseppe Musumeci et al. “Characterization of matrix metalloproteinase-2 and -9, ADAM-10 and N-cadherin expression in human glioblastoma multiforme”. In: *Cell and Tissue Research* 362.1 (Oct. 22, 2015). Publisher: Springer Verlag, pp. 45–60. ISSN: 14320878. DOI: 10.1007/S00441-015-2197-5/TABLES/2. URL: https://link.springer.com/article/10.1007/s00441-015-2197-5 (visited on 01/17/2023).

[71] Eric M. Tam, Todd R. Moore, Georgina S. Butler, and Christopher M. Overall. “Characterization of the Distinct Collagen Binding, Helicase and Cleavage Mechanisms of Matrix Metalloproteinase 2 and 14 (Gelatinase A and MT1-MMP)”. In: *Journal of Biological Chemistry* 279.41 (Oct. 8, 2004). Publisher: Elsevier BV, pp. 43336–43344. ISSN: 00219258. DOI: 10.1074/jbc.m407186200. URL: http://www.jbc.org/article/S0021925820770894/fulltext (visited on 06/02/2023).

[72] Yuntao Zhang, Xiuli Mao, Tyler Schwend, Stacy Littlechild, and Gary W. Conrad. “Resistance of Corneal RFUVA–Cross-Linked Collagens and Small Leucine-Rich Proteoglycans to Degradation by Matrix Metalloproteinases”. In: *Investigative Opthalmology & Visual Science* 54.2 (Feb. 5, 2013), p. 1014. ISSN: 1552-5783. DOI: 10.1167/iovs.12-11277. URL: http://iovs.arvojournals.org/article.aspx?doi=10.1167/iovs.12-11277 (visited on 06/06/2023).

[73] Yoshifumi Itoh and Motoharu Seiki. “MT1-MMP: A potent modifier of pericellular microenvironment”. In: *Journal of Cellular Physiology* 206.1 (Jan. 1, 2006). Publisher: John Wiley & Sons, Ltd, pp. 1–8. ISSN: 1097-4652. DOI: 10.1002/JCP.20431. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/jcp.20431 (visited on 06/09/2023).

[74] J. Varani, Y. Hattori, Y. Chi, T. Schmidt, P. Perone, M. E. Zeigler, D. J. Fader, and T. M. Johnson. “Collagenolytic and gelatinolytic matrix metalloproteinases and their inhibitors in basal cell carcinoma of skin: comparison with normal skin”. In: *British Journal of Cancer 2000 82:3* 82.3 (Jan. 7, 2000). Publisher: Nature Publishing Group, pp. 657–665. ISSN: 1532-1827. DOI: 10.1054/bjoc.1999.0978. URL: https://www.nature.com/articles/6690978 (visited on 06/05/2023).

[75] Laure Laforgue, Arnold Fertin, Yves Usson, Claude Verdier, and Valérie M. Laurent. “Efficient deformation mechanisms enable invasive cancer cells to migrate faster in 3D collagen networks”. In: *Scientific Reports 2022 12:1* 12.1 (May 12, 2022). Publisher: Nature Publishing Group ISBN: 0123456789, pp. 1–15. ISSN: 2045-2322. DOI: 10.1038/s41598-022-11581-2. URL: https://www.nature.com/articles/s41598-022-11581-2 (visited on 02/24/2023).

[76] Shreyas S. Rao, Jessica Dejesus, Aaron R. Short, Jose J. Otero, Atom Sarkar, and Jessica O. Winter. “Glioblastoma behaviors in three-dimensional collagen-hyaluronan composite hydrogels”. In: *ACS Applied Materials and Interfaces* 5.19 (2013), pp. 9276–9284. ISSN: 19448252. DOI: 10.1021/am402097j.

[77] Daniela K. Schlüter, Ignacio Ramis-Conde, and Mark A.J. Chaplain. “Computational modeling of single-cell migration: The leading role of extracellular matrix fibers”. In: *Biophysical Journal* 103.6 (Sept. 19, 2012). Publisher: Elsevier, pp. 1141–1151. ISSN: 00063495. DOI: 10.1016/j.bpj.2012.07.048. URL: http://www.cell.com/article/S0006349512008697/fulltext (visited on 01/20/2023).

[78] Jeffrey D. Orth, Ines Thiele, and Bernhard Ø Palsson. “What is flux balance analysis?” In: *Nature Biotechnology* 28.3 (Mar. 2010). Number: 3 Publisher: Nature Publishing Group, pp. 245–248. ISSN: 1546-1696. DOI: 10.1038/nbt.1614. URL: https://www.nature.com/articles/nbt.1614 (visited on 09/26/2023).

[79] Eberhard O. Voit. “The best models of metabolism”. In: *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* 9.6 (Nov. 1, 2017). Publisher: Wiley-Blackwell, e1391. ISSN: 19395094. DOI: 10.1002/wsbm.1391. URL: http://doi.wiley.com/10.1002/wsbm.1391 (visited on 02/18/2021).

[80] E. T. Denisov, O. M. Sarkisov, and G. I. Likhtenshtein. “Chapter 1 - General ideas of chemical kinetics”. In: *Chemical Kinetics*. Ed. by E. T. Denisov, O. M. Sarkisov, and G. I. Likhtenshtein. Amsterdam: Elsevier Science, Jan. 1, 2003, pp. 1–15. ISBN: 978-0-444-50938-3. DOI: 10.1016/B978-044450938-3/50022-3. URL: https://www.sciencedirect.com/science/article/pii/B9780444509383500223 (visited on 09/26/2023).

[81] Donald Voet and Judith G. Voet. *Biochemistry*. Google-Books-ID: 8mV19Jz_v4AC. John Wiley & Sons, Dec. 1, 2010. 1526 pp. ISBN: 978-0-470-57095-1.

[82] Pramod R. Somvanshi and Kareenhalli V. Venkatesh. “Hill Equation”. In: *Encyclopedia of Systems Biology*. Ed. by Werner Dubitzky, Olaf Wolkenhauer, Kwang-Hyun Cho, and Hiroki Yokota. New York, NY: Springer, 2013, pp. 892–895. ISBN: 978-1-4419-9863-7. DOI: 10.1007/978-1-4419-9863-7_946. URL: https://doi.org/10.1007/978-1-4419-9863-7_946 (visited on 09/26/2023).

[83] Laurence Calzone, Vincent Noël, Emmanuel Barillot, Guido Kroemer, and Gautier Stoll. “Modeling signaling pathways in biology with MaBoSS: From one single cell to a dynamic population of heterogeneous interacting cells”. In: *Computational and Structural Biotechnology Journal* 20 (Jan. 1, 2022), pp. 5661–5671. ISSN: 2001-0370. DOI: 10.1016/j.csbj.2022.10.003. URL: https://www.sciencedirect.com/science/article/pii/S2001037022004512 (visited on 09/27/2023).

[84] Ahmed Abdelmonem Hemedan, Anna Niarakis, Reinhard Schneider, and Marek Ostaszewski. “Boolean modelling as a logic-based dynamic approach in systems medicine”. In: *Computational and Structural Biotechnology Journal* 20 (Jan. 1, 2022), pp. 3161–3172. ISSN: 2001-0370. DOI: 10.1016/j.csbj.2022.06.035. URL: https://www.sciencedirect.com/science/article/pii/S2001037022002495 (visited on 09/27/2023).

[85] Michael Raatz, Saumil Shah, Guranda Chitadze, Monika Brüggemann, and Arne Traulsen. “The impact of phenotypic heterogeneity of tumour cells on treatment and relapse dynamics”. In: *PLoS Computational Biology* 17.2 (Feb. 1, 2021). Publisher: Public Library of Science, e1008702. ISSN: 15537358. DOI: 10.1371/JOURNAL.PCBI.1008702. URL: https://doi.org/10.1371/journal.pcbi.1008702 (visited on 06/16/2021).

[86] Anne-Cécile Lesart. “Modélisation théorique du développement tumoral sous fenêtre dorsale : Vers un outil clinique d’individualisation et d’optimisation de la thérapie”. In: (Nov. 2013). URL: https://theses.hal.science/tel-01061838%20https://theses.hal.science/tel-01061838/document.

[87] Tsuyoshi Hirashima, Elisabeth G. Rens, and Roeland M. H. Merks. “Cellular Potts modeling of complex multicellular behaviors in tissue morphogenesis”. In: *Development, Growth & Differentiation* 59.5 (2017). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/dgd.12358, pp. 329–339. ISSN: 1440-169X. DOI: 10.1111/dgd.12358. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/dgd.12358 (visited on 09/27/2023).

[88] James M. Osborne, Alexander G. Fletcher, Joe M. Pitt-Francis, Philip K. Maini, and David J. Gavaghan. “Comparing individual-based approaches to modelling the self-organization of multicellular tissues”. In: *PLOS Computational Biology* 13.2 (Feb. 13, 2017). Publisher: Public Library of Science, e1005387. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005387. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005387 (visited on 09/27/2023).

[89] Farnoosh Abbas-Aghababazadeh, Qian Li, and Brooke L. Fridley. “Comparison of normalization approaches for gene expression studies completed with highthroughput sequencing”. In: *PLoS ONE* 13.10 (Oct. 1, 2018). Publisher: Public Library of Science. ISSN: 19326203. DOI: 10.1371/journal.pone.0206312. URL: /pmc/articles/PMC6209231/?report=abstract (visited on 11/19/2020).

[90] Angélique Stéphanou. “Une Approche Computationnelle pour l’Etude de Processus Morphogénétiques - de la motilité des cellules à la croissance des vaisseaux”. thesis. Université Joseph-Fourier - Grenoble I, Dec. 15, 2011. URL: https://theses.hal.science/tel-00749373 (visited on 09/27/2023).

[91] Aalpen A. Patel, Edward T. Gawlinski, Susan K. Lemieux, and Robert A. Gatenby. “A cellular automaton model of early tumor growth and invasion: The effects of native tissue vascularity and increased anaerobic tumor metabolism”. In: *Journal of Theoretical Biology* 213.3 (Dec. 7, 2001). Publisher: Academic Press, pp. 315–331. ISSN: 00225193. DOI: 10.1006/jtbi.2001.2385. (Visited on 06/07/2021).

[92] Brenda M. Rubenstein and Laura J. Kaufman. “The Role of Extracellular Matrix in Glioma Invasion: A Cellular Potts Model Approach”. In: *Biophysical Journal* 95.12 (Dec. 15, 2008). Publisher: Cell Press, pp. 5661–5680. ISSN: 0006-3495. DOI: 10.1529/BIOPHYSJ.108.140624. (Visited on 03/27/2023).

[93] Erika Tsingos, Bente Hilde Bakker, Koen A. E. Keijzer, Hermen Jan Hupkes, and Roeland M. H. Merks. *Modelling the mechanical cross-talk between cells and fibrous extracellular matrix using hybrid cellular Potts and molecular dynamics methods*. Pages: 2022.06.10.495667 Section: New Results. July 7, 2022. DOI: 10.1101/2022.06.10.495667. URL: https://www.biorxiv.org/content/10.1101/2022.06.10.495667v3 (visited on 09/27/2023).

[94] Ahmadreza Ghaffarizadeh, Randy Heiland, Samuel H. Friedman, Shannon M. Mumenthaler, and Paul Macklin. “PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems”. In: *PLOS Computational Biology* 14.2 (Feb. 23, 2018). Publisher: Public Library of Science, e1005991. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1005991. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005991 (visited on 09/27/2023).

[95] Mark Robertson-Tessi, Robert J. Gillies, Robert A. Gatenby, and Alexander R.A. Anderson. “Impact of Metabolic Heterogeneity on Tumor Growth, Invasion, and Treatment Outcomes”. In: *Cancer Research* 75.8 (Apr. 15, 2015). Publisher: American Association for Cancer Research Inc., pp. 1567–1579. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-14-1428. URL: http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-14-1428 (visited on 05/19/2021).

[96] Kieran Smallbone, Robert A. Gatenby, and Philip K. Maini. “Mathematical modelling of tumour acidity”. In: *Journal of Theoretical Biology* 255.1 (2008), pp. 106–112. ISSN: 00225193. DOI: 10.1016/j.jtbi.2008.08.002.

[97] A. Stéphanou, A. C. Lesart, J. Deverchère, A. Juhem, A. Popov, and F. Estève. “How tumour-induced vascular changes alter angiogenesis: Insights from a computational model”. In: *Journal of Theoretical Biology* 419 (January 2017). Publisher: Elsevier, pp. 211–226. ISSN: 10958541. DOI: 10.1016/j.jtbi.2017.02.018. URL: http://dx.doi.org/10.1016/j.jtbi.2017.02.018.

[98] Edoardo Milotti and Roberto Chignola. “Emergent Properties of Tumor Microenvironment in a Real-Life Model of Multicell Tumor Spheroids”. In: *PLOS ONE* 5.11 (2010). Publisher: Public Library of Science, e13942. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0013942. arXiv: 1010.1965. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013942 (visited on 01/04/2023).

[99] Miguel AS Cavadas, Lan K. Nguyen, and Alex Cheong. “Hypoxia-inducible factor (HIF) network: insights from mathematical models”. In: *Cell Communication and Signaling* 11.1 (June 10, 2013), p. 42. ISSN: 1478-811X. DOI: 10.1186/1478-811X-11-42. URL: https://doi.org/10.1186/1478-811X-11-42 (visited on 09/28/2023).

[100] Baptiste Bedessem and Angélique Stéphanou. “Role of compartmentalization on HiF-1α degradation dynamics during changing oxygen conditions: A computational approach”. In: *PLoS ONE* 9.10 (Oct. 22, 2014). Publisher: Public Library of Science, p. 110495. ISSN: 19326203. DOI: 10.1371/journal.pone.0110495. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0110495 (visited on 01/13/2021).

[101] Alaa Tafech, Pierre Jacquet, Céline Beaujean, Arnold Fertin, Yves Usson, and Angélique Stéphanou. “Characterization of the Intracellular Acidity Regulation of Brain Tumor Cells and Consequences for Therapeutic Optimization of Temozolomide”. In: *Biology* 12.9 (Sept. 2023). Number: 9 Publisher: Multidisciplinary Digital Publishing Institute, p. 1221. ISSN: 2079-7737. DOI: 10.3390/biology12091221. URL: https://www.mdpi.com/2079-7737/12/9/1221 (visited on 10/19/2023).

[102] Gary R. Mirams et al. “Chaste: An Open Source C++ Library for Computational Physiology and Biology”. In: *PLOS Computational Biology* 9.3 (Mar. 14, 2013). Publisher: Public Library of Science, e1002970. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1002970. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002970 (visited on 09/28/2023).

[103] Seunghwa Kang, Simon Kahan, Jason McDermott, Nicholas Flann, and Ilya Shmulevich. “Biocellion : accelerating computer simulation of multicellular biological system models”. In: *Bioinformatics* 30.21 (Nov. 1, 2014), pp. 3101–3108. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btu498. URL: https://doi.org/10.1093/bioinformatics/btu498 (visited on 09/28/2023).

[104] Maciej H. Swat, Gilberto L. Thomas, Julio M. Belmonte, Abbas Shirinifard, Dimitrij Hmeljak, and James A. Glazier. “Chapter 13 - Multi-Scale Modeling of Tissues Using CompuCell3D”. In: *Methods in Cell Biology*. Ed. by Anand R. Asthagiri and Adam P. Arkin. Vol. 110. Computational Methods in Cell Biology. Academic Press, Jan. 1, 2012, pp. 325–366. DOI: 10.1016/B978-0-12-388403-9.00013-8. URL: https://www.sciencedirect.com/science/article/pii/B9780123884039000138 (visited on 09/28/2023).

[105] Maciej Cytowski and Zuzanna Szymanska. “Large-Scale Parallel Simulations of 3D Cell Colony Dynamics”. In: *Computing in Science & Engineering* 16.5 (Sept. 2014). Conference Name: Computing in Science & Engineering, pp. 86–95. ISSN: 1558-366X. DOI: 10.1109/MCSE.2014.2. URL: https://ieeexplore.ieee.org/document/6728930 (visited on 09/28/2023).

[106] Stefan Hoehme and Dirk Drasdo. “A cell-based simulation software for multi-cellular systems”. In: *Bioinformatics* 26.20 (Oct. 15, 2010), pp. 2641–2642. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btq437. URL: https://doi.org/10.1093/bioinformatics/btq437 (visited on 09/28/2023).

[107] Ahmadreza Ghaffarizadeh, Randy Heiland, Samuel H. Friedman, Shannon M. Mumenthaler, and Paul Macklin. “PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems”. In: *PLOS Computational Biology* 14.2 (Feb. 1, 2018). Publisher: Public Library of Science, e1005991. ISSN: 1553-7358. DOI: 10.1371/JOURNAL.PCBI.1005991. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005991 (visited on 09/16/2021).

[108] Ahmadreza Ghaffarizadeh, Samuel H. Friedman, and Paul Macklin. “BioFVM: an efficient, parallelized diffusive transport solver for 3-D biological simulations”. In: *Bioinformatics* 32.8 (Apr. 15, 2016), pp. 1256–1258. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btv730. URL: https://doi.org/10.1093/bioinformatics/btv730 (visited on 09/08/2023).

[109] Wenbo Li and Jin Wang. “Uncovering the Underlying Mechanisms of Cancer Metabolism through the Landscapes and Probability Flux Quantifications”. In: *iScience* 23.4 (Apr. 24, 2020). Publisher: Elsevier Inc., p. 101002. ISSN: 25890042. DOI: 10.1016/j.isci.2020.101002. URL: https://doi.org/10.1016/j.isci.2020.101002 (visited on 02/12/2021).

[110] Mingyang Lu, Mohit Kumar Jolly, Ryan Gomoto, Bin Huang, José Onuchic, and Eshel Ben-Jacob. “Tristability in cancer-associated microRNA-TF chimera toggle switch”. In: *Journal of Physical Chemistry B* 117.42 (Oct. 24, 2013). Publisher: American Chemical Society, pp. 13164–13174. ISSN: 15205207. DOI: 10.1021/jp403156m. URL: https://pubs.acs.org/doi/abs/10.1021/jp403156m (visited on 03/04/2021).

[111] Geoffrey M Cooper. “The Eukaryotic Cell Cycle”. In: (2000). Publisher: Sinauer Associates. URL: https://www.ncbi.nlm.nih.gov/books/NBK9876/ (visited on 01/24/2022).

[112] EMBL-EBI. *Microarrays | Functional genomics II*. URL: https://www.ebi.ac.uk/training/online/courses/functional-genomics-ii-common-technologies-and-data-analysis-methods/microarrays/ (visited on 09/18/2023).

[113] EMBL-EBI. *RNA sequencing | Functional genomics II*. URL: https://www.ebi.ac.uk/training/online/courses/functional-genomics-ii-common-technologies-and-data-analysis-methods/rna-sequencing/ (visited on 09/18/2023).

[114] Brian J. Haas and Michael C. Zody. "Advancing RNA-Seq analysis". In: *Nature Biotechnology* 28.5 (May 2010). Number: 5 Publisher: Nature Publishing Group, pp. 421–423. ISSN: 1546-1696. DOI: 10.1038/nbt0510-421. URL: https://www.nature.com/articles/nbt0510-421 (visited on 09/18/2023).

[115] Ana Conesa et al. "A survey of best practices for RNA-seq data analysis". In: *Genome Biology* 17.1 (Jan. 26, 2016). Publisher: BioMed Central Ltd. ISSN: 1474760X. DOI: 10.1186/s13059-016-0881-8. URL: /pmc/articles/PMC4728800/?report=abstract (visited on 11/04/2020).

[116] Yingdong Zhao, Ming-Chung Li, Mariam M. Konaté, Li Chen, Biswajit Das, Chris Karlovich, P. Mickey Williams, Yvonne A. Evrard, James H. Doroshow, and Lisa M. McShane. "TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository". In: *Journal of Translational Medicine* 19.1 (Dec. 2021). Number: 1 Publisher: BioMed Central, pp. 1–15. ISSN: 1479-5876. DOI: 10.1186/s12967-021-02936-w. URL: https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-021-02936-w (visited on 09/18/2023).

[117] Anders Lundin et al. "Human iPS-Derived Astroglia from a Stable Neural Precursor State Show Improved Functionality Compared with Conventional Astrocytic Models". In: *Stem Cell Reports* 10.3 (Mar. 13, 2018). Publisher: Cell Press, pp. 1030–1045. ISSN: 2213-6711. DOI: 10.1016/J.STEMCR.2018.01.021. (Visited on 07/18/2022).

[118] Anna Falk et al. "Capture of Neuroepithelial-Like Stem Cells from Pluripotent Stem Cells Provides a Versatile System for In Vitro Production of Human Neurons". In: *PLOS ONE* 7.1 (Jan. 17, 2012). Publisher: Public Library of Science, e29597. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0029597. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029597 (visited on 11/14/2022).

[119] Mark D. Robinson, Davis J. McCarthy, and Gordon K. Smyth. "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data". In: *Bioinformatics* 26.1 (Jan. 1, 2010). Publisher: Oxford Academic, pp. 139–140. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTP616. URL: https://academic.oup.com/bioinformatics/article/26/1/139/182458 (visited on 07/19/2022).

[120] Matthew E. Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. "limma powers differential expression analyses for RNA-sequencing and microarray studies". In: *Nucleic Acids Research* 43.7 (Apr. 20, 2015). Publisher: Oxford Academic, e47–e47. ISSN: 0305-1048. DOI: 10.1093/NAR/GKV007. URL: https://academic.oup.com/nar/article/43/7/e47/2414268 (visited on 09/17/2021).

[121] Michael I. Love, Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* 15.12 (Dec. 5, 2014). Publisher: BioMed Central Ltd., pp. 1–21. ISSN: 1474760X. DOI: 10.1186/s13059-014-0550-8. URL: https://link.springer.com/article/10.1186/s13059-014-0550-8 (visited on 11/04/2020).

[122] Zhengtao Xiao, Ziwei Dai, and Jason W. Locasale. "Metabolic landscape of the tumor microenvironment at single cell resolution". In: *Nature Communications* 10.1 (Dec. 1, 2019). Publisher: Nature Publishing Group. ISSN: 20411723. DOI: 10.1038/s41467-019-11738-0.

[123] Magali Richard, Clémentine Decamps, Florent Chuffart, Elisabeth Brambilla, Sophie Rousseaux, Saadi Khochbin, and Daniel Jost. "PenDA, a rank-based method for personalized differential analysis: Application to lung cancer". In: *PLOS Computational Biology* 16.5 (May 11, 2020). Ed. by Amin Emad. Publisher: Public Library of Science, e1007869. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1007869. URL: https://dx.plos.org/10.1371/journal.pcbi.1007869.

[124] Jüri Reimand et al. "Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap HHS Public Access". In: *Nat Protoc* 14.2 (2019), pp. 482–517. DOI: 10.1038/s41596-018-0103-9. URL: https://www.nature.com/articles/s41596-018-0103-9 (visited on 11/02/2020).

[125] Purvesh Khatri, Marina Sirota, and Atul J. Butte. “Ten years of pathway analysis: Current approaches and outstanding challenges”. In: *PLoS Computational Biology* 8.2 (Feb. 2012). Publisher: Public Library of Science. ISSN: 1553734X. DOI: 10.1371/journal.pcbi.1002375. URL: /pmc/articles/PMC3285573/?report=abstract (visited on 10/30/2020).

[126] Uku Raudvere, Liis Kolberg, Ivan Kuzmin, Tambet Arak, Priit Adler, Hedi Peterson, and Jaak Vilo. “g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)”. In: *Nucleic Acids Research* 47 (W1 July 2, 2019). Publisher: Oxford Academic, W191–W198. ISSN: 0305-1048. DOI: 10.1093/NAR/GKZ369. URL: https://academic.oup.com/nar/article/47/W1/W191/5486750 (visited on 09/20/2021).

[127] Aravind Subramanian et al. “Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles”. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.43 (Oct. 25, 2005). Publisher: National Academy of Sciences, pp. 15545–15550. ISSN: 00278424. DOI: 10.1073/pnas.0506580102. URL: /pmc/articles/PMC1239896/?report=abstract (visited on 10/28/2020).

[128] Minoru Kanehisa, Yoko Sato, Miho Furumichi, Kanae Morishima, and Mao Tanabe. “New approach for understanding genome variations in KEGG”. In: *Nucleic Acids Research* 47 (D1 Jan. 8, 2019). Publisher: Oxford Academic, pp. D590–D595. ISSN: 0305-1048. DOI: 10.1093/NAR/GKY962. URL: https://academic.oup.com/nar/article/47/D1/D590/5128935 (visited on 07/21/2022).

[129] Marc Gillespie et al. “The reactome pathway knowledgebase 2022”. In: *Nucleic Acids Research* 50 (D1 Jan. 7, 2022). Publisher: Oxford Academic, pp. D687–D692. ISSN: 0305-1048. DOI: 10.1093/NAR/GKAB1028. URL: https://academic.oup.com/nar/article/50/D1/D687/6426058 (visited on 07/21/2022).

[130] Sarah Mubeen, Charles Tapley Hoyt, André Gemünd, Martin Hofmann-Apitius, Holger Fröhlich, and Daniel Domingo-Fernández. “The Impact of Pathway Database Choice on Statistical Enrichment Analysis and Predictive Modeling”. In: *Frontiers in Genetics* 10 (Nov. 22, 2019). Publisher: Frontiers Media S.A., p. 1203. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.01203. URL: https://www.frontiersin.org/article/10.3389/fgene.2019.01203/full (visited on 10/14/2022).

[131] Joanna Zyla, Michal Marczyk, January Weiner, and Joanna Polanska. “Ranking metrics in gene set enrichment analysis: do they matter?” In: *BMC Bioinformatics 2017 18:1* 18.1 (May 12, 2017). Publisher: BioMed Central, pp. 1–12. ISSN: 1471-2105. DOI: 10.1186/S12859-017-1674-0. URL: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1674-0 (visited on 09/13/2021).

[132] Clémentine Decamps. “Développements méthodologiques pour la caractérisation de l’hétérogénéité tumorale”. In: (Nov. 2021). URL: https://theses.hal.science/tel-03601942%20https://theses.hal.science/tel-03601942/document.

[133] B. Wei, L. Wang, X. L. Zhao, Y. Jin, D. L. Kong, G. Z. Hu, and Z. G. Sun. “Co-mutated pathways analysis highlights the coordination mechanism in glioblastoma multiforme.” In: *Neoplasma* 61.4 (Jan. 1, 2014). Publisher: SAP - Slovak Academic Press, spol. s.r.o., pp. 424–432. ISSN: 0028-2685. DOI: 10.4149/NEO_2014_052. URL: https://europepmc.org/article/med/25027741 (visited on 02/09/2023).

[134] Zhijun Liu et al. “Induction of store-operated calcium entry (SOCE) suppresses glioblastoma growth by inhibiting the Hippo pathway transcriptional coactivators YAP/TAZ”. In: *Oncogene 2018 38:1* 38.1 (Aug. 6, 2018). Publisher: Nature Publishing Group, pp. 120–139. ISSN: 1476-5594. DOI: 10.1038/s41388-018-0425-7. URL: https://www.nature.com/articles/s41388-018-0425-7 (visited on 08/25/2022).

[135] Upal Basu-Roy, N. Sumru Bayin, Kirk Rattanakorn, Eugenia Han, Dimitris G. Placantonakis, Alka Mansukhani, and Claudio Basilico. “Sox2 antagonizes the Hippo pathway to maintain stemness in cancer cells”. In: *Nature Communications 2015 6:1* 6.1 (Apr. 2, 2015). Publisher: Nature Publishing Group, pp. 1–14. ISSN: 2041-1723. DOI: 10.1038/ncomms7411. URL: https://www.nature.com/articles/ncomms7411 (visited on 08/25/2022).

[136] Genaro R. Villa et al. “An LXR-Cholesterol Axis Creates a Metabolic Co-Dependency for Brain Cancers”. In: *Cancer Cell* 30.5 (2016). Publisher: Elsevier Inc., pp. 683–693. ISSN: 18783686. DOI: 10.1016/j.ccell.2016.09.008. URL: http://dx.doi.org/10.1016/j.ccell.2016.09.008.

[137] Leila Pirmoradi, Nayer Seyfizadeh, Saeid Ghavami, Amir A. Zeki, and Shahla Shojaei. “Targeting cholesterol metabolism in glioblastoma: A new therapeutic approach in cancer therapy”. In: *Journal of Investigative Medicine* 67.4 (2019), pp. 715–719. ISSN: 17088267. DOI: 10.1136/jim-2018-000962.

[138] Gustavo Facchini, Raffaela Silvestre Ignarro, Erika Rodrigues-Silva, André Schwambach Vieira, Iscia Lopes-Cendes, Roger Frigério Castilho, and Fabio Rogerio. “Toxic effects of phytol and retinol on human glioblastoma cells are associated with modulation of cholesterol and fatty acid biosynthetic pathways”. In: *Journal of Neuro-Oncology* 136.3 (2018). Publisher: Springer US ISBN: 0123456789, pp. 435–443. ISSN: 15737373. DOI: 10.1007/s11060-017-2672-9. URL: http://dx.doi.org/10.1007/s11060-017-2672-9.

[139] Rui Liu et al. “Itraconazole suppresses the growth of glioblastoma through induction of autophagy: Involvement of abnormal cholesterol trafficking”. In: *Autophagy* 10.7 (2014), pp. 1241–1255. ISSN: 15548635. DOI: 10.4161/auto.28912.

[140] Pablo Ríos-Marco, Mario Martín-Fernández, Isabel Soria-Bretones, Antonio Ríos, María P. Carrasco, and Carmen Marco. “Alkylphospholipids deregulate cholesterol metabolism and induce cell-cycle arrest and autophagy in U-87 MG glioblastoma cells”. In: *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids* 1831.8 (2013). Publisher: Elsevier B.V., pp. 1322–1334. ISSN: 18792618. DOI: 10.1016/j.bbalip.2013.05.004. URL: http://dx.doi.org/10.1016/j.bbalip.2013.05.004.

[141] Diane M. Kambach, Alan S. Halim, A. Gesine Cauer, Qian Sun, Carlos A. Tristan, Orieta Celiku, Aparna H. Kesarwala, Uma Shankavaram, Eric Batchelor, and Jayne M. Stommel. “Disabled cell density sensing leads to dysregulated cholesterol synthesis in glioblastoma”. In: *Oncotarget* 8.9 (2017), pp. 14860–14875. ISSN: 19492553. DOI: 10.18632/oncotarget.14740.

[142] Maria João Ramalho, Stéphanie Andrade, Manuel Álvaro Neto Coelho, Joana Angélica Loureiro, and Maria Carmo Pereira. “Biophysical interaction of temozolomide and its active metabolite with biomembrane models: The relevance of drug-membrane interaction for Glioblastoma Multiforme therapy”. In: *European Journal of Pharmaceutics and Biopharmaceutics* 136 (August 2018 2019). Publisher: Elsevier, pp. 156–163. ISSN: 18733441. DOI: 10.1016/j.ejpb.2019.01.015. URL: https://doi.org/10.1016/j.ejpb.2019.01.015.

[143] Yutaro Yamamoto et al. “Involvement of intracellular cholesterol in temozolomide-induced glioblastoma cell death”. In: *Neurologia Medico-Chirurgica* 58.7 (2018), pp. 296–302. ISSN: 13498029. DOI: 10.2176/nmc.ra.2018-0040.

[144] Daniela K. Schlüter, Ignacio Ramis-Conde, and Mark A.J. Chaplain. “Multi-scale modelling of the dynamics of cell colonies: insights into cell-adhesion forces and cancer invasion from in silico simulations”. In: *Journal of The Royal Society Interface* 12.103 (Feb. 6, 2015). Publisher: The Royal Society. ISSN: 17425662. DOI: 10.1098/RSIF.2014.1080. URL: https://royalsocietypublishing.org/doi/10.1098/rsif.2014.1080 (visited on 01/20/2023).

[145] Alexander R.A. Anderson. “A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion”. In: *Mathematical medicine and biology : a journal of the IMA* 22.2 (2005). Publisher: Math Med Biol, pp. 163–186. ISSN: 1477-8599. DOI: 10.1093/IMAMMB/DQI005. URL: https://pubmed.ncbi.nlm.nih.gov/15781426/ (visited on 01/10/2022).

[146] Daisuke Hoshino, Naohiko Koshikawa, Takashi Suzuki, Vito Quaranta, Alissa M. Weaver, Motoharu Seiki, and Kazuhisa Ichikawa. “Establishment and Validation of Computational Model for MT1-MMP Dependent ECM Degradation and Intervention Strategies”. In: *PLOS Computational Biology* 8.4 (2012). Publisher: Public Library of Science, e1002479. ISSN: 1553-7358. DOI: 10.1371/JOURNAL.PCBI.1002479. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002479 (visited on 04/06/2023).

[147] Johanna Myllyharju and Ernestina Schipani. “Extracellular matrix genes as hypoxia-inducible targets”. In: *Cell and Tissue Research* 339.1 (Jan. 7, 2010). Publisher: Springer, pp. 19–29. ISSN: 0302766X. DOI: 10.1007/S00441-009-0841-7/FIGURES/3. URL: https://link.springer.com/article/10.1007/s00441-009-0841-7 (visited on 01/06/2023).

[148] Christopher J. Brereton et al. “Pseudohypoxic HIF pathway activation dysregulates collagen structure-function in human lung fibrosis”. In: *eLife* 11 (Feb. 1, 2022). Publisher: eLife Sciences Publications Ltd. ISSN: 2050084X. DOI: 10.7554/ELIFE.69348. (Visited on 01/05/2023).

[149] Jing Lin, Lei Jiang, Xiaogang Wang, Wenxin Wei, Chaoli Song, Yong Cui, Xiaojun Wu, and Guan Zhong Qiu. “P4HA2 Promotes Epithelial-to-Mesenchymal Transition and Glioma Malignancy through the Collagen-Dependent PI3K/AKT Pathway”. In: *Journal of Oncology* 2021 (2021). Publisher: Hindawi Limited. ISSN: 16878469. DOI: 10.1155/2021/1406853. (Visited on 01/10/2023).

[150] M. Aktar Ali, Aimee Reis, Liang Hao Ding, Michael D. Story, Amyn A. Habib, Ansuman Chattopadhyay, and Debabrata Saha. “SNS-032 prevents hypoxia-mediated glioblastoma cell invasion by inhibiting hypoxia inducible factor-1α expression”. In: *International Journal of Oncology* 34.4 (Apr. 1, 2009). Publisher: Spandidos Publications, pp. 1051–1060. ISSN: 10196439. DOI: 10.3892/IJO_00000231/HTML. URL: http://www.spandidos-publications.com/10.3892/ijo_00000231/abstract (visited on 01/06/2023).

[151] Brenda L. Petrella, Jouko Lohi, and Constance E. Brinckerhoff. “Identification of membrane type-1 matrix metalloproteinase as a target of hypoxia-inducible factor-2α in von Hippel–Lindau renal cell carcinoma”. In: *Oncogene* 24.6 (Feb. 2005). Number: 6 Publisher: Nature Publishing Group, pp. 1043–1052. ISSN: 1476-5594. DOI: 10.1038/sj.onc.1208305. URL: https://www.nature.com/articles/1208305 (visited on 10/03/2023).

[152] Jun Wan, Huiping Chai, Zaicheng Yu, Wei Ge, Ningning Kang, Wanli Xia, and Yun Che. “HIF-1α effects on angiogenic potential in human small cell lung carcinoma”. In: *Journal of Experimental & Clinical Cancer Research* 30.1 (Dec. 2011). Number: 1 Publisher: BioMed Central, pp. 1–14. ISSN: 1756-9966. DOI: 10.1186/1756-9966-30-77. URL: https://jeccr.biomedcentral.com/articles/10.1186/1756-9966-30-77 (visited on 10/03/2023).

[153] Takeharu Sakamoto and Motoharu Seiki. “Integrated functions of membrane-type 1 matrix metalloproteinase in regulating cancer malignancy: Beyond a proteinase”. In: *Cancer Science* 108.6 (2017). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cas.13231, pp. 1095–1100. ISSN: 1349-7006. DOI: 10.1111/cas.13231. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cas.13231 (visited on 10/03/2023).

[154] Sandeep Unwith, Hailin Zhao, Lindsay Hennah, and Daqing Ma. “The potential role of HIF on tumour progression and dissemination”. In: *International Journal of Cancer* 136.11 (2015). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijc.28889, pp. 2491–2503. ISSN: 1097-0215. DOI: 10.1002/ijc.28889. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.28889 (visited on 10/03/2023).

[155] Rahimsan K. Ramachandran, Mia D. Sørensen, Charlotte Aaberg-Jessen, Simon K. Hermansen, and Bjarne W. Kristensen. “Expression and prognostic impact of matrix metalloproteinase-2 (MMP-2) in astrocytomas”. In: *PLOS ONE* 12.2 (Feb. 1, 2017). Publisher: Public Library of Science ISBN: 1111111111, e0172234. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0172234. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0172234 (visited on 01/17/2023).

[156] Emmanouil D. Karagiannis and Aleksander S. Popel. “A theoretical model of type I collagen proteolysis by matrix metalloproteinase (MMP) 2 and membrane type 1 MMP in the presence of tissue inhibitor of metalloproteinase 2”. In: *Journal of Biological Chemistry* 279.37 (Sept. 10, 2004). Publisher: Elsevier, pp. 39105–39114. ISSN: 00219258. DOI: 10.1074/jbc.M403627200. URL: http://www.jbc.org/article/S0021925820729674/fulltext (visited on 05/31/2023).

[157] Seth L. Schor, Ana M. Schor, Brian Winn, and Graham Rushton. “The use of three-dimensional collagen gels for the study of tumour cell invasion in vitro: Experimental parameters influencing cell migration into the gel matrix”. In: *International Journal of Cancer* 29.1 (Jan. 15, 1982). Publisher: John Wiley & Sons, Ltd, pp. 57–62. ISSN: 1097-0215. DOI: 10.1002/IJC.2910290110. URL: https://onlinelibrary.wiley.com/doi/full/10.1002/ijc.2910290110 (visited on 02/23/2023).

[158] Shaobo Zhou, Jonathan Salisbury, Victor R. Preedy, and Peter W. Emery. “Increased Collagen Synthesis Rate during Wound Healing in Muscle”. In: *PLOS ONE* 8.3 (Mar. 19, 2013). Publisher: Public Library of Science, e58324. ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0058324. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0058324 (visited on 03/29/2023).

Colophon

This document was typeset using a look-and-feel based on the `classicthesis` developed by André Miede and Ivo Pletikosić, available at https://bitbucket.org/amiede/classicthesis/.